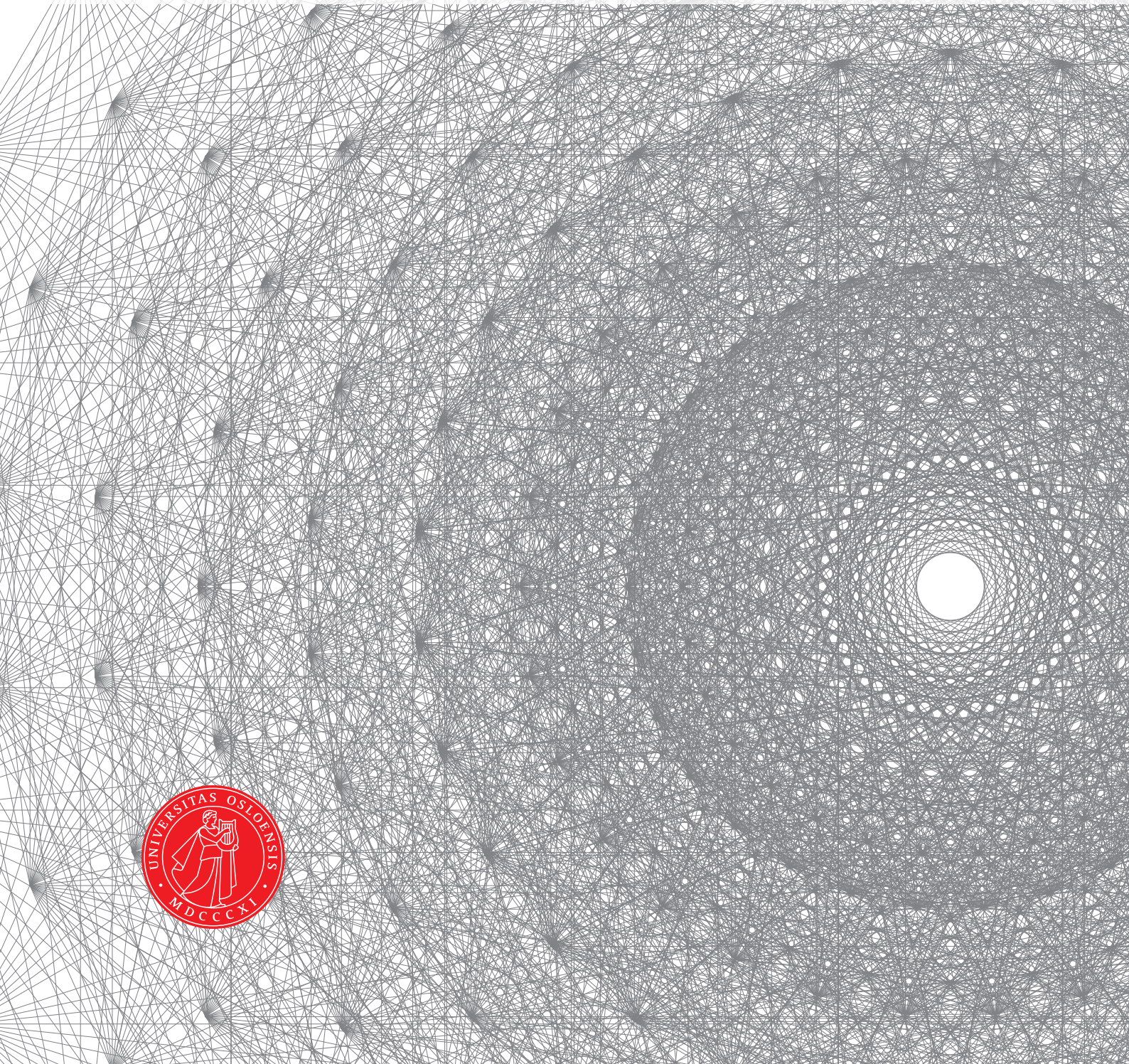
UiO : Department of Mathematics
University of Oslo

Multi-State Models for Interval-Censored Data with Transition Times from Gamma Processes

Marthe Elisabeth Aastveit

Master's Thesis, Spring 2021



This master's thesis is submitted under the master's programme *Stochastic Modelling, Statistics and Risk Analysis*, with programme option *Statistics*, at the Department of Mathematics, University of Oslo. The scope of the thesis is 60 credits.

The front page depicts a section of the root system of the exceptional Lie group E_8, projected into the plane. Lie groups were invented by the Norwegian mathematician Sophus Lie (1842–1899) to express symmetries in differential equations and today they play a central role in various parts of mathematics.

Abstract

Describing progression of a disease or the life history of an individual with multi-state models has been a topic of interest for many years. A challenge with these studies is that the data are often not continuously observed, i.e. the transition times are not recorded precisely and therefore interval-censored. The aim of this thesis is to introduce modeling of transition times as the threshold crossing times for Gamma processes in multi-state models for interval-censored data. To make this possible, we construct a suitable likelihood framework, where we set up a general likelihood for the three-state progressive model, the illness-death model, the four-state progressive model and a four-state illness-death model. The likelihood framework we create is general, meaning the transition times can be modeled by any parametric survival model. The fitting of our parametric models and the large-sample properties of the maximum likelihood estimates are also investigated using simulated data.

Another central theme in this thesis is the Markov property. Multi-state models with interval-censored data often rely on the Markov property, and we therefore investigate the Markov property in our model framework. By calculating the transition probabilities, we prove that our model framework does not necessarily rely on the Markov property. For example, when we model the transition times as the threshold crossing times for Gamma processes, the Markov property does not hold. However, if the transition times are exponentially distributed, the Markov property is satisfied and we end up with a homogeneous Markov model. For application purposes, we consider a dataset on CAV (coronary allograft vasculopathy), a post-transplant complication. The disease progression of CAV is described with a four-state illness-death model. We model the transition times as the threshold crossing times for Gamma processes, and calculate the maximum likelihood estimates. In the end, we compare our results to homogeneous and inhomogeneous Markov models, both with and without covariates. Our findings indicate that the models with Gamma processes are preferred over both the homogeneous and inhomogeneous Markov models. This holds both with and without covariates.

Acknowledgements

First and foremost, I would like to thank my supervisors Céline Cunen and Nils Lid Hjort for their encouragement, interest and investment in this thesis. Their patience, feedback, help and support have been invaluable. Developing this project together has been very interesting. I am grateful for all of our discussions and everything I have learned on the way. All errors in this thesis are surely my own.

I would also like to thank my family and friends for moral support, encouragement and believing in me. A special thanks to Joakim for support, patience, discussions and code help whenever I needed it.

Oslo, August 2021
Marthe Elisabeth Aastveit

Contents

CHAPTER 1

Introduction

Survival analysis is a field in statistics, where one studies lifetime and survival time. Lifetime is the time for an event to occur for individuals in a population. The definition of survival time is the time from an initial event to the event of interest. The event of interest can for example be death or the onset of a disease. Data where the event can only happen once are called survival data (Aalen et al., 2008, p. 2). Survival data are used in a variety of contexts, for example medicine, biology and engineering. An example is the time from a patient gets cancer treatment until death.

There also exist data with multiple events of interest. If there is a possibility of more than one final event of interest, we are in a situation with competing risks (Putter et al., 2007). If several events can happen after each other and an event can happen multiple times, we are in a situation with multiple states. We can then make use of multi-state models (Putter et al., 2007). A multi-state model is a model for a stochastic process, which at any point in time must occupy one of a set of discrete states (Hougaard, 1999). Unlike in standard survival models, there are multiple paths in multi-state models, because the individuals can transition between several states. The time of transition from one state to another is called the transition time. Since there are multiple paths, we do not necessarily know which transitions occurred (Commenges, 2002). By using a multi-state model, we can describe many different events, for example progression of a disease or the life history of an individual. Multi-state models therefore gives a great amount of flexibility for modeling different types of longitudinal data (Hougaard, 1999).

Progression of diseases, like cancer, has been a topic of interest for many years. For example in Armitage and Doll (1954), they study carcinogenesis, which is the process where normal cells are transformed into cancer cells. They test if cancer is always the end-result for different successive changes in the cells. This is done by examining age specific mortality rates of 17 types of cancer. In addition, they obtain a formula where they weight the strengths of these carcinogenic factors at different periods in time. Frank (2004) argues that the results in Armitage and Doll (1954) mark a divide in cancer research because they created mathematical models with principles of cancer progression and epidemiology before one knew the roles of different genes.

Since Armitage and Doll (1954), there has been a great amount of literature on modeling progression of diseases. Progression of cancer is still a frequently studied topic, where recent literature is for example found in Putter et al. (2006), Meira-Machado et al. (2009) and Le-Rademacher et al. (2018). In Putter

et al. (2006) they analyze 2795 patients from a breast cancer trial by applying a multi-state model. They show how the model can be used to predict the development after surgery for a fictitious patient with a given set of prognostic factors and treatment for different intermediate events. For inference in this multi-state model, they use a stratified Cox regression model. In Meira-Machado et al. (2009), they review different modeling approaches for multi-state models. They consider both parametric and nonparametric approaches, and apply the resulting models on breast cancer data. In Le-Rademacher et al. (2018) they study how multi-state models can give a deeper understanding of the effect of treatment in cancer-clinical trials.

A frequently studied multi-state model, regardless of whether the disease of interest is cancer, HIV, dementia or any other irreversible disease, is the three-state illness-death model. A three-state illness-death model, which we from now on refer to as the illness-death model, is illustrated in Figure 1.1. The individual can transition from healthy to diseased, from healthy to dead or from diseased to dead. A variant of this illness-death model is for instance discussed in Fix and Neyman (1951) and Sverdrup (1965). In Fix and Neyman (1951), they present a stochastic model of recovery, relapse and death of cancer patients. To capture those people lost after recovery, they use four states. They define state 0 as being in cancer treatment, state 1 as being dead immediately after cancer, state 2 as recovered, while state 3 is lost after recovery, which means either death from other causes or difficulties tracing the patient. In recent years, illness-death models are often studied with a nonparametric approach. Examples of a nonparametric approach to an illness-death model are for example found in de Uña-Álvarez and Meira-Machado (2015) and Frydman (1995).

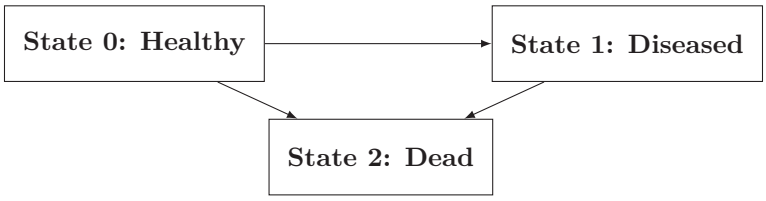


Figure 1.1: Illness-death model

The data we consider in this thesis are interval-censored data where each individual is screened (observed) multiple times. This is often called panel data. Panel data are when we observe the state of a process an arbitrarily set of times, and one do not necessarily know the exact time of transition, meaning the data can be interval-censored (Jackson, 2016). Interval-censored data means that we do not know the exact point an event happened, only that it happened between two time points (Lindsey & Ryan, 1998). The setting is therefore more complex since the data are not continuously observed, i.e. the transition times are not recorded precisely. In this thesis, when we refer to interval-censored data in a multi-state setting, it is the same as panel data where the transition times are not known exactly.

In this thesis, we consider a fully parametric approach when studying multi-state models for interval-censored data. We model the transition times as the threshold crossing times for Gamma processes. A Gamma process is a continuous process in continuous time where the increments follow a Gamma

distribution (Caroni, 2017, p. 76). Modeling the transition times as the threshold crossing times for Gamma processes for multi-state models with interval-censored data, has to the best of our knowledge never been done before. The type of model we just described, will in the rest of this thesis be referred to as the Gamma process model.

We construct a suitable likelihood framework for multi-state models with interval-censored data for the Gamma process models. This likelihood framework is general, meaning the transition times can be modeled using any parametric survival times model. The likelihood framework is related to the idea behind the general models in Hougaard (1999), but our framework is tailored to interval-censored data. In addition, we also define the likelihood in a different way, meaning that we construct our likelihood by dividing it into different contributions, based on the time points for screening and in which states the individual is observed. We call these groups of different likelihood contributions *types*. The number of types required depends on the number of states and possible transitions between the states. In an illness-death model, one individual can for instance be observed in state 0 at all the screening time points, while another individual can first be observed in state 0, then in state 1. These two individuals are different types since they give different likelihood contributions.

The fitting of our parametric models and the properties of the maximum likelihood estimates (MLE) are also investigated by using simulated data. This means we check the large-sample properties of the MLEs by estimation. We find that the large-sample properties in general are satisfied.

A property multi-state models often rely on, is the Markov property. In a multi-state setting, the Markov property means that given the present state and history of an individual, the transition to the next state and the time this occurs, only depends on the present state (Putter et al., 2007). One reason is that when the transition to the next state also depend on when the individual was in all of the previous states, the model becomes much more complicated (Hougaard, 2000, p. 159). We therefore want to investigate the Markov property in our model framework. By calculating the transition probabilities, we prove that the Markov property is not in general fulfilled in our models. For example, the Markov property is not fulfilled in the Gamma process model. However, if the transition times are exponentially distributed, the Markov property is fulfilled. Therefore, by changing the modeling of the transition times, we can adapt to the data based on whether the Markov property is realistic or not.

We apply the likelihood framework we set up on a dataset on CAV (coronary allograft vasculopathy), a complication after a heart transplantation. The dataset CAV is found in the *msm*-package in *R*, see Jackson (2019) and Jackson (2011) for further information. Since the *msm*-package is primarily based on Markov models, the CAV-data has frequently been studied using Markov models. We therefore compare the Gamma process models with the Markov models. We use AIC for model selection, and the model with the lowest AIC is considered to be the preferred model. For the CAV-data, we find that the Gamma process models have a much lower AIC than the homogeneous and inhomogeneous Markov models. This holds for the models with and without covariates. We then discuss various explanation for why the Gamma process models appear to be better than the Markov models for this dataset. In addition, for the models without covariates, we also compare the total survival probability functions from the Gamma process models and the Markov models to a Kaplan-Meier

estimate of the total survival probability.

This thesis is organized as follows. In Chapter 2 we present the preliminaries and background of the thesis. We focus on the theoretical aspects, both relating to interval-censored data, maximum likelihood theory and generalities about first-hitting time regression models. In Chapter 3 we construct the likelihood framework using likelihood contributions which we call types. Then we investigate the model construction through simulations in Chapter 4. We check that the estimated parameters are close to the true parameters and that the maximum likelihood estimates are close to normally distributed. We investigate the Markov property in our model framework in Chapter 5. Further, in Chapter 6, we apply our likelihood construction on a dataset called CAV. Finally, we summarize and discuss future work in Chapter 7.

CHAPTER 2

Preliminaries

2.1 Basic Concepts in Survival Analysis

In Chapter 1, we introduced the terms survival time and lifetime. We defined lifetime as the time for an event to occur for individuals in a population, while survival time is defined as the time from an initial event to the event of interest (Aalen et al., 2008, p. 2). An example of survival time is the time from cancer diagnosis to death for a certain individual.

The survival function gives the probability that the event of interest has not happened by time t (Aalen et al., 2008, p. 5). We define the survival function in Equation 2.1.

$$S(t) = Pr(T > t) = 1 - Pr(T \leq t) = 1 - F(t), \tag{2.1}$$

where $F(t)$ is the cumulative distribution function. The density becomes

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}.$$

An important concept in survival analysis is censoring. Assume we have a study about cervical cancer, where we follow women over time. There are three possibilities for a woman at the end of the study; the woman can be healthy, have cervical cancer or be dead. However, we do not know if one of the healthy women will develop cervical cancer later on. These incomplete observations are therefore censored, and we call them censored survival times (Aalen et al., 2008, p. 3).

In the example of censoring above, an individual may leave the study before it ends or the study ends before the event has occurred. This is the most common type of censoring, and is called right censoring. When we have right censoring, either the event for individual i is observed before the censored time C_i and we observe the lifetime T_i, or the true lifetime is to the right of C_i. This means we either know the true lifetime T_i or the censoring time C_i (Lawless, 2003, p. 52).

In addition to right censoring, we also have left and interval censoring. Left censoring is when an event has already happened before the starting point, but you do not know exactly when it happened (Clark et al., 2003). The true lifetime T_i is then to the left of the censoring time C_i. For example assume we study at which age children learn partial integration. Then we might have left censoring, since some of the children may already know partial integration at the start of the study.

Interval censoring happens if you do not exactly know at which time point an event occurred, only that it happened between two time points. For an individual i, we only observe data which consist of an interval $(U_i, V_i]$. The true lifetime T_i then lies somewhere in between these two timepoints, meaning $U_i < T_i \leq V_i$ (Lawless, 2003, p. 64). An example is relapse of a disease. If you take blood samples every third month and the last sample was normal and the next sample shows relapse, then you do not know exactly at which point in time the relapse happened. The only information you have is that it happened sometime in these three months. Interval-censored data in survival analysis is for example discussed in Lindsey and Ryan (1998). In that tutorial, they argue that there is a lack of well-known statistical methodology for interval-censored data. One therefore often assumes that the event happened at the beginning, midpoint or end of each interval and then use standard methods for time-to-event data. This approach may lead to invalid inferences. Especially, the standard errors will be underestimated. In their paper, they therefore compare and illustrate available methods, both parametric and nonparametric, where they account for the data being interval-censored.

The hazard rate is another relevant concept in survival analysis. The hazard rate $\alpha(t)$ is defined as

$$\alpha(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr(t \leq T < t + \Delta t | T \geq t).$$

$\alpha(t)dt$ is interpreted as the probability that the individuals not having experienced the event by time t, will experience the event in the small time interval $[t, t+dt)$ (Aalen et al., 2008, p. 6). The connection between the survival function and the hazard function is

$$\alpha(t) = -\frac{S'(t)}{S(t)}$$

(Aalen et al., 2008, p. 6).

The cumulative hazard rate, $A(t)$, is defined as

$$A(t) = \int_0^t \alpha(s)ds.$$

The cumulative hazard rate is interpreted as an accumulation of the hazard functions over time. In parametric models, it is also connected to the survival function through

$$A(t) = -\log(S(t))$$

(Aalen et al., 2008, p. 6).

It is not straightforward to estimate the hazard rate, but the cumulative hazard rate can be estimated nonparametrically by the Nelson-Aalen estimator (Aalen et al., 2008, p. 6). The Nelson-Aalen estimator is given by

$$\hat{A}(t) = \sum_{T_j \leq t} \frac{1}{Y(T_j)}.$$

In order to explain the intuition of the Nelson-Aalen estimator, we start by splitting the interval $[0, t]$ into small intervals, for example $[s, s + ds)$. Each

interval contains at most one observed event. The contribution to the cumulative hazard for this interval is $\alpha(s)ds$. $\alpha(s)ds$ is interpreted as the conditional probability that an event occurs in this interval, given that it has not happened before time s. If no event is observed in this time interval, $\alpha(s)ds$ is estimated to be zero. If an event is observed at time $T_j \in [s, s + ds)$, then a natural estimator for $\alpha(s)ds$ will be one divided by the number of individuals still at risk, which is $1/Y(s) = 1/Y(T_j)$. By aggregating these contributions, we get $\hat{A}(t)$, which is a sensible estimator for $A(t)$ (Aalen et al., 2008, p. 72).

The Kaplan-Meier estimator is a nonparametric way of estimating the survival function. We start by giving an intuitive introduction to the Kaplan-Meier estimator. The first step is to divide the interval $[0, t]$ into a number of small time intervals $0 = t_0 < t_1 < \cdots < t_K = t$. Then using the multiplication rule for conditional probabilities

$$S(t) = \prod_{k=1}^{K} S(t_k|t_{k-1}),$$

where $S(v|u)$ for $v > u$ means the conditional probability that an event will occur later than time v, given that it has not happened yet at time u. An important assumption is that there are no tied events, and the time intervals are so small that they contain at most one event. If no event is observed in $(t_{k-1}, t_k]$, we estimate $S(t_k|t_{k-1})$ by 1, but if an event is observed in $T_j \in (t_{k-1}, t_k]$, it is natural to estimate $S(t_k|t_{k-1})$ by $1 - 1/Y(t_{k-1}) = 1 - 1/Y(T_j)$. The Kaplan-Meier estimator becomes

$$\hat{S}(t) = \prod_{T_j \leq t} \Big\{1 - \frac{1}{Y(T_j)}\Big\}$$

(Aalen et al., 2008, pp. 90 - 91).

2.2 Competing Risks

In this section, we give a brief introduction to a special case of multi-state models, called competing risks. In a competing risks situation, there is more than one possible endpoint, which means there is more than one possible cause of failure. Figure 2.1 shows a competing risks situation with three different causes of failure. The causes of failure depend on the research question, and can for example be different causes of death. A criticized assumption for competing risks is that the risk of failure in the remaining causes are unchanged if one cause of failure is removed. This is often true in industrial settings, but not in medical settings (Putter et al., 2007).

Competing risks problems are often formulated by using latent failure times for each type of failure. Let the failure times $Y_1, \ldots, Y_m$ correspond to each type of failure $J = 1, \ldots, m$. We observe the time point T and type of failure J, where $T = \min(Y_1, \ldots, Y_M)$ and $J = \{j|Y_j \leq Y_k, k = 1, \ldots, m\}$ (Prentice et al., 1978). This means that for the observed failure time $T = Y_j$, the individual fails of cause j. The focus is often on the joint distribution of the times to the J events. The joint survival function is then $S(t_1, \ldots, t_J) = Pr(Y_1 > t_1, \ldots, T_J > t_J)$. However, one issue is that without any further assumptions, the joint survival function not identifiable from the observed data (Putter et al., 2007). There

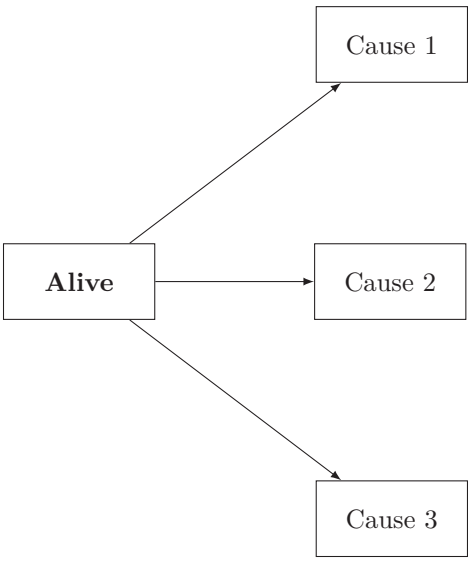


Figure 2.1: Competing risks model

exists a large amount of literature about competing risks, see for example Cox (1959), Prentice et al. (1978), Berman (1963), Nelson (1970) and Putter et al. (2007).

Competing risks can also be analyzed using stochastic thresholds and first-hitting time regression models. Studying semi-competing risks using stochastic thresholds is for example done in Sildnes and Lindqvist (2018). Semi-competing risks means that both a terminal event, for example death, and a non-terminal event, for example disease recurrence, are considered. They present a model, where time to event is a stochastic process. The time to the terminal event is the first passage time to a fixed level c, while for the non-terminal event is a stochastic threshold S. S is independent of the stochastic process. They let the stochastic process be a Gamma process.

2.3 Multi-State Models

Different approaches to multi-state models have been reviewed and analyzed in a variety of settings, for example in Andersen and Keiding (2002), Hougaard (1999) and Putter et al. (2007). A multi-state model is defined as a model for a stochastic process, where an individual at any point in time occupy one of a small set of discrete states (Hougaard, 1999). The states in a multi-state model are divided into initial, intermediate/transient and final/absorbing states. The absorbing state is the endpoint, and the individual can not leave this state when it has been reached. The states in the middle are called intermediate or transient states (Putter et al., 2007).

The complexity of a multi-state model depends on the number of states and whether the process is progressive or not. Two common multi-state models are presented in Figures 2.2 and 2.3. The model in Figure 2.2 is a k-progressive model, while the model in Figure 2.3 is the illness-death model. A process is progressive when each state, except the initial state, has only one possible

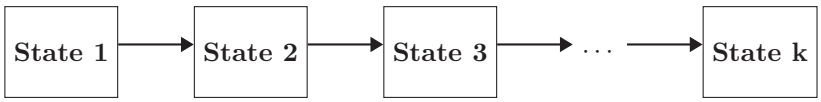


Figure 2.2: k-state progressive model

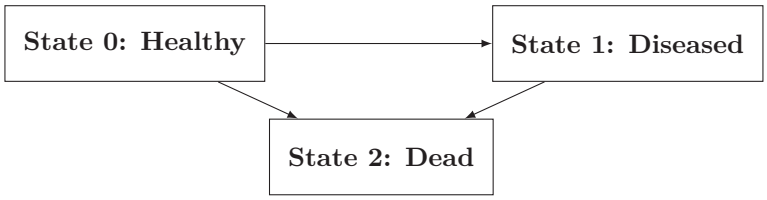


Figure 2.3: Illness-death model

transition into it. In a progressive model, the current state includes information about which states have been visited and in which order. The time of transition is not necessarily included (Hougaard, 1999). An illness-death model is not progressive, since the individual also can transition directly from the initial state to the absorbing state.

Following Hougaard (2000, p. 144) we consider a stochastic process $X_t, t \in [0, \infty)$, where $X_t = \ell$ if the process is in state ℓ at time t. The process is right continuous and piecewise constant, with limits from the left. When we say history or past at time t, we mean the information in the development of the process over the time $[0, t]$. We then have the stochastic process X_s, where $0 \leq s \leq t$. The transition probability is

$$P_\ell(t) = Pr(X_t = \ell),$$

which is the probability of a process X being in state ℓ at time t. Note that if the processes do not start in state 0, the expression should depend on the initial state. The transition probability at time v is defined as

$$P_\ell(v, t) = Pr(X_t = \ell | X_u, u \in [0, v]),$$

where we condition on the development until time point v. From Hougaard (1999), we have that the transition intensity (hazard) for transitioning from state m to state ℓ can be expressed as

$$q_{m\ell}(t|X_u, u \in [0, t)) = \lim_{\Delta t \searrow 0} \frac{Pr\{X_{t+\Delta t} = \ell | X_{t^-} = m\}}{\Delta t}. \tag{2.2}$$

2.3.1 Markov, semi-Markov and extended Markov models

A common property in multi-state models is the Markov property. In a multi-state setting, the Markov property means that given the present state and history of an individual, the time of transition to the next state, only depends on the present state (Putter et al., 2007). More formally, the Markov property can be written as

$$P_{m\ell}(v, t) = Pr(X_t = \ell | X_v = m) = Pr(X_t = \ell | X_v = m, X_u, u \in [0, v)),$$

for $v \leq t$ (Hougaard, 2000, p. 144) .

We define the transition intensities in the same way as in Equation 2.2. Let us consider a four-state illness-death model where it is possible to transition both ways. Figure 2.4 illustrates the possible transitions.

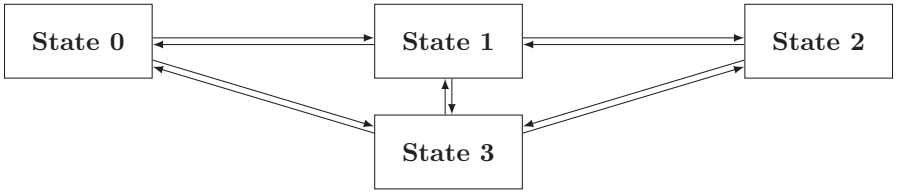


Figure 2.4: Four-state illness-death model with transitions both ways

The possible transitions with corresponding transition intensities, are written in a transition matrix called $\boldsymbol{Q}$. The transition matrix corresponding to Figure 2.4 is

$$\boldsymbol{Q}(t) = \begin{bmatrix} -(q_{12}(t)+q_{13}(t)+q_{14}(t)) & q_{12}(t) & q_{13}(t) & q_{14}(t) \\ q_{21}(t) & -(q_{21}(t)+q_{23}(t)+q_{24}(t)) & q_{23}(t) & q_{24}(t) \\ q_{31}(t) & q_{32}(t) & -(q_{31}(t)+q_{32}(t)+q_{34}(t)) & q_{34}(t) \\ q_{41}(t) & q_{42}(t) & q_{43}(t) & -(q_{41}(t)+q_{42}(t)+q_{43}(t)) \end{bmatrix}.$$

After the transition matrix $\boldsymbol{Q}$ is defined, one is often interested in the transition probabilities. The transition probabilities can be calculated directly, which we do in Section 2.3.1.1. For homogeneous Markov models, one can also use the Kolmogorov backward equation (Jackson, 2011). In a homogeneous Markov model, the transition intensities are constant and do not depend on time (Hougaard, 2000, p. 160). Following Hougaard (2000, p. 160), we consider a continuous-time Markov process with a time-homogeneous $K \times K$ $\boldsymbol{Q}$-matrix, the transition probability $\boldsymbol{P}(t)$ for $t > 0$ is the solution to the Kolmogorov backward equation $\boldsymbol{P}'(t) = \boldsymbol{Q}\boldsymbol{P}(t)$ subject to $\boldsymbol{P}(0) = \boldsymbol{I}_K$, where $\boldsymbol{I}_K$ is the identity matrix. The solution becomes the matrix exponential

$$\boldsymbol{P}(t) = e^{t\boldsymbol{Q}} = \sum_{k=0}^{\infty} \frac{(t\boldsymbol{Q})^k}{k!}.$$

If the $K \times K$ matrix $\boldsymbol{Q}$ has K linearly independent eigenvectors, then we can express this exponential using an eigenvalue decomposition of $\boldsymbol{Q}$ (Van Den Hout, 2017, p. 199). We define

$$\boldsymbol{P}(t) = \boldsymbol{R}e^{\boldsymbol{N}(t)}\boldsymbol{R}^{-1},$$

where $\boldsymbol{R}$ consists of the eigenvectors to $\boldsymbol{Q}$ and $\boldsymbol{N}$ is a matrix with the eigenvalues on the diagonal. An example of complete calculations using eigenvalue decomposition in a four-state illness-death model is found in Klotz and Sharples (1994). In addition, we also do similar calculations in Chapter 5.

The Markov model for panel data, was described for the first time in Kalbfleisch and Lawless (1985) and Kay (1986). They derive the likelihood from the transition probability matrix. The same procedure is used in Jackson (2011) for the *msm*-package. The likelihood is then a product of the probabilities of transitioning between the observed states, for all the individuals i and observation times j. For interval-censored transition times, the likelihood

becomes

$$\mathcal{L}(\boldsymbol{Q}) = \prod_i L_i = \prod_{i,j} L_{i,j} = \prod_{i,j} p_{X(t_{ij})X(t_{i,j+1})}(t_{i,j+1} - t_{i,j}),$$

where each component $L_{i,j}$ is the entry of the transition matrix $\boldsymbol{P}(t)$ for the $X(t_{ij})$th row and $X(t_{i,j+1})$th column evaluated at $t = t_{i,j+1} - t_{ij}$. The likelihood $\mathcal{L}(\mathcal{Q})$ is maximized in terms of $\log(q_{rs})$ and one can use standard optimization algorithms to do this. If time of death is known, we have $X(t_{i,j+1}) = D$, the the likelihood contribution at this time, is summed over the unknown state m at the instant before death

$$\mathcal{L}_{i,j} = \sum_{m \neq D} p_{X(t_{ij}),m}(t_{i,j+1} - t_{ij}) q_{m,D}.$$

Another type of Markov model is the time-inhomogeneous model. In a time-inhomogeneous model, the transition intensities may depend on time. An example is piecewise-constant intensities where one chooses change points for the intensity function (Jackson, 2011). For example, if the transition intensities change after 5 years, then one constructs a model with the time period as a factor. We then get two levels, $(-\infty, 5]$ and $[5, \infty)$, where the first period $(-\infty, 5]$ is the baseline (Jackson, 2011). We discuss the implementations of a time-inhomogeneous Markov model in Chapter 6, where we also consider an example of a time-inhomogeneous Markov model for the CAV data. Another example of an analysis of a time-inhomogeneous Markov model is found in Gil et al. (2007) where they study ALS in an illness-death model. They use a time-inhomogeneous model where the transition intensities are piecewise constant and they consider two periods where the intensities vary between these two periods.

For some type of data, it is also useful to relax the Markov property. An example is the extended Markov models. In the extended Markov models, the hazard functions may depend on the time of the latest transition. A semi-Markov model is a special case of an extended Markov model. In a semi-Markov model, the hazard does not depend on the current time, but on the duration of the current state (Hougaard, 2000, pp. 168-169). An example of a semi-Markov multi-state model is for example found in Foucher et al. (2007). In this paper, they define a semi-Markov model where they allow for interval-censored data. They define parametric hazard functions with a $\cup$- or $\cap$-shape, more specifically the generalized Weibull hazard function, and the initial states are determined according to covariates. The hazard function from state i to j is given by

$$\alpha_{ij}(d_{h,r}) = \frac{1}{\theta_{ij}} \Big(1 + \Big(\frac{d_{h,r}}{\sigma_{ij}}\Big)^{v_{ij}}\Big)^{1/\theta_{ij}-1} \frac{v_{ij}}{\sigma_{ij}} \Big(\frac{d_{h,r}}{\sigma_{ij}}\Big)^{v_{ij}-1},$$

where $d_{h,r} = t_{h,r+1} - t_{h,r} \geq 0$ and $t_{h,r}$ is the time of the rth transition for the hth subject. In addition, $v_{ij} > 0$ is the shape parameter, $\sigma_{ij} > 0$ is the scale parameter and θ_{ij} is the location parameter. Each modeling approach is specific of each transition. They evaluate a multi-state model with several absorbing states, and apply their model on a kidney transplant recipient follow-up.

2.3.1.1 Example: Three-State Progressive Model

Figure 2.5: Three-state progressive model

In this part, we give an example of a three-state progressive model for interval-censored data for a homogeneous Markov model. $q_{m\ell}(t)$ is the hazard for the transition from state m to state ℓ at time t. The cumulative hazard for leaving state 0 in the time interval $(t_1, t_2]$ is

$$A_1(t_1, t_2) = \int_{t_1}^{t_2} q_{01}(u)du,$$

and the same for leaving state 2

$$A_2(t_1, t_2) = \int_{t_1}^{t_2} q_{12}(u)du,$$

We assume that state 3 is the absorbing state. Then $q_{02}(t) = q_{10}(t) = q_{20}(t) = q_{21}(t) = 0$ and $q_{22}(t) = 1$. The corresponding $\boldsymbol{Q}$-matrix is

$$\boldsymbol{Q} = \begin{bmatrix} -q_{01} & q_{01} & 0 \\ 0 & -q_{12} & q_{12} \\ 0 & 0 & 1 \end{bmatrix}.$$

Since this model is quite small, we can find $\boldsymbol{P}(t)$ in two different ways. Either by using eigenvalue decomposition, as we explained in Section 2.3.1 or by finding the probabilities directly. It is easier to calculate the transition probabilities directly than going through the eigenvalue decomposition for this three-state progressive model. We follow Van Den Hout (2017, p. 35), which finds the probabilities directly. These derivations constitute a special case of the general likelihood construction which we present in Chapter 3. Assume an exponential model, where the hazard is constant, which gives $q_{rs}(t) = q_{rs}$. The transition probabilities $p_{m\ell}(t_1, t_2) = Pr(X_{t_2} = \ell | X_{t_1} = m)$ are

$$p_{00}(t_1, t_2) = S_1(t_2 - t_1) = \exp(-A_1(t_1, t_2)) = \exp(-q_{01}(t_2 - t_1))$$

$$\begin{aligned} p_{01}(t_1, t_2) &= \int_{t_1}^{t_2} S_1(u - t_1) q_{01}(u) S_2(t_2 - u) du \\ &= \int_{t_1}^{t_2} \exp(-q_{01}(u - t_1)) q_{01} \exp(-q_{12}(t_2 - u)) du \\ &= \frac{q_{01}}{q_{12} + q_{01}} \Big(\exp(q_{12}(t_1 - t_2)) - \exp(q_{01}(t_1 - t_2)) \Big) \end{aligned}$$

$$\begin{aligned} p_{02}(t_1, t_2) &= 1 - p_{00} - p_{01} \\ &= 1 - \frac{q_{01}}{q_{12} + q_{01}} \exp(-q_{12}(t_2 - t_1)) + \frac{q_{12}}{q_{12} + q_{01}} \exp(-q_{01}(t_2 - t_1)) \end{aligned}$$

$$p_{10}(t_1, t_2) = 0$$
$$p_{11}(t_1, t_2) = S_2(t_2 - t_1) = \exp(-A_2(t_1, t_2)) = \exp(-q_{12}(t_2 - t_1))$$
$$p_{12}(t_1, t_2) = 1 - p_{11}(t_1, t_2) = 1 - \exp(-q_{12}(t_2 - t_1))$$
$$p_{20}(t_1, t_2) = 0$$
$$p_{21}(t_1, t_2) = 0$$
$$p_{22}(t_1, t_2) = 1$$

We get that

$$\boldsymbol{P}(t) = \begin{bmatrix} e^{-q_{01}t} & \frac{q_{01}}{q_{12}+q_{01}}(e^{-q_{12}t} - e^{-q_{01}t}) & 1 - \frac{q_{01}}{q_{12}+q_{01}}e^{-q_{12}t} + \frac{q_{12}}{q_{12}+q_{01}}e^{-q_{01}t} \\ 0 & e^{-q_{12}t} & 1 - e^{-q_{12}t} \\ 0 & 0 & 1 \end{bmatrix}$$

2.3.1.2 Examples: Markov Models for Estimation of Dementia

We present two examples of application and construction of Markov models for an irreversible disease, in this case dementia. The data are interval-censored, except time of death which is assumed to be known exactly. In Jack Jr. et al. (2016), they consider a multi-state Markov model while in Williams et al. (2020), they consider a hidden multi-state Markov model. The dataset consist of at least two biomarkers, amyloid and neurodegeneration. It is considered known in the medical community that amyloid protein buildup in the brain and significant neurodegeneration are associated with dementia (Williams et al., 2020). In Figure 2.6 and Figure 2.7, A^+ means high amyloid protein buildup and N^+ means significant neurodegeneration. In both of these papers, they are especially interested in the relationship between age and dementia.

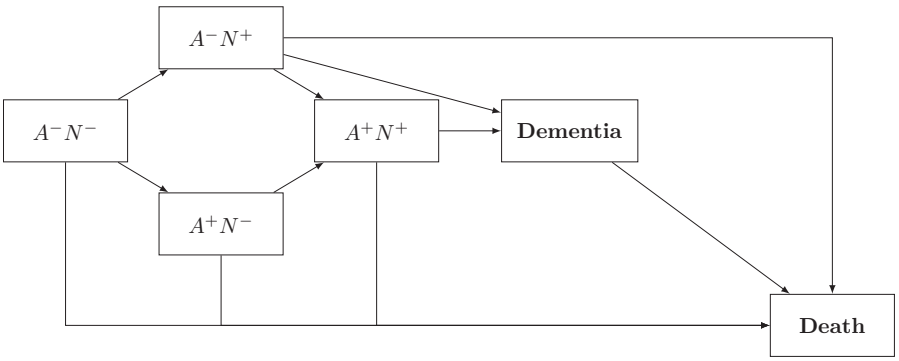


Figure 2.6: The multi-state model in Jack Jr. et al. (2016). A^+ means high amyloid protein buildup and N^+ means significant neurodegeneration

Figure 2.6 illustrates the possible transitions in the Markov model in Jack Jr. et al. (2016). The model consists of six states and an individual can always transition directly to death. They allow the transition probabilities to vary with age. To construct the overall likelihood, they use data from different data sources. In the end, they maximize the likelihood and use the results to calculate the estimated transition rates.

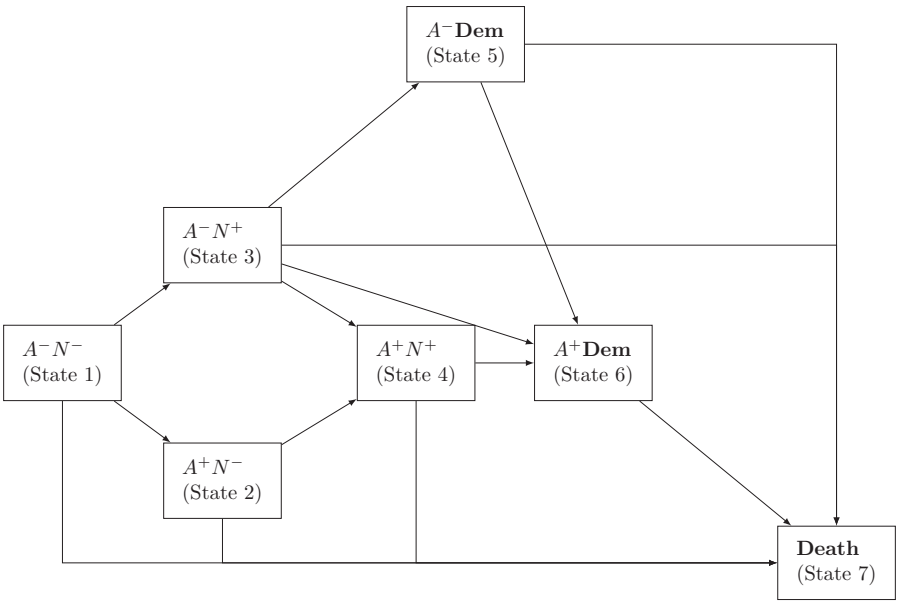


Figure 2.7: The multi-state model in Williams et al. (2020). A^+ means high amyloid protein buildup, N^+ means significant neurodegeneration and Dem means dementia

In Williams et al. (2020) they build on the Markov model in Jack Jr. et al. (2016), where the possible transitions are illustrated in Figure 2.7. They expand the model to seven states, in order to pinpoint the Alzheimer transition, from state 4 (A^+N^+) to state 6 (A^+ **Dem**). In addition, they consider a hidden Markov model (HMM) instead of a Markov model. A hidden Markov model is a double stochastic process. It consists of an underlying stochastic process, which is not observable, but can be observed through another set of stochastic processes (Rabiner & Juang, 1986). They use a HMM in Williams et al. (2020) because the underlying state sequences for the patients are not observed in their data. They therefore use the responses emitted from the underlying process to give information about the underlying state.

The HMM in Williams et al. (2020) consists of seven states where the individual can always transition directly to death from any of the states in the model. They estimate the transition intensities for each of the 13 nonzero transition rates illustrated in Figure 2.7 by q_l for $l \in \{1, \ldots, 13\}$

$$\log(q_l) = \beta_0^{(l)} + \beta_1^{(l)} \cdot \text{age} + \beta_2^{(l)} \cdot \text{male} + \beta_3^{(l)} \cdot \text{educ} + \beta_4^{(l)} \cdot \text{apoe4},$$

where the covariates are age, sex, years of educations and presence of an APOE-ϵ4 allele. APOE-ϵ4 allele is known to increase the risk of $A+$ (Williams et al., 2020). They define the transition matrix as

$$\boldsymbol{Q} = \begin{bmatrix} -(q_1+q_2+q_3) & q_1 & q_2 & 0 & 0 & 0 & q_3 \\ 0 & -(q_4+q_5) & 0 & q_4 & 0 & 0 & q_5 \\ 0 & 0 & -(q_6+q_7+q_8) & q_6 & q_7 & 0 & q_8 \\ 0 & 0 & 0 & -(q_9+q_{10}) & 0 & q_9 & q_{10} \\ 0 & 0 & 0 & 0 & -(q_{11}+q_{12}) & q_{11} & q_{12} \\ 0 & 0 & 0 & 0 & 0 & -q_{13} & q_{13} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

In Williams et al. (2020) they find the transition probabilities using these transition intensities. They use the transition probabilities and the emitted responses to calculate the likelihood for the HMM. The four emitted responses in this analysis are (i) $\log(PIB - 1)$, (ii) thickness, (iii) MMSE and (iv) dementia diagnosis. PIB is the measure of the amyloid buildup (A) and thickness is the measure of neurodegeneration (N). MMSE is a Mini-Mental State Exam, which is a questionnaire-based test to see whether a person has cognitive impairment or not. If a response is missing, it is integrated out of the likelihood (Williams et al., 2020). In order to estimate the parameters in the HMM, they propose a hierarchical Bayesian approach where the model is fitted by Markov Chain Monte Carlo (MCMC).

2.3.2 General Models

In this part, we introduce the general progressive models presented in Hougaard (1999) and Hougaard (2000). These models do not rely on the Markov property, which makes them different from the already presented Markov models.

We start by observing a set of n processes over specified time periods. If the absorbing state is not reached, the end of the observation is a censoring time. However, if the absorbing state is reached, there is no information about the process after this state is reached. We observe E events, where the times of transitions are $T_1, \dots, T_E$ and the states the transitions lead into are called $X_1, \dots, X_E$ (Hougaard, 1999).

Following Hougaard (2000, p.159), we present the transition probabilities in a general progressive model. In a general progressive model, all of the terms depend on the whole history and the transition probabilities are therefore more complicated. The hazard of the transition from state m to state ℓ is defined as $\alpha_{m\ell}(t|T_1, \dots, T_k)$. From this hazard function, we implicitly know the process was in state m immediately before time point t. Since the transition probabilities are only defined for progressive models, state m implicitly informs which states have been visited up to time k. This means that state X_j is known for $j = 0, 1, \dots, k$, where $X_k = m$. The transition probability to state m is then

$$P_m(t) = \int_0^t \int_{t_1}^t \cdots \int_{t_{k-1}}^t \Big[\prod_{j=1}^{k} \alpha_{X_{j-1}X_j}(u_j|\tilde{u}_{j-1}) \exp\Big(- \int_{u_{j-1}}^{u_j} \alpha_{X_{j-1}}(v|\tilde{u}_{j-1})dv \Big) \Big]$$
$$\exp\Big(- \int_{u_k}^{t} \alpha_{s_k}(v|\tilde{u}_k)dv \Big) du_1, \dots du_k,$$

where $\tilde{u}_j = (u_1, \dots u_j)$ is all the time points up to the jth.

Next, we follow Hougaard (2000, p. 182), and present the likelihood function for a general progressive model in the time period 0 to C with E events. The likelihood is

$$\{\prod_{j=1}^{E} \alpha_{X_{j-1},X_j}(T_j|\tilde{T}_{j-1}) \exp\Big\{-\int_{T_{j-1}}^{T_j} \alpha_{X_{j-1}}(v|\tilde{T}_{j-1})dv\Big\}\} \exp\Big\{-\int_{T_E}^{C} \alpha_{s_E}(v|\tilde{T}_E)\Big\},$$

where $\tilde{T}_j = (T_1, \ldots, T_j)$. Since the transitions between the states happen at the observed time points $T_1, \ldots, T_E$, it is not interval-censored. It is possibly right-censoring because the end of the study is at time point C and the last observation happens earlier. This likelihood for the general progressive models, are to some extent related to the likelihood framework we construct in Chapter 3. However, a difference is that the likelihood framework in Chapter 3 is tailored to interval-censored data. The models are therefore used for different types of data.

2.4 Nonparametric Methods for Interval-Censored Survival Data and Panel Data

There exists a large amount of literature on nonparametric methods for multi-state models. For example, in Aalen and Johansen (1978) they introduce nonparametric estimation of the transition probabilities for right-censored observations when there are multiple states. This estimator is called the Aalen-Johansen estimator, and is frequently used for inhomogeneous Markov models. In recent years there has been a big interest in the development of nonparametric estimators for the transition probabilities for multi-state models where one do not assume a Markov model. This is for example done in Meira-Machado et al. (2006) and de Uña-Álvarez and Meira-Machado (2015). In these papers, they mainly focus on estimators for the illness-death model, and argue that their models outperform the Aalen-Johansen estimator when the Markov property is violated.

In this part, we present the construction of the likelihood with a nonparametric approach for interval-censored lifetime data. This is because the likelihood for nonparametric approaches in multi-state models with multiple observations, build on the likelihood and nonparametric approaches for lifetime data. We follow Lawless (2003, p. 124) in the construction of the likelihood using a nonparametric approach. Let the true lifetime be between two points, which means $U_i < T_i \leq V_i$. The likelihood function when the lifetimes for an individual T_i is identically distributed with cumulative distribution function (c.d.f.) $F(t)$, becomes

$$\mathcal{L} = \prod_{i=1}^{n} [F(V_i) - F(U_i)]. \tag{2.3}$$

This means that $F(t)$ only depends on the values through the observation times (U_i, V_i) and the survival function $S(t) = 1 - F(t)$. This can be reformulated, by letting $0 = t_0 < t_1 < \cdots < t_{k-1} < t_k = \infty$ to be the distinct values in the set $\{0, \infty; U_i, V_i : i = 1, \ldots, n\}$, where the exact observation t is considered as $(t-, t]$. Let $p_j = F(t_j) - F(t_{j-1})$ and $\eta_{ij} = I\{(t_{j-1}, t_j) \subseteq (U_i, V_i]\}$. Then

rewriting equation 2.3

$$\mathcal{L}(\boldsymbol{p}) = \prod_{i=1}^{n} \Big[\sum_{j=1}^{k} \eta_{ij} p_j \Big], \tag{2.4}$$

where $\mathcal{L}(\boldsymbol{p})$ is maximized subject to the constraints $p_j \geq 0$ and $\sum p_j = 1$ in order to obtain $\hat{F}$. Many algorithms are proposed to maximize 2.3 and 2.4. For example, the survival distribution was first estimated using a nonparametric method in Peto (1973). He constructed an experimental survival curve by using a suitably constrained Newton-Raphson search algorithm. The idea was developed in Turnbull (1976), where the same estimator was used, but a different approach in the estimation. He also developed an algorithm, which he argues is simpler than the one in Peto (1973).

We now present some relevant publications about nonparametric approaches for multi-state models with interval-censored data. An example is Frydman (1995). She considers nonparametric estimation of the cumulative transition intensities in an illness-death model for a time-inhomogeneous Markov process for interval-censored data. The exact time of death is assumed to be known and right-censored. In Frydman (1995), the transition intensity from state 1 to state 2 is denoted by $\Lambda_{12}(s)$, from state 1 to state 3 it is denoted by $\Lambda_{13}(s)$ and from state 2 to state 3 is is denoted by $\Lambda_{23}(s)$. If the data are right-censored, it is easy to estimate $\Lambda_{12}(s), \Lambda_{13}(s)$ and $\Lambda_{23}(s)$ using the Nelson-Aalen estimator. Since the transition $1 \to 2$ is interval-censored, she has to develop a nonparametric maximum likelihood procedure for estimating $\Lambda_{12}(s), \Lambda_{13}(s)$ and $\Lambda_{23}(s)$. The method primarily consists of two steps. In the first step, she inspects the likelihood functions and find the sets on which the maximum likelihood estimators of $\Lambda_{12}(s), \Lambda_{13}(s)$ and $\Lambda_{23}(s)$ can increase. This is done indirectly for $\Lambda_{12}(s)$ and $\Lambda_{13}(s)$ by characterising the sets of increase of the corresponding subdistribution functions. In the second step, she presents a version of the EM algorithm.

Another example of nonparametric approaches to multi-state models with interval-censored data is found in Leung and Elashoff (1996). They consider a three-state model, where they allow the distribution for the transition times to depend on covariates and time in the previous state. In order to obtain the maximum likelihood estimators, they use the EM-algorithm introduced in Turnbull (1976). They also consider the smoothed EM-algorithm proposed in Silverman et al. (1990). They apply their methodology on data from an AIDS study and a cancer study, more specifically for patients with malignant melanoma.

2.5 Likelihood Theory

In this section, we present some relevant likelihood theory in survival analysis for different types of data and censoring. We start with a brief introduction of the likelihood when we only have one observation per individual and no censoring. Then we present two different approaches for obtaining the same likelihood when the data are censored independently at random. In the end, we consider interval censoring, where we focus on both the likelihood construction for interval-censored lifetime data, and some central results from the likelihood theory for multi-state models with interval-censored data.

Let t be the data observed in the study and $\boldsymbol{\theta}$ is the parameter vector. The likelihood becomes

$$\mathcal{L}(\boldsymbol{\theta}) = Pr(t; \boldsymbol{\theta}),$$

which is the probability density or mass function (Lawless, 2003, p. 545). If we assume that the probability density function $f(t)$ has a specific parametric form $f(t; \boldsymbol{\theta})$ with lifetimes $t_1, \ldots, t_n$ for n independent individuals, then the likelihood function becomes

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(t_i; \boldsymbol{\theta})$$

(Lawless, 2003, p. 546).

2.5.1 Independent Random Censoring

In this part, we give a brief introduction to the likelihood construction when the data are censored independently at random. Independent random censoring happens if the lifetime and censoring time for an individual are independent continuous random variables (Lawless, 2003, p. 54). When we construct the likelihood using likelihood contributions in Chapter 3, we divide the likelihood into likelihood contributions. An individual is then a specific type, which is based on the screening time points and in which states the individual is observed. We therefore want to show that this way of constructing the likelihood is equal to the more traditional way, when the likelihood is not divided into likelihood contributions. In the more traditional way, censoring is used as an indicator function in the final likelihood.

We follow Lawless (2003, pp. 54-55) in the construction of the first likelihood. Let the lifetime T and censoring time C for an individual be independent continuous random variables. When the data are censored independently at random, we assume an individual has a lifetime T and a censoring time C, where T and C are independent continuous random variables. Let $S(t)$ be the survival function when T is observed, and $G(t)$ when we have censoring. The lifetimes are also mutually independent. In addition $t_i = \min(T_i, C_i)$ and $\delta_i = 1$ if $T_i \leq C_i$ and $\delta_i = 0$ if $T_i > C_i$. The data for n individuals come as pairs (t_i, δ_i), where $i = 1, \ldots, n$. We also assume that $f(t)$ and $g(t)$ are the probability density functions for T_i and C_i respectively and we assume they do not contain any of the same parameters. We get

$$Pr(t_i = [t, t+\epsilon), \delta_i = 0) = Pr(C_i = [t, t+\epsilon), T_i > C_i) = g(t)S(t)\epsilon, \quad (2.5)$$

$$Pr(t_i = [t, t+\epsilon), \delta_i = 1) = Pr(T_i = [t, t+\epsilon), T_i \leq C_i) = f(t)G(t)\epsilon. \quad (2.6)$$

Combining these two expressions, we get

$$Pr(t_i = [t, t+\epsilon), \delta_i) = [f(t)G(t)\epsilon]^{\delta_i}[g(t)S(t)\epsilon]^{1-\delta_i}.$$

The distribution for (t_i, δ_i), $i = 1, \ldots, n$ becomes

$$\prod_{i=1}^{n} [f(t_i)G(t_i)]^{\delta_i}[g(t_i)S(t_i)]^{1-\delta_i}.$$

Then

$$\mathcal{L} = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}, \tag{2.7}$$

since $G(t)$ and $g(t)$ do not contain any of the parameters in $f(t)$. The likelihood is therefore defined up to a multiplicative constant.

The same likelihood can be obtained by using a different approach in the construction of the likelihood. With the same data and censoring pattern as above, we have two types of individuals. In the first type, the lifetime T_i for individual i is observed, which means $T_i \leq C_i$ and $\delta_i = 1$. For the second type of individuals, the lifetime is not observed, which means $T_i > C_i$ and $\delta_i = 0$. For the individuals where $\delta_i = 0$, the likelihood contribution for one individual comes from Equation 2.5. For all the individuals where $\delta_i = 1$, the likelihood contribution for one individual comes from Equation 2.6. We get the likelihood by dividing the product into two products

$$\mathcal{L} = \prod_{i:\delta_i=1} f(t_i) \prod_{i:\delta_i=0} S(t_i),$$

since $G(t)$ and $g(t)$ do not contain any of the parameters in $f(t)$. This likelihood is equal to the likelihood in Equation 2.7, but written in a different way. We show the second approach of constructing the likelihood, since this is the approach we tak in Chapter 3.

2.5.2 Interval Censoring

For the rest of this section, we focus on the likelihood theory for parametric models with interval-censored data. We start with presenting the likelihood for interval-censored lifetime data. Following Lawless (2003, p. 64), we assume a framework where each individual $i = 1, \ldots, n$ is observed a specified number of times $0 = t_{i0} < t_{i1} < \cdots < t_{im_i} < \infty$. If an individual fails at time $t_{i,j-1}$, where $j = 1, \ldots, m_i$, we do not observe $t_{i,j}$. However, if the individual did not fail at time $t_{i,j-1}$, we also observe $t_{i,j}$. Therefore, the data consists of an interval $(U_i, V_i]$ for each individual. We know that the true lifetime for individual i, T_i is interval-censored, and therefore $U_i < T_i \leq V_i$. If failure has not occurred by time t_{im_i}, then $V_i = \infty$ and $U_i = t_{im_i}$ is right-censored. The likelihood for the lifetime data then becomes

$$\mathcal{L} = \prod_{i=1}^{n} [F_i(V_i) - F_i(U_i)] = \prod_{i=1}^{n} [(1 - S_i(V_i)) - (1 - S_i(U_i))] = \prod_{i=1}^{n} [S_i(U_i) - S_i(V_i)],$$

where $F_i(t)$ is the distribution function for T_i. Lastly, Lawless states that the inference for interval-censored data for parametric models with this likelihood falls under standard large-sample likelihood theory, which we will present next.

Following Van Den Hout (2017, pp. 65 - 68), we present some central results from the likelihood theory for multi-state models with interval-censored data. This large-sample likelihood theory is therefore a generalization of the large-sample likelihood theory for lifetime data. Consider a maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n$ which depends on the sample size n, and the vector $\boldsymbol{\theta}_0$ with the true values. We have

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{D} N_p(\mathbf{0}, \boldsymbol{\Sigma}),$$

where the arrow indicates convergence in distribution and $\boldsymbol{\Sigma}$ is a well-defined variance-covariance matrix in this limit. $\boldsymbol{\Sigma}$ is estimated from data, which happens when we calculate the inverse of the observed Fisher information matrix. The delta-method for a function g differentiable at $\boldsymbol{\theta}_0$ becomes

$$\sqrt{n}(g(\hat{\boldsymbol{\theta}}_n) - g(\boldsymbol{\theta}_0)) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{O}^T \boldsymbol{\Sigma} \boldsymbol{O}),$$

where $\boldsymbol{O}$ is the gradient of g at $\boldsymbol{\theta}_0$. The proof of this follows from a Taylor series and is found in Casella and Berger (2002, p. 243).

Applying the delta method, we get that the covariance matrix for a function g depending on the maximum likelihood estimates is

$$\hat{\boldsymbol{V}}_{g(\boldsymbol{\theta})} = \Big(\frac{\partial g}{\partial \boldsymbol{\theta}}\Big)^T \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} \Big(\frac{\partial g}{\partial \boldsymbol{\theta}}\Big),$$

where $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}$ is the estimated covariance matrix for the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$. We use the delta-method in Chapter 6 to calculate the pointwise 95% confidence intervals for the survival functions.

To find the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}$ we first need to introduce the score function, the Fisher information matrix and the estimated Fisher information matrix. The first-order derivative of the log-likelihood is called the score function, $\boldsymbol{U}(\boldsymbol{\theta})$. The k-th entry of the score function $\boldsymbol{U}(\boldsymbol{\theta})$ is

$$\boldsymbol{U}_k(\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{j=1}^{J_i} \frac{\partial \log(\mathcal{L}_{ij})}{\partial \theta_k}.$$

The expected information matrix, known as the Fisher information matrix, is given by

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) = E\Big[\boldsymbol{U}(\boldsymbol{\theta})\boldsymbol{U}(\boldsymbol{\theta})^T\Big].$$

The asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ is $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})^{-1}$. The estimated Fisher information matrix $\hat{\boldsymbol{I}}$ is

$$\hat{\boldsymbol{I}} = -\frac{\partial^2 \log(\mathcal{L}(\hat{\boldsymbol{\theta}}))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T},$$

which is often used as the estimated covariance matrix for the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$. In summary, we have that

$$\hat{\boldsymbol{\theta}} \approx_d N_p\Big(\boldsymbol{\theta}_{\mathbf{0}}, \hat{\boldsymbol{I}}^{-1}\Big),$$

where $\boldsymbol{\theta}_0$ is the true parameter and $\hat{\boldsymbol{I}}$ is the estimated Fisher information matrix.

For this standard likelihood theory to hold, we assume for the Fisher information matrix $\hat{\boldsymbol{I}}$ and the sample size n, that $\hat{\boldsymbol{I}}/n$ converges to a positive definite matrix (P. Hougaard, personal communication, July, 2021).

Lastly, we give a brief introduction to the likelihood theory for a nonparametric approach. For interval-censored survival data, relevant theory is for example presented in Huang and Wellner (1997) and Gentleman and Geyer (1994). In Huang and Wellner (1997) they discuss both the theory for nonparametric and semi-parametric models. For the nonparametric setting, they describe the asymptotic properties of the nonparametric maximum likelihood estimator. They also discuss the theory for the semi-parametric models, where

they focus on proportional hazards, proportional odds and accelerated failure time regression models. For example, the compute the Fisher information and the regression parameter estimators by maximizing the semi-parametric estimators.

Next we present a requirement for the standard likelihood theory to hold for a nonparametric approach. Let us consider a lifetime situation, where $U_i < T_i \leq V_i$, and T_i is the lifetime for individual i. Assume we have a model with piecewise constant transition intensities, where each interval has length 1. In this case, we must ensure that the distribution of U_i and V_i covers all the intervals. For example if all U_i and V_i are values lower than 20.5, then we can only decide the intensities until 21. However, if all U_i are lower than 5 and all V_i are higher than 10, then we can only decide the sum of the intensities of the interval from 5 to 10 (P. Hougaard, personal communication, July, 2021).

2.6 Model Selection and Goodness-of-Fit

In this thesis, we use the Akaike Information Criterion (AIC) for model selection. AIC was introduced in Akaike (1973) and Akaike (1974). Formally, $AIC = 2k - 2\log(\mathcal{L})$, where k is the number of parameters and $\mathcal{L}$ is the likelihood function evaluated at the MLE. The idea of AIC is to correct the maximum likelihood estimate by adding a function of the number of model parameters k (Vrieze, 2012).

AIC was derived as an estimate of expected relative Kullback-Leibler (K-L) divergence. K-L measures the distance between the candidate model and the true model. From Vrieze (2012), the formula for the K-L divergence is

$$KL(g\|f) = \int g(\boldsymbol{y}) \log \frac{g(\boldsymbol{y})}{f(\boldsymbol{y})} d\boldsymbol{y},$$

where $g(\boldsymbol{y})$ is the probability density function (p.d.f.) of the true model, while $f(\boldsymbol{y})$ is the p.d.f. of the candidate model. In order to calculate the exact value of the K-L divergence, the true distribution $g(\boldsymbol{y})$ must be known. Often, the true distribution $g(\boldsymbol{y})$ is unknown. For comparing models, this is inconsequential because $g(\boldsymbol{y})$ is the same for all of the candidate models. The relative differences between the candidate models are the same whether $g(\boldsymbol{y})$ is known or not (Vrieze, 2012). The K-L divergence from the true model to the candidate model is implicitly estimated by AIC. Even though the true model is unknown, we can still use the relative differences between the models to rank the models. A smaller distance means the candidate model is closer to the truth. We therefore have that the preferred model is the one with the lowest AIC, since the this model gives the lowest expected K-L divergence (Vrieze, 2012).

For AIC to be a consistent estimator for the K-L divergence, the true model must be in the candidate set. The reason is that for k to be a correct penalty for the log-likelihood function evaluated at the MLE, the true model must be in the candidate set. If this is not fulfilled, then k is biased (Vrieze, 2012). In this thesis, we assume the true model is in the candidate set.

We also want to evaluate the fit of our models and therefore consider goodness-of-fit. According to Van Den Hout (2017), it is difficult to find a suitable measure of goodness-of-fit for multi-state models with censoring or

when there is variation in observation times between and within individuals. The main problem with interval-censored data is that the process is latent between the observation times. We can therefore not compare estimated time of transition and actual time of transition because the time of transition is not observed exactly.

According to Titman and Sharples (2010) and Gentleman et al. (1994) it is common to use the Kaplan-Meier estimates as an informal way of validating a Markov model for data when the time to the absorbing state is known. However, if the entry to the absorbing state is also interval-censored, one can not use the Kaplan-Meier estimates. In this case, one can use an analogous method which uses a nonparametric survival estimate for interval-censored data. In summary, if all the subjects start in the same state at time zero, progress to the absorbing state and the assumptions in the parametric model are correct, then there should be close agreement between the empirical survival curve and the survival curve from the fitted parametric model (Titman & Sharples, 2010). A common way is to plot a 95% confidence interval of the Kaplan-Meier estimate, which we introduced in Section 2.1. If the estimated survival curve goes outside the confidence limits, then it can be considered as an informal evidence of lack of fit (Titman & Sharples, 2010). We use this informal way of assessing the fit for the total survival probability from the first state to the absorbing state for the analysis of the CAV-data without covariates in Chapter 6.

2.7 First-Hitting Time Models

Modeling lifetime as a first passage time of a threshold for a stochastic process is often convenient in survival analysis. Such models are for example reviewed in Aalen and Gjessing (2001), where they particularly study the Wiener process as the underlying stochastic process. In this section, we present two different first-hitting time regression models, the Wiener process and the Gamma process. A first-hitting time (FHT) model consists of two basic components. The first component is a parent stochastic process $\{Z(t), t \in \mathcal{T}, z \in \mathcal{Z}\}$, where $Z(0) = z_0$. The process $Z(t)$ can either be an observable or an unobservable, latent process. The second component is a boundary set or a threshold, $\mathcal{B} \subset \mathcal{Z}$. It can be fixed, $\mathcal{B}$, or it can depend on time, $\mathcal{B}(t)$. We assume that the process starts a time zero outside the boundary set. The first passage time is defined as the time elapsed from zero until the process enters $\mathcal{B}$.

$$S(t) = \inf\{t : Z(t) \in \mathcal{B}\},$$

where $S(t)$ is the survival function. The event of reaching the boundary is not guaranteed to happen, depending on the type of process and boundary (Caroni, 2017, p. 58)

Let us consider the case where the final event be an observable outcome of an underlying process. This can for example be a disease diagnosis or death. Then, the underlying process can be modeled as a stochastic process $Z(t)$, where t is the time variable and $Z(0) = z_0 > 0$. Let r be the time of failure. This means that the first time $Z(t) \leq 0$ happens, is at time $t = r$. The lifetime is defined as the time it takes for the process to reach the threshold zero for the first time. This is a natural choice for some stochastic processes, but for other stochastic

processes it may be more appropriate to assume a fixed starting point at zero and a variable threshold above zero (Caroni, 2017, p. 61).

2.7.1 Wiener Process

An example of an underlying stochastic process is the Wiener process. Let $Z(t)$ be a random walk in continuous time and space, also called Brownian motion with drift. Following Caroni (2017, p. 61), the Wiener process can be defined as

1. Z(t) has independent increments, which means that $Z(t_2) - Z(t_1)$ and $Z(t_4) - Z(t_3)$ are independent for any pair of non-overlapping intervals (t_1, t_2) and (t_3, t_4).
2. For any interval (t_1, t_2),
$$Z(t_2) - Z(t_1) \sim N(\mu(t_2 - t_1), \sigma^2(t_2 - t_1)).$$

If we then assume $\mu \leq 0$, and from the Wiener process setup we have that the lifetime T follow an inverse Gaussian distribution

$$f(t|z_0, \mu, \sigma^2) = \frac{z_0}{(2\pi\sigma^2 t^3)^{1/2}} \exp\Big[- \frac{(z_0 + \mu t)^2}{2\sigma t}\Big],$$

where one usually assumes $\sigma = 1$ (Caroni, 2017, p. 60-61).

2.7.2 Gamma Process

Following Sildnes and Lindqvist (2018), we define the Gamma process as a continuous time stochastic process $Z(t) = \{Z(t) : t \geq 0\}$ with shape parameter $a(t) > 0$ and scale parameter $\rho > 0$. Let

1. $Z(0) = 0$ with probability 1
2. $\{Z(t) : t \geq 0\}$ has independent increments
3. $Z(t) - Z(s)$ is gamma distributed with shape parameter $a(t) - a(s)$ and scale parameter ρ for every $0 < s < t$.

In this part, we prove that we can set the scale parameter ρ to be 1 without loss of generality. Let $X \sim Gam(a, \rho)$ be a Gamma distributed variable. We prove that ρX is Gamma distributed with shape parameter a and scale parameter 1 by using the moment-generating function (mgf).

$$E(\exp(t\rho x)) = \int_0^\infty \frac{\rho^a}{\Gamma(a)} \exp(-x(\rho - t\rho)) x^{a-1} dx.$$

Using substitution, we have $y = x(\rho - \rho t)$, so $x = \frac{1}{\rho(1-t)} y$, so $dx = \frac{1}{\rho(1-t)} dy$. Then

$$\begin{aligned}
E(\exp(t\rho x)) &= \frac{\rho^a}{\Gamma(a)} \int_0^\infty \exp(-y) \Big(\frac{1}{\rho(1-t)} y \Big)^{a-1} \frac{1}{\rho(1-t)} dy \\
&= \frac{\rho^a}{\Gamma(a)} \Big(\frac{1}{\rho(1-t)} \Big)^a \int_0^\infty y^{a-1} \exp(-y) dy \\
&= \frac{\rho^a}{\Gamma(a)} \Big(\frac{1}{\rho(1-t)} \Big)^a \Gamma(a) \\
&= \Big(\frac{\rho}{\rho - \rho t} \Big)^a = \Big(\frac{1}{1-t} \Big)^a,
\end{aligned}$$

which is the mgf for a Gamma distributed function with shape a and scale 1. ρX is therefore Gamma distributed with shape parameter a and scale parameter 1, which implies

$$\int_0^c \frac{\rho^a}{\Gamma(a)} x^{a-1} \exp(-dx)dx = \int_0^{\rho c} \frac{1^a}{\Gamma(a)} x^{a-1} \exp(-x)dx. \tag{2.8}$$

This proves that the scale parameter ρ only appears together with the threshold c, and we can without loss of generality set $\rho = 1$. Since we set the scale parameter to 1, we get one less parameter to estimate. This means that that computational burden when we estimate all the parameters in a multi-state setting is lower.

We now present the Gamma processes in a general survival analysis setting. For $a > 0$ and $c > 0$, let $Z_0(t) = \{Z_0(t) : t \geq 0\}$ be a Gamma process, where $Z_0(t) \sim Gam(at, \rho)$, with the same properties as stated above. Using the information from Equation 2.8, we set $\rho = 1$. This means $Z_0(t) \sim Gam(at, 1)$. In addition, we can consider any time dependent nondecreasing function $M(t)$. Using a general motor function $M(t)$ has previously been worked on in Claeskens and Hjort (2008, pp. 88-90). The Gamma process with a general motor function becomes

$$Z(t) = Z_0(M(t)) \sim Gam(aM(t), 1).$$

In this thesis, we use $M(t) = t^b$. This is a common motor function and is also considered in Sildnes and Lindqvist (2018). We consider two different versions of the motor function $M(t)$, where either b is a parameter or we assume $b = 1$. First, we let $b = 1$. The survival function is

$$\begin{aligned} S_0(t, a, c) &= Pr(T_0 \geq t) = Pr(Z_0(t) < c) \\ &= G(c, at, 1) = \int_0^c g(x, at, 1)dx \\ &= \int_0^c \frac{1}{\Gamma(at)} x^{at-1} \exp(-x)dx. \end{aligned}$$

The density corresponding to $S_0(t, a, c)$ is

$$\begin{aligned} f_0(t, a, c) = -\frac{\partial S_0(t)}{\partial t} &= -\int_0^c g(x, at, 1)\{-a\psi(at) + a \log x\}dx \\ &= a\psi(at)G(c, at, 1) - a\int_0^t \log x g(x, at, 1)dx \end{aligned},$$

where $\psi(x) = \frac{\partial \log \Gamma(x)}{\partial x}$. We use numerical approximations to compute this. The hazard rate becomes

$$\alpha_0(t, a, c) = -\frac{S_0'(t)}{S_0(t)} = \frac{f_0(t, a, c)}{G(c, at, 1)}.$$

Now, let $M(t) = t^b$, where $b > 0$. We have $Z(t) = Z_0(t^b) \sim Gam(at^b, 1)$. Then the survival function is

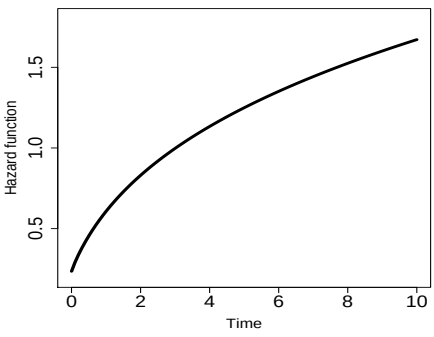
$$S(t, a, c, b) = S_0(t^b, a, c) = \int_0^c \frac{1}{\Gamma(at^b)} x^{at^b-1} \exp(-x)dx,$$

where at^b decides how fast the Gamma process increases with time and c is the threshold (Caroni, 2017, p. 76). This means that both at^b and c together decides how fast the survival function decreases. For example for the same a and c, if $0 < b < 1$, then the survival function is slowly decreasing with time. If $b > 1$ the survival function decreases faster with time, because the individuals fail earlier. The difference for the survival time function between $b < 1$ or $b > 1$ depends on a and c. The density corresponding to $S(t)$ is

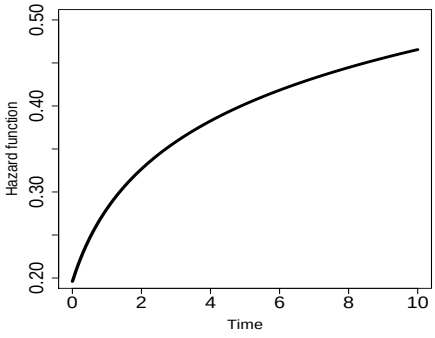
$$f(t, a, c, b) = -\frac{\partial S(t, a, c, b)}{\partial t} = -\frac{\partial G(c, at^b, 1)}{\partial t}.$$

We use numerical approximations to compute this. The hazard rate becomes

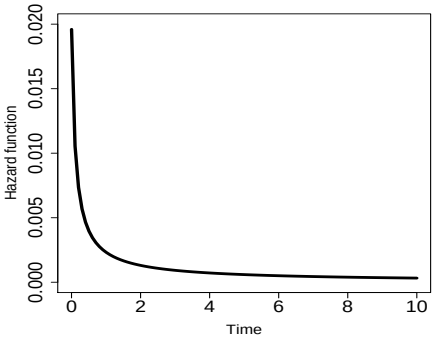
$$\alpha(t, a, c, b) = \frac{f(t, a, c, b)}{G(c, at^b, 1)}.$$



(a) $c = 0.1, a = 0.1, b = 1.2$



(b) $c = 0.1, a = 0.1, b = 1$



(c) $c = 0.1, a = 0.1, b = 0.1$

Figure 2.8: Illustrations of hazard functions for different values of b

The shape of the hazard function varies for different values of a, b and c. Often, the hazard function is increasing if b is close to 1. The shape varies a bit when b is higher than 1 or b is close to 0. In these situations, both a and c contributes to the nature of the hazard function. Typically, if $b > 1$, the hazard

function is increasing. It may decrease as well, but at some point it increases. If b is close to (but larger than) 0, the opposite is true. The hazard function in this situation is often decreasing, but it may increase sometimes as well.

Typical examples of the shape of the hazard functions are illustrated in Figure 2.8. For all of the different hazard functions, $c = 0.1$ and $a = 0.1$, but we have different values of b. The shape of the hazard functions are quite equal in Figure 2.8 (a) and (b). They are both increasing and concave. However, the shape of the hazard function in Figure 2.8 (c) is quite different. It is decreasing and convex. We see in Chapter 6, that it may be an advantage of letting b be a parameter and not fixing it at 1.

Finally, we want to make a remark about the flexibility of the Gamma processes. For any given survival function $S(t)$, and a given threshold c, a motor function $M(t)$ can be found, numerically, to make

$$S(t) = G(c, aM(t), 1)$$

which is our Gamma process threshold crossing model with that motor function. This makes it possible to construct many variants of the Gamma process models.

Traditionally, Gamma processes have been used in engineering, while Wiener processes have been more popular in medical applications (Sildnes & Lindqvist, 2018). This can for example be seen in Qiu and Cui (2019) and van Noortwijk (2009). In Qiu and Cui (2019), they consider safety-critical systems, such as aircrafts, submarines and space stations. Here, missions are performed continuously, and in order for the systems to survive, a mission with problems must be aborted. By using a two-stage gamma process, they find the optimal mission abort policy. A mission can for example be aborted if the degeneration level is above a threshold in the Gamma process. Moreover in van Noortwijk (2009) the application of Gamma processes in maintenance is surveyed. Gamma processes are much used in maintenance because they are well suited for modeling temporal variability of deterioration. It has especially been successful when determining optimal inspection and maintenance decisions. More specifically, the expected deterioration, $E(Z(t)) = aM(t)/b$ often follows a power function in t (van Noortwijk, 2009). An example may be $M(t) = t^b$, where $b > 0$, which we also use later on in this thesis.

2.7.3 Example

Consider survival data $(t_1, \delta_1), \ldots, (t_n, \delta_n)$ where t_i is the possibly censored lifetime and δ_i is an indicator for non-censoring. We define the survival function as $S(t_i) = G(c, at_i, 1)$ and the density is defined in the same way as as previously, $f(t_i, a, c) = -\frac{\partial S(a,c,t_i)}{\partial t_i}$.

The log-likelihood then becomes

$$\begin{aligned}\ell(a, c) &= \sum_{\delta_i=0} \log S(t_i, a, c) + \sum_{\delta_i=1} \log f_0(t_i, a, c) \\ &= \sum_{\delta_i=0} \log G(c, at, 1) + \sum_{\delta_i=1} \log \Big(- \frac{\partial S(t, a, c)}{\partial t} \Big),\end{aligned}$$

and numerically this will typically be computed as

$$\approx \sum_{\delta_i=0} \log G(c, at, 1) + \sum_{\delta_i=1} \log \Big(\frac{G(c, at, 1) - G(c, a(t+\epsilon), 1)}{\epsilon} \Big),$$

which can be computed and maximized.

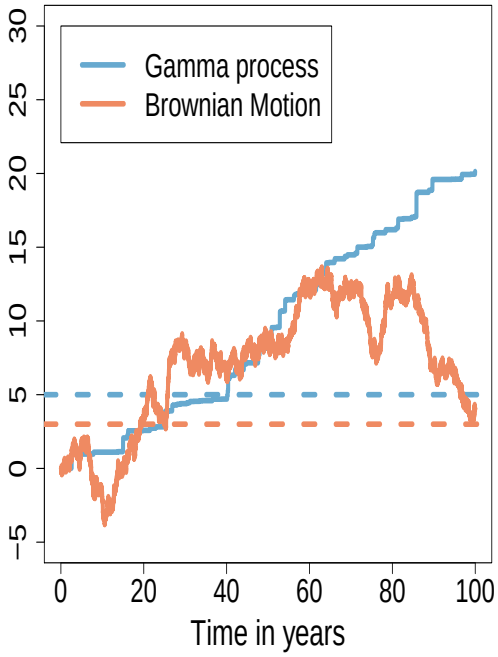


Figure 2.9: Simulated Gamma process where $a = 0.2$ and $c = 5$, and Brownian Motion where $\mu = 0$ and $\sigma = 1$

Figure 2.9 shows a simulated Gamma process, where $a = 0.2$ and $c = 5$, and a simulated Brownian motion, where $\mu = 0$ and $\sigma = 1$. The Brownian motion starts at 0 and we consider the time it takes for the process to reach $c = 3$ for the first time.

For the Gamma process, the individual crosses the threshold $c = 5$ after around 40 years. If c is higher in Figure 2.9, the threshold is crossed later on. For example if $c = 10$, the individual crosses the threshold after around 60 years. The a-parameter is included in deciding the shape of the Gamma process. If a is very low, then the Gamma process is quite flat and it takes much longer time for the individual to reach the threshold. If a is higher, then the Gamma process is very steep and the threshold is reached faster. When it comes to the Brownian motion, it reaches the threshold 3 for the first time after around 20 years. Both the mean and the variance controls the shape of the process and how long it takes for the process to reach the threshold.

A difference between a Gamma process and a Brownian motion is that a Gamma process is always positive and increasing. This is also illustrated in Figure 2.9, where the Brownian motion is for example decreasing after 10 years and also becomes negative. Gamma processes are therefore well-suited in situations where we consider the cumulative risk to only increase. An example is in multi-state models where one can only transition one way.

CHAPTER 3

General Likelihood Construction

3.1 Motivation

In this chapter, we construct a general likelihood for four different multi-state models with interval-censored data. These four multi-state models are the three-state progressive model, the illness-death model, the four-state progressive model and a four-state illness-death model, but the construction can be extended to other multi-state models too. In order to capture a bigger variety of data, we construct a likelihood for the case where the entrance into the absorbing state is observed exactly, and for the case where the entrance into the observing state is not observed exactly. Since we include the possibility of not exactly observing the transition to the last state, we can also study data where the transition to the absorbing state is interval-censored. This may for instance happen if the absorbing state is not death, but something else, for example a disease.

To the best of our knowledge, we have not found anyone constructing a likelihood in this way for these multi-state models for interval-censored data. As we mentioned in Chapter 2, the likelihood construction is to some extent related the ideas for the general progressive model in Hougaard (2000, p. 159). The main differences are the our framework is tailored to interval-censored data and we construct the likelihood by dividing it into likelihood contributions. The formulation of the likelihood in Jackson (2011), which is based on the formulations in Kalbfleisch and Lawless (1985) and Kay (1986), are to some extent similar to this likelihood. For example, we find in Chapter 5 that if the transition times are modeled by exponential distributions in the four-state illness-death model, the likelihood construction we present in this chapter is equal to the one in Jackson (2011). However, there are at least two differences between the likelihood construction we present and the likelihood construction in Jackson (2011). The first difference is that they use the transition probability matrix when they construct the likelihood and the second difference is that their likelihood relies on the Markov property.

We start the construction of the likelihood by determining the different types of individuals one may observe. From these types of individuals, we create the likelihood contributions. Which type an individual is, depends on which states are visited and the observation times. If we are in a simple survival case with independent random censoring, as discussed in Chapter 2, one way of writing the likelihood is

$$\mathcal{L} = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \tag{3.1}$$

(Lawless, 2003, pp. 54-55). If we want to write this likelihood using likelihood contributions, we divide the individuals into two types. Type 1 means death at time t_i, while type 2 means the individual is alive at time t_i. Equation 3.1 is then equal to

$$\prod_{type1, i:\delta_i=1} f(t_i) \prod_{type2, i:\delta_i=0} S(t_i). \tag{3.2}$$

These two likelihoods are equal, and one can choose the preferred approach of constructing the likelihood. We choose the second way of constructing the likelihood.

In order to construct a likelihood with different types in a multi-state model, we consider all the different states and the possible transitions. If a person is observed in state 0 at time t_1 and in state 1 at the rest of the screening times, we have one likelihood type. However, if a person is observed in state 0 at all the screening times, this is another likelihood type. The contributions to the likelihood are different, and the individuals are therefore different types.

An advantage of dividing the likelihood into likelihood contributions is the amount of information we get from constructing the likelihood. For example, we get an overview of all the transitions each individual makes over time and the share of individuals in each type. This information can be used to find the most suitable survival time model for the transition times. A disadvantage of dividing the likelihood into likelihood contributions is that each contribution becomes more complex when the complexity of the multi-state models increases. With additional states and possible transitions, we get additional types and each likelihood contribution becomes more complex. In addition, there is always a possibility of forgetting a type. This can to some extent be solved by checking that all the individuals are included as one type in the likelihood in a real or simulated dataset, we minimize the risk of forgetting a type.

Since a more complex multi-state model with many observation times means more different types, we start by considering the three-state progressive model with only one screening. In a three-state progressive model there are only three states and two transition times. For comparison, our final multi-state model is the four-state illness-death model, where there are four states and five transition times. The reason for ending up with this four-state illness-death model, is to be able to use our likelihood construction on the CAV-dataset discussed in Jackson (2011). Our analysis of this dataset is found in Chapter 6.

3.2 Assumptions

We start by assuming a multi-state setting with interval-censored data. The screening times t_i, are predetermined, and therefore not stochastic. The transition times to the different states are assumed to be independent and have parametric densiites. The entrance to the last state is either observed exactly or not exactly.

Further, we assume a meaningful starting point. At this starting point, all of the individuals are not necessarily screened, but are considered to be in state 0. For example, if one study the development of dementia, a meaningful starting point is 40 years. However, the first screening may be conducted five years later, when the individual is 45 years old. Then if an individual at 45 is diagnosed with dementia, we assume it happened between 40 and 45 years. If everybody is screened at the starting point, then everybody starts in state 0. In the analysis in Chapter 6, everybody starts in state 0.

For the rest of this chapter, we define $\boldsymbol{\theta}$ as a vector of all the parameters. To make notation easier, we only write $\boldsymbol{\theta}$ in the final likelihood. In addition, when we write (I) under the product sign in the likelihood, we mean all the individuals which are type 1. Further (II) means all the individuals which are type 2 and so on.

3.3 Three-State Progressive Model

In this section, we analyze the three-state progressive model. In a three-state progressive model, the individuals from a population can transition from state 0 to state 1 to state 2. The individuals cannot transition directly from state 0 to state 2 without going through state 1. We illustrate the three-state progressive model in Figure 3.1.

Figure 3.1: Three-state progressive model

We define T_0 and T_1 respectively as the transition times from state 0 to state 1 and from state 1 to state 2. These transition times are assumed to be independent, with parametric densities f_0 and f_1, survival time functions $S_0 = 1 - F_0$ and $S_1 = 1 - F_1$ and hazard rate functions $\alpha_0 = f_0/S_0$ and $\alpha_1 = f_1/S_1$. The total time from state 0 to state 2 is $T_0 + T_1$. However, we also want knowledge about T_0 and T_1 separately, which is demanding since there is interval censoring.

3.3.1 One Screening

Assume we only have one screening. In addition, we observe the exact time of death for those individuals that die during the study period. With this observation scheme, we have five types of patients.

1. Suppose an individual is screened at time t. This screening shows that the individual is in state 0. The likelihood contribution is

$$Pr(T_0 > t) = S_0(t).$$

2. Suppose an individual is screened at time t. This screening shows that the individual is in state 1. The likelihood contribution is

$$\begin{aligned} Pr(T_0 < t, T_0 + T_1 > t) &= \int_0^t f_0(s) Pr(T_1 > t - T_0 | T_0 = s) ds \\ &= \int_0^t f_0(s) S_1(t-s) ds. \end{aligned}$$

3. Suppose an individual dies at time t, with no intermittent screening. The likelihood contribution is

$$\begin{aligned} Pr(T_0 + T_1 \in [t, t+\epsilon)) &= Pr(0 < T_0 < t, T_0 + T_1 \in [t, t+\epsilon)) \\ &= \int_0^t f_0(s) f_1(t-s) ds\epsilon. \end{aligned}$$

4. Suppose an individual dies at time t, and the individual was screened once at time point u. At time point u, the individual was in state 0. The likelihood contribution is

$$\begin{aligned} &Pr(T_0 > u, T_0 + T_1 \in [t, t+\epsilon)) \\ &= Pr(u < T_0 < t, T_0 + T_1 \in [t, t+\epsilon)) \\ &= \int_u^t f_0(s) Pr(T_0 + T_1 \in [t, t+\epsilon) | T_0 = s) ds \\ &= \int_u^t f_0(s) f_1(t-s) ds\epsilon. \end{aligned}$$

5. Suppose an individual dies at time t, and the individual was screened once at time point u. At time point u, the individual was in state 1. The likelihood contribution is

$$\begin{aligned} &Pr(T_0 < u, T_0 + T_1 \in [t, t+\epsilon)) \\ &= Pr(0 < T_0 < u, T_0 + T_1 \in [t, t+\epsilon)) \\ &= \int_0^u f_0(s) Pr(T_0 + T_1 \in [t, t+\epsilon) | T_0 = s) ds \\ &= \int_0^u f_0(s) f_1(t-s) ds\epsilon. \end{aligned}$$

The full likelihood for all the individuals $p = 1, \ldots, m$ becomes

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) = &\prod_{(I)} S_0(t_p, \boldsymbol{\theta} | x_p) \prod_{(II)} \int_0^{t_p} f_0(s, \boldsymbol{\theta} | x_p) S_1(t_p - s, \boldsymbol{\theta} | x_p) ds \\ &\prod_{(III)} \int_0^{t_p} f_0(s, \boldsymbol{\theta} | x_p) f_1(t_p - s, \boldsymbol{\theta} | x_p) ds \\ &\prod_{(IV)} \int_{u_p}^{t_p} f_0(s, \boldsymbol{\theta} | x_p) f_1(t_p - s, \boldsymbol{\theta} | x_p) ds \\ &\prod_{(V)} \int_0^{u_p} f_0(s, \boldsymbol{\theta} | x_p) f_1(t_p - s, \boldsymbol{\theta} | x_p) ds. \end{aligned}$$

The log-likelihood all the individuals $p = 1, \ldots, m$ becomes

$$\begin{aligned}\ell(\boldsymbol{\theta}) = &\sum_{(I)} \log(S_0(t_p, \boldsymbol{\theta}|x_p)) \\ &+ \sum_{(II)} \log(\int_0^{t_p} f_0(s, \boldsymbol{\theta}|x_p) S_1(t_p - s, \boldsymbol{\theta}|x_p) ds) \\ &+ \sum_{(III)} \log(\int_0^{t_p} f_0(s, \boldsymbol{\theta}|x_p) f_1(t_p - s, \boldsymbol{\theta}|x_p) ds) \\ &+ \sum_{(IV)} \log(\int_{u_p}^{t_p} f_0(s, \boldsymbol{\theta}|x_p) f_1(t_p - s, \boldsymbol{\theta}|x_p) ds) \\ &+ \sum_{(V)} \log(\int_0^{u_p} f_0(s, \boldsymbol{\theta}|x_p) f_1(t_p - s, \boldsymbol{\theta}|x_p) ds).\end{aligned}$$

3.3.2 Multiple Screenings

We are still in a three-state progressive model, but the individuals are screened multiple times. Therefore, we have more than five types of patients and the likelihood is updated.

3.3.2.1 Exact Time of Entry into the Absorbing State is Known

The individuals are screened $t_1, t_2, \ldots, t_n$ times.

1. Suppose an individual is only observed in state 0 at all the screening time points, where t_n is the last screening. Then

$$Pr(T_0 > t_n) = S_0(t_n).$$

2. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is still in state 1 at the last screening point, t_n. Then

$$\begin{aligned}&Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 < t_n, T_0 + T_1 > t_{i+1}, T_0 + T_1 > t_n) \\ &= Pr(t_i < T_0 < t_{i+1}, T_0 + T_1 > t_n) \\ &= \int_{t_i}^{t_{i+1}} f_0(s) Pr(T_0 + T_1 > t_n | T_0 = s) \\ &= \int_{t_i}^{t_{i+1}} f_0(s) Pr(T_1 > t_n - s) ds \\ &= \int_{t_i}^{t_{i+1}} f_0(s) S_1(t_n - s) ds.\end{aligned}$$

3. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1} the individual is observed in state 1. We observe that the individual dies at

exact time point t_{i+k}, where $k > 1$. Then

$$\begin{aligned}
&Pr(T_0 > t_i, T_0 > t_{i+1}, T_0 + T_1 \in [t_{i+k}, t_{i+k} + \epsilon)) \\
&= Pr(t_i < T_0 < t_{i+1}, t_{i+k} < T_0 + T_1 < t_{i+k} + \epsilon) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k} < T_0 + T_1 < t_{i+k} + \epsilon | T_0 = s) ds \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k} < s + T_1 < t_{i+k} + \epsilon) ds \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k} + \epsilon - s < T_1 < t_{i+k} + \epsilon - s) ds \\
&= \int_{t_i}^{t_{i+1}} f_0(s) \frac{F_1(t_{i+k} + \epsilon - s) - F_1(t_{i+k} - s)}{\epsilon} ds\epsilon \\
&= \int_{t_i}^{t_{i+1}} f_0(s) f_1(t_{i+k} - s) ds\epsilon.
\end{aligned}$$

4. Suppose an individual is only observed in state 1 at all the screening time points, where t_n is the last screening. Then

$$\begin{aligned}
&Pr(T_0 < t_1, T_0 < t_n, T_0 + T_1 > t_1, T_0 + T_1 > t_n) \\
&= Pr(T_0 < t_1, T_0 + T_1 > t_n) \\
&= \int_0^{t_1} f_0(s) S_1(t_n - s) ds
\end{aligned}$$

.

5. Suppose an individual is observed in state 1 from t_1 to t_i. At the exact time point t_{i+1}, the individual is observed in state 2. Then

$$\begin{aligned}
&Pr(T_0 < t_1, T_0 + T_1 \in [t_{i+1}, t_{i+1} + \epsilon)) \\
&= Pr(T_0 < t_1, t_{i+1} < T_0 + T_1 < t_{i+1} + \epsilon) \\
&= \int_0^{t_1} f_0(s) Pr(t_{i+1} < T_0 + T_1 < t_{i+1} + \epsilon | T_0 = s) ds \\
&= \int_0^{t_1} f_0(s) Pr(t_{i+1} - s < T_1 < t_{i+1} + \epsilon - s) ds \\
&= \int_0^{t_1} f_0(s) f_1(t_{i+1} - s) ds\epsilon.
\end{aligned}$$

6. Suppose an individual is observed in state 0 from t_1 to t_i. At the exact time point t_{i+1}, the individual is observed in state 2. Then

$$\begin{aligned}
&Pr(T_0 > t_i, T_0 + T_1 \in [t_{i+1}, t_{i+1} + \epsilon)) \\
&= Pr(t_i < T_0 < t_{i+1}, t_{i+1} < T_0 + T_1 < t_{i+1} + \epsilon) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) f_1(t_{i+1} - s) ds\epsilon.
\end{aligned}$$

7. Suppose an individual is observed in state 2 at the exact time point t_1, without any intermittent screening. Then

$$\begin{aligned} & Pr(t_1 < T_0 + T_1 < t_1 + \epsilon) \\ &= \int_0^{t_1} f_0(s) Pr(t_1 - s < T_1 < t_1 + \epsilon - s) ds \\ &= \int_0^{t_1} f_0(s) f_1(t_1 - s) ds\epsilon. \end{aligned}$$

Consequently, the full likelihood for dataset where the individual has been screened as we described above for the individuals $p = 1, \ldots, m$, becomes

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) = & \prod_{(I)} S_0(t_{n,p}, \boldsymbol{\theta}|x_p) \prod_{(II)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_1(t_{n,p} - s, \boldsymbol{\theta}|x_p) ds \\ & \prod_{(III)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p) f_1(t_{i+k,p} - s, \boldsymbol{\theta}|x_p) ds \\ & \prod_{(IV)} \int_0^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_1(t_{n,p} - s, \boldsymbol{\theta}|x_p) ds \\ & \prod_{(V)} \int_0^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p) f_1(t_{i+1,p} - s, \boldsymbol{\theta}|x_p) ds \\ & \prod_{(VI)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p) f_1(t_{i+1,p} - s, \boldsymbol{\theta}|x_p) ds \\ & \prod_{(VII)} \int_0^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p) f_1(t_{1,p} - s, \boldsymbol{\theta}|x_p) ds. \end{aligned}$$

3.3.2.2 Exact Time of Entry into the Absorbing State is not Known

Suppose we screened an individual $t_1, t_2, \ldots, t_n$ times. The likelihood contributions for patients of type 1, 2 and 4 are unchanged. If the exact time of entry into the absorbing state is not known and the patient is type 3, then the likelihood contribution is

$$\begin{aligned} & Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 + T_1 > t_{i+k-1}, T_0 + T_1 < t_{i+k}) \\ &= Pr(t_i < T_0 < t_{i+1}, t_{i+k-1} < T_0 + T_1 < t_{i+k}) \\ &= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k-1} < T_0 + T_1 < t_{i+k} | T_0 = s) ds \\ &= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k-1} < s + T_1 < t_{i+k}) ds \\ &= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k-1} - s < T_1 < t_{i+k} - s) ds \\ &= \int_{t_i}^{t_{i+1}} f_0(s) (F_1(t_{i+k} - s) - F_1(t_{i+k-1} - s)) ds \\ &= \int_{t_i}^{t_{i+1}} f_0(s) (S_1(t_{i+k-1} - s) - S_1(t_{i+k} - s)) ds. \end{aligned}$$

The difference is that $f_1(t_{i+k} - s)$ is replaced by $(S_1(t_{i+k-1} - s) - S_1(t_{i+k} - s))$. If the patient is type 6, and the exact exact time of entry into the absorbing state is not known, then the likelihood contribution is

$$Pr(T_0 > t_i, T_0 + T_1 < t_{i+1}) = Pr(t_1 < T_0 < t_{i+1}, T_0 + T_1 < t_{i+1})$$
$$= \int_{t_i}^{t_{i+1}} f_0(s)(1 - S_1(t_{i+1} - s))ds.$$

Similar changes happens to all of the equations where the patient is observed in the absorbing state. The full likelihood for the individuals $p = 1, \ldots, m$ is

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{(I)} S_0(t_{n,p}, \boldsymbol{\theta}|x_p) \prod_{(II)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_1(t_{n,p} - s, \boldsymbol{\theta}|x_p) ds$$
$$\prod_{(III)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p)(S_1(t_{i+k-1,p} - s, \boldsymbol{\theta}|x_p) - S_1(t_{i+k,p} - s, \boldsymbol{\theta}|x_p)) ds$$
$$\prod_{(IV)} \int_{0}^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_1(t_{n,p} - s, \boldsymbol{\theta}|x_p) ds$$
$$\prod_{(V)} \int_{0}^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p)(S_1(t_{i,p} - s, \boldsymbol{\theta}|x_p) - S_1(t_{i+1,p} - s, \boldsymbol{\theta}|x_p)) ds$$
$$\prod_{(VI)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p)(1 - S_1(t_{i+1,p} - s, \boldsymbol{\theta}|x_p)) ds$$
$$\prod_{(VII)} \int_{0}^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p)(1 - S_1(t_{1,p} - s, \boldsymbol{\theta}|x_p)) ds.$$

3.4 Illness-Death Model

In the three-state progressive model, the individuals cannot transition directly from the first state, state 0, to the absorbing state, state 2. However, in an illness-death model, the individuals can transition directly from state 0 to state 2. Let two independent parametric survival times models competing with each other start at the predetermined starting point. Which threshold is crossed first, decides the transition for the patient.

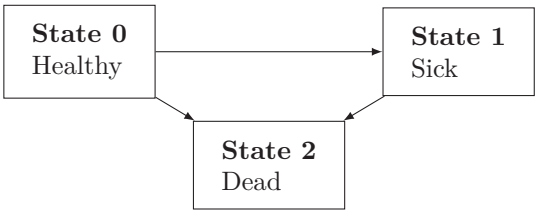


Figure 3.2: Illness-death model

T_0 and T_1 are defined in the same way as in the three-state progressive model. With a meaningful start point, we define an additional transition time, T_{02},

which is the transition time from state 0 to state 2. T_0, T_1 and T_{02} are independent by assumption. The density for T_{02} is f_{02}, with survival time function $S_{02} = 1 - F_{02}$ and hazard function $\alpha_{02} = f_{02}/S_{02}$. In addition, if $T_0 > T_{02}$, then the individual goes from state 0 to state 2. If $T_0 < T_{02}$, then the individual goes from state 0 to state 1.

3.4.1 Exact Time of Entry into the Absorbing State is Known

Suppose the individuals are screened $t_1, t_2, \ldots, t_n$ times.

1. Suppose an individual is only observed in state 0 at all the screening time points, where t_n is the last screening. Then

$$Pr(T_0 > t_n, T_{02} > t_n) = S_0(t_n)S_{02}(t_n).$$

2. Suppose an individual is seen to be in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is still in state 1 at the last screening point t_n. We also have that $T_{02} > T_0$. Then

$$\begin{aligned}
&Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 < t_n, T_0 + T_1 > t_{i+1}, T_0 + T_1 > t_n, T_{02} > T_0) \\
&= Pr(t_i < T_0 < t_{i+1}, T_0 + T_1 > t_n, T_{02} > T_0) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(T_0 + T_1 > t_n | T_0 = s) Pr(T_{02} > T_0 | T_0 = s) ds \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(T_1 > t_n - s) S_{02}(s) ds \\
&= \int_{t_i}^{t_{i+1}} f_0(s) S_1(t_n - s) S_{02}(s) ds.
\end{aligned}$$

3. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. At the exact time point t_{i+k}, where $k > 1$, the individual is observed in state 2. We also have that $T_{02} > T_0$. Then

$$\begin{aligned}
&Pr(T_0 > t_i, T_0 > t_{i+1}, T_0 + T_1 \in [t_{i+k}, t_{i+k} + \epsilon), T_{02} > T_0) \\
&= Pr(t_i < T_0 < t_{i+1}, t_{i+k} < T_0 + T_1 < t_{i+k} + \epsilon, T_{02} > T_0) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k} < T_0 + T_1 < t_{i+k} + \epsilon | T_0 = s) Pr(T_{02} > T_0 | T_0 = s) ds \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k} < s + T_1 < t_{i+k} + \epsilon) S_{02}(s) ds \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k} + \epsilon - s < T_1 < t_{i+k} + \epsilon - s) S_{02}(s) ds \\
&= \int_{t_i}^{t_{i+1}} f_0(s) \frac{F_1(t_{i+k} + \epsilon - s) - F_1(t_{i+k} - s)}{\epsilon} S_{02}(s) ds\epsilon \\
&= \int_{t_i}^{t_{i+1}} f_0(s) f_1(t_{i+k} - s) S_{02}(s) ds\epsilon.
\end{aligned}$$

4. Suppose an individual is observed in state 1 at all the screening time points, where t_n is the last screening. We also have that $T_{02} > T_0$. Then

$$\begin{aligned}
&Pr(T_0 < t_1, T_0 < t_n, T_0 + T_1 > t_1, T_0 + T_1 > t_n, T_{02} > T_0) \\
&= Pr(T_0 < t_1, T_0 + T_1 > t_n, T_{02} > T_0) \\
&= \int_0^{t_1} f_0(s) S_1(t_n - s) S_{02}(s) ds.
\end{aligned}$$

5. Suppose an individual is observed in state 1 from t_1 to t_i. At the exact time point t_{i+1}, the individual is observed in state 2. We also have that $T_{02} > T_0$. Then

$$\begin{aligned}
&Pr(T_0 < t_1, T_0 + T_1 \in [t_{i+1}, t_{i+1} + \epsilon), T_{02} > T_0) \\
&= Pr(T_0 < t_1, t_{i+1} < T_0 + T_1 < t_{i+1} + \epsilon, T_{02} > T_0) \\
&= \int_0^{t_1} f_0(s) Pr(t_{i+1} < T_0 + T_1 < t_{i+1} + \epsilon | T_0 = s) Pr(T_{02} > T_0 | T_0 = s) ds \\
&= \int_0^{t_1} f_0(s) Pr(t_{i+1} - s < T_1 < t_{i+1} + \epsilon - s) S_{02}(s) ds \\
&= \int_0^{t_1} f_0(s) f_1(t_{i+1} - s) S_{02}(s) ds \epsilon.
\end{aligned}$$

6. Suppose an individual is observed in state 0 from t_1 to t_i. At the exact time point t_{i+1}, the individual is observed in state 2. We also have that $T_{02} > T_0$. Then

$$\begin{aligned}
&Pr(T_0 > t_i, T_0 + T_1 \in [t_{i+1}, t_{i+1} + \epsilon), T_{02} > T_0) \\
&= Pr(t_i < T_0 < t_{i+1}, t_{i+1} < T_0 + T_1 < t_{i+1} + \epsilon, T_{02} > T_0) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) f_1(t_{i+1} - s) S_{02}(s) ds \epsilon.
\end{aligned}$$

7. Suppose an individual is observed in state 2 at the exact time point t_1 without any intermittent screening. We also have that $T_{02} > T_0$. Then

$$\begin{aligned}
&Pr(t_1 < T_0 + T_1 < t_1 + \epsilon, T_{02} > T_0) \\
&= \int_0^{t_1} f_0(s) Pr(t_1 - s < T_1 < t_1 + \epsilon - s) Pr(T_{02} > T_0 | T_0 = s) ds \\
&= \int_0^{t_1} f_0(s) f_1(t_1 - s) S_{02}(s) ds \epsilon.
\end{aligned}$$

8. Suppose an individual is observed in state 2 at the exact time point t_1 without any intermittent screening. We also have that $T_{02} < T_0$. Then

$$\begin{aligned}
Pr(T_{02} \in [t_1, t_1 + \epsilon), T_0 > T_{02}) &= Pr(t_1 < T_{02} < t_1 + \epsilon, T_0 > t_1) \\
&= f_{02}(t_1) S_{02}(t_1) \epsilon.
\end{aligned}$$

9. Suppose an individual is observed in state 0 from t_1 to t_i. At the exact time point t_{i+1}, the individual is observed in state 2. We also have that

$T_{02} < T_0$. Then

$$\begin{aligned}
&Pr(T_{02} > t_i, T_{02} \in [t_{i+1}, t_{i+1} + \epsilon), T_0 > T_{02}) \\
&= Pr(t_{i+1} < T_{02} < t_{i+1} + \epsilon, T_0 > t_{i+1}) \\
&= f_{02}(t_{i+1}) S_0(t_{i+1}) \epsilon.
\end{aligned}$$

We then get the likelihood for the individuals $p = 1, \ldots, m$

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) = &\prod_{(I)} S_0(t_{n,p}, \boldsymbol{\theta}|x_p) S_{02}(t_{n,p}, \boldsymbol{\theta}|x_p) \prod_{(II)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_1(t_{n,p} - s, \boldsymbol{\theta}|x_p) S_{02}(s, \boldsymbol{\theta}|x_p) ds \\
&\prod_{(III)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p) f_1(t_{i+k,p} - s, \boldsymbol{\theta}|x_p) S_{02}(s, \boldsymbol{\theta}|x_p) ds \\
&\prod_{(IV)} \int_{0}^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_1(t_{n,p} - s, \boldsymbol{\theta}|x_p) S_{02}(s, \boldsymbol{\theta}|x_p) ds \\
&\prod_{(V)} \int_{0}^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p) f_1(t_{i+1,p} - s, \boldsymbol{\theta}|x_p) S_{02}(s, \boldsymbol{\theta}|x_p) ds \\
&\prod_{(VI)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p) f_1(t_{i+1,p} - s, \boldsymbol{\theta}|x_p) S_{02}(s|x_p) ds \\
&\prod_{(VII)} \int_{0}^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p) f_1(t_{1,p} - s, \boldsymbol{\theta}|x_p) S_{02}(s, \boldsymbol{\theta}|x_p) ds \\
&\prod_{(VIII)} f_{02}(t_{1,p}, \boldsymbol{\theta}|x_p) S_0(t_{1,p}, \boldsymbol{\theta}|x_p) \\
&\prod_{(IX)} f_{02}(t_{i+1,p}, \boldsymbol{\theta}|x_p) S_0(t_{i+1,p}, \boldsymbol{\theta}|x_p).
\end{aligned}$$

3.4.2 Exact Time of Entry into the Absorbing State is not Known

The individuals are still screened $t_1, t_2, \ldots, t_n$ times. The likelihood contributions for type 1, 2 and 4 are unchanged. The differences between when the exact time of entry into the absorbing state is known or not known are similar to what we explained in Section 3.3.2.2. The full likelihood for the individuals $p = 1, \ldots, m$ is

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) = &\prod_{(I)} S_0(t_{n,p}, \boldsymbol{\theta}|x_p) S_{02}(t_{n,p}, \boldsymbol{\theta}|x_p) \prod_{(II)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_1(t_{n,p} - s, \boldsymbol{\theta}|x_p) S_{02}(s, \boldsymbol{\theta}|x_p) ds \\
&\prod_{(III)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p) (S_1(t_{i+k-1,p} - s, \boldsymbol{\theta}|x_p) - S_1(t_{i+k,p} - s, \boldsymbol{\theta}|x_p)) S_{02}(s, \boldsymbol{\theta}|x_p) ds \\
&\prod_{(IV)} \int_{0}^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_1(t_{n,p} - s, \boldsymbol{\theta}|x_p) S_{02}(s, \boldsymbol{\theta}|x_p) ds \\
&\prod_{(V)} \int_{0}^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p) (S_1(t_{i,p} - s, \boldsymbol{\theta}|x_p) - S_1(t_{i+1,p} - s, \boldsymbol{\theta}|x_p)) S_{02}(s, \boldsymbol{\theta}|x_p) ds \\
&\prod_{(VI)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p) (1 - S_1(t_{i+1,p} - s, \boldsymbol{\theta}|x_p)) S_{02}(s, \boldsymbol{\theta}|x_p) ds
\end{aligned}$$

$$\prod_{(VII)} \int_0^{t_{1,p}} f_0(s,\boldsymbol{\theta}|x_p)(1-S_1(t_{1,p}-s,\boldsymbol{\theta}|x_p))S_{02}(s,\boldsymbol{\theta}|x_p)ds$$

$$\prod_{(VIII)} \int_0^{t_{1,p}} f_{02}(s,\boldsymbol{\theta}|x_p)S_0(s,\boldsymbol{\theta}|x_p)ds$$

$$\prod_{(IX)} \int_{t_{i,p}}^{t_{i+1,p}} f_{02}(s,\boldsymbol{\theta}|x_p)S_0(s,\boldsymbol{\theta}|x_p)ds.$$

3.5 Four-State Progressive Model

Suppose we have a four-state progressive model, where the individuals can move from state 0 to state 1, from state 1 to state 2 and from state 2 to state 3. The individuals can not move directly from state 0 to state 3 or from state 1 to state 3. The four-state progressive model with the possible transitions is illustrated in Figure 3.3.

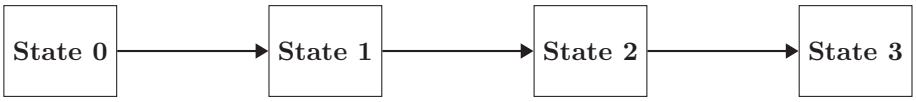


Figure 3.3: Four-state progressive model

T_0 and T_1 are defined in the same way as in the three-state progressive model. In addition, we define T_2 to be the transition time from state 2 to state 3. The transition times are still assumed to be independent. The transition time T_2 has density f_2, survival function $S_2 = 1 - F_2$ and hazard rate $\alpha_2 = f_2/S_2$.

3.5.1 Exact Time of Entry into the Absorbing State is Known

The individuals are screened $t_1, t_2, \ldots, t_n$ times.

1. Suppose an individual is only observed in state 0 at all the screening time points, where t_n is the last screening. Then

$$Pr(T_0 > t_n) = S_0(t_n).$$

2. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is still in state 1 at the last screening point t_n. Then

$$\begin{aligned}
&Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 < t_n, T_0 + T_1 > t_{i+1}, T_0 + T_1 > t_n) \\
&= Pr(t_i < T_0 < t_{i+1}, T_0 + T_1 > t_n) = \int_{t_i}^{t_{i+1}} f_0(s) Pr(T_0 + T_1 > t_n | T_0 = s) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(T_1 > t_n - s) ds \\
&= \int_{t_i}^{t_{i+1}} f_0(s) S_1(t_n - s) ds.
\end{aligned}$$

3. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is observed in state 1 until t_{i+k-1}, where $k > 1$. At t_{i+k}, the individual is observed in state 2. The individual is still in state 2 at the last screening point t_n. Then

$$
\begin{aligned}
&Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 + T_1 > t_{i+1}, T_0 + T_1 < t_{i+k},\\
&T_0 + T_1 + T_2 > t_{i+k}, T_0 + T_1 + T_2 > t_n)\\
&= Pr(t_i < T_0 < t_{i+1}, t_{i+k-1} < T_0 + T_1 < t_{i+k}, T_0 + T_1 + T_2 > t_n)\\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k-1} < T_0 + T_1 < t_{i+k} | T_0 = s) Pr(T_0 + T_1 + T_2 > t_n | T_0 = s) ds\\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) Pr(T_0 + T_1 + T_2 > t_n | T_0 = s, T_1 = u) du ds\\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) Pr(T_2 > t_n - u - s) du ds\\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) S_2(t_n - u - s) du ds.
\end{aligned}
$$

4. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is observed in state 1 until t_{i+k-1}, where $k > 1$. The individual is observed in state 2 at t_{i+k}. The individual is observed in state 2 until $t_{i+k+l-1}$, where $l > 1$. The individual is observed in state 3 at the exact time point t_{i+k+l}. Then

$$
\begin{aligned}
&Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 + T_1 > t_{i+1}, T_0 + T_1 < t_{i+k},\\
&T_0 + T_1 + T_2 \in [t_{i+k+l}, t_{i+k+l} + \epsilon))\\
&= Pr(t_i < T_0 < t_{i+1}, t_{i+k-1} < T_0 + T_1 < t_{i+k}, t_{i+k+l} < T_0 + T_1 + T_2 < t_{i+k+l} + \epsilon)\\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k-1} < T_0 + T_1 < t_{i+k} | T_0 = s)\\
&Pr(t_{i+k+l} < T_0 + T_1 + T_2 < t_{i+k+l} + \epsilon | T_0 = s) ds\\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) Pr(t_{i+k+l} < T_0 + T_1 + T_2 < t_{i+k+l} + \epsilon | T_0 = s, T_1 = u) du ds\\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) Pr(t_{i+k+l} - s - u < T_2 < t_{i+k+l} + \epsilon - s - u) du ds\\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) \frac{(F_2(t_{i+k+l} + \epsilon - s - u) - F_2(t_{i+k+l} - s - u))}{\epsilon} du ds \epsilon\\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) f_2(t_{i+k+l} - s - u) du ds \epsilon.
\end{aligned}
$$

5. Suppose an individual is observed in state 0 from t_1 to t_i. At the exact time point t_{i+1}, the individual is observed in state 3. Then

$$
\begin{aligned}
&Pr(T_0 > t_i, T_0 + T_1 + T_2 \in [t_{i+1}, t_{i+1} + \epsilon)\\
&= Pr(t_i < T_0 < t_{i+1}, T_0 + T_1 + T_2 \in [t_{i+1}, t_{i+1} + \epsilon))\\
&= Pr(t_i < T_0 < t_{i+1}, t_i < T_0 + T_1 < t_{i+1}, t_{i+1} < T_0 + T_1 + T_2 < t_{i+1} + \epsilon)\\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_i < T_0 + T_1 < t_{i+1} | T_0 = s) Pr(t_{i+1} < T_0 + T_1 + T_2 < t_{i+1} + \epsilon | T_0 = s) ds
\end{aligned}
$$

$$= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) Pr(t_{i+1} < T_0 + T_1 + T_2 < t_{i+1} + \epsilon | T_0 = s, T_1 = u) duds$$
$$= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) Pr(t_{i+1} - s - u < T_2 < t_{i+1} + \epsilon - s - u) duds$$
$$= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) f_2(t_{i+1} - u - s) duds\epsilon.$$

6. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. The individual is still in state 2 at the last screening point t_n. Then

$$Pr(T_0 > t_i, T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_n)$$
$$= Pr(t_i < T_0 < t_{i+1}, t_i < T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_n)$$
$$= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_i < T_0 + T_1 < t_{i+1} | T_0 = s) Pr(T_0 + T_1 + T_2 > t_n | T_0 = s)$$
$$= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) Pr(T_0 + T_1 + T_2 > t_n | T_0 = s, T_1 = u) duds$$
$$= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) S_2(t_n - s - u) duds.$$

7. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. The individual is observed in state 2 until t_{i+k-1}, where $k > 1$. At the exact time point t_{i+k}, the individual is observed in state 3. Then

$$Pr(T_0 > t_i, T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 \in [t_{i+k}, t_{i+k} + \epsilon))$$
$$= Pr(t_i < T_0 < t_{i+1}, t_i < T_0 + T_1 < t_{i+1}, t_{i+k} < T_0 + T_1 + T_2 < t_{i+k} + \epsilon)$$
$$= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) Pr(t_{i+k} - u - s < T_2 < t_{i+k} + \epsilon - u - s) duds$$
$$= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) f_2(t_{i+k} - u - s) duds\epsilon.$$

8. Suppose an individual is only observed in state 1 at all the screening time points, where t_n is the last screening. Then

$$Pr(T_0 < t_1, T_0 < t_n, T_0 + T_1 > t_1, T_0 + T_1 > t_n)$$
$$= Pr(T_0 < t_1, T_0 + T_1 > t_n)$$
$$= \int_0^{t_1} f_0(s) S_1(t_n - s) ds.$$

9. Suppose an individual is observed in state 1 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. The individual is still in state 2 at the

last screening point t_n. Then

$$\begin{aligned}
&Pr(T_0 < t_1, T_0 + T_1 > t_i, T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_n) \\
&= Pr(T_0 < t_1, t_i < T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_n) \\
&= \int_0^{t_1} f_0(s) Pr(t_i < T_0 + T_1 < t_{i+1} | T_0 = s) Pr(T_0 + T_1 + T_2 > t_n | T_0 = s) ds \\
&= \int_0^{t_1} \int_{t_i - s}^{t_{i+1} - s} f_0(s) f_1(u) Pr(T_0 + T_1 + T_2 > t_n | T_0 = s, T_1 = u) du ds \\
&= \int_0^{t_1} \int_{t_i - s}^{t_{i+1} - s} f_0(s) f_1(u) S_2(t_n - s - u) du ds.
\end{aligned}$$

10. Suppose an individual is observed in state 1 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. The individual is observed in state 2 until t_{i+k-1}, where $k > 1$. At the exact time point t_{i+k}, the individual is observed in state 3. Then

$$\begin{aligned}
&Pr(T_0 < t_1, T_0 + T_1 > t_i, T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 \in [t_{i+k}, t_{i+k} + \epsilon)) \\
&= Pr(T_0 < t_1, t_i < T_0 + T_1 < t_{i+1}, t_{i+k} < T_0 + T_1 + T_2 < t_{i+k} + \epsilon) \\
&= \int_0^{t_1} f_0(s) Pr(t_i < T_0 + T_1 < t_{i+1} | T_0 = s) Pr(t_{i+k} < T_0 + T_1 + T_2 < t_{i+k} + \epsilon | T_0 = s) ds \\
&= \int_0^{t_1} \int_{t_i - s}^{t_{i+1} - s} f_0(s) f_1(u) Pr(t_{i+k} < T_0 + T_1 + T_2 < t_{i+k} + \epsilon | T_0 = s, T_1 = u) du ds \\
&= \int_0^{t_1} \int_{t_i - s}^{t_{i+1} - s} f_0(s) f_1(u) f_2(t_{i+k} - s - u) du ds \epsilon.
\end{aligned}$$

11. Suppose an individual is observed in state 1 from t_1 to t_i. At the exact time point t_{i+1}, the individual is observed in state 3. Then

$$\begin{aligned}
&Pr(T_0 < t_1, T_0 + T_1 > t_i, T_0 + T_1 + T_2 \in [t_{i+1}, t_{i+1} + \epsilon)) \\
&= Pr(T_0 < t_1, t_i < T_0 + T_1 < t_{i+1}, t_{i+1} < T_0 + T_1 + T_2 < t_{i+1} + \epsilon) \\
&= \int_0^{t_1} \int_{t_i - s}^{t_{i+1} - s} f_0(s) f_1(u) Pr(t_{i+1} < T_0 + T_1 + T_2 < t_{i+1} + \epsilon | T_0 = s, T_1 = u) du ds \\
&= \int_0^{t_1} \int_{t_i - s}^{t_{i+1} - s} f_0(s) f_1(u) f_2(t_{i+1} - s - u) du ds \epsilon.
\end{aligned}$$

12. Suppose an individual is only observed in state 2 at all the screening time points, where t_n is the last screening. Then

$$\begin{aligned}
&Pr(T_0 < t_1, T_0 < t_n, T_0 + T_1 < t_1, T_0 + T_1 < t_n, T_0 + T_1 + T_2 > t_n) \\
&= Pr(T_0 < t_1, T_0 + T_1 < t_1, T_0 + T_1 + T_2 > t_n) \\
&= \int_0^{t_1} \int_0^{t_1 - s} f_0(s) f_1(u) S_2(t_n - u - s) du ds.
\end{aligned}$$

13. Suppose an individual is observed in state 2 from t_1 to t_i. At t_{i+1}, the individual is observed in state 3. Then

$$\begin{aligned}
&Pr(T_0 < t_1, T_0 + T_1 < t_1, t_{i+1} < T_0 + T_1 + T_2 < t_{i+1} + \epsilon) \\
&= \int_0^{t_1} f_0(s) Pr(T_0 + T_1 < t_1 | T_0 = s) Pr(t_{i+1} < T_0 + T_1 + T_2 < t_{i+1} + \epsilon | T_0 = s) ds \\
&= \int_0^{t_1} \int_0^{t_1 - s} f_0(s) f_1(u) f_2(t_{i+1} - s - u) du ds \epsilon.
\end{aligned}$$

14. Suppose an individual is observed in state 3 at the exact time point t_1, without any intermittent screening. Then

$$\begin{aligned}
&Pr(T_0 + T_1 + T_2 \in [t_1, t_1 + \epsilon)) \\
&= Pr(0 < T_0 < t_1, 0 < T_0 + T_1 < t_1, t_1 < T_0 + T_1 + T_2 < t_1 + \epsilon) \\
&= \int_0^{t_1} \int_0^{t_1 - s} f_0(s) f_1(u) f_2(t_1 - s - u) du ds \epsilon.
\end{aligned}$$

15. Suppose an individual is observed in state 0 from t_1 to t_i. The individual is observed in state 1 at t_{i+1}. The individual is observed in state 1 until t_{i+k-1}, where $k > 1$. At the exact time point t_{i+k}, the individual is observed in state 3. Then

$$\begin{aligned}
&Pr(t_i < T_0 < t_{i+1}, T_0 + T_1 > t_{i+k-1}, T_0 + T_1 + T_2 \in [t_{i+k}, t_{i+k} + \epsilon)) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k-1} < T_0 + T_1 < t_{i+k} | T_0 = s) \\
&Pr(t_{i+k} < T_0 + T_1 + T_2 < t_{i+k} + \epsilon | T_0 = s) ds \\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1} - s}^{t_{i+k} - s} f_0(s) f_1(u) f_2(t_{i+k} - s - u) du ds \epsilon.
\end{aligned}$$

The full likelihood for the individuals $p = 1, \ldots, m$ becomes

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) = &\prod_{(I)} S_0(t_{n,p}, \boldsymbol{\theta} | x_p) \prod_{(II)} \int_{t_{i,p}}^{t_{i+1}} f_0(s, \boldsymbol{\theta} | x_p) S_1(t_{n,p} - s, \boldsymbol{\theta} | x_p) ds \\
&\prod_{(III)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p} - s}^{t_{i+k,p} - s} f_0(s, \boldsymbol{\theta} | x_p) f_1(u, \boldsymbol{\theta} | x_p) S_2(t_{n,p} - u - s, \boldsymbol{\theta} | x_p) du ds \\
&\prod_{(IV)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p} - s}^{t_{i+k,p} - s} f_0(s, \boldsymbol{\theta} | x_p) f_1(u, \boldsymbol{\theta} | x_p) f_2(t_{i+k+l,p} - s - u, \boldsymbol{\theta} | x_p) du ds \\
&\prod_{(V)} \int_{t_{i,p}}^{t_{i+1,p}} \int_0^{t_{i+1,p} - s} f_0(s, \boldsymbol{\theta} | x_p) f_1(u, \boldsymbol{\theta} | x_p) f_2(t_{i+1,p} - s - u, \boldsymbol{\theta} | x_p) du ds \\
&\prod_{(VI)} \int_{t_{i,p}}^{t_{i+1,p}} \int_0^{t_{i+1,p} - s} f_0(s, \boldsymbol{\theta} | x_p) f_1(u, \boldsymbol{\theta} | x_p) S_2(t_{n,p} - s - u, \boldsymbol{\theta} | x_p) du ds \\
&\prod_{(VII)} \int_{t_{i,p}}^{t_{i+1,p}} \int_0^{t_{i+1,p} - s} f_0(s, \boldsymbol{\theta} | x_p) f_1(u, \boldsymbol{\theta} | x_p) f_2(t_{i+k,p} - u - s, \boldsymbol{\theta} | x_p) du ds \\
&\prod_{(VIII)} \int_0^{t_{1,p}} f_0(s, \boldsymbol{\theta} | x_p) S_1(t_{n,p} - s, \boldsymbol{\theta} | x_p) ds \\
&\prod_{(IX)} \int_0^{t_{1,p}} \int_{t_{i,p} - s}^{t_{i+1,p} - s} f_0(s, \boldsymbol{\theta} | x_p) f_1(u, \boldsymbol{\theta} | x_p) S_2(t_{n,p} - s - u, \boldsymbol{\theta} | x_p) du ds \\
&\prod_{(X)} \int_0^{t_{1,p}} \int_{t_{i,p} - s}^{t_{i+1,p} - s} f_0(s, \boldsymbol{\theta} | x_p) f_1(u, \boldsymbol{\theta} | x_p) f_2(t_{i+k,p} - s - u, \boldsymbol{\theta} | x_p) du ds \\
&\prod_{(XI)} \int_0^{t_{1,p}} \int_{t_{i,p} - s}^{t_{i+1,p} - s} f_0(s, \boldsymbol{\theta} | x_p) f_1(u, \boldsymbol{\theta} | x_p) f_2(t_{i+1,p} - s - u, \boldsymbol{\theta} | x_p) du ds
\end{aligned}$$

$$\prod_{(XII)} \int_0^{t_{1,p}} \int_0^{t_{1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) S_2(t_{n,p}-u-s,\boldsymbol{\theta}|x_p) duds$$

$$\prod_{(XIII)} \int_0^{t_{1,p}} \int_0^{t_{1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) f_2(t_{i+1,p}-s-u,\boldsymbol{\theta}|x_p) duds$$

$$\prod_{(XIV)} \int_0^{t_{1,p}} \int_0^{t_{1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) f_2(t_{1,p}-s-u,\boldsymbol{\theta}|x_p) duds$$

$$\prod_{(XV)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p}-s}^{t_{i+k,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) f_2(t_{i+k,p}-s-u,\boldsymbol{\theta}|x_p) duds.$$

3.5.2 Exact Time of Entry into the Absorbing State is not Known

The individuals are still screened $t_1, t_2, \ldots, t_n$ times. In this case, we assume that we do not observe the exact time of death, but only in which interval the transition happened. The likelihood contributions for type 1, 2 3, 6, 8, 9 and 12 are unchanged. The differences between when the exact time of entry into the absorbing state is known or not known are similar to what we explained in Section 3.3.2.2. The full likelihood for the individuals $p = 1, \ldots, m$ is

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) = & \prod_{(I)} S_0(t_{n,p}|x_p) \prod_{(II)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s,\boldsymbol{\theta}|x_p) S_1(t_n-s,\boldsymbol{\theta}|x_p) ds \\
& \prod_{(III)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p}-s}^{t_{i+k,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) S_2(t_{n,p}-u-s,\boldsymbol{\theta}|x_p) duds \\
& \prod_{(IV)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p}-s}^{t_{i+k,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) (S_2(t_{i+k+l-1,p}-s-u,\boldsymbol{\theta}|x_p) \\
& - S_2(t_{i+k+l,p}-s-u,\boldsymbol{\theta}|x_p)) duds \\
& \prod_{(V)} \int_{t_{i,p}}^{t_{i+1,p}} \int_0^{t_{i+1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) (1 - S_2(t_{i+1,p}-s-u,\boldsymbol{\theta}|x_p)) duds \\
& \prod_{(VI)} \int_{t_{i,p}}^{t_{i+1,p}} \int_0^{t_{i+1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) S_2(t_{n,p}-s-u,\boldsymbol{\theta}|x_p) duds \\
& \prod_{(VII)} \int_{t_{i,p}}^{t_{i+1,p}} \int_0^{t_{i+1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) (S_2(t_{i+k-1,p}-u-s,\boldsymbol{\theta}|x_p) \\
& - S_2(t_{i+k,p}-u-s,\boldsymbol{\theta}|x_p)) duds \\
& \prod_{(VIII)} \int_0^{t_{1,p}} f_0(s,\boldsymbol{\theta}|x_p) S_1(t_{n,p}-s,\boldsymbol{\theta}|x_p) ds \\
& \prod_{(IX)} \int_0^{t_{1,p}} \int_{t_{i,p}-s}^{t_{i+1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) S_2(t_{n,p}-s-u,\boldsymbol{\theta}|x_p) duds \\
& \prod_{(X)} \int_0^{t_{1,p}} \int_{t_{i,p}-s}^{t_{i+1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) (S_2(t_{i+k-1,p}-s-u,\boldsymbol{\theta}|x_p) - \\
& S_2(t_{i+k,p}-s-u,\boldsymbol{\theta}|x_p)) duds
\end{aligned}$$

$$\prod_{(XI)} \int_0^{t_{1,p}} \int_{t_{i,p}-s}^{t_{i+1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p)(1 - S_2(t_{i+1,p} - s - u, \boldsymbol{\theta}|x_p))duds$$

$$\prod_{(XII)} \int_0^{t_{1,p}} \int_0^{t_{1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) S_2(t_{n,p} - u - s, \boldsymbol{\theta}|x_p)duds$$

$$\prod_{(XIII)} \int_0^{t_{1,p}} \int_0^{t_{1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p)(S_2(t_{i,p} - s - u, \boldsymbol{\theta}|x_p) -$$

$$S_2(t_{i+1,p} - s - u, \boldsymbol{\theta}|x_p))duds$$

$$\prod_{(XIV)} \int_0^{t_1} \int_0^{t_{1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p)(1 - S_2(t_{1,p} - s - u, \boldsymbol{\theta}|x_p))duds$$

$$\prod_{(XV)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p}-s}^{t_{i+k,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p)(1 - S_2(t_{i+k,p} - s - u, \boldsymbol{\theta}|x_p))duds.$$

3.6 Four-State Illness-Death Model

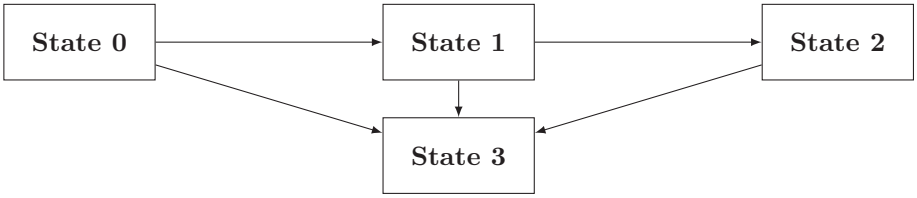


Figure 3.4: Four-state illness-death model

In this section, we consider a four-state illness-death model. It differs from the progressive four-state model, since an individual can move directly from state 0 to state 3 or from state 1 to state 3. The four-state illness-death model with the possible transitions is illustrated in Figure 3.4. T_0, T_1 and T_2 are defined in the same was as for the four-state progressive model. However, we now have two additional transition times, T_{03} and T_{13}. T_{03} is the transition time directly from state 0 to state 3, while T_{13} is the transition time directly from state 1 to state 3. T_{03} and T_{13} have densities f_{03} and f_{13}, survival time functions $S_{03} = 1 - F_{03}$ and $S_{13} = 1 - F_{13}$ and hazard rate functions $\alpha_{03} = f_{03}/S_{03}$ and $\alpha_{13} = f_{13}/S_{13}$. We still assume independence between T_0, T_1, T_2, T_{03}, T_{13}. We also have that if $T_0 > T_{03}$, then the individual goes from state 0 to state 3. If $T_0 < T_{03}$, then the individual goes from state 0 to state 1. In addition, if $T_1 > T_{13}$, the individual goes from state 1 to state 3. If $T_1 < T_{13}$, then the individual goes from state 1 to state 2.

3.6.1 Exact Time of Entry into the Absorbing State is Known

The individuals are screened $t_1, t_2, \ldots, t_n$ times.

1. Suppose an individual is only observed in state 0 at all the screening time points, where t_n is the last screening. Then

$$Pr(T_0 > t_n, T_{03} > t_n) = S_0(t_n)S_{03}(t_n).$$

2. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is still observed in state 1 at the last screening point t_n. We also have that $T_{03} > T_0$ and $T_{13} + T_1 > t_n$. Then

$$\begin{aligned}
&Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 < t_n, T_0 + T_1 > t_{i+1}, T_0 + T_1 > t_n, \\
&T_{03} > T_0, T_{13} + T_0 > t_n) \\
&= Pr(t_i < T_0 < t_{i+1}, T_0 + T_1 > t_n, T_{03} > T_0, T_{13} + T_0 > t_n) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(T_0 + T_1 > t_n | T_0 = s) Pr(T_{03} > T_0 | T_0 = s) \\
&Pr(T_{13} + T_0 > t_n | T_0 = s) ds \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(T_1 > t_n - s) Pr(T_{03} > s) Pr(T_{13} > t_n - s) ds \\
&= \int_{t_i}^{t_{i+1}} f_0(s) S_1(t_n - s) S_{03}(s) S_{13}(t_n - s) ds.
\end{aligned}$$

3. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is observed in state 1 until t_{i+k-1}, where $k > 1$. At t_{i+k}, the individual is observed in state 2. The individual is still in state 2 at the last screening point t_n. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$\begin{aligned}
&Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 + T_1 > t_{i+1}, T_0 + T_1 < t_{i+k}, \\
&T_0 + T_1 + T_2 > t_{i+k}, T_0 + T_1 + T_2 > t_n, T_{03} > T_0, T_{13} > T_1) \\
&= Pr(t_i < T_0 < t_{i+1}, t_{i+k-1} < T_0 + T_1 < t_{i+k}, T_0 + T_1 + T_2 > t_n, \\
&T_{03} > T_0, T_{13} > T_1) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k-1} < T_0 + T_1 < t_{i+k} | T_0 = s) \\
&Pr(T_0 + T_1 + T_2 > t_n | T_0 = s) Pr(T_{03} > T_0 | T_0 = s) Pr(T_{13} > T_1) ds \\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) Pr(T_0 + T_1 + T_2 > t_n | T_0 = s, T_1 = u) \\
&Pr(T_{03} > s) Pr(T_{13} > T_1 | T_1 = u) du ds \\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) Pr(T_2 > t_n - u - s) S_{03}(s) S_{13}(u) du ds \\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) S_2(t_n - u - s) S_{03}(s) S_{13}(u) du ds.
\end{aligned}$$

4. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is observed in state 1 until t_{i+k-1}, where $k > 1$. The individual is observed in state 2 at t_{i+k}. The individual is observed in state 2 until $t_{i+k+l-1}$, where $l > 1$. The individual is observed in state 3 at the exact time point t_{i+k+l}. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 + T_1 > t_{i+1}, T_0 + T_1 < t_{i+k},$
$T_0 + T_1 + T_2 \in [t_{i+k+l}, t_{i+k+l} + \epsilon), T_{03} > T_0, T_{13} > T_1)$
$= Pr(t_i < T_0 < t_{i+1}, t_{i+k-1} < T_0 + T_1 < t_{i+k}, t_{i+k+l} < T_0 + T_1 + T_2 < t_{i+k+l} + \epsilon,$
$T_{03} > T_0, T_{13} > T_1)$
$= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k-1} < T_0 + T_1 < t_{i+k} | T_0 = s)$
$Pr(t_{i+k+l} < T_0 + T_1 + T_2 < t_{i+k+l} + \epsilon | T_0 = s) Pr(T_{03} > T_0 | T_0 = s) Pr(T_{13} > T_1) ds$
$= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) Pr(t_{i+k+l} < T_0 + T_1 + T_2 < t_{i+k+l} + \epsilon | T_0 = s, T_1 = u)$
$S_{03}(s) Pr(T_{13} > T_1 | T_1 = u) du ds$
$= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) Pr(t_{i+k+l} - s - u < T_2 < t_{i+k+l} + \epsilon - s - u)$
$S_{03}(s) S_{13}(u) du ds$
$= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) f_2(t_{i+k+l} - s - u) S_{03}(s) S_{13}(u) du ds \epsilon.$

5. Suppose an individual is observed in state 0 from t_1 to t_i. At the exact time point t_{i+1}, the individual is observed in state 3. We also have that $T_{03} > T_0$ and $T_{13} > T_1$.

$Pr(T_0 > t_i, T_0 + T_1 + T_2 \in [t_{i+1}, t_{i+1} + \epsilon), T_{03} > T_0, T_{13} > T_1)$
$= Pr(t_i < T_0 < t_{i+1}, T_0 + T_1 + T_2 \in [t_{i+1}, t_{i+1} + \epsilon), T_{03} > T_0, T_{13} > T_1)$
$= Pr(t_i < T_0 < t_{i+1}, t_i < T_0 + T_1 < t_{i+1}, t_{i+1} < T_0 + T_1 + T_2 < t_{i+1} + \epsilon, T_{03} > T_0, T_{13} > T_1)$
$= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_i < T_0 + T_1 < t_{i+1} | T_0 = s) Pr(t_{i+1} < T_0 + T_1 + T_2 < t_{i+1} + \epsilon | T_0 = s)$
$Pr(T_{03} > T_0 | T_0 = s) Pr(T_{13} > T_1) ds$
$= \int_{t_i}^{t_{i+1}} \int_{0}^{t_{i+1}-s} f_0(s) f_1(u) Pr(t_{i+1} < T_0 + T_1 + T_2 < t_{i+1} + \epsilon | T_0 = s, T_1 = u)$
$S_{03}(s) Pr(T_{13} > T_1 | T_1 = u) du ds$
$= \int_{t_i}^{t_{i+1}} \int_{0}^{t_{i+1}-s} f_0(s) f_1(u) Pr(t_{i+1} - s - u < T_2 < t_{i+1} + \epsilon - s - u) S_{03}(s) S_{13}(u) du ds$
$= \int_{t_i}^{t_{i+1}} \int_{0}^{t_{i+1}-s} f_0(s) f_1(u) f_2(t_{i+1} - u - s) S_{03}(s) S_{13}(u) du ds \epsilon.$

6. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. The individual is still in state 2 at the last screening point t_n. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$Pr(T_0 > t_i, T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_n, T_{03} > T_0, T_{13} > T_1)$
$= Pr(t_i < T_0 < t_{i+1}, t_i < T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_n, T_{03} > T_0, T_{13} > T_1)$
$= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_i < T_0 + T_1 < t_{i+1} | T_0 = s) Pr(T_0 + T_1 + T_2 > t_n | T_0 = s)$
$Pr(T_{03} > T_0 | T_0 = s) Pr(T_{13} > T_1) ds$
$= \int_{t_i}^{t_{i+1}} \int_{0}^{t_{i+1}-s} f_0(s) f_1(u) Pr(T_0 + T_1 + T_2 > t_n | T_0 = s, T_1 = u) S_{03}(s) S_{13}(u) du ds$
$= \int_{t_i}^{t_{i+1}} \int_{0}^{t_{i+1}-s} f_0(s) f_1(u) S_2(t_n - s - u) S_{03}(s) S_{13}(u) du ds.$

7. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. The individual is observed in state 2 until t_{i+k-1}, where $k > 1$. At the exact time point t_{i+k}, the individual is observed in state 3. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$\begin{aligned}
&Pr(T_0 > t_i, T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 \in [t_{i+k}, t_{i+k} + \epsilon), T_{03} > T_0, T_{13} > T_1) \\
&= Pr(t_i < T_0 < t_{i+1}, t_i < T_0 + T_1 < t_{i+1}, t_{i+k} < T_0 + T_1 + T_2 < t_{i+k} + \epsilon, T_{03} > T_0, T_{13} > T_1) \\
&= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) Pr(t_{i+k} - u - s < T_2 < t_{i+k} + \epsilon - u - s) S_{03}(s) S_{13}(u) du ds \\
&= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) f_2(t_{i+k} - u - s) S_{03}(s) S_{13}(u) du ds \epsilon.
\end{aligned}$$

8. Suppose an individual is only observed in state 1 at all the screening time points, where t_n is the last screening. We also have that $T_{03} > T_0$ and $T_{13} + T_0 > t_n$. Then

$$\begin{aligned}
&Pr(T_0 < t_1, T_0 < t_n, T_0 + T_1 > t_1, T_0 + T_1 > t_n, T_{03} > T_0, T_0 + T_{13} > t_n) \\
&= Pr(T_0 < t_1, T_0 + T_1 > t_n, T_{03} > T_0, T_{13} + T_0 > t_n) \\
&= \int_0^{t_1} f_0(s) S_1(t_n - s) S_{03}(s) S_{13}(t_n - s) ds.
\end{aligned}$$

9. Suppose an individual is observed in state 1 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. The individual is still in state 2 at the last screening point t_n. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$\begin{aligned}
&Pr(T_0 < t_1, T_0 + T_1 > t_i, T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_n, T_{03} > T_0, T_{13} > T_1) \\
&= Pr(T_0 < t_1, t_i < T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_n, T_{03} > T_0, T_{13} > T_1) \\
&= \int_0^{t_1} f_0(s) Pr(t_i < T_0 + T_1 < t_{i+1} | T_0 = s) Pr(T_0 + T_1 + T_2 > t_n | T_0 = s) \\
&Pr(T_{03} > T_0 | T_0 = s) Pr(T_{13} > T_1) ds \\
&\quad = \int_0^{t_1} \int_{t_i - s}^{t_{i+1}-s} f_0(s) f_1(u) Pr(T_0 + T_1 + T_2 > t_n | T_0 = s, T_1 = u) \\
&\quad S_{03}(s) S_{13}(u) du ds \\
&\quad = \int_0^{t_1} \int_{t_i - s}^{t_{i+1}-s} f_0(s) f_1(u) S_2(t_n - s - u) S_{03}(s) S_{13}(u) du ds.
\end{aligned}$$

10. Suppose an individual is observed in state 1 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. The individual is still observed in state 2 until t_{i+k-1}, where $k > 1$. At the exact time point t_{i+k}, the individual is observed in state 3. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$\begin{aligned}
&Pr(T_0 < t_1, T_0 + T_1 > t_i, T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 \in [t_{i+k}, t_{i+k} + \epsilon), T_{03} > T_0, T_{13} > T_1) \\
&= Pr(T_0 < t_1, t_i < T_0 + T_1 < t_{i+1}, t_{i+k} < T_0 + T_1 + T_2 < t_{i+k} + \epsilon, T_{03} > T_0, T_{13} > T_1) \\
&= \int_0^{t_1} f_0(s) Pr(t_i < T_0 + T_1 < t_{i+1} | T_0 = s) Pr(t_{i+k} < T_0 + T_1 + T_2 < t_{i+k} + \epsilon | T_0 = s) \\
&Pr(T_{03} > T_0 | T_0 = s) Pr(T_{13} > T_1) ds \\
&= \int_0^{t_1} \int_{t_i - s}^{t_{i+1}-s} f_0(s) f_1(u) Pr(t_{i+k} < T_0 + T_1 + T_2 < t_{i+k} + \epsilon | T_0 = s, T_1 = u) \\
&S_{03}(s) Pr(T_{13} > T_1 | T_1 = u) du ds \\
&= \int_0^{t_1} \int_{t_i - s}^{t_{i+1}-s} f_0(s) f_1(u) f_2(t_{i+k} - s - u) S_{03}(s) S_{13}(u) du ds \epsilon.
\end{aligned}$$

11. Suppose an individual is observed in state 1 from t_1 to t_i. At the exact time point t_{i+1}, the individual is observed in state 3. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$
\begin{aligned}
&Pr(T_0 < t_1, T_0 + T_1 > t_i, T_0 + T_1 + T_2 \in [t_{i+1}, t_{i+1} + \epsilon), T_{03} > T_0, T_{13} > T_1) \\
&= Pr(T_0 < t_1, t_i < T_0 + T_1 < t_{i+1}, t_{i+1} < T_0 + T_1 + T_2 < t_{i+1} + \epsilon, T_{03} > T_0, T_{13} > T_1) \\
&= \int_0^{t_1} \int_{t_i - s}^{t_{i+1} - s} f_0(s) f_1(u) Pr(t_{i+1} < T_0 + T_1 + T_2 < t_{i+1} + \epsilon | T_0 = s, T_1 = u) \\
&Pr(T_{03} > T_0 | T_0 = s) Pr(T_{13} > T_1 | T_1 = u) du ds \\
&= \int_0^{t_1} \int_{t_i - s}^{t_{i+1} - s} f_0(s) f_1(u) f_2(t_{i+1} - s - u) S_{03}(s) S_{13}(u) du ds \epsilon.
\end{aligned}
$$

12. Suppose an individual is only observed in state 2 at all the screening time points , where t_n is the last screening. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$
\begin{aligned}
&Pr(T_0 < t_1, T_0 < t_n, T_0 + T_1 < t_1, T_0 + T_1 < t_n, T_0 + T_1 + T_2 > t_n, T_{03} > T_0, T_{13} > T_1) \\
&= Pr(T_0 < t_1, T_0 + T_1 < t_1, T_0 + T_1 + T_2 > t_n, T_{03} > T_0, T_{13} > T_1) \\
&= \int_0^{t_1} \int_0^{t_1 - s} f_0(s) f_1(u) S_2(t_n - u - s) S_{03}(s) S_{13}(u) du ds.
\end{aligned}
$$

13. Suppose an individual is observed in state 2 from t_1 to t_i. At t_{i+1}, the individual is observed in state 3. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$
\begin{aligned}
&Pr(T_0 < t_1, T_0 + T_1 < t_1, t_{i+1} < T_0 + T_1 + T_2 < t_{i+1} + \epsilon, T_{03} > T_0, T_{13} > T_1) \\
&= \int_0^{t_1} f_0(s) Pr(T_0 + T_1 < t_1 | T_0 = s) Pr(t_{i+1} < T_0 + T_1 + T_2 < t_{i+1} + \epsilon | T_0 = s) \\
&Pr(T_{03} > T_0 | T_0 = s) Pr(T_{13} > T_1) ds \\
&= \int_0^{t_1} \int_0^{t_1 - s} f_0(s) f_1(u) f_2(t_{i+1} - s - u) S_{03}(s) S_{13}(u) du ds \epsilon.
\end{aligned}
$$

14. Suppose an individual is observed in state 3 at the exact time point t_1, without any intermittent screening. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$
\begin{aligned}
&Pr(T_0 + T_1 + T_2 \in [t_1, t_1 + \epsilon), T_{03} > T_0, T_{13} > T_1) \\
&= Pr(0 < T_0 < t_1, 0 < T_0 + T_1 < t_1, t_1 < T_0 + T_1 + T_2 < t_1 + \epsilon, T_{03} > T_0, T_{13} > T_1) \\
&= \int_0^{t_1} \int_0^{t_1 - s} f_0(s) f_1(u) f_2(t_1 - s - u) S_{03}(s) S_{13}(u) du ds \epsilon.
\end{aligned}
$$

15. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is observed in state 1 until t_{i+k-1}. At the exact time point t_{i+k}, the individual is observed in state 3. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$
\begin{aligned}
&Pr(t_i < T_0 < t_{i+1}, T_0 + T_1 > t_{i+k-1}, T_0 + T_1 + T_2 \in [t_{i+k}, t_{i+k} + \epsilon), T_{03} > T_0, T_{13} > T_1) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k-1} < T_0 + T_1 < t_{i+k} | T_0 = s) Pr(t_{i+k} < T_0 + T_1 + T_2 < t_{i+k} + \epsilon | T_0 = s) \\
&Pr(T_{03} > T_0 | T_0 = s) Pr(T_{13} > T_1) ds \\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1} - s}^{t_{i+k} - s} f_0(s) f_1(u) f_2(t_{i+k} - s - u) S_{03}(s) S_{13}(u) du ds \epsilon.
\end{aligned}
$$

16. Suppose an individual is observed in state 0 from t_1 to t_i. At the exact time point t_{i+1}, the individual is observed in state 3. We also have that $T_{03} < T_0$. Then

$$Pr(t_{i+1} < T_{03} < t_{i+1} + \epsilon, T_{03} < T_0) = Pr(t_{i+1} < T_{03} < t_{i+1} + \epsilon, T_0 > t_{i+1}$$
$$= f_{03}(t_{i+1})S_0(t_{i+1})\epsilon.$$

17. Suppose an individual is observed in state 3 at the exact time point t_1, without any intermittent screening. We also have that $T_{03} < T_0$. Then

$$Pr(t_1 < T_{03} < t_{i+1}, T_{03} < T_0) = Pr(t_1 < T_{03} < t_{i+1}, t_1 < T_0) = f_{03}(t_1)S_0(t_1)\epsilon.$$

18. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is observed in state 1 until t_{i+k-1}. At the exact time point t_{i+k}, the individual is observed in state 3. We also have that $T_{03} > T_0$ and $T_{13} < T_1$. Then

$$Pr(t_i < T_0 < t_{i+1}, t_{i+k} < T_0 + T_{13} < t_{i+k} + \epsilon, T_{03} > T_0, T_0 + T_1 > t_{i+k})$$
$$= \int_{t_i}^{t_{i+1}} f_0(s)S_{03}(s)Pr(t_{i+k} - s < T_{13} < t_{i+k} + \epsilon - s)Pr(T_1 > t_{i+k} - s)ds$$
$$= \int_{t_i}^{t_{i+1}} f_0(s)S_{03}(s)f_{13}(t_{i+k} - s)S_1(t_{i+k} - s)ds\epsilon.$$

19. Suppose an individual is observed in state 1 from t_1 to t_i. At the exact time point t_{i+1}, the individual is observed in state 3. We also have that $T_{03} > T_0$ and $T_{13} < T_1$. Then

$$Pr(T_0 < t_1, t_{i+1} < T_0 + T_{13} < t_{i+1} + \epsilon, T_{03} > T_0, T_0 + T_1 > t_{i+1})$$
$$= \int_{0}^{t_1} f_0(s)S_{03}(s)Pr(t_{i+1} - s < T_{13} < t_{i+1} + \epsilon - s)Pr(T_1 > t_{i+1} - s)ds$$
$$= \int_{0}^{t_1} f_0(s)S_{03}(s)f_{13}(t_{i+1} - s)S_1(t_{i+1} - s)ds\epsilon.$$

20. Suppose an individual is observed in state 3 at the the exact time point t_1, without any intermittent screening. We also have that $T_{03} > T_0$ and $T_{13} < T_1$. Then

$$Pr(t_1 + \epsilon > T_{13} + T_0 > t_1, T_{03} > T_0, T_0 + T_1 > t_1)$$
$$= \int_{0}^{t_1} f_0(s)S_{03}(s)f_{13}(t_1 - s)S_1(t_1 - s)ds\epsilon.$$

21. Suppose an individual is observed in state 0 from t_1 to t_i. At the exact time point t_{i+1}, the individual is observed in state 3. We also have that $T_{03} > T_0$ and $T_{13} < T_1$. Then

$$Pr(t_i < T_0 < t_{i+1}, T_{03} > T_0, T_1 + T_0 > t_{i+1}, t_{i+1} < T_0 + T_{13} < t_{i+1} + \epsilon)$$
$$= \int_{t_i}^{t_{i+1}} f_0(s)S_{03}(s)f_{13}(t_{i+1} - s)S_1(t_{i+1} - s)ds\epsilon.$$

The full likelihood for the individuals $p = 1, \ldots, m$ is

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) = & \prod_{(I)} S_0(t_{n,p}, \boldsymbol{\theta}|x_p) S_{03}(t_{n,p}, \boldsymbol{\theta}|x_p) \\
& \prod_{(II)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_1(t_{n,p} - s, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(t_{n,p} - s, \boldsymbol{\theta}|x_p) ds \\
& \prod_{(III)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p}-s}^{t_{i+k,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) S_2(t_{n,p} - u - s, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(IV)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p}-s}^{t_{i+k,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) f_2(t_{i+k+l,p} - s - u, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(V)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{0}^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) f_2(t_{i+1,p} - u - s, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(VI)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{0}^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) S_2(t_{n,p} - s - u, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(VII)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{0}^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) f_2(t_{i+k,p} - u - s, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(VIII)} \int_{0}^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_1(t_{n,p} - s, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(t_{n,p} - s, \boldsymbol{\theta}|x_p) ds \\
& \prod_{(IX)} \int_{0}^{t_{1,p}} \int_{t_{i,p}-s}^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) S_2(t_{n,p} - s - u, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(X)} \int_{0}^{t_{1,p}} \int_{t_{i,p}-s}^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) f_2(t_{i+k,p} - s - u, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(XI)} \int_{0}^{t_{1,p}} \int_{t_{i,p}-s}^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) f_2(t_{i+1,p} - s - u, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(XII)} \int_{0}^{t_{1,p}} \int_{0}^{t_{1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) S_2(t_{n,p} - u - s, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(XIII)} \int_{0}^{t_{1,p}} \int_{0}^{t_{1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) f_2(t_{i+1,p} - s - u, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(XIV)} \int_{0}^{t_{1,p}} \int_{0}^{t_{1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) f_2(t_{1,p} - s - u, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(XV)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p}-s}^{t_{i+k,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) f_2(t_{i+k,p} - s - u, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(XVI)} f_{03}(t_{i+1,p}|x_p) S_0(t_{i+1,p}, \boldsymbol{\theta}|x_p) \\
& \prod_{(XVII)} f_{03}(t_{1,p}|x_p) S_0(t_{1,p}, \boldsymbol{\theta}|x_p) \\
& \prod_{(XVIII)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) f_{13}(t_{i+1,p} - s, \boldsymbol{\theta}|x_p) S_1(t_{i+1,p} - s, \boldsymbol{\theta}|x_p) ds \\
& \prod_{(XIX)} \int_{0}^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) f_{13}(t_{i+1,p} - s, \boldsymbol{\theta}|x_p) S_1(t_{i+1,p} - s, \boldsymbol{\theta}|x_p) ds \\
& \prod_{(XX)} \int_{0}^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) f_{13}(t_{1,p} - s, \boldsymbol{\theta}|x_p) S_1(t_{1,p} - s, \boldsymbol{\theta}|x_p) ds \\
& \prod_{(XXI)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) f_{13}(t_{i+1,p} - s, \boldsymbol{\theta}|x_p) S_1(t_{i+1,p} - s, \boldsymbol{\theta} v|x_p) ds.
\end{aligned}
$$

3.6.2 Exact Time of Entry into the Absorbing State is not Known

The individuals are still screened $t_1, t_2, \ldots, t_n$ times. In this case, we assume that we do not observe the exact time of death, but only in which interval the transition happened. The likelihood contributions for type 1, 2 3, 6, 8, 9 and 12 are unchanged. The differences between when the exact time of entry into the absorbing state is known or not known are similar to what we explained in Section 3.3.2.2. The full likelihood for the individuals $p = 1, \ldots, m$ is

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) = & \prod_{(I)} S_0(t_{n,p}, \boldsymbol{\theta}|x_p) S_{03}(t_{n,p}, \boldsymbol{\theta}|x_p) \\
& \prod_{(II)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_1(t_{n,p} - s, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(t_{n,p} - s, \boldsymbol{\theta}|x_p) ds \\
& \prod_{(III)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p}-s}^{t_{i+k,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) S_2(t_{n,p} - u - s, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(IV)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p}-s}^{t_{i+k,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) (S_2(t_{i+k+l-1,p} - s - u, \boldsymbol{\theta}|x_p) - S_2(t_{i+k+l} - s - u, \boldsymbol{\theta}|x_p)) \\
& S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(V)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{0}^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) (1 - S_2(t_{i+1,p} - s - u, \boldsymbol{\theta}|x_p)) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(VI)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{0}^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) S_2(t_{n,p} - s - u, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(VII)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{0}^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) (S_2(t_{i+k-1,p} - u - s, \boldsymbol{\theta}|x_p) - S_2(t_{i+k,p} - u - s, \boldsymbol{\theta}|x_p)) \\
& S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(VIII)} \int_{0}^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_1(t_{n,p} - s, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(t_{n,p} - s, \boldsymbol{\theta}|x_p) ds \\
& \prod_{(IX)} \int_{0}^{t_{1,p}} \int_{t_{i,p}-s}^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) S_2(t_{n,p} - s - u, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(X)} \int_{0}^{t_{1,p}} \int_{t_{i,p}-s}^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) (S_2(t_{i+k-1,p} - s - u, \boldsymbol{\theta}|x_p) - S_2(t_{i+k,p} - s - u, \boldsymbol{\theta}|x_p)) \\
& S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(XI)} \int_{0}^{t_{1,p}} \int_{t_{i,p}-s}^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) (1 - S_2(t_{i+1,p} - s - u, \boldsymbol{\theta}|x_p)) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(XII)} \int_{0}^{t_{1,p}} \int_{0}^{t_{1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) S_2(t_{n,p} - u - s, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(XIII)} \int_{0}^{t_{1,p}} \int_{0}^{t_{1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) (S_2(t_{i,p} - s - u, \boldsymbol{\theta}|x_p) - S_2(t_{i+1,p} - s - u, \boldsymbol{\theta}|x_p)) \\
& S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(XIV)} \int_{0}^{t_{1,p}} \int_{0}^{t_{1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) (1 - S_2(t_{1,p} - s - u, \boldsymbol{\theta}|x_p)) \\
& S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(XV)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1}}^{t_{i+k,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) (1 - S_2(t_{i+k,p} - s - u, \boldsymbol{\theta}|x_p)) \\
& S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
& \prod_{(XVI)} \int_{t_{i,p}}^{t_{i+1,p}} f_{03}(s, \boldsymbol{\theta}|x_p) S_0(s, \boldsymbol{\theta}|x_p) ds
\end{aligned}
$$

$$\prod_{(XVII)} \int_0^{t_{1,p}} f_{03}(s,\boldsymbol{\theta}|x_p)S_0(s,\boldsymbol{\theta}|x_p)ds$$

$$\prod_{(XVIII)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p}-s}^{t_{i+k,p}-s} f_0(s,\boldsymbol{\theta}|x_p)S_{03}(s,\boldsymbol{\theta}|x_p)f_{13}(u,\boldsymbol{\theta}|x_p)S_1(u,\boldsymbol{\theta}|x_p)duds$$

$$\prod_{(XIX)} \int_0^{t_{1,p}} \int_{t_{i,p}-s}^{t_{i+1,p}-s} f_0(s,\boldsymbol{\theta}|x_p)S_{03}(s,\boldsymbol{\theta}|x_p)f_{13}(u,\boldsymbol{\theta}|x_p)S_1(u,\boldsymbol{\theta}|x_p)duds$$

$$\prod_{(XX)} \int_0^{t_{1,p}} \int_0^{t_{1,p}-s} f_0(s,\boldsymbol{\theta}|x_p)S_{03}(s,\boldsymbol{\theta}|x_p)f_{13}(u,\boldsymbol{\theta}|x_p)S_1(u,\boldsymbol{\theta}|x_p)duds$$

$$\prod_{(XXI)} \int_{t_{i,p}}^{t_{i+1,p}} \int_0^{t_{i+1,p}-s} f_0(s,\boldsymbol{\theta}|x_p)S_{03}(s,\boldsymbol{\theta}|x_p)f_{13}(u,\boldsymbol{\theta}|x_p)S_1(u,\boldsymbol{\theta}|x_p)duds.$$

3.7 Gamma Process Models

In this section we present the survival and density formulas for Gamma process models. Let $Z(t) \sim Gam(at, 1)$, where the survival function for an individual is

$$S(t,a,c) = P(T \geq t) = P(Z(t) < c) = G(c, at, 1) \tag{3.3}$$

and the density function

$$f(t,a,c) = -\frac{\partial S(t,a,c)}{\partial t}. \tag{3.4}$$

This is often computed numerically as

$$\begin{aligned} f(t,a,c) &\approx \frac{F_0(t+\epsilon,a,c) - F_0(t,a,c)}{\epsilon} \\ &\approx \frac{S_0(t,a,c) - S_0(t+\epsilon,a,c)}{\epsilon} \end{aligned} \tag{3.5}$$

In the three-state progressive model, we consider two Gamma processes, one for the transition from state 0 to state 1, $Z_0(t) \sim Gam(a_0 t, 1)$, and one for the transition from state 1 to state 2, $Z_1(t) \sim Gam(a_1 t, 1)$. In an illness-death model we have three Gamma processes. The two Gamma processes for state 0 to state 1 and from state 1 to state 2 are formulated in the same way as when we have a three-state progressive model. The third Gamma process is from state 0 to state 2. It is defined as $Z_{02} \sim Gam(a_{02}t, 1)$. Z_0 and Z_{02} competes with each other. This means if $Z_{02}(t)$ crosses c_{02} before $Z_0(t)$ crosses c_0, then the patient will jump straight to state 2 without going through state 1.

In the four-state progressive model model we also have three Gamma processes. The two Gamma processes for the transition from state 0 to state 1 and for the transition from state 1 to state 2 are formulated in the same way as when we have a three-state progressive model. The third Gamma process is from state 2 to state 3. It is defined as $Z_2 \sim Gam(a_2 t, 1)$. In a four-state illness-death model we have five Gamma processes. The three Gamma processes for the transition from state 0 to state 1, for the transition from state 1 to state

2 and for the transition from state 2 to state 3 are formulated in the same way as for a four-state progressive model. In addition, we have a Gamma process for the transition from state 0 to state 3 and one for the transition from state 1 to state 3. The transition from state 0 to state 3 is defined as $Z_{03} \sim Gam(a_{03}t, 1)$, while from state 1 to state 3 is defined as $Z_{13} \sim Gam(a_{13}t, 1)$. Z_0 and Z_{03} competes with each other, which means if $Z_{03}(t)$ crosses c_{03} before $Z_0(t)$ crosses c_0, then the patient will jump straight to state 3 without going through state 1 and state 2. If an individual reaches state 1, we also have two process which starts at the same time and competes with each other, Z_1 and Z_{13}. If $Z_{13}(t)$ crosses c_{13} before $Z_1(t)$ crosses c_1, then the individual will jump straight to state 3 without going through state 2.

CHAPTER 4

Simulations

4.1 Background and Motivation

In this chapter, we start by simulating the transition times from a known parametric survival times model, in this case a Gamma process model. Next, we simulate the time points and place the individuals into the correct likelihood type. In the end, we use the log-likelihood to find the maximum likelihood estimates by numerical optimization. The reason for simulating in this fashion is to get a confirmation that we have the right construction of the likelihood and confirm the large-sample properties of the maximum likelihood estimators. We expect the large-sample properties to hold.

We start with presenting the recipe for simulating the transition times. For each patient $p = 1, \ldots, m$ we generate the transition times $T_{k,p}$. In a three-state progressive model, $k = 0, 1$, for an illness-death model $k = 0, 1, 02$, for a four-state progressive model $k = 0, 1, 2$ and for a four-state illness-death model $k = 0, 1, 2, 03, 13$. We find $T_{k,p}$ by solving the equation for x

$$\int_0^{c_k} \frac{1}{\Gamma(a_k T_{k,p})} x^{a_k T_{k,p}-1} \exp(-x) dx = y,$$

where y is a random number between 0 and 1. Thus, we draw a random number from a standard uniform distribution and solve for $T_{k,p}$. We use the different values for $T_{k,p}$ and the screening time points to determine of which type the individual is.

In Chapter 2, we discussed the likelihood theory and the large-sample properties. The aim in this chapter is to check the large-sample properties. We do this by checking two different properties. The first one is that for each simulation, we check that the parameter estimates are close to the true estimates. The second one is we use the simulations of each parameter to check that the densities of the $\mathcal{Z}_{\theta_i}$-values, defined in Equation 4.1, are close to a $N(0, 1)$-distribution. The estimated covariance matrix is the inverse Hessian matrix. The relationship between the Hessian matrix, $\hat{\boldsymbol{H}}$, and the observed Fisher information, $\hat{\boldsymbol{I}}$ is $\hat{\boldsymbol{H}} = -\hat{\boldsymbol{I}}$. The diagonal of the inverse Hessian matrix is an approximation of the variance for the estimated parameters. The estimated variance for parameter θ_i in simulation w is called $\hat{\kappa}_{i,w}$. Then the $\mathcal{Z}$-score for parameter θ_i in simulation w becomes

$$\mathcal{Z}_{\theta_i,w} = \frac{\hat{\theta}_{i,w} - \theta_i}{\hat{\kappa}_{i,w}}, \tag{4.1}$$

where its density should be approximately standard normal. We calculate the $\mathcal{Z}$-score for all the different parameters for each simulation $w = 1, \ldots, r$.

The rest of the chapter is organized as follows. We start by presenting the simulations for one screening and multiple screenings in the three-state progressive model in Section 4.2. The results for the simulations of multiple screenings for the illness-death model are presented in Section 4.3. In the case of multiple screenings, we divide the section into when the time into the absorbing state is and is not observed exactly. In Section 4.4, we consider multiple screenings, where the transition into the absorbing state both is observed exactly and not exactly. In Section 4.5 we only consider when time of death is known. This is because we find that the results for when the transition to the absorbing state is or is not exactly known are very similar. To find the maximum likelihood estimates and their hessian matrix, we use the *optim* function in R. Lastly, in Section 4.6, we illustrate how much information is lost from the fact that the transition times are not observed exactly.

4.2 Three-State Progressive Model

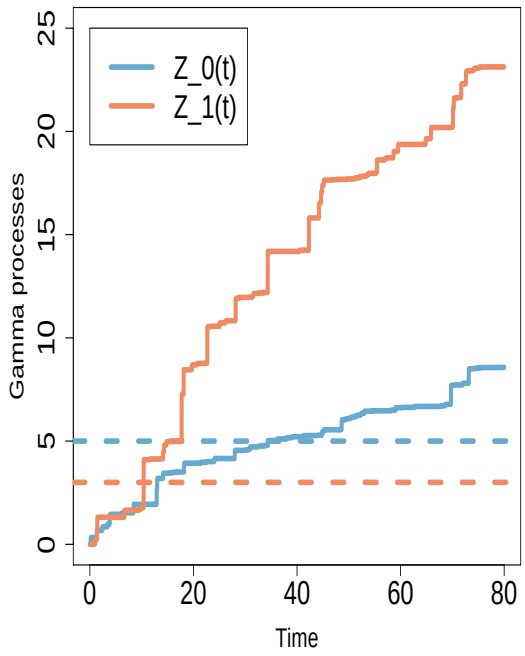


Figure 4.1: Simulation of a Gamma process from state 0 to state 1 (Z_0) and a Gamma process from state 1 to state 2 (Z_1).

In this section, we let $c_0 = 5$, $a_0 = 0.2$, $c_1 = 3$ and $a_1 = 0.2$. We illustrate in Figure 4.1 a simulation of a Gamma process from state 0 to state 1 (Z_0) and a simulation of a Gamma process from state 1 to state 2 (Z_1) using these

parameter values. From the processes illustrated in this plot, the individual moves from state 0 to state 1 after around 35 years and from state 1 to state 2 around 18 years after. The total time from state 0 to state 2 is around 53 years.

4.2.1 One Screening

We start by defining the simulating scheme when we only have one screening, where we in addition may observe time of death.

1. Simulate $T_{0,p}$ and $T_{1,p}$ as described in section 4.1
2. Draw the time points from screening t and u from a uniform distribution

$$t \sim U[0.5, 120],$$

$$u \sim U[0.5, 90].$$

   Then

   a) if $t < T_{0,p}$ and $t < T_{0,p} + T_{1,p}$ then the patient is type 1 at time t.
   b) if $t > T_{0,p}$ and $t < T_{0,p} + T_{1,p}$, then the patient in type 2 at time t.
   c) if $t > T_{0,p} + T_{1,p}$ and $u > T_{0,p} + T_{1,p}$, then the patient is type 3 at time $t = T_{0,p} + T_{1,p}$.
   d) if $t > T_{0,p} + T_{1,p}$, $u < T_{0,p}$ and $u < T_{0,p} + T_{1,p}$, then the patient is type 4. We then have the time points u and $t = T_{0,p} + T_{1,p}$.
   e) if $t > T_{0,p} + T_{1,p}$, $u > T_{0,p}$ and $u < T_{0,p} + T_{1,p}$, then the patient is type 5. We then have the time points u and $t = T_{0,p} + T_{1,p}$.

3. Use these data to optimize the log-likelihood function and find the maximum likelihood estimates.

We consider 500 patients and we do the simulations 1000 times. In each simulation, we estimate the maximum likelihood parameters.

Parameter	**True Value**	**Mean of Estimates**
c_0	5.000	5.089
a_0	0.200	0.203
c_1	3.000	3.192
a_1	0.200	0.211

Table 4.1: True value and mean of the estimated parameters in a three-state progressive model with one screening.

In Table 4.1 we present the true values of the parameters and the mean of the estimates. The mean of the maximum likelihood estimates are close to their true values.

The second large-sample property we check is the density of

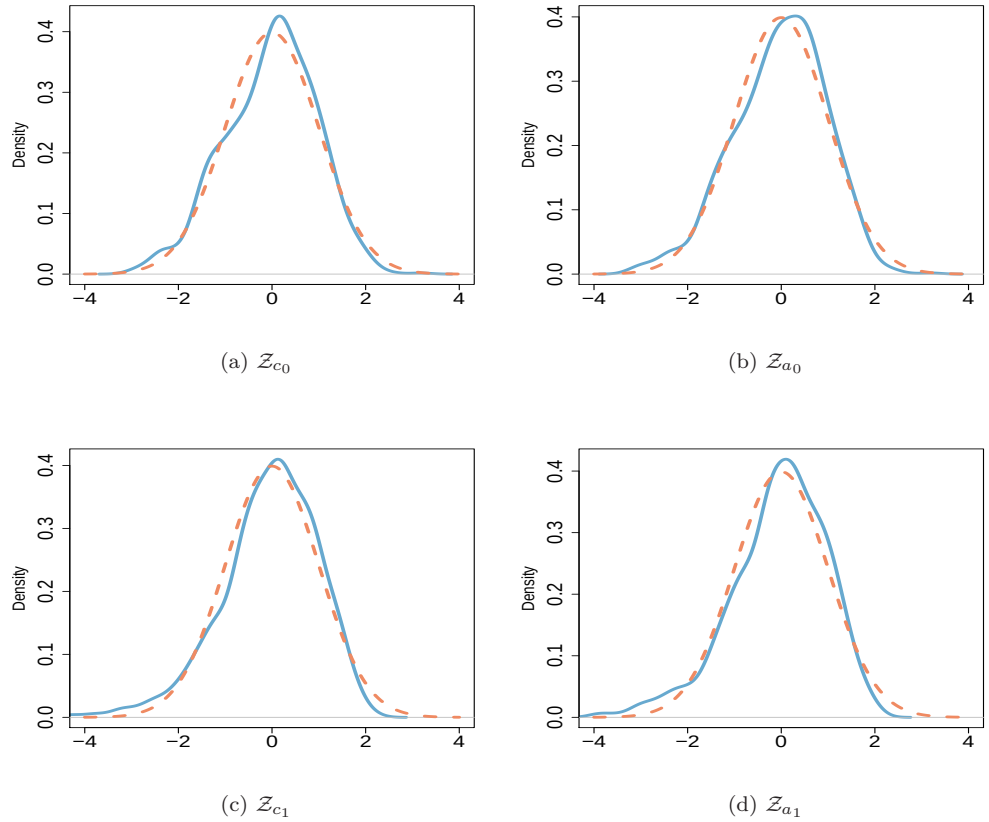
$$\mathcal{Z}_{c_0,w} = \frac{\hat{c}_{0,w} - c_0}{\hat{\kappa}_{c_0,w}}, \text{ where } w = 1 \dots r,$$

	Mean	Variance
$\mathcal{Z}_{c_0}$	-0.0312	0.959
$\mathcal{Z}_{a_0}$	-0.0314	0.979
$\mathcal{Z}_{c_1}$	-0.0529	1.060
$\mathcal{Z}_{a_1}$	-0.0555	1.043

Table 4.2: Mean and variance of the $\mathcal{Z}_{\theta_i}$-values in a three-state progressive model with one screening.

where $\hat{\kappa}_{c_0,w}$ is the square root of the variance for c_0. The variance is from the diagonal of the inverse Hessian at simulation number w.

We present the mean and variance of the $\mathcal{Z}_{\theta_i}$-values in Table 4.2. All of the $\mathcal{Z}_{\theta_i}$-values have a mean close to 0 and a variance close to 1.



(a) $\mathcal{Z}_{c_0}$ (b) $\mathcal{Z}_{a_0}$

(c) $\mathcal{Z}_{c_1}$ (d) $\mathcal{Z}_{a_1}$

Figure 4.2: Density of $\mathcal{Z}_{c_0}, \mathcal{Z}_{a_0}, \mathcal{Z}_{c_1}$ and $\mathcal{Z}_{a_1}$ in a three-state progressive model with one screening

Figure 4.2 shows the density plots for $\mathcal{Z}_{c_0}$, $\mathcal{Z}_{a_0}$, $\mathcal{Z}_{c_1}$ and $\mathcal{Z}_{a_1}$. In each plot, we include a red dashed line which is the density of the standard normal distribution. The densities for the $\mathcal{Z}_{\theta_i}$-values seem to be quite close to the

standard normal distribution.

As expected, we have that the mean of the estimates are close to the true values and the density of the $\mathcal{Z}_{\theta_i}$-values are close to the standard normal distribution. We have that for 500 patients, our conclusion is that the large-sample theory provide fully adequate approximations to the relevant distributions.

4.2.2 Multiple Screenings

In this part, we consider the situation where the individuals are screened between 2 and 15 times. The simulations are done in the same way as in Section 4.2.1, with small modifications. The exact recipe for the simulations is found in Appendix B. We do the simulations 1000 times for 500 patients.

4.2.2.1 Exact Time of Transition to the Absorbing State is Known

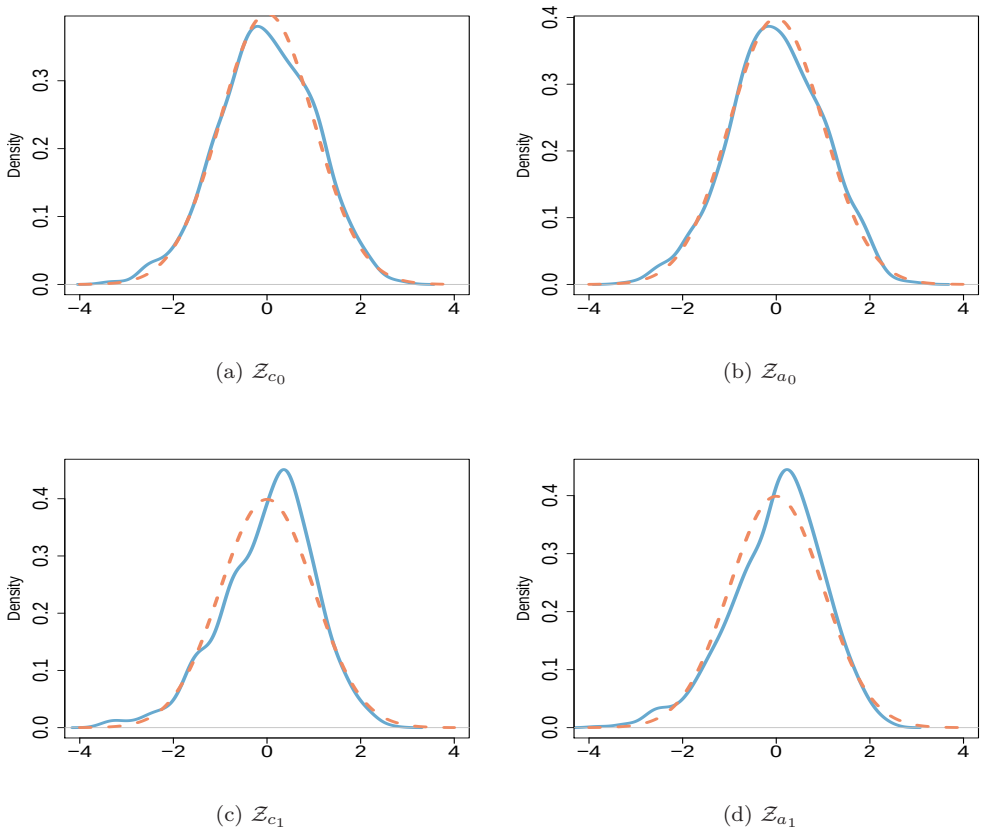
When the exact time of transition to the absorbing state is known, we know that the exact time of death is $T_0 + T_1$. Table 4.3 shows the mean of the

Parameter	**True Value**	**Mean of Estimates**
c_0	5.000	5.043
a_0	0.200	0.202
c_1	3.000	3.084
a_1	0.200	0.205

Table 4.3: True value and mean of the estimated parameters in a three-state progressive model for multiple screenings when the exact time of transition to the absorbing state is known.

estimated parameters and the true values when the exact time of transition to the absorbing state is known. The mean of the estimates are close to the true values.

Figure 4.3 shows the densities for $\mathcal{Z}_{c_0}$, $\mathcal{Z}_{a_0}$, $\mathcal{Z}_{c_1}$ and $\mathcal{Z}_{a_1}$ when we have observed the exact time of death in a three-state progressive model with multiple screenings for each person. The dashed red line in each plot is the density for the standard normal distribution. The densities for the $\mathcal{Z}_{\theta_i}$-values are centered around 0 and they seem to follow the shape of the standard normal distribution. There are some small differences between the densities for the $\mathcal{Z}_{\theta_i}$-values and the standard normal distribution. Our conclusion is still that the large-sample theory provide fully adequate approximations to the relevant distributions.



(a) $\mathcal{Z}_{c_0}$ (b) $\mathcal{Z}_{a_0}$

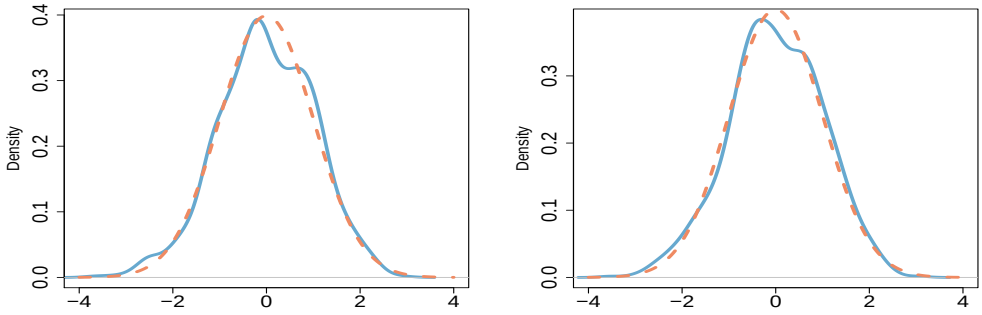
(c) $\mathcal{Z}_{c_1}$ (d) $\mathcal{Z}_{a_1}$

Figure 4.3: Density of $\mathcal{Z}_{c_0}, \mathcal{Z}_{a_0}, \mathcal{Z}_{c_1}$ and $\mathcal{Z}_{a_1}$ for multiple screenings in a three-state progressive model when the exact time of transition to the absorbing state is known

4.2.2.2 Exact Time of Transition to the Absorbing State is not Known

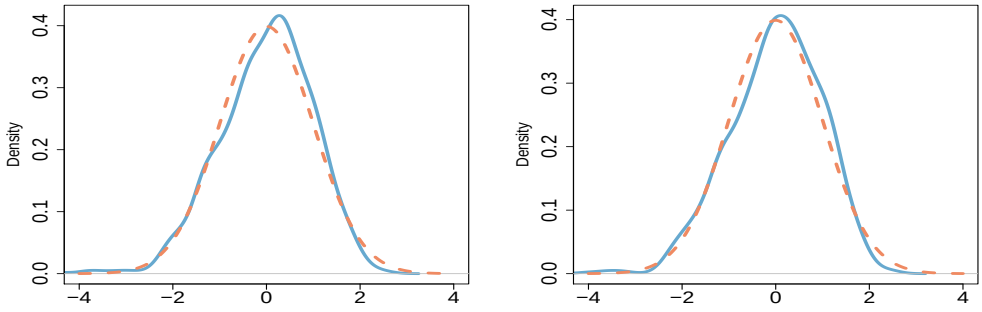
Parameter	True Value	Mean of Estimates
c_0	5.000	5.049
a_0	0.200	0.202
c_1	3.000	3.130
a_1	0.200	0.207

Table 4.4: True value of parameters and mean of the estimated parameters in a three-state progressive model for multiple screenings when the exact time of transition to the absorbing state is not known



(a) $\mathcal{Z}_{c_0}$

(b) $\mathcal{Z}_{a_0}$



(c) $\mathcal{Z}_{c_1}$

(d) $\mathcal{Z}_{a_1}$

Figure 4.4: Density of $\mathcal{Z}_{c_0}, \mathcal{Z}_{a_0}, \mathcal{Z}_{c_1}$ and $\mathcal{Z}_{a_1}$ for multiple screenings in a three-state progressive model when the exact time of transition to the absorbing state is not known.

Table 4.4 presents the true values and mean of the maximum likelihood estimates in a three-state progressive model when the exact time of transition to the absorbing state is not known. There are small differences between the true values and the mean of the estimates. The differences between the true values and mean of the estimates are a bit higher in Table 4.4 than when the exact time of transition to the absorbing state is known, which we presented in Table 4.3.

Figure 4.4 shows the densities for $\mathcal{Z}_{c_0}, \mathcal{Z}_{a_0}, \mathcal{Z}_{c_1}$ and $\mathcal{Z}_{a_1}$ when we have not observed the exact time of transition to the absorbing state in a three-state progressive model with multiple screenings for each person. In addition, the dashed red lines are the density of the standard normal distribution. From the plots, it seems like the densities are quite close to a standard normal distribution. Our conclusion is that the large-sample theory provide fully adequate approximations to the relevant distributions.

4.3 Illness-Death Model

In this section, we assume $c_0 = 5$, $a_0 = 0.2$, $c_1 = 2$, $a_1 = 0.2$, $c_{02} = 4$ and $a_{02} = 0.15$. We illustrate in Figure 4.5 a simulation of three different Gamma processes, one from state 0 to state 1 (Z_0), one from state 1 to state 2 (Z_1) and one from state 0 to state 2 (Z_{02}) with these parameter values. In this plot, the individual moves from state 0 to state 1 after around 38 years and from state 1 to state 2 around 10 years after the individual enters state 1. This means that the total time from state 0 to state 2 is around 48 years. However, this individual will go directly to state 2 without going through state 1, since $Z_{02}(t)$ crosses c_{02} before $Z_0(t)$ crosses c_0. This takes around 25 years.

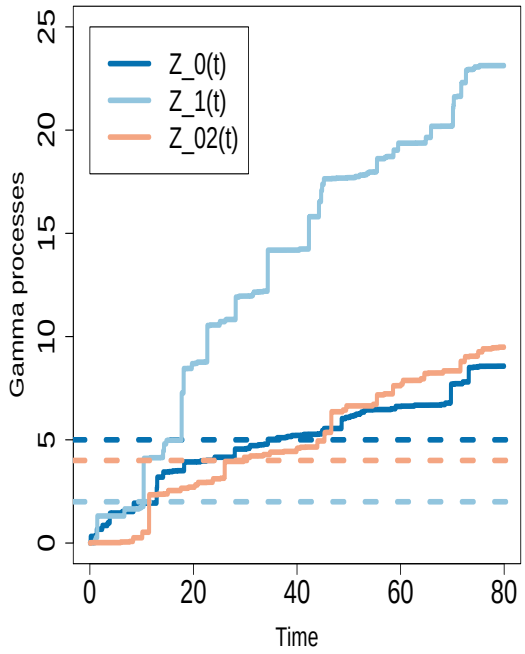


Figure 4.5: Simulation of a Gamma process from state 0 to state 1 (Z_0), a Gamma process from state 1 to state 2 (Z_1) and a Gamma process from state 0 to state 2 (Z_{02})

We only consider multiple screenings, and the individuals are screened between 2 and 15 times. The simulations are done in the same way as in Section 4.2.1, with some modifications. The exact recipe for the simulations is found in Appendix B. We do the simulations 1000 times for 500 patients

4.3.1 Exact Time of Transition to the Absorbing State is Known

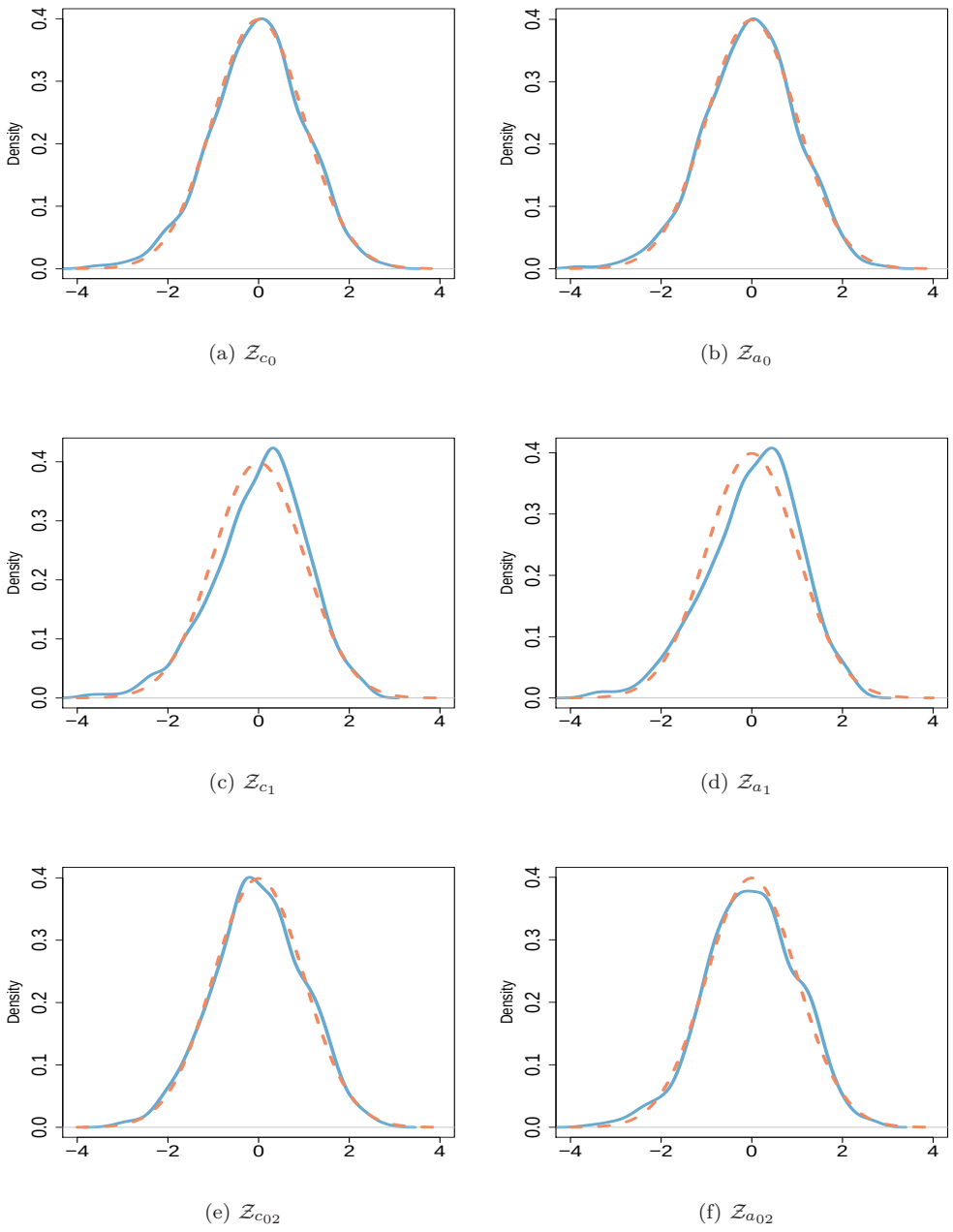
When the exact time of transition to the absorbing state is know, we know that the exact time of death is $T_0 + T_1$ or T_{02}. In Table 4.5, we present the true values and mean of the maximum likelihood estimates in this illness-death

Parameter	**True Value**	**Mean of Estimates**
c_0	5.000	5.049
a_0	0.200	0.202
c_1	2.000	2.119
a_1	0.200	0.210
c_{02}	4.000	4.037
a_{02}	0.150	0.152

Table 4.5: True value and mean of the estimated parameters in an illness-death model for multiple screenings when the exact time of transition to the absorbing state is known.

model when the exact transition to the absorbing state is known. It seems like the mean of the estimates for all of the parameters are close to the true values of the parameters.

Figure 4.6 shows the densities for $\mathcal{Z}_{c_0}$, $\mathcal{Z}_{a_0}$, $\mathcal{Z}_{c_1}$, $\mathcal{Z}_{a_1}$, $\mathcal{Z}_{c_{02}}$ and $\mathcal{Z}_{a_{02}}$. In each plot we include a the dashed red line which is the density of the standard normal distribution. The densities of the $\mathcal{Z}_{\theta_i}$-values seem to correspond well with the standard normal distribution. All of the densities are centered around 0 and the densities follow the standard normal distribution relatively closely. Our conclusion is therefore that the large-sample theory provide fully adequate approximations to the relevant distributions.



(a) $\mathcal{Z}_{c_0}$ (b) $\mathcal{Z}_{a_0}$

(c) $\mathcal{Z}_{c_1}$ (d) $\mathcal{Z}_{a_1}$

(e) $\mathcal{Z}_{c_{02}}$ (f) $\mathcal{Z}_{a_{02}}$

Figure 4.6: Density of $\mathcal{Z}_{c_0}, \mathcal{Z}_{a_0}, \mathcal{Z}_{c_1}, \mathcal{Z}_{a_1}, \mathcal{Z}_{c_{02}}$ and $\mathcal{Z}_{a_{02}}$ for multiple screenings in an illness-death model when the exact time of transition to the absorbing state is known

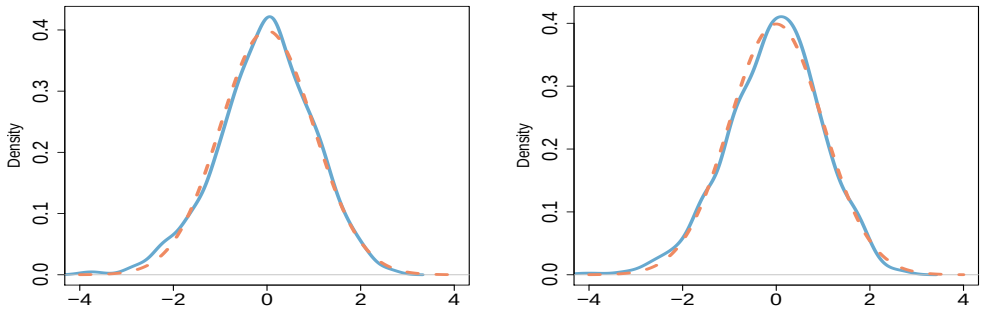
4.3.2 Exact Time of Transition to the Absorbing State is not Known

Parameter	True Value	Mean of Estimates
c_0	5.000	5.055
a_0	0.200	0.202
c_1	2.000	2.136
a_1	0.200	0.211
c_{02}	4.000	4.043
a_{02}	0.150	0.152

Table 4.6: True value and mean of the estimated parameters in an illness-death model for multiple screenings when the exact time of transition to the absorbing state is not known

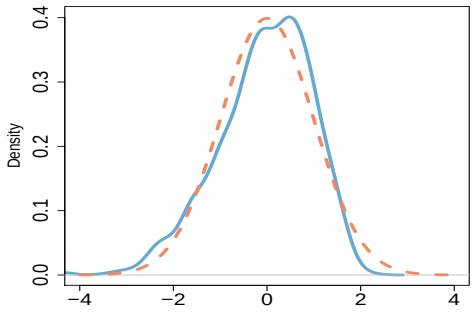
In Table 4.6, we report the true values and the maximum likelihood estimates for the illness-death model when the exact transition to the absorbing state is not known. The estimates are quite close to the true values. The mean of the estimates in Table 4.6 are quite close to the results in Table 4.5, when the exact time of transition to the absorbing state is known.

Figure 4.7 shows the plots of the densities for $\mathcal{Z}_{c_0}$, $\mathcal{Z}_{a_0}$, $\mathcal{Z}_{c_1}$, $\mathcal{Z}_{a_1}$, $\mathcal{Z}_{c_{02}}$ and $\mathcal{Z}_{a_{02}}$. In addition, we include a dashed red line in each plot, which is the density of the standard normal distribution. The densities for the $\mathcal{Z}_{\theta_i}$-values seem to follow the dashed red lines quite closely, which means that the densities correspond well with the standard normal distribution. Our conclusion is therefore that the large-sample theory provide fully adequate approximations to the relevant distributions.

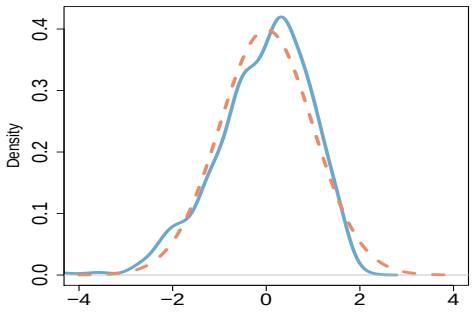


(a) $\mathcal{Z}_{c_0}$

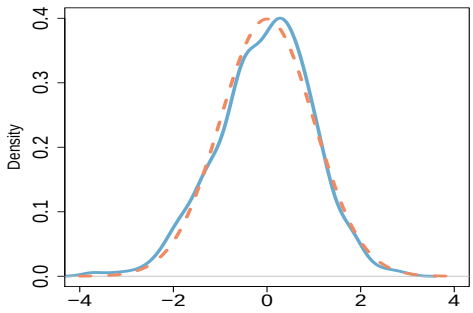
(b) $\mathcal{Z}_{a_0}$



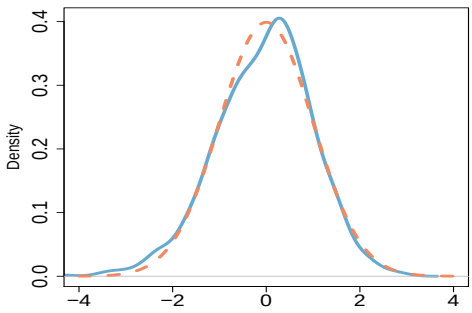
(c) $\mathcal{Z}_{c_1}$



(d) $\mathcal{Z}_{a_1}$



(e) $\mathcal{Z}_{c_{02}}$



(f) $\mathcal{Z}_{a_{02}}$

Figure 4.7: Density of $\mathcal{Z}_{c_0}, \mathcal{Z}_{a_0}, \mathcal{Z}_{c_1}, \mathcal{Z}_{a_1}, \mathcal{Z}_{c_{02}}$ and $\mathcal{Z}_{a_{02}}$ for multiple screenings in an illness-death model when the exact time of transition to the absorbing state is not known

4.4 Four-State Progressive Model

In this section, we assume $c_0 = 5$, $a_0 = 0.2$, $c_1 = 3$, $a_1 = 0.2$, $c_2 = 4$ and $a_2 = 0.1$. We illustrate in Figure 4.8 a simulation of three different Gamma processes, one from state 0 to state 1 (Z_0), one from state 1 to state 2 (Z_1) and one from state 2 to state 3 (Z_2) with these parameter values. In this plot, the individual moves from state 0 to state 1 after around 38 years, from state 1 to state 2 around 10 years after and from state 2 to state 3 after 25 years. This means that the total time from state 0 to state 3 is around 73 years.

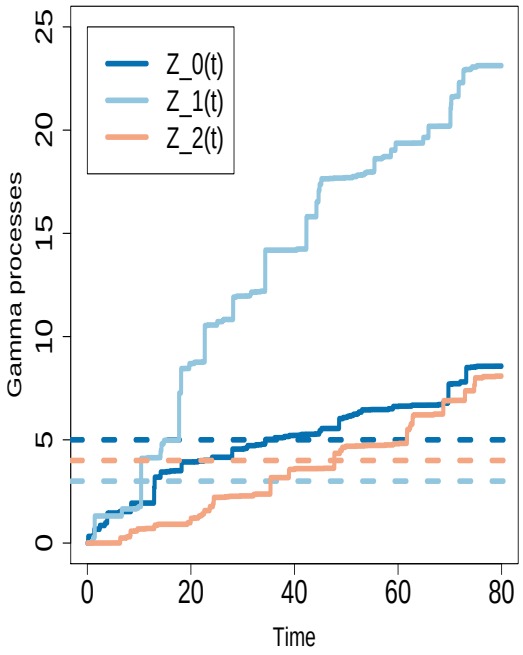


Figure 4.8: Simulation of a Gamma process from state 0 to state 1 (Z_0), a Gamma process from state 1 to state 2 (Z_1) and a Gamma process from state 2 to state 3 (Z_2)

We let the individuals be screened between 2 and 15 times. The simulations are done in the same way as in Section 4.2.1 with modifications. The exact recipe for the simulations is found in Appendix B. In this section, we consider 500 patients, and the simulations are done 100 times.

4.4.1 Exact Time of Transition to the Absorbing State is Known

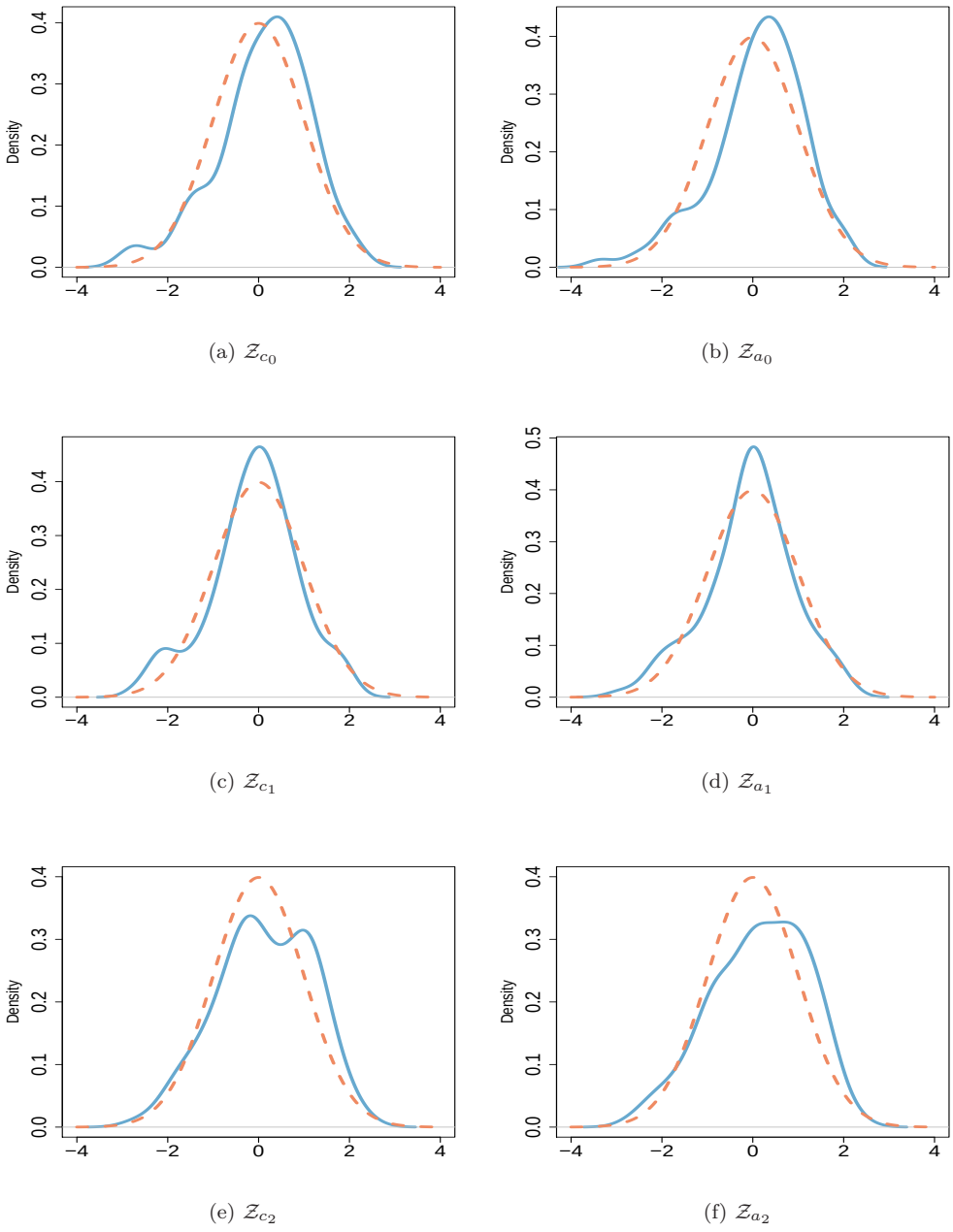
When the exact time of transition to the absorbing state is known, we know that the exact time of death is $T_0 + T_1 + T_2$. In Table 4.7 we present the true values and mean of the maximum likelihood estimates. The mean of the estimates are for most of the parameters close to the true values. The mean of the estimated

Parameter	**True Value**	**Mean of Estimates**
c_0	5.000	5.139
a_0	0.200	0.205
c_1	3.000	3.091
a_1	0.200	0.206
c_2	4.000	4.273
a_2	0.100	0.108

Table 4.7: True value and mean of the estimated parameters in a four-state progressive model for multiple screening when the exact time of transition to the absorbing state is known.

parameter c_2 seems to be a bit further away from its true value than the other parameters. However, it is still quite close to the true value.

We present plots of the densities for $\mathcal{Z}_{c_0}$, $\mathcal{Z}_{a_0}$, $\mathcal{Z}_{c_1}$, $\mathcal{Z}_{a_1}$, $\mathcal{Z}_{c_2}$ and $\mathcal{Z}_{a_2}$ in Figure 4.9. In each plot, we include a dashed red line, which is the density for the standard normal distribution. The densities for $\mathcal{Z}_{c_2}$ and $\mathcal{Z}_{a_2}$ seem to differ a bit from the standard normal distribution around the mean. The reason may be that there are fewer people transitioning from state 2 to state 3 compared to the other states. In addition, the mean of $\mathcal{Z}_{c_0}$ and $\mathcal{Z}_{a_0}$ seem to be a bit skewed to the right, and therefore with a mean above 0, but not by much. In the end, our conclusion is still that the large-sample theory provide fully adequate approximations to the relevant distributions.



(a) $\mathcal{Z}_{c_0}$ (b) $\mathcal{Z}_{a_0}$

(c) $\mathcal{Z}_{c_1}$ (d) $\mathcal{Z}_{a_1}$

(e) $\mathcal{Z}_{c_2}$ (f) $\mathcal{Z}_{a_2}$

Figure 4.9: Densities for $\mathcal{Z}_{c_0}$, $\mathcal{Z}_{a_0}$, $\mathcal{Z}_{c_1}$, $\mathcal{Z}_{a_1}$, $\mathcal{Z}_{c_2}$ and $\mathcal{Z}_{a_2}$ for multiple screenings in a four-state progressive model when the exact time of transition to the absorbing state is known

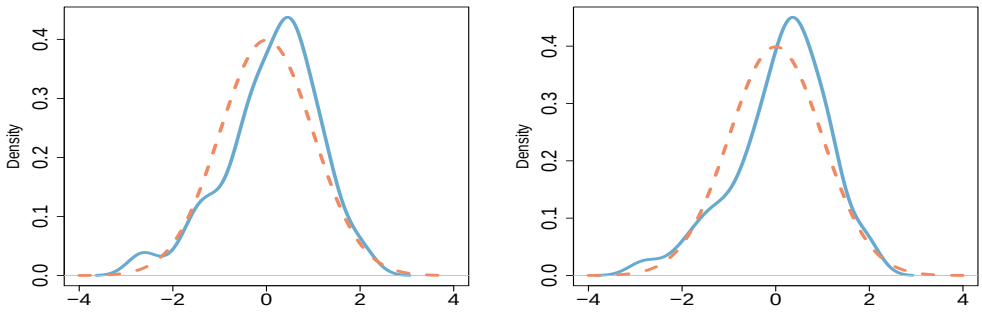
4.4.2 Exact Time of Transition to the Absorbing State is not Known

Parameter	True Value	Mean of Estimates
c_0	5.000	5.140
a_0	0.200	0.205
c_1	3.000	3.095
a_1	0.200	0.206
c_2	4.000	4.340
a_2	0.100	0.110

Table 4.8: True value and mean of the estimated parameters in a four-state progressive model for multiple screenings when the exact time of transition to the absorbing state is not known

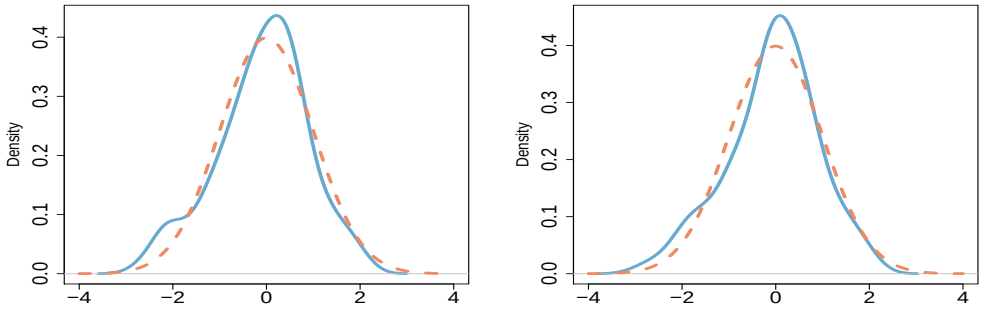
We present the true values and mean of the maximum likelihood estimates in Table 4.8. The mean of the estimated parameters are often close to the true values. However, the mean of the estimated parameter c_2 seems to be a bit further away from its true value. As we explained previously, this may have something to do with fewer people transitioning from state 2 to state 3. The mean of the estimated values are a bit further away from the true value compared to when the exact transition to the absorbing state is known.

In Figure 4.10, we present plots of the densities for $\mathcal{Z}_{c_0}$, $\mathcal{Z}_{a_0}$, $\mathcal{Z}_{c_1}$, $\mathcal{Z}_{a_1}$, $\mathcal{Z}_{c_2}$ and $\mathcal{Z}_{a_2}$. In each plot, we include a red dashed line which is the density of the standard normal distribution. The densities for $\mathcal{Z}_{c_2}$ and $\mathcal{Z}_{a_2}$ seem to differ a bit from the standard normal distribution around the mean. For these transitions, the densities are a bit wider and lower than for the standard normal distribution. This means the variances are a bit higher than 1, and this is confirmed by calculating the variances. We get that the variance for $\mathcal{Z}_{c_2}$ is 1.322 and the variance for $\mathcal{Z}_{a_2}$ is 1.404. This may have something to do with fewer people transitioning from state 2 to state 3 compared to the transitions between the other states. Since we have few transitions, we may get that the spread of the $\mathcal{Z}_{\theta_i}$-values is too big. Even though there are some differences, they are not very big and our conclusion is still that the large-sample theory provide fully adequate approximations to the relevant distributions.



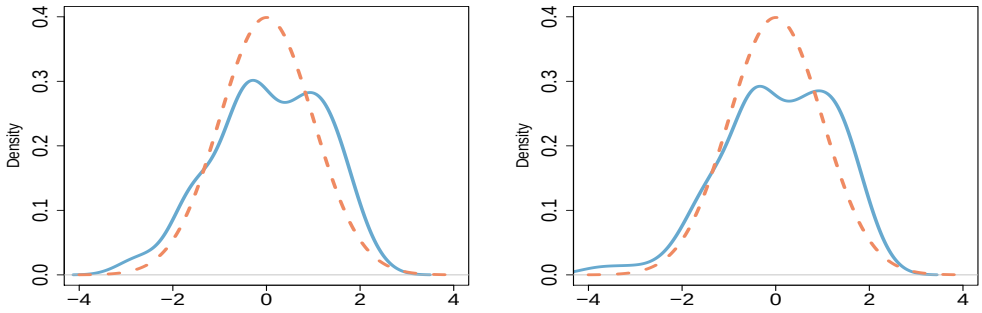
(a) $\mathcal{Z}_{c_0}$

(b) $\mathcal{Z}_{a_0}$



(c) $\mathcal{Z}_{c_1}$

(d) $\mathcal{Z}_{a_1}$



(e) $\mathcal{Z}_{c_2}$

(f) $\mathcal{Z}_{a_2}$

Figure 4.10: Density of $\mathcal{Z}_{c_0}, \mathcal{Z}_{a_0}, \mathcal{Z}_{c_1}, \mathcal{Z}_{a_1}, \mathcal{Z}_{c_2}$ and $\mathcal{Z}_{a_2}$ for multiple screenings when the exact time of transition to the absorbing state is not known

4.5 Four-State Illness-Death Model

In this section, we assume $c_0 = 5$, $a_0 = 0.2$, $c_1 = 3$, $a_1 = 0.2$, $c_2 = 4$, $a_2 = 0.1$, $c_{03} = 6$, $a_{03} = 0.15$, $c_{13} = 4$ and $a_{13} = 0.15$. We illustrate in Figure 4.11 five different Gamma processes, one from state 0 to state 1 (Z_0), one from state 1 to state 2 (Z_1), one from state 2 to state 3 (Z_2), one from state 0 to state 3 (Z_{03}) and one from state 1 to state 3 (Z_{13}) with these parameter values. The transition time from state 0 to state 1 is around 50 years, from state 1 to state 2 around 25 years, from state 2 to state 3 around 55 years, from state 0 to state 1 around 40 years and from state 1 to state 3 around 35 years.

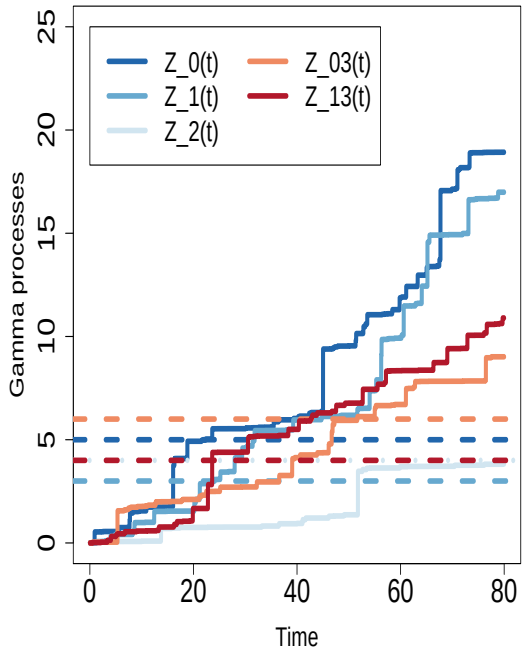


Figure 4.11: Simulation of a Gamma process from state 0 to state 1 (Z_0), a Gamma process from state 1 to state 2 (Z_1), a Gamma process from state 2 to state 3 (Z_2), a Gamma process from state 0 to state 3 (Z_{03}) and a Gamma process from state 1 to state 3 (Z_{13})

We let the individuals be screened between 2 and 15 times. The simulations are done in the same way as in Section 4.2.1 with modifications, where the exact recipe for the simulations is found in Appendix B and the code is found in Appendix D. We do the simulations 100 times with 500 patients.

4.5.1 Exact Time of Transition to the Absorbing State is Known

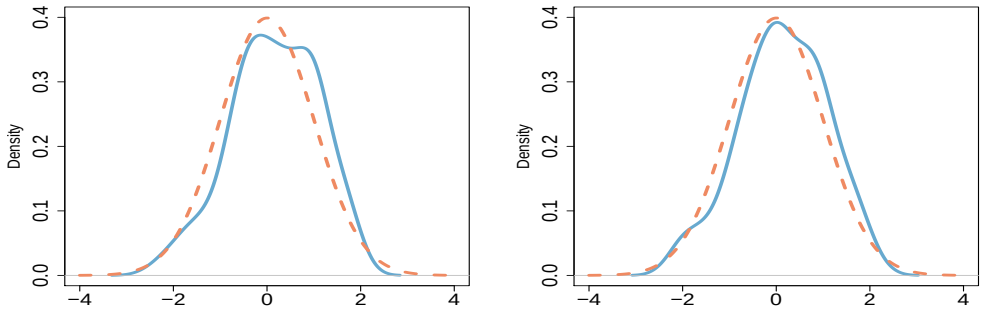
When the exact time of transition to the absorbing state is known, we know that the exact time of death is either $T_0 + T_1 + T_2$, T_{03} or $T_0 + T_{13}$.

Parameter	**True Value**	**Mean of Estimates**
c_0	5.000	5.175
a_0	0.200	0.206
c_1	3.000	3.140
a_1	0.200	0.209
c_2	4.000	4.147
a_2	0.100	0.103
c_{03}	6.000	6.195
a_{03}	0.150	0.155
c_{13}	4.000	4.186
a_{13}	0.150	0.159

Table 4.9: True value and mean of the estimated parameters in a four-state illness-death model for multiple screenings when the exact time of transition to the absorbing state is known

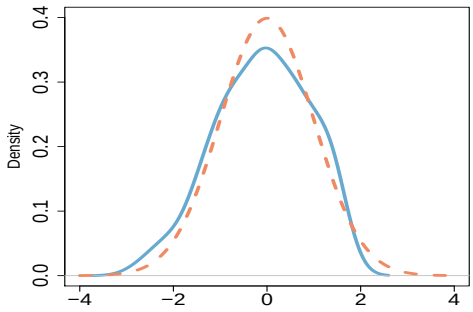
We present the results for the true value of the parameters and the mean of the maximum likelihood estimates in Table 4.9. Mostly, the mean of the estimated parameters are close to their corresponding true value. However, they are a bit further away compared to the models with fewer states and transitions, for example the three-state progressive model. One reason is that there are fewer people making the different transitions when we have more states and possible transitions.

In Figure 4.12, we present plots of the densities for $\mathcal{Z}_{c_0}$, $\mathcal{Z}_{a_0}$, $\mathcal{Z}_{c_1}$, $\mathcal{Z}_{a_1}$, $\mathcal{Z}_{c_2}$, $\mathcal{Z}_{a_2}$, $\mathcal{Z}_{c_{03}}$, $\mathcal{Z}_{a_{03}}$, $\mathcal{Z}_{c_{13}}$ and $\mathcal{Z}_{a_{13}}$. In each plot there is also a red dashed line, which is the density for the standard normal distribution. In Figure 4.12 (a), (b), (c), (d), (e) and (f), the densities are relatively close to the true standard normal distribution. (g) and (h) also seem to be ok, but a bit skewed to the right. (i) and (j) seem to be very high around 0 and have a too low variance, which is confirmed when we calculate the variances. The variance for $\mathcal{Z}_{c_{13}}$ is 0.625 and the variance for $\mathcal{Z}_{a_{13}}$ is 0.679. This will probably be improved if we include more individuals than 500. Even though some of the densities are a bit skewed or give a too low or high variance, and therefore do not fit the standard normal distribution perfectly, they are still quite close to the standard normal distribution. Our conclusion is still that the large-sample theory provide fully adequate approximations to the relevant distributions.

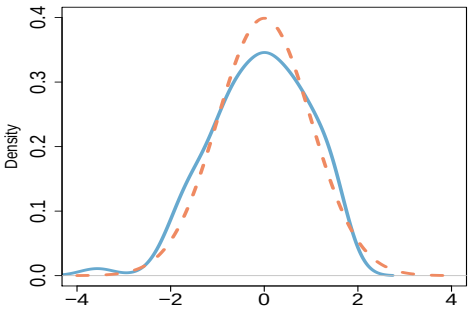


(a) $\mathcal{Z}_{c_0}$

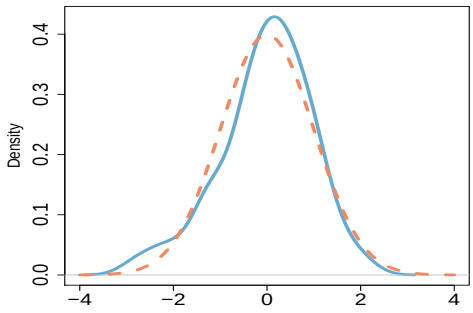
(b) $\mathcal{Z}_{a_0}$



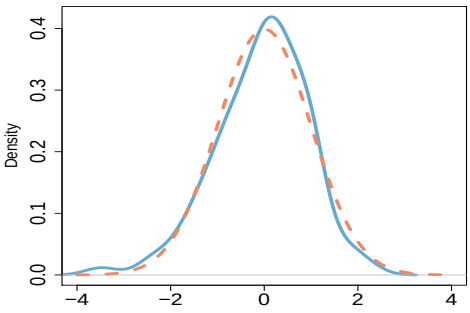
(c) $\mathcal{Z}_{c_1}$



(d) $\mathcal{Z}_{a_1}$

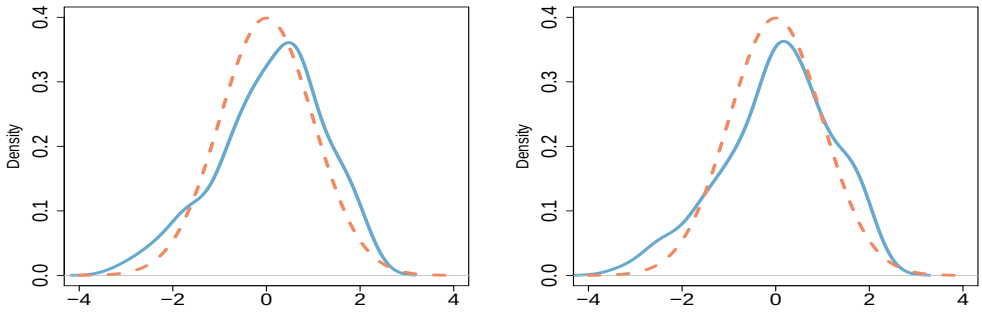


(e) $\mathcal{Z}_{c_2}$

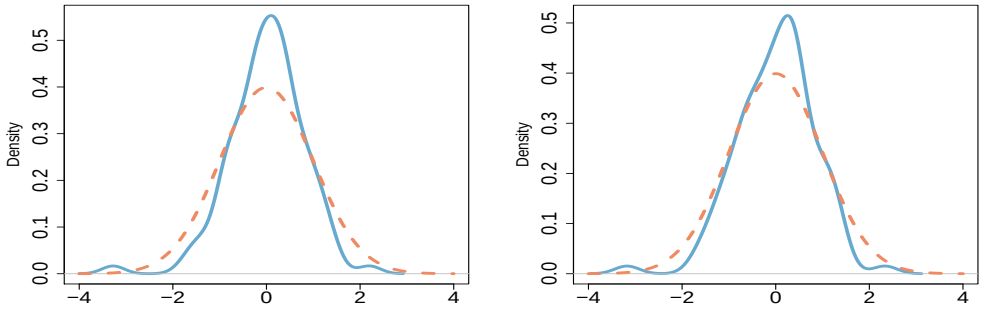


(f) t_{a_2}

Figure 4.12: Density of $\mathcal{Z}_{c_0}, \mathcal{Z}_{a_0}, \mathcal{Z}_{c_1}, \mathcal{Z}_{a_1}, \mathcal{Z}_{c_2}, \mathcal{Z}_{a_2}, \mathcal{Z}_{c_{03}}, \mathcal{Z}_{a_{03}}, \mathcal{Z}_{c_{13}}$ and $\mathcal{Z}_{a_{13}}$ for multiple screenings in a four-state illness-death model when the exact time of transition to the absorbing state is known



(g) $\mathcal{Z}_{c_{03}}$ (h) $\mathcal{Z}_{a_{03}}$



(i) $\mathcal{Z}_{c_{13}}$ (j) $\mathcal{Z}_{a_{13}}$

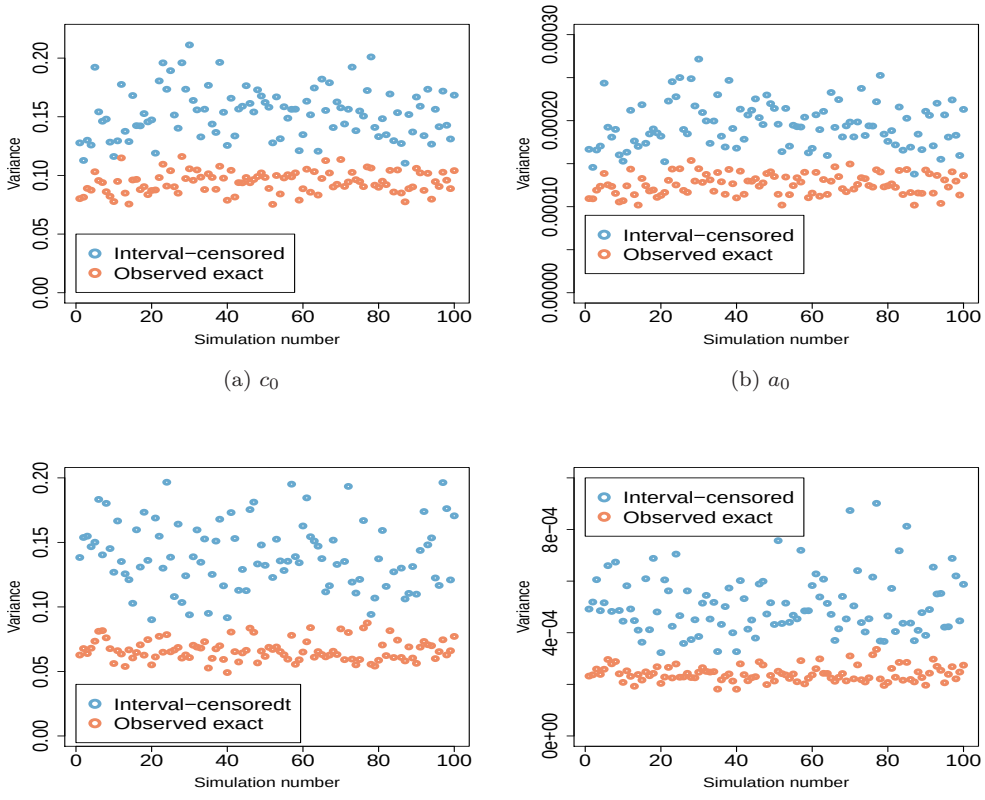
Figure 4.12: Density of $\mathcal{Z}_{c_0}, \mathcal{Z}_{a_0}, \mathcal{Z}_{c_1}, \mathcal{Z}_{a_1}, \mathcal{Z}_{c_2}, \mathcal{Z}_{a_2}, \mathcal{Z}_{c_{03}}, \mathcal{Z}_{a_{03}}, \mathcal{Z}_{c_{13}}$ and $\mathcal{Z}_{a_{13}}$ for multiple screenings in a four-state illness-death model when the exact time of transition to the absorbing state is known

4.6 How Much Information is Lost from not Observing the Transition Times Exact?

In this thesis, we focus on interval-censored data. However, one could consider a situation where the transition times are observed exactly. This means we observe T_0 and T_1 in a three-state progressive model. We construct the likelihood when T_0 and T_1 is observed exactly, by dividing it into likelihood contributions. Type 1 is when the individual is only observed in state 0. Type 2 is when the individual is observed in state 0 and state 1. Type 3 is when the individual is observed in state 0, state 1 and state 2. The likelihood then becomes

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{(I)} S_0(t_{n,p}, \boldsymbol{\theta}|x_p) \prod_{(II)} f_0(T_{0,p}, \boldsymbol{\theta}|x_p) S_1(t_{n,p} - T_{0,p}, \boldsymbol{\theta}|x_p)$$
$$\prod_{(III)} f_0(T_{0,p}, \boldsymbol{\theta}|x_p) f_1(T_{1,p}, \boldsymbol{\theta}|x_p).$$

A relevant question is how much information is lost when the data is observed exactly compared to when it is interval-censored? In order to investigate this question in a three-state progressive model, we look at the variances of the estimated parameters, which is the diagonal of the inverse Hessian matrix.



(a) c_0 (b) a_0

(c) c_1 (d) a_1

Figure 4.13: Variances for c_0, a_0, c_1 and a_1 in a three-state progressive model when T_0 and T_1 are interval-censored or observed exactly

For each simulation, we present the variances for c_0, a_0, c_1 and a_1 in a three-state progressive model in Figure 4.13. The blue dots is when only the transition to the absorbing state is observed exactly, which means T_0 is interval-censored, while the red dots is when both of the the transition times are observed exactly.

There is a clear difference between the red dots and the blue dots. For example in Figure 4.13 (c), the variance when the transition times are not observed exactly is in many cases almost twice as big as when the transition times are observed exactly. We also note that the spread of the variances when the transition times are observed exactly is smaller than when the transition times are not observed exactly. In conclusion, we loose quite a lot of information when the transition times are interval-censored compared to when they are observed exactly. This is just as we expect.

CHAPTER 5

The Markov Property

In Chapter 1 and Chapter 2, we thoroughly discussed the Markov property. As we presented earlier, the Markov property means that given the present state and history of an individual, the transition to the next state and the time this occurs, only depends on the present state (Putter et al., 2007).

In this chapter, we investigate whether the Markov property holds in our models. We consider the Markov property in a four-state illness-death model where the individuals start in state 0 at time 0. In the beginning, we do not assume a specific form of the survival time model for the transition time. Then we investigate whether the Markov property holds for Gamma process model or when the transition times are exponentially distributed. Then in Chapter 5.2, we study the relationship between the exponential distribution model and a homogeneous Markov model.

5.1 The Markov Property in a Four-State Illness-Death Model

Our aim in this section is to prove Lemma 5.1.1.

Lemma 5.1.1. *In the proposed model framework for a four-state illness-death model, when the transition times are exponentially distributed, the Markov property is fulfilled. When the transition times are modeled as the threshold crossing times for Gamma processes, the Markov property does not hold.*

We investigate the Markov property in a four-state illness-death model when the exact time of transition to the absorbing state is known and the individuals start in state 0 at time 0. For each possible transition, we compute two different transition probabilities. In the first version of the transition probability, we condition on the information about the previous state. This means the transition probability is of the form $P_{m\ell}(a,b) = Pr(X_a = \ell | X_b = m)$. In the second version of the transition probability, we condition on the whole state history of the individual. This means $P_{m\ell}(a,b) = Pr(X_a = \ell | X_b = m, X_u, u \in [0,b))$. In the end, we compare these transition probabilities. If they are equal in all of the possible transitions, then the Markov property is fulfilled. Early on, we detect that the Markov property is not necessarily fulfilled in our models. We still calculate the rest of the transition probabilities because we use these probabilities later on, for example in Section 5.1.1.

Assume an individual is in state 0 at time point v, in state 1 at time point t, state 2 at time point r and in state 3 at time point q. $P^M_{m\ell}$ means the transition probability from state m to ℓ, when we only condition on the information in the previous state. For example, with the specified time points above, $P^M_{23} = Pr(X_q = 3|X_r = 2)$. $P^G_{m\ell}$ means that the transition probability from state m to ℓ at certain time points associated with the visited states when we condition on the whole state history of the individual. For example, with the specified time points above, $P^G_{23} = Pr(X_q = 3|X_r = 2, X_t = 1, X_v = 0)$.

Transition from 0 to 1

These transition probabilities are always equal, since we only condition on the individual being in state 0 at time point v.

$$\begin{aligned} P^M_{01} &= Pr(X_t = 1|X_v = 0) = \frac{Pr(X_t = 1, X_v = 0)}{Pr(X_v = 0)} \\ &= \frac{Pr(v < T_0 < t, T_0 + T_1 > t, T_{03} > T_0, T_0 + T_{13} > t)}{Pr(T_0 > v, T_{03} > v)} \\ &= \frac{\int_v^t f_0(s)S_1(t-s)S_{03}(s)S_{13}(t-s)ds}{S_0(v)S_{03}(v)} \\ &= P^G_{01}. \end{aligned}$$

Transition from 0 to 2

These transition probabilities are always equal, since we only condition on the individual being in state 0 at time point v.

$$\begin{aligned} P^M_{02} &= Pr(X_r = 2|X_v = 0) = \frac{Pr(X_r = 2, X_v = 0)}{Pr(X_v = 0)} \\ &= \frac{Pr(v < T_0 < r, v < T_0 + T_1 < r, T_0 + T_1 + T_2 > r, T_{03} > T_0, T_{13} > T_1)}{Pr(T_0 > v, T_{03} > v)} \\ &= \frac{\int_v^r \int_0^{r-s} f_0(s)f_1(u)S_2(r-s-u)S_{03}(s)S_{13}(u)duds}{S_0(v)S_{03}(v)} \\ &= P^G_{02}. \end{aligned}$$

Transition from 1 to 2

We first consider the transition probability when we only condition on the individual being in state 1 at time point t.

$$P^M_{12} = Pr(X_r = 2|X_t = 1) = \frac{Pr(X_r = 2, X_t = 1)}{Pr(X_t = 1)}$$

$$\begin{aligned} &= \frac{Pr(0 < T_0 < t, t < T_0 + T_1 < r, T_0 + T_1 + T_2 > r, T_{03} > T_0, T_{13} > T_1)}{Pr(0 < T_0 < t, T_0 + T_1 > t, T_{03} > T_0, T_0 + T_{13} > t)} \\ &= \frac{\int_0^t \int_{t-s}^{r-s} f_0(s)f_1(u)S_2(r-s-u)S_{03}(s)S_{13}(u)duds}{\int_0^t f_0(s)S_1(t-s)S_{03}(s)S_{13}(t-s)ds}. \end{aligned}$$

Then we consider the transition probability when we also condition on the individual being in state 0 at time point v

$$\begin{aligned}P_{12}^G &= Pr(X_r = 2|X_t = 1, X_v = 0) = \frac{Pr(X_r = 2, X_t = 1, X_v = 0)}{Pr(X_t = 1, X_v = 0)} \\ &= \frac{Pr(v < T_0 < t, t < T_0 + T_1 < r, T_0 + T_1 + T_2 > r, T_{03} > T_0, T_{13} > T_1)}{Pr(v < T_0 < t, T_0 + T_1 > t, T_{03} > T_0, T_0 + T_{13} > t)} \\ &= \frac{\int_v^t \int_{t-s}^{r-s} f_0(s) f_1(u) S_2(r-s-u) S_{03}(s) S_{13}(u) duds}{\int_v^t f_0(s) S_1(t-s) S_{03}(s) S_{13}(t-s) ds}.\end{aligned}$$

P_{12}^M is not necessarily equal to P_{12}^G. The difference lies in the integrals. In the case where we only condition on the time in state 1, the numerator is $\int_0^t \int_{t-s}^{r-s} f_0(s) f_1(u) S_2(r-s-u) S_{03}(s) S_{13}(u) duds$. When we also condition on the time the individual was in state 0, the numerator is $\int_v^t \int_{t-s}^{r-s} f_0(s) f_1(u) S_2(r-s-u) S_{03}(s) S_{13}(u) duds$. When we condition on the individual being in state 1 at time point t, the first integral has a lower limit of 0 and upper limit of t. However, when we also use the information that the individual was in state 0 at time v, the integral has a lower limit of v and an upper limit of t. This also happens in the denominator.

Transition from 2 to 3

We start by looking at the transition probability when we only condition on the individual being in state 2 at time point r.

$$\begin{aligned}P_{23}^M &= Pr(X_q = 3|X_r = 2) = \frac{Pr(X_q = 3, X_r = 2)}{Pr(X_r = 2)} \\ &= \frac{Pr(0 < T_0 < r, 0 < T_0 + T_1 < r, q < T_0 + T_1 + T_2 < q + \epsilon, T_{03} > T_0, T_{13} > T_1)}{Pr(0 < T_0 < r, 0 < T_0 + T_1 < r, T_0 + T_1 + T_2 > r, T_{03} > T_0, T_{13} > T_1)} \\ &= \frac{\int_0^r \int_0^{r-s} f_0(s) f_1(u) f_2(q-s-u) S_{03}(s) S_{13}(u) duds}{\int_0^r \int_0^{r-s} f_0(s) f_1(u) S_2(r-s-u) S_{03}(s) S_{13}(u) duds}.\end{aligned}$$

Then we consider the transition probability when we also condition on the individual being in state 1 at time point t and state 0 at time point v.

$$\begin{aligned}P_{23}^G &= Pr(X_q = 3|X_r = 2, X_t = 1, X_v = 0) = \frac{Pr(X_q = 3, X_r = 2, X_t = 1, X_v = 0)}{Pr(X_r = 2, X_t = 1, X_v = 0)} \\ &= \frac{Pr(v < T_0 < t, t < T_0 + T_1 < r, q < T_0 + T_1 + T_2 < q + \epsilon, T_{03} > T_0, T_{13} > T_1)}{Pr(v < T_0 < t, t < T_0 + T_1 < r, T_0 + T_1 + T_2 > r, T_{03} > T_0, T_{13} > T_1)} \\ &= \frac{\int_v^t \int_{t-s}^{r-s} f_0(s) f_1(u) f_2(q-s-u) S_{03}(s) S_{13}(u) duds}{\int_v^t \int_{t-s}^{r-s} f_0(s) f_1(u) S_2(r-u-s) S_{03}(s) S_{13}(u) duds}.\end{aligned}$$

P_{23}^M is not necessarily equal to P_{23}^G. The difference lies in whether or not we include the time points where the individual left state 0 and state 1.

Transition from 0 to 3

If an individual transitions from state 0 to state 3, then the person can go directly from state 0 to state 3 or from state 0 to state 1 to state 3 or from

state 0 to state 1 to state 2 to state 3. Then

$$\begin{aligned}
P_{03}^M &= Pr(X_q = 3|X_v = 0) = P_{03,0} + P_{03,1} + P_{03,2} \\
&= \frac{Pr(q < T_{03} < q + \epsilon, T_0 > q, T_{03} > v, T_0 > v)}{Pr(T_0 > v, T_{03} > v)} \\
&+ \frac{Pr(v < T_0 < q, q < T_0 + T_{13} < q + \epsilon, T_{03} > T_0, T_0 + T_1 > q)}{Pr(T_0 > v, T_{03} > v)} \\
&+ \frac{Pr(v < T_0 < q, q < T_0 + T_1 + T_2 < q + \epsilon, T_{03} > T_0, T_{13} > T_1)}{Pr(T_0 > v, T_{03} > v)} \\
&= P_{03}^M.
\end{aligned}$$

We start by looking at the setting where we go directly from state 0 to state 3. The two transition probabilities are always equal, since we only condition on the individual being in state 0 at time point v. So

$$\begin{aligned}
P_{03,0}^M &= \frac{Pr(q < T_{03} < q + \epsilon, T_0 > q, T_{03} > v, T_0 > v)}{Pr(T_0 > v, T_{03} > v)} \\
&= \frac{f_{03}(q)S_0(q)S_{03}(v)S_0(v)}{S_0(v)S_{03}(v)} = f_{03}(q)S_0(q).
\end{aligned}$$

In the next step, we consider the case where the individual goes from state 0 to state 1 and then directly to state 3. The two transition probabilities are also equal here, since we only condition on the individual being in state 0 at time point v. So

$$\begin{aligned}
P_{03,1}^M &= \frac{Pr(v < T_0 < q, q < T_0 + T_{13} < q + \epsilon, T_{03} > T_0, T_0 + T_1 > q)}{Pr(T_0 > v, T_{03} > v)} \\
&= \frac{\int_v^q f_0(s)S_{03}(s)f_{13}(q-s)S_1(q-s)ds}{S_0(v)S_{03}(v)}.
\end{aligned}$$

Lastly, we consider the case when the individual goes from state 0 to state 1 to state 2 and then to state 3. These transition probabilities are always equal, since we only condition on the individual being in state 0 at time point v.

$$\begin{aligned}
P_{03,2}^M &= \frac{Pr(v < T_0 < q, q < T_0 + T_1 + T_2 < q + \epsilon, T_{03} > T_0, T_{13} > T_1)}{Pr(T_0 > v, T_{03} > v)} \\
&= \frac{\int_v^q \int_0^{q-s} f_0(s)f_1(u)f_2(q-u-s)S_{03}(s)S_{13}(u)duds}{S_0(v)S_{03}(v)}.
\end{aligned}$$

Transition from 1 to 3

We now consider when an individual transfers from state 1 to state 3. Either the individual goes directly from state 1 to state 3 or the individual goes from state 1 to state 2 to state 3. Then

$$P_{13}^M = P_{13,0}^M + P_{13,1}^M,$$

and

$$P_{13}^G = P_{13,0}^G + P_{13,1}^G.$$

We start by looking at the probability of going directly from state 1 to state 3. The transition probability when we only condition on the individual being in state 1 at time point t becomes

$$\begin{aligned} P^M_{13,0} &= Pr(X_q = 3 | X_t = 1) = \frac{Pr(X_q = 3, X_t = 1)}{Pr(X_t = 1)} \\ &= \frac{Pr(0 < T_0 < t, q < T_0 + T_{13} < q + \epsilon, T_{03} > T_0, T_0 + T_1 > q)}{Pr(0 < T_0 < t, T_0 + T_1 > t, T_0 + T_{13} > t, T_{03} > T_0)} \\ &= \frac{\int_0^t f_0(s) f_{13}(q-s) S_{03}(s) S_1(q-s) ds}{\int_0^t f_0(s) S_1(t-s) S_{13}(t-s) S_{03}(s) ds}. \end{aligned}$$

Then we consider the transition probability when we also condition on the individual being in state 0 at time point v.

$$\begin{aligned} P^G_{13,0} &= Pr(X_q = 3 | X_t = 1, X_v = 0) = \frac{Pr(X_q = 3, X_t = 1, X_v = 0)}{Pr(X_t = 1, X_v = 0)} \\ &= \frac{Pr(v < T_0 < t, q < T_0 + T_{13} < q + \epsilon, T_{03} > T_0, T_0 + T_1 > q)}{Pr(0 < T_0 < t, T_0 + T_1 > t, T_0 + T_{13} > t, T_{03} > T_0)} \\ &= \frac{\int_v^t f_0(s) f_{13}(q-s) S_{03}(s) S_1(q-s) ds}{\int_v^t f_0(s) S_1(t-s) S_{13}(t-s) S_{03}(s) ds}. \end{aligned}$$

In this step, we consider the case when the individual goes through state 2. The transition probability when we only condition on the individual being in state 1 at time point t becomes

$$\begin{aligned} P^M_{13,1} &= Pr(X_q = 3 | X_t = 1) = \frac{Pr(X_q = 3, X_t = 1)}{Pr(X_t = 1)} \\ &= \frac{Pr(0 < T_0 < t, t < T_0 + T_1 < q, q < T_0 + T_1 + T_2 < q + \epsilon, T_{03} > T_0, T_{13} > T_1)}{Pr(0 < T_0 < t, T_0 + T_1 > t, T_0 + T_{13} > t, T_{03} > T_0)} \\ &= \frac{\int_0^t \int_{t-s}^{q-s} f_0(s) f_1(u) f_2(q-s-u) S_{03}(s) S_{13}(u) du ds}{\int_0^t f_0(s) S_1(t-s) S_{13}(t-s) S_{03}(s) ds}. \end{aligned}$$

Then we consider the transition probability when we also condition on the individual being in state 0 at time point v.

$$\begin{aligned} P^G_{13,1} &= Pr(X_q = 3 | X_t = 1, X_v = 0) = \frac{Pr(X_q = 3, X_t = 1, X_v = 1)}{Pr(X_t = 1, X_v = 0)} \\ &= \frac{Pr(v < T_0 < t, t < T_0 + T_1 < q, q < T_0 + T_1 + T_2 < q + \epsilon, T_{03} > T_0, T_{13} > T_1)}{Pr(v < T_0 < t, T_0 + T_1 > t, T_{03} > T_0, T_0 + T_{13} > t)} \\ &= \frac{\int_v^t \int_{t-s}^{v-s} f_0(s) f_1(u) f_2(q-s-u) S_{03}(s) S_{13}(u) du ds}{\int_v^t f_0(s) S_1(t-s) S_{13}(t-s) S_{03}(s) ds}. \end{aligned}$$

P^M_{13} is not necessarily equal to P^G_{13}, for the same reasons as we previously have described.

Staying in state 0

We assume that a person stays in state 0 from time point 0 to v. The transition probabilities are always equal, since we only condition on the individual being

in state 0 at time $v-1$.

$$P_{00}^M = Pr(X_v = 0|X_{v-1} = 0) = \frac{Pr(T_0 > v, T_{03} > v, T_0 > v-1, T_{03} > v-1)}{Pr(T_0 > v-1, T_{03} > v-1)}$$
$$= S_0(v)S_{03}(v) = P_{00}^G.$$

Staying in state 1

We assume that a person stays in state 0 from time point 0 to v and in state 1 in time points k and t. The transition probability when we only condition on the individual being in state 1 at time point k becomes

$$P_{11}^M = Pr(X_t = 1|X_k = 1) = \frac{Pr(X_t = 1, X_k = 1)}{Pr(X_k = 1)}$$
$$= \frac{Pr(T_0 < k, T_0 + T_1 > t, T_{03} > T_0, T_0 + T_{13} > t)}{Pr(T_0 < k, T_0 + T_1 > t, T_{03} > T_0, T_0 + T_{13} > k)}$$
$$= \frac{\int_0^k f_0(s)S_1(t-s)S_{03}(s)S_{13}(t-s)ds}{\int_0^k f_0(s)S_1(k-s)S_{03}(s)S_{13}(k-s)ds}.$$

Then the transition probability when we also condition on the individual being in state 0 at time point v becomes

$$P_{11}^G = Pr(X_t = 1|X_k = 1, X_v = 0) = \frac{Pr(X_t = 1, X_k = 1, X_v = 9)}{Pr(X_k = 1, X_v = 9)}$$
$$= \frac{Pr(v < T_0 < k, T_{03} > T_0, T_0 + T_1 > t, T_0 + T_{13} > t)}{Pr(v < T_0 < k, T_{03} > T_0, T_0 + T_1 > k, T_0 + T_{13} > k)}$$
$$= \frac{\int_v^k f_0(s)S_{03}(s)S_1(t-s)S_{13}(t-s)ds}{\int_v^k f_0(s)S_{03}(s)S_1(k-s)S_{13}(k-s)ds}.$$

P_{11}^M is not necessarily equal to P_{11}^G. This comes from the same reason as we previously have described.

Staying in state 2

We assume that a person stays in state 0 from time point 0 to v, in state 1 at time points k and t and in state 2 at the time points w and r. The transition probability when we only condition on the individual being in state 2 at time point w becomes

$$P_{22}^M = Pr(X_r = 2|X_w = 2) = \frac{Pr(X_r = 2, X_w = 2)}{Pr(X_w = 2)}$$
$$= \frac{Pr(0 < T_0 < w, 0 < T_0 + T_1 < w, T_0 + T_1 + T_2 > r, T_{03} > T_0, T_{13} > T_1)}{Pr(0 < T_0 < w, 0 < T_0 + T_1 < w, T_0 + T_1 + T_2 > w, T_{03} > T_0, T_{13} > T_1)}$$
$$= \frac{\int_0^w \int_0^{w-s} f_0(s)f_1(u)S_2(r-s-u)S_{03}(s)S_{13}(u)duds}{\int_0^w \int_0^{w-s} f_0(s)f_1(u)S_2(w-s-u)S_{03}(s)S_{13}(u)duds}.$$

Then the transition probability when we also condition on the individual being in state 0 at time point v and state 1 at time points k and t becomes

$$\begin{aligned}
P_{22}^G &= Pr(X_r = 2 | X_w = 2, X_v = 0, X_k = 1, X_t = 1) \\
&= \frac{Pr(X_r = 2, X_w = 2, X_v = 0, X_k = 1, X_t = 1)}{Pr(X_w = 2, X_v = 0, X_k = 1, X_t = 1)} \\
&= \frac{Pr(v < T_0 < k, t < T_0 + T_1 < w, T_0 + T_1 + T_2 > r, T_{03} > T_0, T_{13} > T_1)}{Pr(v < T_0 < k, t < T_0 + T_1 < w, T_0 + T_1 + T_2 > w, T_{03} > T_0, T_{13} > T_1)} \\
&= \frac{\int_v^k \int_{t-s}^{w-s} f_0(s) f_1(u) S_2(r - s - u) S_{03}(s) S_{13}(u) du ds}{\int_v^k \int_{t-s}^{w-s} f_0(s) f_1(u) S_2(w - s - u) S_{03}(s) S_{13}(u) du ds}.
\end{aligned}$$

P_{22}^M is not necessarily equal to P_{22}^G. This comes from the same reason as we previously have described.

5.1.1 Example: Exponential Distribution

We want to investigate whether the Markov property is fulfilled in our general model construction when having different distributional assumptions on the transition times. In order to do so, we use the formulas derived in Section 5.1 and check whether the transition probabilities, $P_{m\ell}$, when we only condition on the information about the previous state are equal to the transition probabilities when we condition on the whole state history. In our first example, we consider exponentially distributed transition times.

Transition from 1 to 2

$$\begin{aligned}
P_{12}^M &= Pr(X_r = 2 | X_t = 1) \\
&= \frac{a_0 a_1 \exp(-a_2 r) \int_0^t \exp(s(-a_0 + a_2 - a_{03})) \int_{t-s}^{r-s} \exp(u(-a_1 + a_2 - a_{13})) du ds}{a_0 \exp(t(-a_1 - a_{13})) \int_0^t \exp(s(-a_0 + a_1 - a_{03} + a_{13})) ds} \\
&= \frac{\frac{a_0 a_1}{a_2 - a_1 - a_{13}} \exp(-a_2 r)(\exp(r(a_2 - a_1 - a_{13})) - \exp(t(a_2 - a_1 - a_{13}))) \int_0^t \exp(s(-a_0 - a_{03} + a_1 + a_{13})) ds}{a_0 \exp(t(-a_1 - a_{13})) \int_0^t \exp(s(-a_0 + a_1 - a_{03} + a_{13})) ds} \\
&= \frac{a_1}{a_2 - a_1 - a_{13}} \exp(-a_2 r + t(a_1 + a_{13}))(-\exp(r(a_2 - a_1 - a_{13})) + \exp(t(a_2 - a_1 - a_{13}))).
\end{aligned}$$

$$\begin{aligned}
P_{12}^G &= Pr(X_r = 2 | X_t = 1, X_v = 0) \\
&= \frac{a_0 a_1 \exp(-a_2 r) \int_v^t \exp(s(-a_0 + a_2 - a_{03})) \int_{t-s}^{r-s} \exp(u(-a_1 + a_2 - a_{13})) du ds}{a_0 \exp(t(-a_1 - a_{13})) \int_v^t \exp(s(-a_0 + a_1 - a_{03} + a_{13})) ds} \\
&= \frac{\frac{a_0 a_1}{a_2 - a_1 - a_{13}} \exp(-a_2 r)(\exp(r(a_2 - a_1 - a_{13})) - \exp(t(a_2 - a_1 - a_{13}))) \int_v^t \exp(s(-a_0 - a_{03} + a_1 + a_{13})) ds}{a_0 \exp(t(-a_1 - a_{13})) \int_v^t \exp(s(-a_0 + a_1 - a_{03} + a_{13})) ds} \\
&= \frac{a_1}{-a_2 + a_1 + a_{13}} \exp(-a_2 r + t(a_1 + a_{13}))(-\exp(r(a_2 - a_1 - a_{13})) + \exp(t(a_2 - a_1 - a_{13}))).
\end{aligned}$$

These transition probabilities are equal.

Transition from 2 to 3

$$\begin{aligned}
P_{23}^M &= Pr(X_q = 3 | X_r = 2) \\
&= \frac{a_0 a_1 a_2 \exp(-a_2 q) \int_0^r \exp(s(a_2 - a_0 - a_{03})) \int_0^{r-s} \exp(u(a_2 - a_1 - a_{13})) du ds}{a_0 a_1 \exp(-a_2 r) \int_0^r \exp(s(-a_0 + a_2 - a_{03})) \int_0^{r-s} \exp(u(-a_1 + a_2 - a_{13})) du ds} \\
&= a_2 \exp(a_2(r - q)),
\end{aligned}$$

$$
\begin{aligned}
P^G_{23} &= Pr(X_q = 3 | X_r = 2, X_t = 1, X_v = 0) \\
&= \frac{-a_0 a_1 a_2 \exp(-a_2 q) \int_v^t \exp(s(a_2 - a_0 - a_{03})) \int_{t-s}^{r-s} \exp(u(a_2 - a_1 - a_{13})) du ds}{a_0 a_1 \exp(-a_2 r) \int_v^t \exp(s(a_2 - a_0 - a_{03})) \int_{t-s}^{r-s} \exp(u(a_2 - a_1 - a_{13})) du ds} \\
&= a_2 \exp(a_2(r - q)).
\end{aligned}
$$

These transition probabilities are equal.

Transition from 1 to 3

$$
\begin{aligned}
P^M_{13,0} &= \frac{a_0 a_{13} \exp(q(-a_1 - a_{13})) \int_0^t \exp(s(a_1 - a_0 - a_{03} + a_{13})) ds}{a_0 \exp(t(-a_1 - a_{13})) \int_0^t \exp(s(a_1 - a_0 - a_{03} + a_{13})) ds} \\
&= a_{13} \exp((t - q)(a_1 + a_{13})),
\end{aligned}
$$

$$
\begin{aligned}
P^G_{13,0} &= \frac{a_0 a_{13} \exp(q(-a_1 - a_{13})) \int_v^t \exp(s(a_1 - a_0 - a_{03} + a_{13})) ds}{a_0 \exp(t(-a_1 - a_{13})) \int_v^t \exp(s(a_1 - a_0 - a_{03} + a_{13})) ds} \\
&= a_{13} \exp((t - q)(a_1 + a_{13})).
\end{aligned}
$$

These transition probabilities are equal.

$$
\begin{aligned}
P^M_{13,1} &= \frac{\int_0^t \int_{t-s}^{q-s} a_0 a_1 a_2 \exp(s(a_2 - a_0 - a_{03})) \exp(u(a_2 - a_1 - a_{13})) \exp(-a_2 q) du ds}{\int_0^t a_0 \exp(s(a_1 + a_{13} - a_0 - a_{03})) \exp(-t(a_1 + a_{13})) ds} \\
&= \frac{\frac{a_0 a_1 a_2}{a_2 - a_1 - a_0} \int_0^t \exp(s(a_1 + a_{13} - a_0 - a_{03})) [\exp(q(a_2 - a_1 - a_{13})) - \exp(t(a_2 - a_1 - a_{13}))] \exp(-a_2 q) ds}{\int_0^t a_0 \exp(s(a_1 + a_{13} - a_0 - a_{03})) \exp(-t(a_1 + a_{13})) ds} \\
&= \frac{a_1 a_2 [\exp(q(a_2 - a_1 - a_{13})) - \exp(t(a_2 - a_1 - a_{13}))] \exp(-a_2 q)}{(a_2 - a_1 - a_{13}) \exp(-t(a_1 + a_{13}))} \\
&= \frac{a_1 a_2}{a_2 - a_1 - a_{13}} [\exp(q(a_2 - a_1 - a_{13})) - \exp(t(a_2 - a_1 - a_{13}))] \exp(-a_2 q) \exp(t(a_1 + a_{13})).
\end{aligned}
$$

$$
\begin{aligned}
P^G_{13,1} &= \frac{\int_v^t \int_{t-s}^{q-s} a_0 a_1 a_2 \exp(s(a_2 - a_0 - a_{03})) \exp(u(a_2 - a_1 - a_{13})) \exp(-a_2 q) du ds}{\int_v^t a_0 \exp(s(a_1 + a_{13} - a_0 - a_{03})) \exp(-t(a_1 + a_{13})) ds} \\
&= \frac{a_1 a_2 [\exp(t(a_1 + a_{13} - a_0 - a_{03})) - \exp(v(a_1 + a_{13} - a_0 - a_{03}))] [\exp(q(a_2 - a_1 - a_{13})) - \exp(t(a_2 - a_1 - a_{13}))] \exp(-a_2 q)}{[\exp(t(a_1 + a_{13} - a_0 - a_{03})) - \exp(v(a_1 + a_{13} - a_0 - a_{03}))] \exp(-t(a_1 + a_{13}))} \\
&= \frac{a_1 a_2}{a_2 - a_1 - a_{13}} [\exp(q(a_2 - a_1 - a_{13})) - \exp(t(a_2 - a_1 - a_{13}))] \exp(-a_2 q) \exp(t(a_1 + a_{13})).
\end{aligned}
$$

These transition probabilities are equal.

Staying in state 1

$$
\begin{aligned}
P^M_{11} &= \frac{\exp(-t(a_1 + a_{13})) \int_0^k a_0 \exp(s(a_1 + a_{13} - a_0 - a_{03}))}{\exp(-k(a_1 + a_{13})) \int_0^k a_0 \exp(s(a_1 + a_{13} - a_0 - a_{03}))} \\
&= \exp((k - t)(a_1 + a_{13})),
\end{aligned}
$$

$$
\begin{aligned}
P^G_{11} &= \frac{\exp(-t(a_1 + a_{13})) \int_v^k a_0 \exp(s(a_1 + a_{13} - a_0 - a_{03}))}{\exp(-k(a_1 + a_{13})) \int_v^k a_0 \exp(s(a_1 + a_{13} - a_0 - a_{03}))} \\
&= \exp((k - t)(a_1 + a_{13})).
\end{aligned}
$$

These transition probabilities are equal.

Staying in state 2

$$P_{22}^{M} = \frac{\exp(-ra_2)\int_0^m \int_0^{m-s} a_0 a_1 \exp(s(a_2 - a_1 - a_{13}))\exp(u(a_2 - a_0 - a_{03}))}{\exp(-ma_2)\int_0^m \int_0^{m-s} a_0 a_1 \exp(s(a_2 - a_1 - a_{13}))\exp(u(a_2 - a_0 - a_{03}))}$$
$$= \exp((m-r)a_2),$$
$$P_{22}^{G} = \frac{\exp(-ra_2)\int_v^k \int_{t-s}^{m-s} a_0 a_1 \exp(s(a_2 - a_1 - a_{13}))\exp(u(a_2 - a_0 - a_{03}))}{\exp(-ma_2)\int_v^k \int_{t-s}^{m-s} a_0 a_1 \exp(s(a_2 - a_1 - a_{13}))\exp(u(a_2 - a_0 - a_{03}))}$$
$$= \exp((m-r)a_2).$$

These transition probabilities are equal.

In conclusion, we find that when the transition times follow an exponential distribution, the Markov property is satisfied. This follows from the calculations of the transition probabilities.

5.1.2 Example: Gamma Process Models

In our second example, we consider the Gamma process model. Since it is difficult to calculate the exact formulas for the transition probabilities, we calculate the transition probabilities numerically. We find that the transition probabilities $P_{m\ell}^{M}$, when we condition on the previous state, in general are not equal to $P_{m\ell}^{G}$, when we condition on the whole state history. The Markov property is therefore not fulfilled. Assume for the transition time from state 0 to state 1 that $Pr(T_0 \geq t) = G(c_0, a_0 t, 1)$, with similar formula for the other transitions. Let $c_0 = 0.208, a_0 = 0.0486, c_1 = 0.00315, a_1 = 0.0380, c_2 = 1.108, a_2 = 0.323, c_{03} = 0.398, a_{03} = 0.0616, c_{13} = 1.939, a_{13} = 0.452$. The time points are $v = 2$, $t = 5$, $r = 8$ and $q = 10$.

	$\boldsymbol{P_{m\ell}^{M}}$	$\boldsymbol{P_{m\ell}^{G}}$
$m = 1, \ell = 2$	0.312	0.323
$m = 2, \ell = 3$	0.162	0.152

Table 5.1: Transition probabilities when the transition probabilities we condition on the previous state or the whole state history for alternative 1 of the Gamma process model

In Table 5.1, we present two examples of the transition probabilities from the formulas in Section 5.1. From Table 5.1, we see that $P_{m\ell}^{M}$ is not equal to $P_{m\ell}^{G}$. However, the differences between $P_{m\ell}^{M}$ and $P_{m\ell}^{G}$ are quite small.

	$\boldsymbol{P_{m\ell}^{M}}$	$\boldsymbol{P_{m\ell}^{G}}$
$m = 1, \ell = 2$	0.00493	0.00626
$m = 2, \ell = 3$	0.187	0.157

Table 5.2: Transition probabilities when the transition probabilities we condition on the previous state or the whole state history for alternative 2 of the Gamma process model

Let us consider a second alternative with another Gamma process, where $Pr(T_0 \geq t) = G(c_0, a_0 t^{b_0}, 1)$ and $c_0 = 1, b_0 = 1, a_0 = 1, c_1 = 0.1, b_1 = 0.1, a_1 =$

$0.1, c_2 = 0.8, b_2 = 0.5, a_2 = 1, c_{03} = 1, b_{03} = 0.7, a_{03} = 0.1, c_{13} = 1, b_{13} = 0.5, a_{13} = 1$. The time points are $v = 2$, $t = 5$, $r = 8$ and $q = 10$. In Table 5.2, we present the transition probabilities with this Gamma process. The differences between $P^M_{m\ell}$ and $P^G_{m\ell}$ are bigger for these transition probabilities. P^M_{12} is clearly lower than P^G_{12}, and P^M_{23} is clearly higher than P^G_{23}.

Depending on the parameters in the Gamma process models, the transition probabilities can be close or not close to satisfying the Markov property. This makes the Gamma process models flexible. Depending on the data and research question, it is possible to find suitable parameters where the hazard function is and is not constant.

5.2 The Relationship Between the Exponentially Distributed Transition Times and a Homogeneous Markov Model

Lemma 5.2.1. *In the proposed model framework, the four-state illness-death model where the transition times are exponentially distributed, is equal to the homogeneous Markov model in Jackson (2011).*

The aim of this section is to prove Lemma 5.2.1. Firstly, we show how the likelihood calculations are done in Jackson (2011). They calculate their transition probabilities using eigenvalue decomposition. Then, we show that our likelihood construction using exponentially distributed transition times is equal to the likelihood in Jackson (2011). This is for data where all the individuals start in state 0 at time 0.

In order to construct a Markov model, the transition intensities must be defined. The transition intensities for moving from one state to another in a multi-state model is equal to the hazard functions (Meira-Machado et al., 2009). From Chapter 2, we have that the intensity of moving from state m to state ℓ is

$$q_{m\ell}(t|X_u, u \in [0,t)) = \lim_{\Delta t \searrow 0} \frac{Pr\{X_{t+\Delta t} = \ell | X_{t^-} = m\}}{\Delta t},$$

which then corresponds to the hazard function for the same transition.

We are considering the homogeneous process where $\boldsymbol{P}(u, t+u) = \boldsymbol{P}(t)$, where

$$\boldsymbol{P}(t) = Exp(t\boldsymbol{Q}),$$

The matrix exponential $Exp()$ is difficult to calculate directly. We can use eigenvalue decomposition. Then

$$Exp(t\boldsymbol{Q}) = \boldsymbol{R} e^{\boldsymbol{N}} \boldsymbol{R}^{-1},$$

where $\boldsymbol{R}$ consists of the eigenvectors and $\boldsymbol{N}$ is a matrix with the eigenvalues on the diagonal. We then need to find the eigenvalues for

$$t\boldsymbol{Q} = \begin{bmatrix} -(q_0+q_{03})t & q_0 t & 0 & q_{03}t \\ 0 & -(q_1+q_{13})t & q_1 t & q_{13}t \\ 0 & 0 & -q_2 t & q_2 t \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

Since this matrix is an upper triangular matrix, we have that the eigenvalues are on the diagonal. The eigenvalues then becomes $-(q_0+q_{03})t$, $-(q_1+q_{13})t$,

$-q_2t$ and 0. We then get

$$e^{\boldsymbol{N}} = \begin{bmatrix} e^{-(q_0+q_{03})t} & 0 & 0 & 0 \\ 0 & e^{-(q_1+q_{13})t} & 0 & 0 \\ 0 & 0 & e^{-q_2t} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

We must also find the eigenvectors. We start with the eigenvector for $\lambda_1 = -(q_0 + q_{03})t$:

$$(t\boldsymbol{Q} - \lambda_1 \boldsymbol{I}_4) = \begin{bmatrix} 0 & q_0t & 0 & q_{03}t \\ 0 & -(q_1 + q_{13})t + q_0t + q_{03}t & q_1t & q_{13}t \\ 0 & 0 & -q_2t + q_0t + q_{03}t & q_2t \\ 0 & 0 & 0 & (q_0 + q_{03})t \end{bmatrix}$$

We must solve the equation $(t\boldsymbol{Q} - \lambda \boldsymbol{I}_4) \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$. Solving this equation, gives for example the eigenvector

$$\boldsymbol{R}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

For the eigenvalue $\lambda_2 = -(q_1 + q_{13})t$.

$$(t\boldsymbol{Q} - \lambda_2 \boldsymbol{I}_4) = \begin{bmatrix} -(q_0 + q_{03})t + q_1t + q_{13}t & q_0t & 0 & q_{03}t \\ 0 & 0 & q_1t & q_{13}t \\ 0 & 0 & -q_2t + q_1t + q_{13}t & q_2t \\ 0 & 0 & 0 & q_1t + q_{13}t \end{bmatrix}$$

Solving in the same way, we end up with the eigenvector

$$\boldsymbol{R}_2 = \begin{bmatrix} 1 \\ \frac{q_0+q_{03}-q_1-q_{13}}{q_0} \\ 0 \\ 0 \end{bmatrix},$$

For the eigenvalue $\lambda_3 = -q_2t$

$$(t\boldsymbol{Q} - \lambda_3 \boldsymbol{I}_4) = \begin{bmatrix} -(q_0 + q_{03})t + q_2t & q_0t & 0 & q_{03}t \\ 0 & -(q_1 + q_{13})t + q_2t & q_1t & q_{13}t \\ 0 & 0 & 0 + q_{13}t & q_2t \\ 0 & 0 & 0 & q_2t \end{bmatrix}$$

And solving in the same way, we end up with the eigenvector

$$\boldsymbol{R}_3 = \begin{bmatrix} \frac{q_0}{q_0+q_{03}-q_2} \\ 1 \\ \frac{q_1+q_{13}-q_2}{q_1} \\ 0 \end{bmatrix},$$

And for the eigenvalue $\lambda_4 = 0$, an eigenvector may be

$$\boldsymbol{R}_4 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix},$$

We then get

$$\boldsymbol{P}(t) = Exp(t\boldsymbol{Q}) = \begin{bmatrix} 1 & 1 & \frac{q_0}{q_0+q_{03}-q_2} & 1 \\ 0 & \frac{q_0+q_{03}-q_1-q_{13}}{q_0} & 1 & 1 \\ 0 & 0 & \frac{q_1+q_{13}-q_2}{q_1} & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e^{-(q_0+q_{03})t} & 0 & 0 & 0 \\ 0 & e^{-(q_1+q_{13})t} & 0 & 0 \\ 0 & 0 & e^{-q_2 t} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 1 & \frac{q_0}{q_0+q_{03}-q_2} & 1 \\ 0 & \frac{q_0+q_{03}-q_1-q_{13}}{q_0} & 1 & 1 \\ 0 & 0 & \frac{q_1+q_{13}-q_2}{q_1} & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1}$$
$$= \begin{bmatrix} c_{00} & c_{01} & c_{02} & c_{03} \\ 0 & c_{11} & c_{12} & c_{13} \\ 0 & 0 & c_{22} & c_{23} \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where

$$c_{00} = e^{-(q_0+q_{03})t},$$
$$c_{01} = \frac{q_0}{q_0+q_{03}-q_1-q_{13}}(e^{-(q_1+q_{13})t} - e^{-(q_0+q_{03})t}),$$
$$c_{02} = \frac{q_0 q_1}{(q_0-q_2+q_{03})(q_0-q_1+q_{03}-q_{13})} e^{-(q_0+q_{03})t},$$
$$+\frac{q_0 q_1}{(q_1-q_2+q_{13})(-q_0+q_1-q_{03}+q_{13})} e^{-(q_1+q_{13})t},$$
$$+\frac{q_0 q_1}{(q_1-q_2+q_{13})(q_0+q_{03}-q_2)} e^{-q_2 t},$$
$$c_{03} = \frac{q_1(q_{03}-q_2)-(q_0-q_2+q_{03})(q_{03}-q_{13})}{(q_0-q_2+q_{03})(q_0-q_1+q_{03}-q_{13}} e^{-(q_0+q_{03})t}$$
$$-\frac{q_0(q_2-q_{13})}{(q_1-q_2+q_{13})(-q_0+q_1-q_{03}+q_{13})} e^{-(q_1+q_{13})t}$$
$$-\frac{q_0 q_1}{(q_0+q_{03}-q_2)(q_1-q_2+q_{13})} e^{-q_2 t} + 1,$$
$$c_{11} = \frac{q_0+q_{03}}{q_0}\frac{q_0}{q_0+q_{03}} e^{-(q_1+q_{13})t} = e^{-(q_1+q_{13})t},$$
$$c_{12} = \frac{q_1}{q_1+q_{13}-q_2}(-e^{-(q_1+q_{13})t} + e^{-q_2 t}),$$
$$c_{13} = \frac{q_2-q_{13}}{q_1+q_{13}-q_2} e^{-(q_1+q_{13})t} - \frac{q_1}{q_1+q_{13}-q_2} e^{-q_2} + 1,$$
$$c_{22} = e^{-q_2 t}\frac{q_1}{q_1+q_{13}-q_2}\frac{q_1+q_{13}-q_2}{q_1} = e^{-q_2 t},$$
$$c_{23} = -\frac{q_1+q_{13}-q_2}{q_1} e^{-q_2 t}\frac{q_1}{q_1+q_{13}-q_2} + 1 = -e^{-q_2 t} + 1,$$

We also want to find the $\boldsymbol{P}(t)$ matrix when the transition times are exponentially distributed in our model. Then

$$\boldsymbol{P}(t)^e = \begin{bmatrix} p_{00} & p_{01} & p_{02} & p_{03} \\ 0 & p_{11} & p_{12} & p_{13} \\ 0 & 0 & p_{22} & p_{23} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

By using the results from Section 5.1, where the time points are 0 and t, we get

$$p_{00} = P_{00}^M(0,t) = S_0(t)S_{03}(t) = e^{-a_0 t}e^{-a_{03}t} = e^{-t(a_0+a_{03})},$$

$$\begin{aligned} p_{01} = P_{01}^M(0,t) &= \frac{\int_0^t a_0 e^{-a_0 s}e^{-a_1(t-s)}e^{-a_{03}s}e^{-a_{13}(t-s)}}{e^0 e^0 ds} \\ &= \frac{a_0}{a_1 + a_{13} - a_0 - a_{03}}(e^{-t(a_0+a_{03})} - e^{-t(a_1+a_{13})}), \end{aligned}$$

$$\begin{aligned} p_{02} = P_{02}^M(0,t) &= \frac{\int_0^t \int_0^{t-s} a_0 a_1 e^{s(a_2-a_0-a_{03})}e^{u(a_2-a_1-a_{13})}e^{-a_2 t}duds}{e^0 e^0} \\ &= \frac{a_0 a_1}{(a_2 - a_0 - a_{03})(a_1 + a_{13} - a_0 - a_{03})}e^{-t(a_0+a_{03})} \\ &+ \frac{a_0 a_1}{(-a_2 + a_1 + a_{13})(a_1 + a_{13} - a_0 - a_{03})}e^{-t(a_1+a_{13})} \\ &+ \frac{a_0 a_1}{(a_2 - a_1 - a_{13})(a_2 - a_0 - a_{03})}e^{-a_2 t}. \end{aligned}$$

In Section 5.1, we assumed to know the exact time of death. This is not the case in the $\boldsymbol{P}(t)$-matrix. However, we can use the fact that

$$\begin{aligned} p_{03} &= 1 - p_{00} - p_{01} - p_{02} \\ &= \frac{a_1(a_{03} - a_2) - (a_0 - a_2 + a_{03})(a_{03} - a_{13})}{(a_0 - a_2 + a_{03})(a_0 - a_1 + a_{03} - a_{13})}e^{-(a_0+a_{03})t} \\ &- \frac{a_0(a_2 - a_{13})}{(a_1 - a_2 + a_{13})(-a_0 + a_1 - a_{03} + a_{13})}e^{-(a_1+a_{13})} \\ &- \frac{a_0 a_1}{(a_0 + a_{03} - a_2)(a_1 - a_2 + a_{13})}e^{-a_2 t} + 1. \end{aligned}$$

We have from Section 5.1 and 5.1.1

$$p_{11} = P_{11}^M(0,t) = e^{-t(a_1+a_{13})},$$

$$\begin{aligned} p_{12} = P_{12}^M(0,t) &= \frac{a_1}{a_1 + a_{13} - a_2}e^{-a_2 t}(-e^{-t(a_2-a_1-a_{13})} + 1) \\ &= \frac{a_1}{a_1 + a_{13} - a_2}(e^{-a_2 t} - e^{-(a_1+a_{13})t}). \end{aligned}$$

$$\begin{aligned} p_{13} &= 1 - p_{11} - p_{12} = 1 - e^{-t(a_1+a_{13})} - \frac{a_1}{a_1 + a_{13} - a_2}(e^{-a_2 t} - e^{-(a_1+a_{13})t}) \\ &= \frac{a_2 - a_{13}}{a_1 + a_{13} - a_2}e^{-(a_1+a_{13})t} - \frac{a_1}{a_1 + a_{13} - a_2}e^{-a_2} + 1. \end{aligned}$$

From Section 5.1 and 5.1.1

$$p_{22} = P_{22}^{M}(0,t) = e^{-a_2 t},$$

$$p_{23} = 1 - p_{22} = 1 - e^{-a_2 t},$$

In addition, we have that $\boldsymbol{P}(t)^e = \boldsymbol{P}(t)$ if $q_0 = a_0$, $q_1 = a_1$, $q_2 = a_2$, $q_{03} = a_{03}$ and $q_{13} = a_{13}$.

The next step is to calculate the hazard functions. In the Markov model, this means $\alpha_0^{MM} = q_0$, $\alpha_1^{MM} = q_1$, $\alpha_2^{MM} = q_2$, $\alpha_{03}^{MM} = q_{03}$ and $\alpha_{13}^{MM} = q_{13}$. For the exponentially distributed transition times, the hazard functions are $\alpha_0^e = \frac{f_0(t)}{S_0(t)} = \frac{a_0 \exp(-a_0 t)}{\exp(-a_0 t)} = a_0$, $\alpha_1^e = a_1$, $\alpha_2^e = a_2$, $\alpha_{03}^e = a_{03}$ and $\alpha_{13}^e = a_{13}$.

We calculate the likelihood for the Markov model in the same way as presented in Chapter 2. In the CAV-dataset that we analyze in depth in Chapter 6, all the patients start in state 0. Then we only include the likelihood types from Chapter 3, more specifically Section 3.6, where the individuals start in state 0. These are types 1, 2, 3, 4, 5, 6, 7, 15, 16, 18 and 21.

An individual observed in state 0 and then in state 3 is either type 5, 16 or 21. If the individual is type 5, then the individual transfers through state 1 and state 2 before it is observed in state 3. However, if the individual is type 16, the individual transfers directly from state 0 to state 3. Finally, if the individual transfers through state 1, before transferring directly to state 3, then the individual is type 21. Therefore, we consider a sum of the likelihood contributions for these types as the final likelihood contribution for those individuals observed in state 0, then in state 3.

When an individual is observed in state 0, then state 1 and finally state 3, the individual is either type 15 or 18. If the individual is type 15, then the individual transfers through state 2 before state 3 is observed. However, if the individual is type 18, then the individual transfers directly from state 1 to state 3. Therefore we consider the sum of the likelihood contributions of type 15 and 18.

We start by assuming an individual is only observed in state 0, and is therefore type 1. The likelihood contribution is

$$\mathcal{L}_{\exp,(I)} = S_0(t_n) S_{03}(t_n) = \exp(-t_n(a_0 + a_{03})).$$

In the homogeneous Markov model

$$\mathcal{L}_{MM,(I)} = c_{00}(t_n - 0) = \exp(-t_n(q_0 + q_{03})).$$

If an individual is observed in state 0 until time point t_i, before the individual is observed in state 1 from t_{i+1}, where the individual stays, the individual is type 2. The likelihood contribution is

$$\begin{aligned}
\mathcal{L}_{\exp,(II)} &= \int_{t_i}^{t_{i+1}} f_0(s) S_1(t_n - s) S_{03}(s) S_{13}(t_n - s) ds \\
&= \int_{t_i}^{t_{i+1}} a_0 \exp(s(a_1 + a_{13} - a_0 - a_{03})) \exp(-t_n(a_1 + a_{13})) ds \\
&= \frac{a_0}{a_1 + a_{13} - a_0 - a_{03}} \exp(-t_n(a_1 + a_{13}))[\exp(t_{i+1}(a_1 + a_{13} - a_0 - a_{03})) \\
&\quad - \exp(t_i(a_1 + a_{13} - a_0 - a_{03}))].
\end{aligned}$$

In the homogeneous Markov model

$$\begin{aligned}\mathcal{L}_{MM,(II)} &= c_{00}(t_i - 0)c_{01}(t_{i+1} - t_i)c_{11}(t_n - t_{i+1}) \\ &= \exp(-t_i(q_0 + q_{03}))\frac{q_0}{q_0 + q_{03} - q_1 - q_{13}}[\exp(-(q_1 + q_{13})(t_{i+1} - t_i)) \\ &- \exp(-(q_0 + q_{03})(t_{i+1} - t_i))]\exp(-(t_n - t_{i+1})(q_1 + q_{13})) \\ &= \frac{q_0}{q_1 + q_{13} - q_0 - q_{03}}\exp(-t_n(q_1 + q_{13}))[\exp(t_{i+1}(q_1 + q_{13} - q_0 - q_{03})) \\ &- \exp(t_i(q_1 + q_{13} - q_0 - q_{03}))].\end{aligned}$$

If an individual is observed in state 0 until time point t_i, in state 1 from t_{i+1} to t_{i+k-1}, and then in state 2 at t_{i+k}, where the individuals stays, then the individual is type 3. The likelihood contribution is

$$\begin{aligned}\mathcal{L}_{\exp,(III)} &= \int_{t_i}^{t_{i+1}}\int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s)f_1(u)S_2(t_n - s - u)S_{03}(s)S_{13}(u)duds \\ &= \int_{t_i}^{t_{i+1}}\int_{t_{i+k-1}-s}^{t_{i+k}-s} a_0a_1\exp(s(a_2 - a_0 - a_{03})) \\ &\exp(u(a_2 - a_1 - a_{13}))\exp(-a_2t_n)duds \\ &= \frac{a_0a_1}{(a_2 - a_1 - a_{13})(a_1 + a_{13} - a_0 - a_{03})}\exp(-a_2t_n) \\ &[\exp(t_{i+1}(a_1 + a_{13} - a_0 - a_{03})) \\ &- \exp(t_i(a_1 + a_{13} - a_0 - a_{03}))][\exp(t_{i+k}(a_2 - a_1 - a_{13})) \\ &- \exp(t_{i+k-1}(a_2 - a_1 - a_{13}))].\end{aligned}$$

In the homogeneous Markov model

$$\begin{aligned}\mathcal{L}_{MM,(III)} &= c_{00}(t_i)c_{01}(t_{i+1} - t_i)c_{11}(t_{i+k-1} - t_{i+1})c_{12}(t_{i+k} - t_{i+k-1}) \\ &c_{22}(t_n - t_{i+k}) \\ &= \frac{q_0q_1}{(q_2 - q_1 - q_{13})(q_1 + q_{13} - q_0 - q_{03})}\exp(-q_2t_n) \\ &[\exp(t_{i+1}(q_1 + q_{13} - q_0 - q_{03})) \\ &- \exp(t_i(q_1 + q_{13} - q_0 - q_{03}))][\exp(t_{i+k}(q_2 - q_1 - q_{13})) \\ &- \exp(t_{i+k-1}(q_2 - q_1 - q_{13}))].\end{aligned}$$

Assume an individual is observed in state 0 until time point t_i, in state 1 from t_{i+1} to t_{i+k-1}, in state 2 from t_{i+k} to $t_{i+k+l-1}$, before the individual dies at the exact time point t_{i+k+l}. This individual is type 4. The likelihood contribution is

$$\begin{aligned}\mathcal{L}_{\exp,(IV)} &= \int_{t_i}^{t_{i+1}}\int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s)f_1(u)f_2(t_{i+k+l} - s - u)S_{03}(s)S_{13}(u)duds \\ &= \int_{t_i}^{t_{i+1}}\int_{t_{i+k-1}-s}^{t_{i+k}-s} a_0a_1a_2\exp(s(a_2 - a_0 - a_{03})) \\ &\exp(u(a_2 - a_1 - a_{13}))\exp(-a_2t_{i+k+l})duds\end{aligned}$$

.

$$
\begin{aligned}
&= \frac{a_0 a_1 a_2}{(a_2 - a_1 - a_{13})(a_1 + a_{13} - a_0 - a_{03})} \exp(-a_2 t_{i+k+l}) \\
&[\exp(t_{i+1}(a_1 + a_{13} - a_0 - a_{03})) \\
&- \exp(t_i(a_1 + a_{13} - a_0 - a_{03}))][\exp(t_{i+k}(a_2 - a_1 - a_{13})) \\
&- \exp(t_{i+k-1}(a_2 - a_1 - a_{13}))].
\end{aligned}
$$

In the homogeneous Markov model

$$
\begin{aligned}
\mathcal{L}_{MM,(IV)} &= c_{00}(t_i) c_{01}(t_{i+1} - t_i) c_{11}(t_{i+k-1} - t_{i+1}) c_{12}(t_{i+k} - t_{i+k-1}) \\
&\quad c_{22}(t_{i+k+l} - t_{i+k}) q_2 \\
&= \frac{q_0 q_1 q_2}{(q_2 - q_1 - q_{13})(q_1 + q_{13} - q_0 - q_{03})} \exp(-q_2 t_{i+k+l}) \\
&\quad [\exp(t_{i+1}(q_1 + q_{13} - q_0 - q_{03})) \\
&\quad - \exp(t_i(q_1 + q_{13} - q_0 - q_{03}))][\exp(t_{i+k}(q_2 - q_1 - q_{13})) \\
&\quad - \exp(t_{i+k-1}(q_2 - q_1 - q_{13}))].
\end{aligned}
$$

Assume an individual is observed in state 0 until time point t_i, then in state 2 from t_{i+1}, where the individual stays. Then the individual is type 6. The likelihood contribution becomes

$$
\begin{aligned}
\mathcal{L}_{\exp,(VI)} &= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) S_2(t_n - s - u) S_{03}(s) S_{13}(u) du ds \\
&= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} a_0 a_1 \exp(s(a_2 - a_0 - a_{03})) \\
&\quad \exp(u(a_2 - a_1 - a_{13})) \exp(-a_2 t_n) du ds \\
&= \frac{a_0 a_1}{a_2 - a_1 - a_{13}} \exp(-a_2 t_n) \\
&\quad \Big[\frac{a_2 - a_1 - a_{13}}{(a_1 + a_{13} - a_0 - a_{03})(a_2 - a_0 - a_{03})} \exp(t_{i+1}(a_2 - a_0 - a_{03})) \\
&\quad - \frac{1}{a_1 + a_{13} - a_0 - a_{03}} \exp(t_{i+1}(a_2 - a_1 - a_{13})) \exp(t_i(a_1 + a_{13} - a_0 - a_{03})) \\
&\quad + \frac{1}{a_2 - a_0 - a_{03}} \exp(t_i(a_2 - a_0 - a_{03})) \Big]
\end{aligned}
$$

In the homogeneous Markov model

$$
\begin{aligned}
\mathcal{L}_{MM,(VI)} &= c_{00}(t_i) c_{02}(t_{i+1} - t_i) c_{22}(t_n - t_{i+1}) \\
&= \exp(-q_2 t_n) \Big[\frac{q_0 q_1}{(q_0 - q_2 + q_{03})(q_0 - q_1 + q_{03} - q_{13})} \exp(t_{i+1}(q_0 + q_{03} - q_2)) \\
&\quad + \frac{q_0 q_1}{(q_1 - q_2 + q_{13})(q_0 - q_1 + q_{03} - q_{13})} \exp(t_{i+1}(q_2 - q_1 - q_{13})) \\
&\quad \exp(t_i(q_1 + q_{13} - q_0 - q_{03}) \\
&\quad + \frac{q_0 q_1}{(q_0 - q_2 + q_{03})(-q_2 + q_{13} + q_1)} \exp(t_i(q_2 - q_0 - q_{03})) \Big],
\end{aligned}
$$

and the equations for the homogeneous Markov model and the likelihood contribution for the exponentially distributed transition times for type 6 are equal.

Assume an individual is observed in state 0 until time point t_i, then in state 2 from t_{i+1} to t_{i+k-1}, before the individual dies at t_{i+k}. Then the individual is

type 7. The likelihood contribution is

$$\begin{aligned}
\mathcal{L}_{\exp,(VII)} &= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) f_2(t_{i+k}-s-u) S_{03}(s) S_{13}(u) du ds \\
&= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} a_0 a_1 a_2 \exp(s(a_2-a_0-a_{03})) \\
&\exp(u(a_2-a_1-a_{13})) \exp(-a_2 t_{i+k}) du ds \\
&= \frac{a_0 a_1}{a_2-a_1-a_{13}} \exp(-a_2 t_{i+k}) \\
&\Big[\frac{a_2-a_1-a_{13}}{(a_1+a_{13}-a_0-a_{03})(a_2-a_0-a_{03})} \exp(t_{i+1}(a_2-a_0-a_{03})) \\
&- \frac{1}{a_1+a_{13}-a_0-a_{03}} \exp(t_{i+1}(a_2-a_1-a_{13})) \exp(t_i(a_1+a_{13}-a_0-a_{03})) \\
&+ \frac{1}{a_2-a_0-a_{03}} \exp(t_i(a_2-a_0-a_{03})) \Big].
\end{aligned}$$

In the homogeneous Markov model

$$\begin{aligned}
\mathcal{L}_{MM,(VII)} &= c_{00}(t_i) c_{02}(t_{i+1}-t_i) c_{22}(t_{i+k}-t_{i+1}) q_2 \\
&= \exp(-q_2 t_{i+k}) \Big[\frac{q_0 q_1 q_2}{(q_0-q_2+q_{03})(q_0-q_1+q_{03}-q_{13})} \exp(t_{i+1}(q_0+q_{03}-q_2)) \\
&+ \frac{q_0 q_1 q_2}{(q_1-q_2+q_{13})(q_0-q_1+q_{03}-q_{13})} \exp(t_{i+1}(q_2-q_1-q_{13})) \\
&\exp(t_i(q_1+q_{13}-q_0-q_{03}) \\
&+ \frac{q_0 q_1 q_2}{(q_0-q_2+q_{03})(-q_2+q_{13}+q_1)} \exp(t_i(q_2-q_0-q_{03})) \Big].
\end{aligned}$$

Assume an individual is observed in state 0 until t_i, before the individual dies at time t_{i+1}. The individual can either be type 5, 16 or 21. We start by considering the likelihood contribution for type 5

$$\begin{aligned}
\mathcal{L}_{\exp,(V)} &= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) f_2(t_{i+1}-s-u) S_{03}(s) S_{13}(u) du ds \\
&= \frac{a_0 a_1 a_2}{a_2-a_1-a_{13}} \exp(-a_2 t_{i+1}) \\
&\Big[\frac{a_2-a_1-a_{13}}{(a_1+a_{13}-a_0-a_{03})(a_2-a_0-a_{03})} \exp(t_{i+1}(a_2-a_0-a_{03})) \\
&- \frac{1}{a_1+a_{13}-a_0-a_{03}} \exp(t_{i+1}(a_2-a_1-a_{13})) \exp(t_i(a_1+a_{13}-a_0-a_{03})) \\
&+ \frac{1}{a_2-a_0-a_{03}} \exp(t_i(a_2-a_0-a_{03})) \Big].
\end{aligned}$$

Then from type 16

$$\mathcal{L}_{\exp,(XVI)} = f_{03}(t_{i+1}) S_0(t_{i+1}) = a_{03} \exp(-t_{i+1}(a_0+a_{03})),$$

Then from type 21

$$\begin{aligned}
\mathcal{L}_{\exp,(XXI)} &= \int_{t_i}^{t_{i+1}} f_0(s) S_{03}(s) f_{13}(t_{i+1}-s) S_1(t_{i+1}-s) ds \\
&= \frac{a_0 a_{13}}{a_1+a_{13}-a_0-a_{03}} \exp(-t_{i+1}(a_{13}+a_1)) \Big[\exp(t_{i+1}(a_1+a_{13}-a_0-a_{03})) \\
&- \exp(t_i(a_1+a_{13}-a_0-a_{03})) \Big].
\end{aligned}$$

Then, the total likelihood contribution becomes

$$
\begin{aligned}
\mathcal{L}_{\exp,0\to3} = \prod_{i:0\to3} \Bigg(& \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) f_2(t_{i+1}-u-s) S_{03}(s) S_{13}(u) du ds \\
& + f_{03}(t_{i+1}) S_0(t_{i+1}) \\
& + \int_{t_i}^{t_{i+1}} f_0(s) S_{03}(s) f_{13}(t_{i+1}-s) S_1(t_{i+1}-s) \Bigg) \\
= \prod_{i:0\to3} \Bigg(& \frac{a_0 a_1 a_2}{a_2 - a_1 - a_{13}} \exp(-a_2 t_{i+1}) \\
& \Big[\frac{a_2 - a_1 - a_{13}}{(a_1 + a_{13} - a_0 - a_{03})(a_2 - a_0 - a_{03})} \exp(t_{i+1}(a_2 - a_0 - a_{03})) \\
& - \frac{1}{a_1 + a_{13} - a_0 - a_{03}} \exp(t_{i+1}(a_2 - a_1 - a_{13})) \exp(t_i(a_1 + a_{13} - a_0 - a_{03})) \\
& + \frac{1}{a_2 - a_0 - a_{03}} \exp(t_i(a_2 - a_0 - a_{03})) \Big] \\
& + a_{03} \exp(-t_{i+1}(a_0 + a_{03})) \\
& + \frac{a_0 a_{13}}{a_1 + a_{13} - a_0 - a_{03}} \exp(-t_{i+1}(a_{13} + a_1)) \Big[\exp(t_{i+1}(a_1 + a_{13} - a_0 - a_{03})) \\
& - \exp(t_i(a_1 + a_{13} - a_0 - a_{03})) \Big] \Bigg).
\end{aligned}
$$

The total likelihood contribution for a homogeneous Markov model becomes

$$
\begin{aligned}
\mathcal{L}_{MM,0\to3} = \prod_{i:0\to3} \Bigg(& c_{00}(t_{i+1}) q_{03} + c_{00}(t_i) c_{01}(t_{i+1} - t_i) q_{13} \\
& + c_{00}(t_i) c_{02}(t_{i+1} - t_i) q_2 \Bigg) \\
= \prod_{i:0\to3} \Bigg(& q_{03} \exp(-t_{i+1}(q_{03} + q_0)) \\
& + \frac{q_0 q_{13}}{q_0 + q_{03} - q_1 - q_{13}} \exp(-t_i(q_{03} + q_0)) \\
& \Big[\exp(-(t_{i+1} - t_i)(q_1 + q_{13})) - \exp(-(t_{i+1} - t_i)(q_0 + q_{03})) \Big] \\
& + \exp(-t_i(q_{03} + q_0)) \Big[\frac{q_0 q_1 q_2}{(q_0 - q_2 + q_{03})(q_0 - q_1 + q_{03} - q_{13})} \exp(-(q_0 + q_{03})(t_{i+1} - t_i)) \\
& + \frac{q_0 q_1 q_2}{(q_1 - q_2 + q_{13})(-q_0 + q_1 - q_{03} + q_{13})} \exp(-(q_1 + q_{13}(t_{i+1} - t_i))) \\
& + \frac{q_0 q_1 q_2}{(q_0 - q_2 + q_{03})(q_1 - q_2 + q_{13})} \exp(-q_2(t_{i+1} - t_i)) \Big] \Bigg) \\
= \prod_{i:0\to3} \Bigg(& \frac{q_0 q_1 q_2}{q_2 - q_1 - q_{13}} \exp(-q_2 t_{i+1}) \\
& \Big[\frac{q_2 - q_1 - q_{13}}{(q_1 + q_{13} - q_0 - q_{03})(q_2 - q_0 - q_{03})} \exp(t_{i+1}(q_2 - q_0 - q_{03})) \\
& - \frac{1}{q_1 + q_{13} - q_0 - q_{03}} \exp(t_{i+1}(q_2 - q_1 - q_{13})) \exp(t_i(q_1 + q_{13} - q_0 - q_{03})) \\
& + \frac{1}{q_2 - q_0 - q_{03}} \exp(t_i(q_2 - q_0 - q_{03})) \Big] \\
& + q_{03} \exp(-t_{i+1}(q_0 + q_{03}))
\end{aligned}
$$

$$+\frac{q_0q_{13}}{q_1+q_{13}-q_0-q_{03}}\exp(-t_{i+1}(q_{13}+q_1))\Big[\exp(t_{i+1}(q_1+q_{13}-q_0-q_{03}))$$
$$-\exp(t_i(q_1+q_{13}-q_0-q_{03}))\Big]\Bigg).$$

Assume an individual is observed in state 0 until t_i, then the individual is observed in state 1 from t_{i+1} to t_{i+k-1}, before the individual dies at time t_{i+k}. Then the individual can either bet type 15 or 18. Then for type 15

$$\begin{aligned}\mathcal{L}_{\exp,(XV)} &= \int_{t_i}^{t_{i+1}}\int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s)f_1(u)f_2(t_{i+k}-s-u)S_{03}(s)S_{13}(u)duds\\ &= \frac{a_0a_1a_2}{(a_2-a_1-a_{13})(a_1+a_{13}-a_0-a_{03})}\exp(-a_2t_{i+k})\\ &\Big[\exp(t_{i+k}(a_2-a_1-a_{13}))-\exp(t_{i+k-1}(a_2-a_1-a_{13}))\Big]\\ &\Big[\exp(t_{i+1}(a_1+a_{13}-a_0-a_{03}))-\exp((a_1+a_{13}-a_0-a_{03})t_i)\Big].\end{aligned}$$

For type 18

$$\begin{aligned}\mathcal{L}_{\exp,(XVIII)} &= \int_{t_i}^{t_{i+1}} f_0(s)S_{03}(s)f_{13}(t_{i+k}-s)S_1(t_{i+k}-s)ds\\ &= \frac{a_0a_{13}}{a_1+a_{13}-a_0-a_{03}}\exp(t_{i+k}(a_1+a_{13})\\ &\Big[\exp(t_{i+1}(a_{13}+a_1-a_0-a_{03}))-\exp(t_i(a_{13}+a_1-a_0-a_{03}))\Big].\end{aligned}$$

The total likelihood contribution becomes

$$\begin{aligned}\mathcal{L}_{\exp,1\to3} &= \prod_{i:0\to1\to3}\Bigg(\int_{t_i}^{t_{i+1}}\int_{t_{i+k-1}-s}^{t_{i+k}} f_0(s)f_1(u)f_2(t_{i+k}-s-u)S_{03}(s)S_{13}(u)duds\\ &+\int_{t_i}^{t_{i+1}} f_0(s)S_{03}(s)f_{13}(t_{i+k}-s)S_1(t_{i+k}-s)ds\Bigg)\\ &= \prod_{i:0\to1\to3}\Bigg(\frac{a_0a_1a_2}{(a_2-a_1-a_{13})(a_1+a_{13}-a_0-a_{03})}\exp(-a_2t_{i+k})\\ &\Big[\exp(t_{i+k}(a_2-a_1-a_{13}))-\exp(t_{i+k-1}(a_2-a_1-a_{13}))\Big]\\ &\Big[\exp(t_{i+1}(a_1+a_{13}-a_0-a_{03}))-\exp((a_1+a_{13}-a_0-a_{03})t_i)\Big]\\ &+\frac{a_0a_{13}}{a_1+a_{13}-a_0-a_{03}}\exp(t_{i+k}(a_1+a_{13}))\\ &\Big[\exp(t_{i+1}(a_{13}+a_1-a_0-a_{03}))-\exp(t_i(a_{13}+a_1-a_0-a_{03}))\Big]\Bigg).\end{aligned}$$

The total likelihood contribution for a homogeneous Markov model becomes

$$
\begin{aligned}
\mathcal{L}_{MM,1\to3} = &\prod_{i:0\to1\to3}\Bigg(c_{00}(t_i)c_{01}(t_{i+1}-t_i)c_{11}(t_{i+k}-t_{i+1})q_{13}\\
&+ c_{00}(t_i)c_{01}(t_{i+1}-t_i)c_{11}(t_{i+k-1}-t_{i+1})c_{12}(t_{i+k}-t_{i+k-1})q_2\\
&= \prod_{i:0\to1\to3}\Bigg(\frac{q_0q_{13}}{q_0+q_{03}-q_1-q_{13}}\exp(-t_i(q_0+q_{03}))\exp(-(t_{i+k}-t_{i+1})(q_1+q_{13}))\\
&\Big[\exp(-(t_{i+1}-t_i)(q_1+q_{13}))-\exp(-(q_0+q_{03})(t_{i+1}-t_i))\Big]\\
&+\frac{q_0q_1q_2}{(q_0+q_{03}-q_1-q_{13})(q_1+q_{13}-q_2)}\exp(-t_i(q_0+q_{03}))\\
&\exp(-(t_{i+k-1}-t_{i+1})(q_1+q_{13}))\Big[\exp(-(t_{i+1}-t_i)(q_1+q_{13}))-\exp(-(t_{i+1}-t_i)(q_0+q_{03}))\Big]\\
&\Big[\exp(-q_2(t_{i+k}-t_{i+k-1}))-\exp(-(t_{i+k}-t_{i+k-1})(q_1+q_{13}))\Big]\Bigg)\\
&= \prod_{i:0\to1\to3}\Bigg(\frac{q_0q_1q_2}{(q_2-q_1-q_{13})(q_1+q_{13}-q_0-q_{03})}\exp(-q_2t_{i+k})\\
&\Big[\exp(t_{i+k}(q_2-q_1-q_{13}))-\exp(t_{i+k-1}(q_2-q_1-q_{13}))\Big]\\
&\Big[\exp(t_{i+1}(q_1+q_{13}-q_0-q_{03}))-\exp((q_1+q_{13}-q_0-q_{03})t_i)\Big]\\
&+\frac{q_0q_{13}}{q_1+q_{13}-q_0-q_{03}}\exp(t_{i+k}(q_1+q_{13}))\\
&\Big[\exp(t_{i+1}(q_{13}+q_1-q_0-q_{03}))-\exp(t_i(q_{13}+q_1-q_0-q_{03}))\Big]\Bigg).
\end{aligned}
$$

The likelihoods are equal, and we have therefore shown that our likelihood construction is equal to a homogeneous Markov model in Jackson (2011) in the special case where the transition times are exponentially distributed and all the individuals start at the same time in state 0.

CHAPTER 6

Application: CAV

6.1 Description of the Data

In this chapter, we analyze a dataset called CAV. The data come from a study of the progression of coronary allograft vasculopathy (CAV), a post-transplant complication where there is a deterioration of the arterial walls (Jackson, 2011). Previously, the dataset has been used in different publications on multi-state models, for example in Van Den Hout (2017) and Williams et al. (2020). The data are obtained from a package in R called *msm*, where the manuals are found in Jackson (2011) and Jackson (2019).

The dataset consists of 2816 state observations from 614 individuals. The youngest person to get a transplant is around 6 years old and the oldest person to get a transplant is around 64 years old. In this analysis the starting point is the time of transplantation, which means that all the individuals in the study start at time point 0 in state 0. The individuals then get a transplant at different ages.

In the dataset, there are 4 states:

- State 0: No CAV
- State 1: Mild/moderate CAV
- State 2: Severe CAV
- State 3: Death

The possible transitions are illustrated in Figure 6.1. In this chapter, we therefore consider the four-state illness-death model when analyzing the CAV-data. An analysis of the illness-death model for a modified version of the CAV-data is found in Appendix C.

Some diseases are an irreversible process, which means it is not possible to recover from the diseased state. Progression of coronary allograft vasculopathy is also considered as an irreversible process (Jackson, 2011). This means that the subjects which transfer the opposite way are considered as errors. Therefore, we exclude these subjects from the dataset and we get a total of 2398 observations for 556 individuals. Table 6.1 shows the transitions between the states for the individuals. This means, the number of times an observation in state m was followed by an observation in state ℓ (Jackson, 2011). For example, we have 138 deaths from state 0, 36 deaths from state 1 and 50 deaths from state 2.

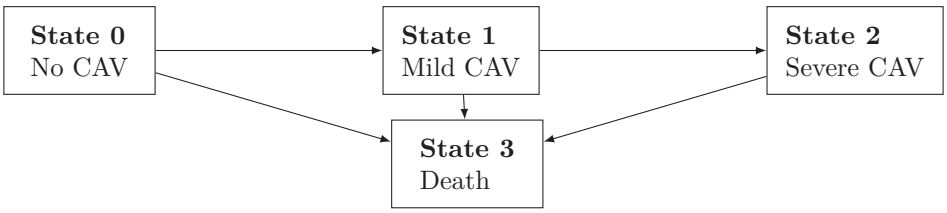


Figure 6.1: Four-state illness-death model for the CAV-data

	State 0	State 1	State 2	State 3
State 0	1233	136	30	138
State 1	0	91	42	36
State 2	0	0	86	50

Table 6.1: Observations of the transitions in each state for the individuals in the CAV-dataset.

The CAV-data consist of yearly examinations up to almost 20 years. Some patiens skipped one or more scheduled examinations. The transitions between the states where the patient is alive are interval-censored. However, if the patient died during the follow-up period, then the time of death is exactly known (Van Den Hout, 2017, p. 5). The patient number, age of the transplant, years since transplant, age of the heart transplant donor, reason for transplantation, sex, cumulative number of rejection states and state at different time points are included in the data. The covariates which are assumed to affect the rate of the progression of CAV is the age of the heart transplant donor (variable dage) and the reason for transplantation (pdiag) (Jackson, 2011).

6.2 Intention

The goal in this chapter is to illustrate how our likelihood construction can be applied on a real dataset. We are especially interested in the shapes of the survival and hazard functions, and how they change for the different models. We analyze the CAV-data using the Gamma process models, which means that the transition times are modeled as the threshold crossing times for Gamma processes. In addition, we consider when the transition times are exponentially distributed. The results of using the exponential distribution in our likelihood construction is, as we know from Chapter 5, equivalent to the homogeneous Markov model. In Section 6.5, we compare the Gamma process models with the Markov models studied in Jackson (2011). We compare the models both with and without covariates by using the AIC-values. In addition, for the models without covariates, we also compare the different survival curves for the total survival probability with the empirical Kaplan-Meier survival curve.

When we construct a Gamma process model, we consider one of two different Gamma processes. When modeling the transition times in the first alternative, we have $Pr(T \geq t) = G(c, at, 1) = G(\exp(\tau), \exp(\nu)t, 1)$, while in the second alternative we have $Pr(T \geq t) = G(c, at^b, 1) = G(\exp(\tau), \exp(\nu)t^b, 1)$. We consider the exponentials, since this ensures that $c > 0$ and $a > 0$. The hazard

functions in Gamma process model alternative 1 can be very different than for Gamma process model alternative 2. How big the differences are, depend on the data.

The analysis is done both with and without covariates. We consider two covariates, where the first one is whether the patient was initially diagnosed with *ihd* or not and the second one is the age of the donor, which we standardize. By calculating and testing whether the parameters for the covariates are significantly different from 0, we are able to find the significant parameters for the covariates. For example, a parameter for a covariate can be significant in the transition from state 1 to state 2, but not in the transition from state 2 to state 3.

We make plots for the different survival functions with pointwise 95% confidence intervals. When we create the confidence intervals, we do a transformation to make sure the lower confidence interval never falls below 0. We start with

$$S^{\log} = -\log(S). \tag{6.1}$$

The variance is calculated using the delta-method from Equation 6.1. The lower confidence band is

$$\exp(-S^{\log} - 1.96\sigma_{S^{\log}}),,$$

and the upper confidence band

$$\exp(-S^{\log} + 1.96\sigma_{S^{\log}}).$$

If $-S^{\log} + 1.96\sigma_{S^{\log}} > 0$ then the upper band will be higher than 1. We often have high variance if there are few individuals making this transition.

6.3 Analysis of the CAV-Data Using Different Parametric Survival Time Models

In this section, we analyze the four-state illness-death model using different parametric survival time models both with and without covariates. More specifically, we consider Gamma process models and when the transition times are exponentially distributed.

As we discussed in Chapter 5, if an individual is only observed in state 0 and state 3, then the individual can transfer directly from state 0 to state 3, the individual can transfer from state 0 to state 1 and then to state 3 or the individual can transfer from state 0 to state 1 to state 2 to state 3. If an individual is observed in states 0, 1 and 3, the individual can either transfer directly from state 0 to state 1 to state 3 or transfer from state 0 to state 1 to state 2 and then to state 3. The different likelihood types are type 1, 2, 3, 4, 6, 7, combination of 5, 16 and 21, and a combination of 15 and 18. The log-likelihood becomes

$$\begin{aligned}
\ell = &\sum_{(I)} \log\Big(S_0(t_n, \boldsymbol{\theta}|x_m)) + \log(S_{03}(t_n, \boldsymbol{\theta}|x_m)\Big) \\
&+ \sum_{(II)} \log\Big(\int_{t_i}^{t_{i+1}} f_0(s, \boldsymbol{\theta}|x_m) S_1(t_n - s, \boldsymbol{\theta}|x_m) S_{03}(s, \boldsymbol{\theta}|x_m) S_{13}(t_n - s, \boldsymbol{\theta}|x_m) ds\Big) \\
&+ \sum_{(III)} \log\Big(\int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s, \boldsymbol{\theta}|x_m) f_1(u, \boldsymbol{\theta}|x_m) S_2(t_n - u - s, \boldsymbol{\theta}|x_m) S_{03}(s, \boldsymbol{\theta}|x_m) S_{13}(u, \boldsymbol{\theta}|x_m) du ds\Big)
\end{aligned}$$

$$+\sum_{(IV)}\log\Big(\int_{t_i}^{t_{i+1}}\int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s,\boldsymbol{\theta}|x_m)f_1(u,\boldsymbol{\theta}|x_m)f_2(t_{i+k+l}-s-u,\boldsymbol{\theta}|x_m)S_{03}(s,\boldsymbol{\theta}|x_m)S_{13}(u,\boldsymbol{\theta}|x_m)duds\Big)$$

$$+\sum_{(V)}\log\Big(\int_{t_i}^{t_{i+1}}\int_{0}^{t_{i+1}-s} f_0(s,\boldsymbol{\theta}|x_m)f_1(u,\boldsymbol{\theta}|x_m)S_2(t_n-s-u,\boldsymbol{\theta}|x_m)S_{03}(s,\boldsymbol{\theta}|x_m)S_{13}(u,\boldsymbol{\theta}|x_m)duds\Big)$$

$$+\sum_{(VI)}\log\Big(\int_{t_i}^{t_{i+1}}\int_{0}^{t_{i+1}-s} f_0(s|x_m)f_1(u|x_m)f_2(t_{i+k}-s-u|x_m)S_{03}(s|x_m)S_{13}(u|x_m)duds\Big)$$

$$+\sum_{(VII)}\log\Big(\int_{t_i}^{t_{i+1}}\int_{0}^{t_{i+1}-s} f_0(s|x_m)f_1(u|x_m)f_2(t_{i+1}-s-u|x_m)S_{03}(s|x_m)S_{13}(u|x_m)duds$$

$$+f_{03}(t_{i+1}|x_m)S_0(t_{i+1}|x_m)+\int_{t_i}^{t_{i+1}} f_0(s|x_m)S_{03}(s|x_m)f_{13}(t_{i+1}-s|x_m)S_1(t_{i+1}-s|x_m)ds\Big)$$

$$+\sum_{(VIII)}\log\Big(\int_{t_i}^{t_{i+1}}\int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s|x_m)f_1(u|x_m)f_2(t_{i+k}-s-u|x_m)S_{03}(s|x_m)S_{13}(u|x_m)duds$$

$$+\int_{t_i}^{t_{i+1}} f_0(s|x_m)S_{03}(s|x_m)f_{13}(t_{i+k}-s|x_m)S_1(t_{i+k}-s|x_m)ds\Big).$$

6.3.1 Without Covariates

We start this analysis by considering models without covariates.

6.3.1.1 Gamma Process Model, Alternative 1

The survival function from state 0 to 1 is of the form

$$S_0(t,\tau_0,\nu_0)=G(\exp(\tau_0),t\exp(\nu_0),1).$$

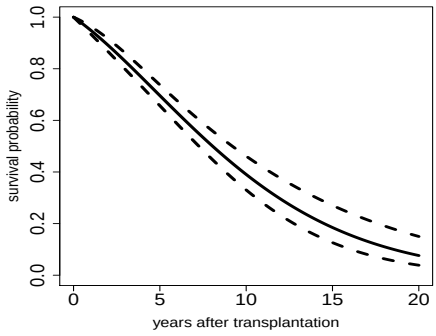
$S_1(t,\tau_1,\nu_1)$, $S_2(t,\tau_2,\nu_2)$, $S_{03}(t,\tau_{03},\nu_{03})$ and $S_{13}(t,\tau_{13},\nu_{13})$ have the same form. The code for this analysis is found in Appendix D.

Parameter	**Estimate** (exp(estimate))	**Standard error**
$\hat{\tau}_0$	-0.369 (0.691)	0.317
$\hat{\nu}_0$	-2.088 (0.124)	0.246
$\hat{\tau}_1$	-1.917 (0.147)	2.005
$\hat{\nu}_1$	-1.768 (0.171)	0.922
$\hat{\tau}_2$	-3.488 (0.0306)	4.349
$\hat{\nu}_2$	-2.480 (0.0837)	1.169
$\hat{\tau}_{03}$	-5.112 (0.00602)	10.478
$\hat{\nu}_{03}$	-4.601 (0.0100)	2.251
$\hat{\tau}_{13}$	1.507 (4.513)	0.588
$\hat{\nu}_{13}$	-0.0116 (0.988)	0.576

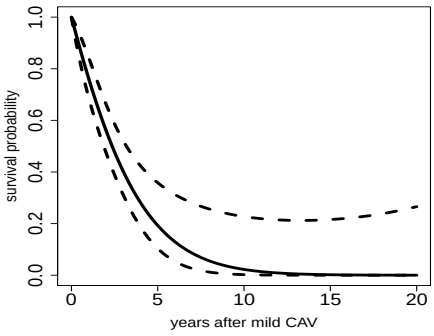
Table 6.2: Estimates and standard errors in a Gamma process model without covariates, alternative 1

We present the maximum likelihood estimates and the corresponding standard errors in Table 6.2. In the parenthesis, we report the exponential of the estimated parameters. We find that the estimated parameters are mostly below zero, except $\hat{\tau}_{13}$. $\exp(\hat{\tau}_{13})$ is above 4, which is much higher than the other parameters. That $\exp(\hat{\tau}_{13})$ is high means that the threshold for reaching state 3 from state 1 is high. However, since $\exp(\hat{\nu}_{13})$ is also relatively high, there

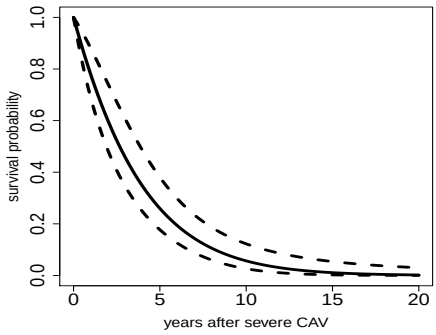
is a steep decrease in the survival function compared to if $\exp(\hat{\nu}_{13})$ was much smaller, for example 0.1.



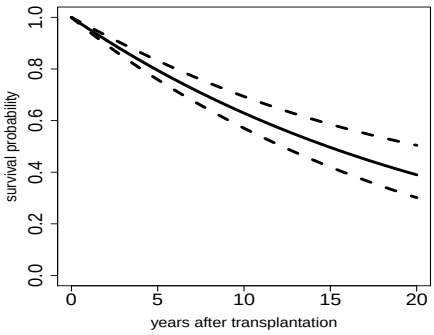
(a) State 0 to state 1



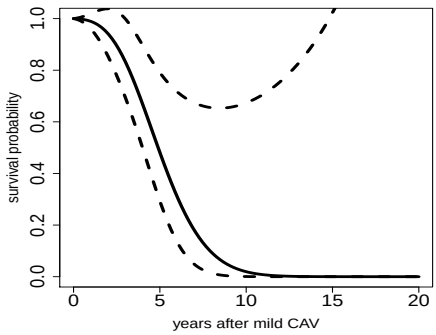
(b) State 1 to state 2



(c) State 2 to state 3



(d) State 0 to state 3



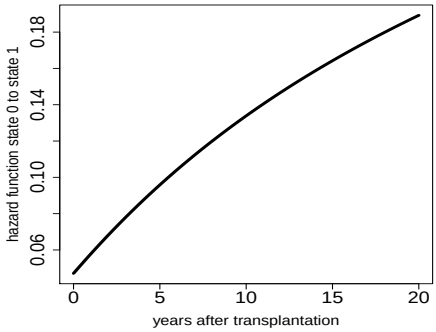
(e) State 1 to state 3

Figure 6.2: Survival functions with pointwise 95% confidence intervals in a Gamma process model without covariates, alternative 1

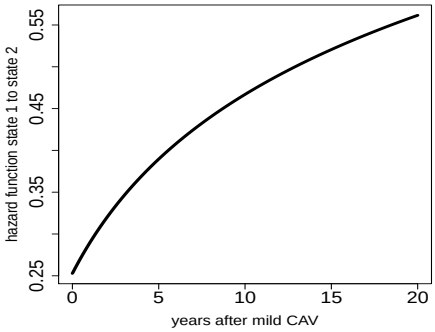
Figure 6.2 shows the plots of the survival functions S_0, S_1, S_2, S_{03} and S_{13} with pointwise 95% confidence intervals. Survival for S_0 means not entering state

1, survival for S_1 means not entering state 2, survival for S_2, S_{03} and S_{13} means not entering the absorbing state 3. A 10-year survival probability is around 0.4 for S_0, 0.05 for S_1, S_2 and S_{13} and 0.6 for S_{03}. The confidence intervals mostly follow the estimated survival function for each of the transitions. The exception is the upper confidence bands for S_{13}. Here, $1.96\sigma_{S^{\log}} > S^{\log}$, which means that $\exp(-S^{\log} + 1.96\sigma_{S^{\log}}) > 1$. Since the variance is high, probably because of few individuals making this transition, the upper confidence band becomes very high. We know that few people are transitioning, because from Table 6.1, we have that only 36 people transition from state 1 to state 3. In addition, we do not know how many of these 36 people have transitioned directly without going through state 2. This may be the reason for the high variance for the estimated parameters of this transition.

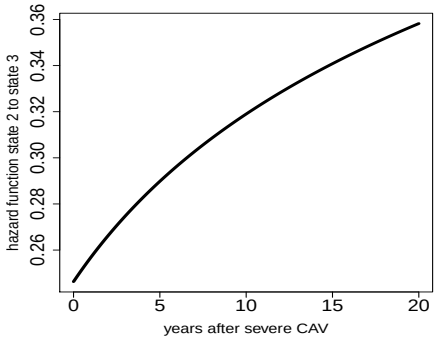
We show the plots for the resulting hazard functions in Figure 6.3. All of the hazard functions are increasing, where the increase mostly seem to wear off with time. The hazard function for $0 \to 3$ is low and almost linear. In addition, this hazard does not seem to change much with time. Of the possible transitions, the hazard function for $0 \to 3$ is the closest one to a constant hazard.



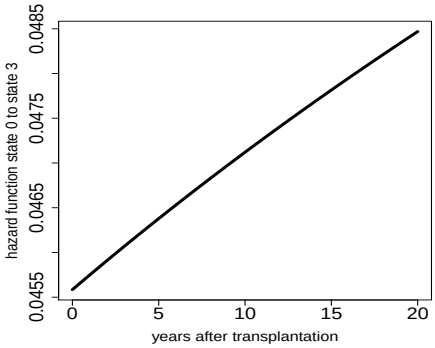
(a) State 0 to state 1



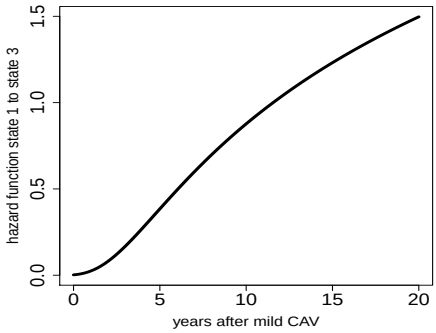
(b) State 1 to state 2



(c) State 2 to state 3



(d) State 0 to state 3



(e) State 1 to state 3

Figure 6.3: Hazard functions in a Gamma process model without covariates, alternative 1

6.3.1.2 Gamma Process Model, Alternative 2

In this part, we consider alternative 2 of the Gamma process model. For alternative 2, the survival functions are of the form

$$S_0(t, \tau_0, \nu_0, b_0) = G(\exp(\tau_0), \exp(\nu_0)t^{b_0}, 1),$$

and similar for $S_1(t, \tau_1, \nu_1, b_1)$, $S_2(t, \tau_2, \nu_2, b_2)$, $S_{03}(t, \tau_{03}, \nu_{03}, b_{03})$ and $S_{13}(t, \tau_{13}, \nu_{13}, b_{13})$.

We present the maximum likelihood estimates and their standard errors in Table 6.3. In the parenthesis of the maximum likelihood estimates, we report the exponential of the maximum likelihood estimates. The estimated τ_k and ν_k parameters, where $k = \{0, 1, 2, 03, 13\}$, are negative, and their exponential are therefore between 0 and 1. These maximum likelihood estimates are also different from the estimated parameters in alternative 1. In addition, we test the hypothesis $H_0 : \hat{b} = 1$. We can only reject this null hypothesis at a 1%-level for $\hat{b}_{03}$. The other p-values are much higher, and we can therefore not reject the null hypothesis for any of the other b-values.

Figure 6.4, shows the plots of the survival functions S_0, S_1, S_2, S_{03} and S_{13} with pointwise 95% confidence intervals. Survival for S_0 means not entering state 1, survival for S_1 means not entering state 2, survival for S_2, S_{03} and S_{13} means not entering the absorbing state 3. A 10-year survival probability is around 0.4 for S_0, 0.05 for S_1, S_{13} and S_2 and 0.7 for S_{03}.

The confidence bands in Figure 6.4 (a), (c) and (d) seem to follow the shape of their corresponding survival function closely. In Figure 6.4 (b) and (e) the upper confidence bands start to increase after some time. After 10 years, the upper confidence band in Figure 6.4 (b) increases slightly. However, the increase in the upper confidence band in Figure 6.4 (e) is bigger and happens earlier. A possible explanation is, as presented in Section 6.2 and for alternative 1, that if the variance for the survival curve exceeds the value of the survival probability, we are in a situation where $-S^{\log} + 1.96\sigma_{S^{\log}} > 0$. Then $\exp(-S^{\log} + 1.96\sigma_{S^{\log}})$ becomes high and the pointwise upper confidence interval for the survival curve may exceed 1.

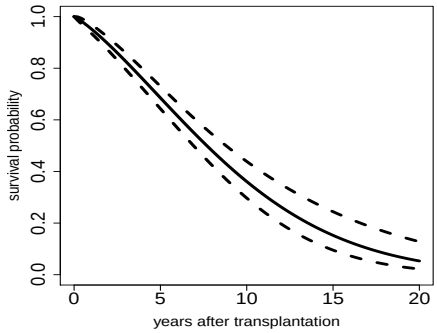
In Figure 6.5, we show the plots of the hazard functions for the four-state illness-death Gamma process model alternative 2. The plotted hazard functions in Figure 6.5 (a) and (b) are increasing and concave. For the plotted hazard function in Figure 6.5(c) there is a steep decrease in the beginning, before there is a steady increase. Also in Figure 6.5 (d), we see from the plot that the hazard function has a steep decrease in the beginning, before the hazard function slowly decreases toward 0. The hazard in Figure 6.5 (e) is increasing and convex. These hazard functions are quite different compared to alternative 1. This is because of the b-parameters, which makes it possible for the hazard function to have a different shape.

The interpretation of the hazard function for the transition $2 \to 3$ is that the instantaneous risk of moving from state 2 to state 3 in a small time interval is very high in the beginning. Then the hazard function drops, before it slowly starts to increase again. This means if you have severe CAV, then the instantaneous risk of dying in a small time interval is high when you first get the diagnose, but then it drops to be smaller. After this drop, the instantaneous risk of dying in a small time interval slowly increases. The hazard function for state 0 to state 3 has a big drop around 0, and then it slowly goes toward 0.

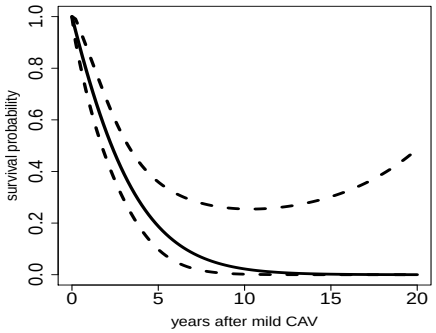
This means that the instantaneous risk of dying in a very small time interval after a transplant is big in the beginning, but after a couple of years, it is almost 0. The hazard function in Figure 6.5 (e) is almost zero in the beginning, before it increases relatively slow until 5 years after mild CAV. Then the hazard function increases quite fast until 10 years. After 10 years, the survival function and the density is almost 0.

Parameter	**Estimate** (exp(estimate) for $\hat{\nu}_k$ and $\hat{\tau}_k$)	**Standard error**	**p-value** ($H_0 : \hat{b}_k = 1$)
$\hat{\tau}_0$	-0.887 (0.412)	1.407	
$\hat{b}_0$	1.144	0.244	0.555
$\hat{\nu}_0$	-2.684 (0.0683)	1.360	
$\hat{\tau}_1$	-2.140 (0.118)	5.588	
$\hat{b}_1$	0.998	0.305	0.995
$\hat{\nu}_1$	-1.839 (0.159)	2.619	
$\hat{\tau}_2$	-2.486 (0.0832)	5.060	
$\hat{b}_2$	0.930	0.312	0.822
$\hat{\nu}_2$	-2.017 (0.133)	2.280	
$\hat{\tau}_{03}$	-0.0992 (0.906)	0.586	
$\hat{b}_{03}$	0.514	0.0845	8.847×10^{-9}
$\hat{\nu}_{03}$	-1.464 (0.231)	0.687	
$\hat{\tau}_{13}$	-0.303 (0.739)	1.847	
$\hat{b}_{13}$	1.683	1.175	0.561
$\hat{\nu}_{13}$	-2.601 (0.0742)	2.921	

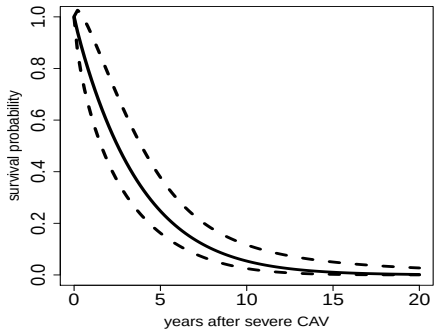
Table 6.3: Estimates, standard errors and p-values testing the null the null hypothesis $H_0 : \hat{b}_k = 1$ in a Gamma process model without covariates, alternative 2



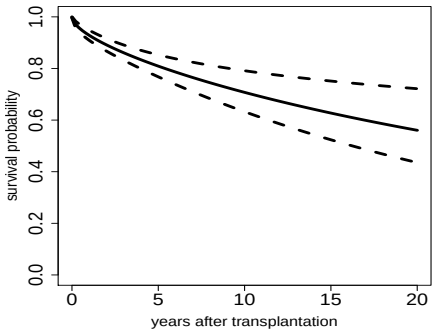
(a) State 0 to state 1



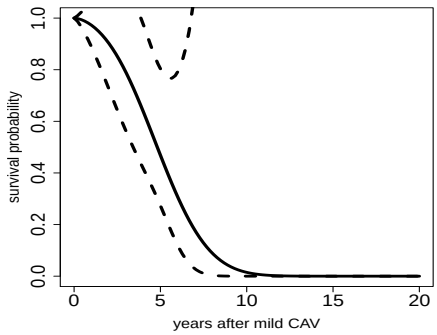
(b) State 1 to state 2



(c) State 2 to state 3

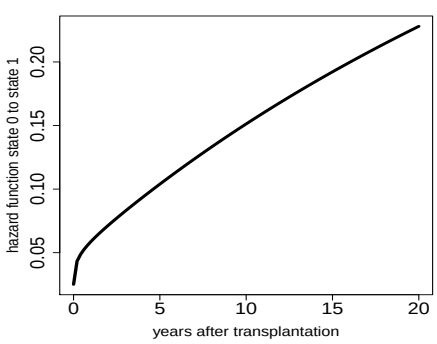


(d) State 0 to state 3

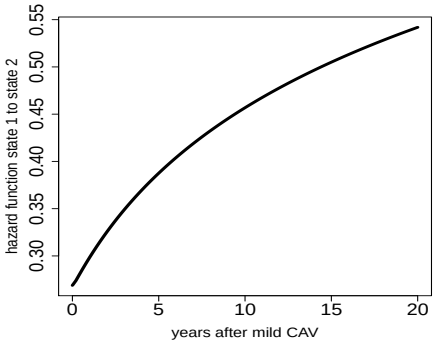


(e) State 1 to state 3

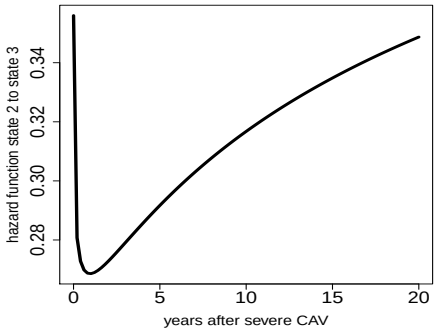
Figure 6.4: Survival functions with pointwise 95% confidence intervals in a Gamma process model without covariates, alternative 2



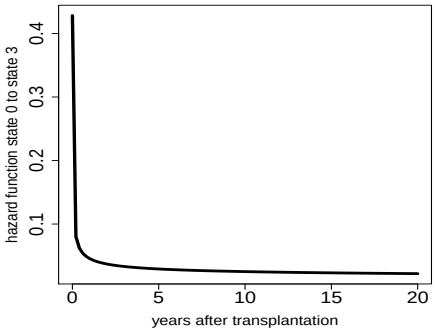
(a) State 0 to state 1



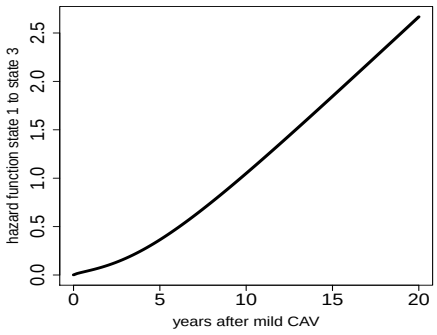
(b) State 1 to state 2



(c) State 2 to state 3



(d) State 0 to state 3



(e) State 1 to state 3

Figure 6.5: Hazard functions in a Gamma process model without covariates, alternative 2

6.3.1.3 Exponential Distribution

Another possibility is to consider the transition times to follow an exponential distribution.

$$S_0(t, a_0) = Pr(T_0 \geq t) = 1 - (1 - \exp(-a_0 t)) = \exp(-a_0 t)$$

and

$$f_0(t, a_0) = -\frac{\partial S(t, a_0)}{\partial t} = a_0 \exp(-a_0 t)$$

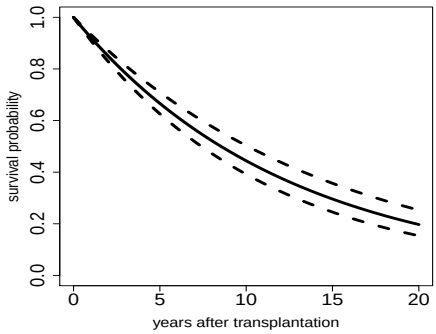
and corresponding survival and density functions $S_1(t, a_1)$, $f_1(t, a_1)$, $S_2(t, a_2)$, $f_2(t, a_2)$, $S_{03}(t, a_{03})$, $f_{03}(t, a_{03})$, $S_{13}(t, a_{13})$ and $f_{13}(t, a_{13})$.

We present the maximum likelihood estimates and the standard errors in Table 6.4. As discussed in Chapter 5, the hazards are equal to the a-parameter, which means the hazards are constant.

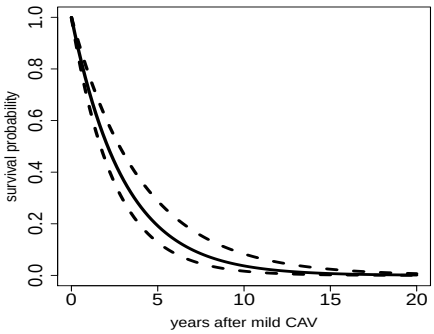
In Figure 6.6 we show the plots of the survival functions S_0, S_1, S_2, S_{03} and S_{13} with pointwise 95% confidence intervals. The survival functions can be interpreted in the same way as for the Gamma process models. The 10-year survival probability for S_0 is around 0.5, for S_1 and S_2 it is around 0.05, 0.7 for S_{03} and around 0.5 for S_{13}. The confidence intervals are mostly narrow. The exception is S_{13}, where the confidence bands are wide. As we discussed for the Gamma process models alternative 1 and 2, we have few transitions from state 1 to state 3 and therefore a higher variance.

Parameter	**Estimate**	**Standard error**
$\hat{a}_0$	0.0812	0.00637
$\hat{a}_1$	0.330	0.0415
$\hat{a}_2$	0.289	0.0358
$\hat{a}_{03}$	0.0445	0.00486
$\hat{a}_{13}$	0.0635	0.0295

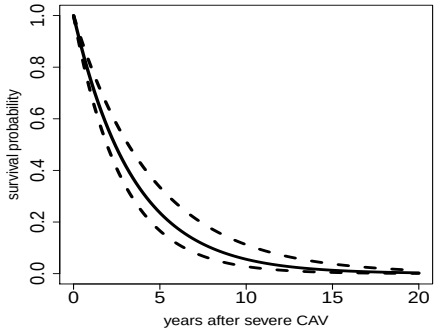
Table 6.4: Estimates and standard errors in a exponential distribution model without covariates



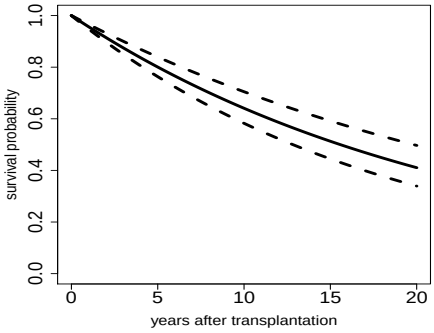
(a) State 0 to state 1



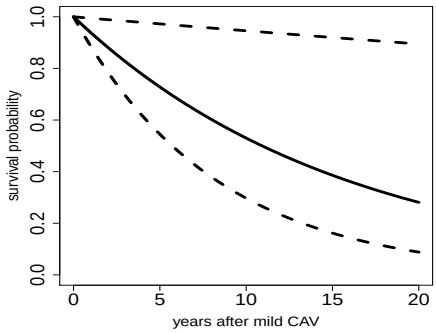
(b) State 1 to state 2



(c) State 2 to state 3



(d) State 0 to state 3



(e) State 1 to state 3

Figure 6.6: Survival functions with pointwise 95% confidence intervals in a exponential distribution model without covariates

6.3.2 With Covariates

According to Jackson (2011), the age of the heart transplant donor and the primary diagnosis/reason for transplantation affect the progression of CAV.

In Jackson (2011), they create a binary variable *ihd* representing ischaemic heart disease. Other covariates which may be included are sex and cumulative number of rejection episodes. We include the variable *ihd*, as factor with levels 0 and 1, and age of the heart transplant donor as covariates. The age of the heart transplant donor is standardized, so

$$\text{dage}_{\text{st}} = \frac{\text{dage} - \mu_{\text{dage}}}{\sigma_{\text{dage}}} = \frac{\text{dage} - 30.622}{12.280}.$$

6.3.2.1 Gamma Process Model, Alternative 1

We include the covariates in the threshold $\exp(\tau)$ for the Gamma process model alternative 1 without covariates. The survival function from state 0 to state 1 becomes

$$S_0(t, \boldsymbol{\theta}) = G(\exp(\beta_{0,0} + \beta_{1,0} ihd + \beta_{2,0} \text{dage}_{\text{st}}), t \exp(\nu_0), 1),$$

and the same form for $S_1(t, \boldsymbol{\theta})$, $S_2(t, \boldsymbol{\theta})$, $S_{03}(t, \boldsymbol{\theta})$ and $S_{13}(t, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of all the parameters.

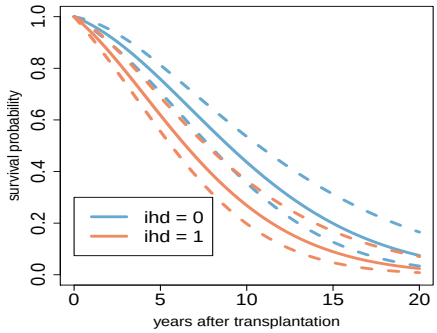
We present the estimated maximum likelihood parameters, their standard errors and the p-values from testing $H_0 : \beta = 0$ in Table 6.5. At a 1%-level, only $\hat{\beta}_{1,0}$ and $\hat{\beta}_{2,0}$ are significant.

Figure 6.7 shows the plots for the survival functions S_0, S_1, S_2, S_{03} and S_{13} with pointwise 95% confidence intervals for the case where $dage_{st} = -0.132$, which is the median of the standardized donor age, and $ihd = 0$ or $ihd = 1$, where $ihd = 1$ means that the patient was initially diagnosed with *ihd*. Survival for S_0 means not entering state 1, survival for S_1 means not entering state 2, survival for S_2, S_{03} and S_{13} means not entering the absorbing state 3. A 10-year survival probability when $ihd = 0$ is around 0.5 for S_0, 0.05 for S_1, 0.05 for S_2, 0.7 for S_{03} and 0.3 for S_{13}. A 10-year survival probability when $ihd = 1$ is around 0.3 for S_0, 0.05 for S_1, 0.1 for S_2, 0.6 for S_{03} and 0.05 for S_{13}. For the same donor age, having *ihd* as the reason for transplantation decreases all the survival probabilities except S_2. Note the upper confidence band for the transition $1 \to 3$ is above 1, which is probably because of the same reasons discussed in Section 6.3.1, where $\exp(-S^{\log} + 1.96\sigma_{S^{\log}}) > 1$.

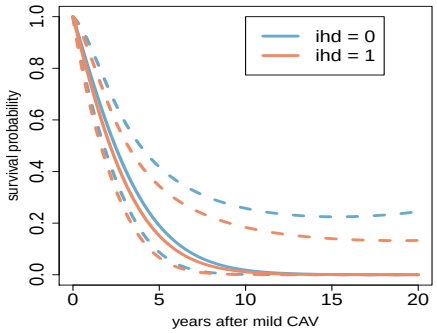
We show the plots for the hazard functions in Figure 6.8. The shape of the hazard functions in Figure 6.8 are increasing. When $ihd = 0$, the hazard functions are mostly lower than when $ihd = 1$. The exception is the hazard function for the transition $2 \to 3$ illustrated in Figure 6.8 (c). This means that the instantaneous risk in a very small time interval of transitioning from state 2 to 3 is a bit higher when $ihd = 0$ than when $ihd = 1$. In all the other hazard functions, the opposite applies. Note that the hazard function for the transition $0 \to 3$ is relatively constant.

Parameter	**Estimate**	**Standard error**	**p-value** ($H_0 : \beta = 0$)
$\hat{\beta}_{0,0}$	0.160	0.243	
$\hat{\beta}_{1,0}$	-0.422	0.135	0.00177
$\hat{\beta}_{2,0}$	-0.252	0.0731	0.000566
$\hat{\nu}_0$	-1.767	0.214	
$\hat{\beta}_{0,1}$	-1.322	1.414	
$\hat{\beta}_{1,1}$	-0.238	0.443	0.593
$\hat{\beta}_{2,1}$	0.340	0.284	0.231
$\hat{\nu}_1$	-1.525	0.771	
$\hat{\beta}_{0,2}$	-3.144	2.710	
$\hat{\beta}_{1,2}$	1.331	1.053	0.206
$\hat{\beta}_{2,2}$	0.221	0.376	0.556
$\hat{\nu}_2$	-2.084	0.760	
$\hat{\beta}_{0,03}$	-3.063	7.763	
$\hat{\beta}_{1,03}$	-0.896	2.174	0.680
$\hat{\beta}_{2,03}$	-1.352	3.297	0.682
$\hat{\nu}_{03}$	-4.162	2.743	
$\hat{\beta}_{0,13}$	1.412	0.910	
$\hat{\beta}_{1,13}$	-0.664	0.534	0.214
$\hat{\beta}_{2,13}$	0.466	0.296	0.115
$\hat{\nu}_{13}$	-0.536	0.805	

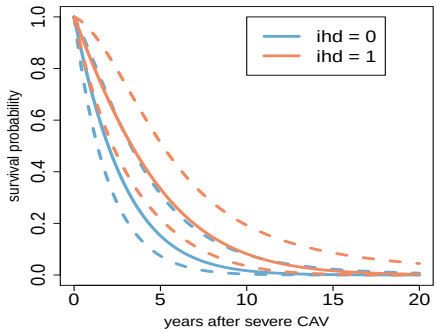
Table 6.5: Estimates, standard errors and p-values testing the null hypothesis $H_0 : \beta = 0$ in a Gamma process model with covariates, alternative 1



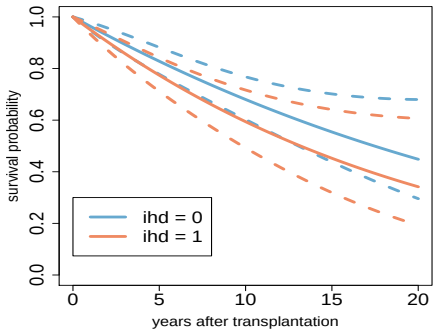
(a) State 0 to state 1



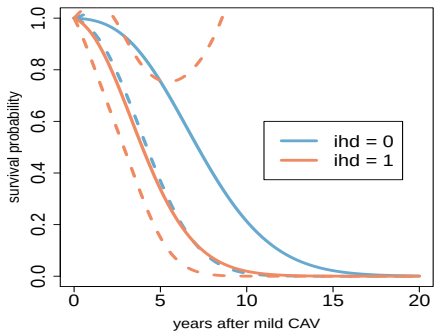
(b) State 1 to state 2



(c) State 2 to state 3

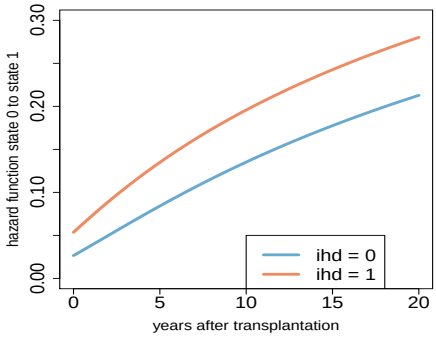


(d) State 0 to state 3

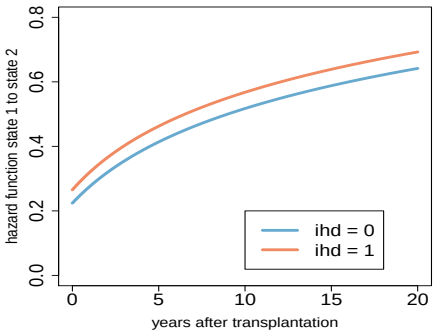


(e) State 1 to state 3

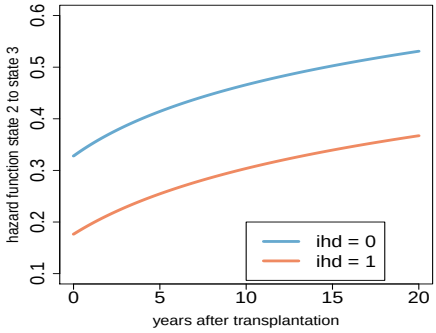
Figure 6.7: Survival functions with pointwise 95% confidence intervals in a Gamma process model with covariates, alternative 1. The covariate values are $dage_{st} = -0.132$ and $ihd = 0$ or $ihd = 1$



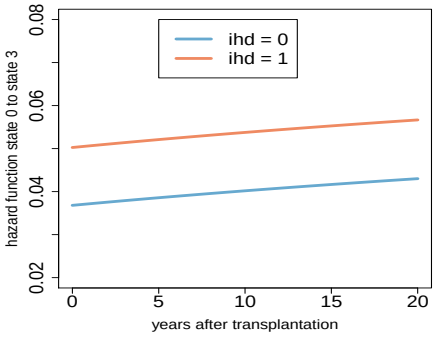
(a) State 0 to state 1



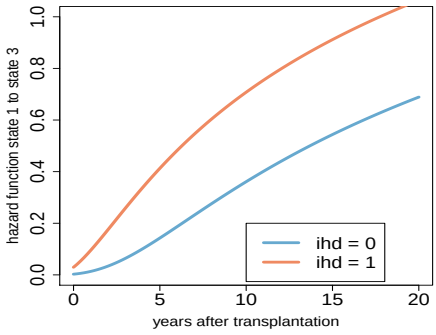
(b) State 1 to state 2



(c) State 2 to state 3

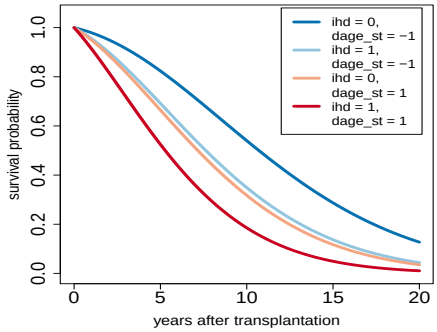


(d) State 0 to state 3

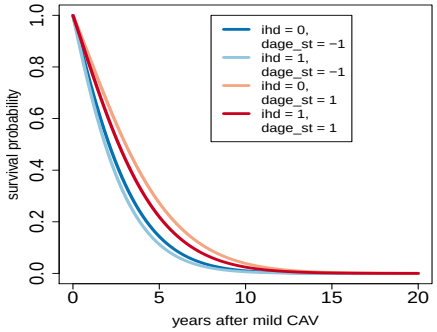


(e) State 1 to state 3

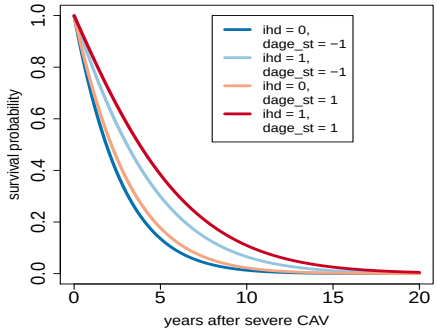
Figure 6.8: Hazard functions in a Gamma process model with covariates, alternative 1. The covariate values are $dage_{st} = -0.132$ and $ihd = 0$ or $ihd = 1$



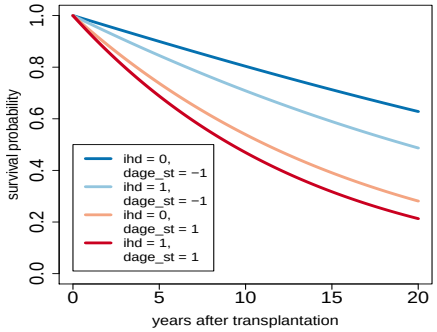
(a) State 0 to state 1



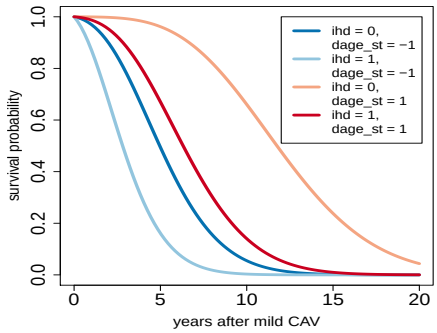
(b) State 1 to state 2



(c) State 2 to state 3



(d) State 0 to state 3



(e) State 1 to state 3

Figure 6.9: Survival functions for different values of $dage_{st}$ and ihd in a Gamma process model with covariates, alternative 1

We show the plots of the survival functions for $dage_{st} = -1$ or $dage_{st} = 1$ and $ihd = 0$ or $ihd = 1$ in Figure 6.9. This means that the age of the donor is around around 18 when $dage_{st} = -1$ or 43 years when $dage_{st} = 1$. In Figure 6.9 (a) and (d), the survival probability is lowest if the donor is older and the person

was initially diagnosed with *ihd*. In Figure 6.9 (e), the survival probability from state 1 to state 3 is highest if the individual has a donor which is a bit older.

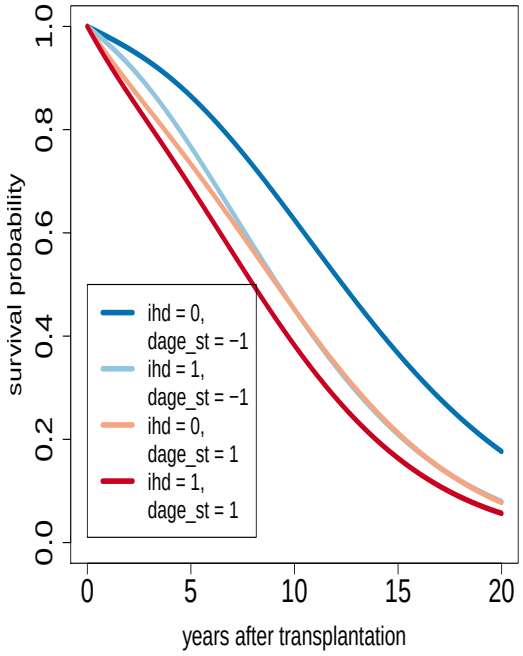


Figure 6.10: Total survival probability for different values of $dage_{st}$ and *ihd* in a Gamma process model with covariates, alternative 1.

Figure 6.10 shows the total survival probability function. How we derive the formula for the total survival probability is found in Section 6.5. The total survival probability is lowest when the donor is older and the person was initially diagnosed with *ihd*. If the donor is older and the person did not initially have *ihd* or if the donor is younger but the person was initially diagnosed with *ihd*, have around the same survival probability. As expected, the total survival probability is higher when the individual was not initially diagnosed with *ihd* and the donor is younger.

6.3.2.2 Gamma Process Model, Alternative 2

It is also possible to do an analysis with covariates using the same motor function as in Gamma process model alternative 2 without covariates. The computational burden then increases, since we end up with optimizing 25 parameters. Because of the computational burden, we consider a simplified model. Specifically, we only include covariates in the first transition and we only include the motor function t^b for the transition from state 0 to state 3. For the other transitions, we use the Gamma processes from the Gamma process model alternative 1 without covariates. We choose to include covariates in the transition $0 \to 1$ because it was only for this transition that the covariates appeared to have a

significant effect in the Gamma process model alternative 1. Also, we include b_{03} in the transition $0 \to 3$ because it was the only b-parameter significantly different from 1 in the Gamma process model alternative 2 without covariates.

For the transition $0 \to 1$, we include the covariates in the threshold in the following way

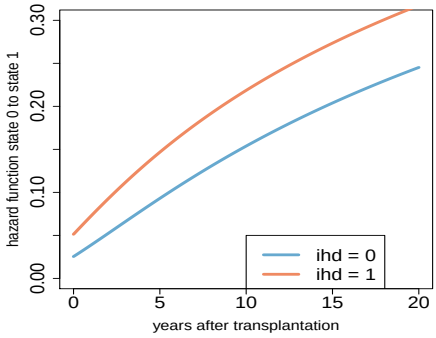
$$S_0(t, \boldsymbol{\theta}) = G(\exp(\beta_{0,0} + \beta_{1,0}\text{ihd} + \beta_{2,0}\text{dage}_{\text{st}}), t\exp(\nu_0), 1).$$

$S_1(t, \boldsymbol{\theta})$, $S_2(t, \boldsymbol{\theta})$, and $S_{13}(t, \boldsymbol{\theta})$ have the same form as in the Gamma process model alternative 1 without covariates and $S_{03}(t, \boldsymbol{\theta})$ has the same form as in the Gamma process model alternative 2 without covariates.

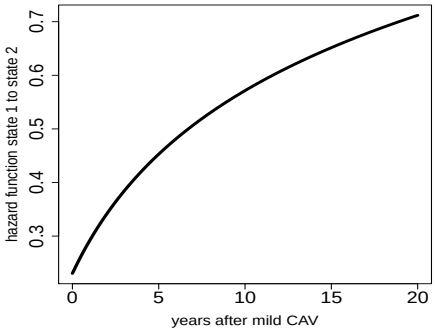
We present the estimated maximum likelihood parameters, their standard errors and the p-values for testing the hypothesis $H_0 : \beta = 0$ in Table 6.6. Both of the parameters for the covariates are significant at a 1%-level.

Parameter	**Estimate**	**Standard error**	**p-value** ($H_0 : \beta = 0$)
$\hat{\beta}_{0,0}$	0.245	0.217	
$\hat{\beta}_{1,0}$	-0.394	0.120	0.00103
$\hat{\beta}_{2,0}$	-0.271	0.0669	0.0000510
$\hat{\nu}_0$	-1.638	0.194	
$\hat{\tau}_1$	-1.264	1.263	
$\hat{\nu}_1$	-1.418	0.713	
$\hat{\tau}_2$	-3.489	4.203	
$\hat{\nu}_2$	-2.454	1.129	
$\hat{\tau}_{03}$	-0.319	0.909	
$\hat{b}_{03}$	0.486	0.0860	
$\hat{\nu}_{03}$	-1.727	1.000	
$\hat{\tau}_{13}$	-0.582	2.741	
$\hat{\nu}_{13}$	-1.762	1.973	

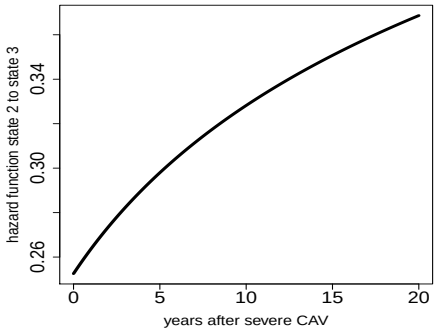
Table 6.6: Estimates, standard errors and p-values testing the null hypothesis $H_0 : \beta = 0$ in a Gamma process model with covariates, alternative 2



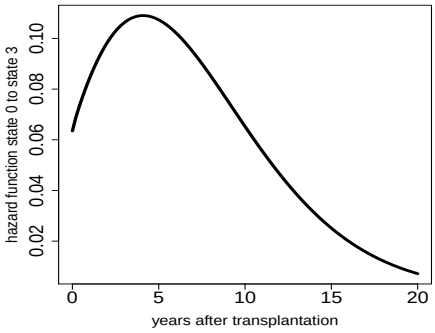
(a) State 0 to state 1



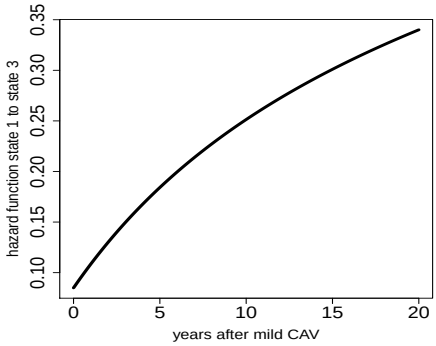
(b) State 1 to state 2



(c) State 2 to state 3



(d) State 0 to state 3



(e) State 1 to state 3

Figure 6.11: Hazard functions in a Gamma process model with covariates, alternative 2. The covariate values for the transition $0 \to 1$ are $dage_{st} = -0.132$ and $ihd = 0$ or $ihd = 1$

Figure 6.11 shows the hazard functions. All of the hazard functions are increasing, except the hazard function for the transition $0 \to 3$. For this transition, the hazard function increases in the beginning, before it reaches a

maximum point. Then the hazard starts to decrease again. This means the instantaneous risk of transitioning from state 0 to state 3 increases from 0 to 5 years, but then the instantaneous risk start to decrease.

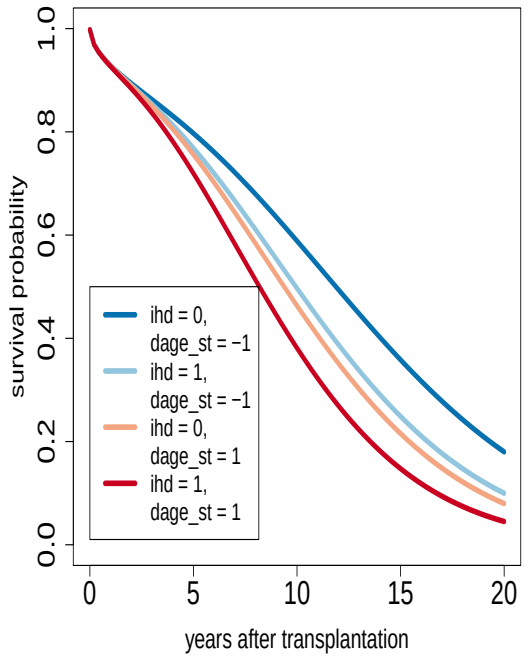


Figure 6.12: Total survival probability for different values of $dage_{st}$ and *ihd* in a Gamma process model with covariates, alternative 2

The interpretations of the survival functions do not change much from previous models, but we include a plot of the total survival probability for different values of the covariates. Figure 6.12 shows this plot of the total survival probability function for the combinations $dage_{st} = -1$ or $dage_{st} = 1$ and *ihd* $= 0$ or *ihd* $= 1$. How we derive the formula for the total survival probability is found in Section 6.5. The total survival probability is lowest when the donor is older and the person was initially diagnosed with *ihd*. The survival probability when the individual was initially diagnosed with *ihd* and the donor is younger is almost equal to the survival probability when the donor is older and the person did not initially have *ihd*. As we expected, the highest survival probability is when the individual was not initially diagnosed with *ihd* and the donor is young.

6.4 Analysis of the CAV-Data Using the Markov Models in Jackson (2011)

In this section, we do some of the same analysis as in Jackson (2011), with small modifications. We start with the homogeneous Markov model, before

we continue with the inhomogeneous model. In the end, we consider the inhomogeneous Markov model with covariates.

6.4.1 Homogeneous Markov Model

Previously, the CAV-dataset has been analyzed with a Markov model in Jackson (2011). We therefore construct a Markov model based on the Markov model Jackson (2011) in order to compare with the Gamma process models. First, we define the initial transition matrix for the analysis of the CAV-data. In Jackson (2011) they define an initial transition matrix and a misclassification matrix for the observations going the wrong way. In our analysis, we have removed the observations going the wrong way, but we still use the same initial transition matrix, in this case called $\boldsymbol{Q}_0$.

$$\boldsymbol{Q}_0 = \begin{bmatrix} 0 & 0.1 & 0 & 0.04 \\ 0 & 0 & 0.3 & 0.05 \\ 0 & 0 & 0 & 0.3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

This means we initially assume the instantaneous risk of moving from state 0 to state 1 in a very small time interval to be 0.1. This initial transition matrix is used as the start matrix for the optimization of the log-likelihood.

	Well	**Mild**	**Severe**	**Death**
Well	-0.126	0.0812	0	0.0445
Mild	0	-0.393	0.330	0.0635
Severe	0	0	-0.289	0.289

Table 6.7: Estimated transition intensities in a time-homogeneous Markov model

Table 6.7 presents the transition matrix after the analysis is done. For instance, the estimated instantaneous risk of moving from state 0, **Well**, to state 1, **Mild**, is 0.081.

Figure 6.13 shows the plot of the total survival probability function from state 0 to state 3 in a time-homogeneous Markov model. The empirical line is the Kaplan-Meier estimate and gives an estimate of the "observed" survival probability (Jackson, 2011). We discussed the use of a Kaplan-Meier estimate as a measure of fit in Chapter 2. The fitted Markov model is quite close to the empirical model until 10 years after transplantation. However, after 10 years, the survival probability in the Markov model and the empirical probability differs.

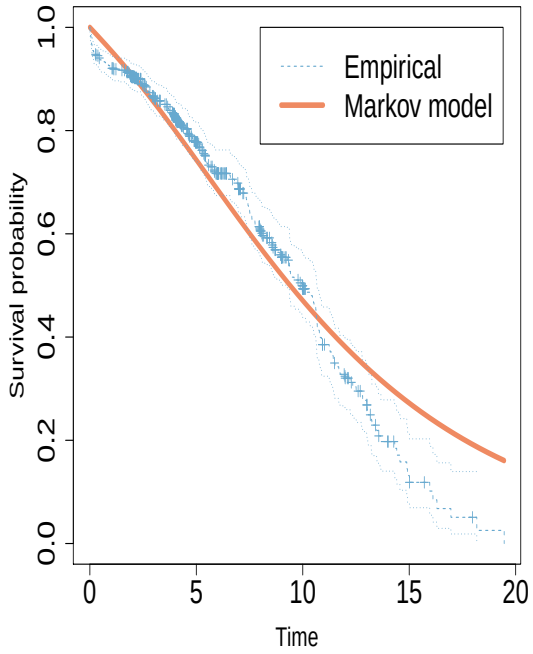


Figure 6.13: Total survival probability from Kaplan-Meier estimates and from a time-homogeneous Markov model

6.4.2 Time-Inhomogeneous Markov Model

A possible time-inhomogeneous model is a model where the transition matrix $\boldsymbol{Q}$ is piecewise-constant. An example is if a covariate varies continuously through time, but is only observed at the same times as the state of the Markov process. This means that the piecewise-constant covariate can change at other times than $(t_{i1}, \ldots, t_{in_i})$. The solution is to take the sum of the likelihood over the unknown observed state when the covariates change in time (Jackson, 2011).

A way of creating a time-inhomogeneous model, is to change the intensities at the same time for all of the individuals. In this analysis, we construct an inhomogeneous model to the CAV-data by letting the intensities change 5 years after transplantation. In the *msm*-package, a binary covariate called "timeperiod" is created as a factor. The levels are then the baseline $(-\infty, 5]$ and the second time period is $[5, \infty)$ (Jackson, 2011). The probability of getting CAV or dying changes as time goes by, which makes this change realistic.

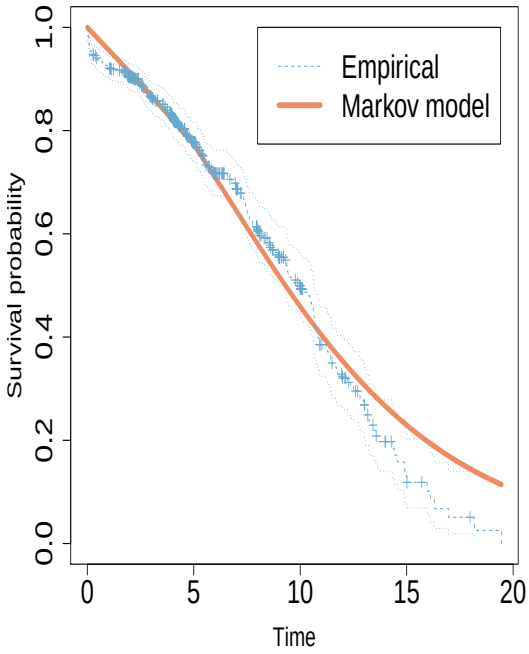


Figure 6.14: Total survival probability from Kaplan-Meier estimates and from a time-inhomogeneous Markov model

Figure 6.14 shows the total survival probability function from state 0 to state 3 in a time-inhomogeneous Markov model. The fitted model is quite close to the empirical model until around 13 years after transplantation. However, after 13 years, the fitted and empirical probability differs, but is closer to the empirical results than the time-homogeneous model.

6.4.3 With Covariates

We include the same covariates as for the Gamma process models. These are the primary diagnosis and the standardized age of the donor. We replace $q_{m\ell}$ with $q_{m\ell}(ihd, dage_{st})$ and according to Jackson (2011) it is on the form

$$q_{m\ell}(ihd, dage_{st}) = q_{m\ell}^{(0)} \exp(\beta_{0,m\ell} ihd + \beta_{1,m\ell} dage_{st}).$$

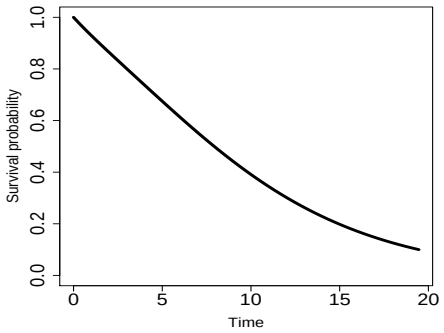
We present the transition intensities with hazard ratios for each covariate in the homogeneous Markov model in Table 6.8, while we present the results for the inhomogeneous Markov model in Table 6.9. The baselines are when the covariates are set to their means. *ihd* is included as a factor, where the factor levels are 0 and 1. When we present the plots of the survival functions, $ihd = 0$ means the patient was not initially diagnosed with *ihd*.

	Baseline	**dage**$_{st}$	***ihd***
Well → Well	-0.120		
Well → Mild	0.0808	1.238	1.651
Well → Death	0.0388	1.570	1.263
Mild → Mild	-0.391		
Mild → Severe	0.336	0.818	1.212
Mild → Death	0.0558	0.339	2.988
Severe → Severe	-0.308		
Severe → Death	0.308	0.909	0.647

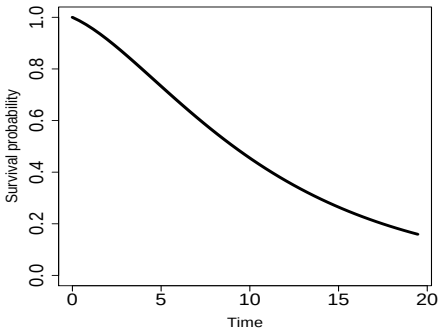
Table 6.8: Estimated transition intensities in a time-homogeneous Markov model with covariates

	Baseline	**dage**$_{st}$	***ihd***	**Timeperiod** $[5, \infty)$
Well → Well	-0.120			
Well → Mild	0.0854	1.306	1.638	2.442
Well → Death	0.0348	1.586	1.294	0.768
Mild → Mild	-0.393			
Mild → Severe	0.336	0.803	1.101	0.772
Mild → Death	0.0574	0.449	3.071	2.753
Severe → Severe	-0.259			
Severe → Death	0.259	0.936	0.608	1.435

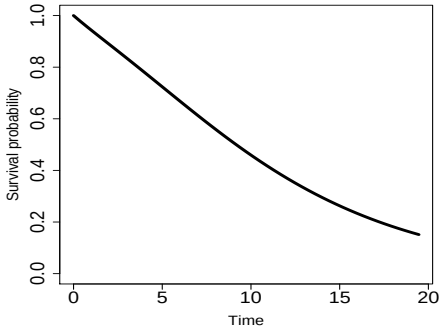
Table 6.9: Estimated transition intensities in a time-inhomogeneous Markov model with covariates



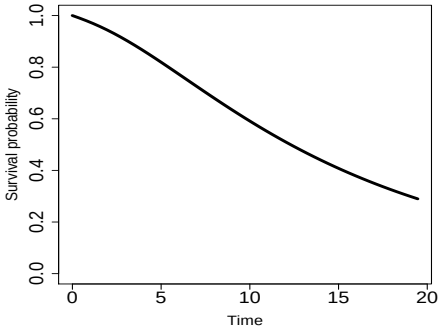
(a) $dage_{st} = 1, ihd = 1$



(b) $dage_{st} = -1, ihd = 1$

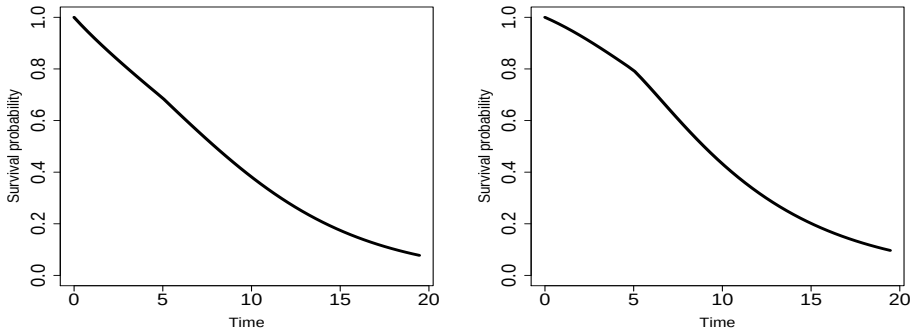


(c) $dage_{st} = 1, ihd = 0$



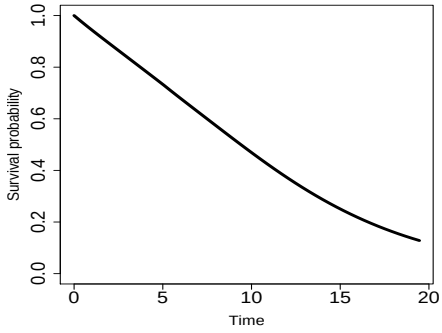
(d) $dage_{st} = -1, ihd = 0$

Figure 6.15: Survival functions for different values of *ihd* and $dage_{st}$ in a time-homogeneous Markov model with covariates

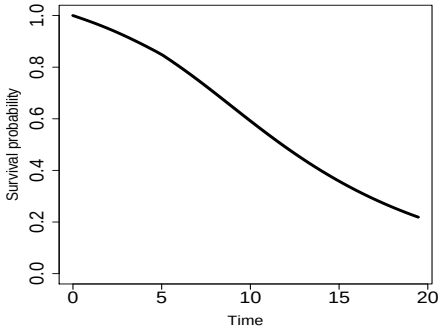


(a) $dage_{st} = 1, ihd = 1$

(b) $dage_{st} = -1, ihd = 1$



(c) $dage_{st} = 1, ihd = 0$



(d) $dage_{st} = -1, ihd = 0$

Figure 6.16: Survival functions for different values of *ihd* and $dage_{st}$ in a time-inhomogeneous Markov model with covariates

Figures 6.15 and 6.16 show the total survival probability for different values of $dage_{st}$ and *ihd* for a homogeneous and inhomogeneous Markov model. The 10-year survival probability from state 0 in both of the Markov models is highest when $dage_{st} = -1$ and $ihd = 0$. This means that the 10-year survival probability from state 0 is highest when the donor is younger and the individual was not initially diagnosed with *ihd*.

6.5 Comparison of the Gamma Process Models to the Markov models

In this section, we compare the Gamma process models to the homogeneous and inhomogeneous Markov models. We compare them both with and without covariates. In general, we find that the Gamma process models have a lower AIC than the Markov models.

6.5.1 Without Covariates

As discussed in Chapter 5, the transition intensities can be compared to the hazard functions. We summarize the transition intensities for the Markov model/exponential distribution model and the hazard functions for the Gamma process models in Table 6.10. Note that the hazard functions for the two alternatives of the Gamma process models are not independent of time.

	GP alternative 1 (**min**, **max**)	**GP alternative 2** (**min**, **max**)	**MM/Exp.**
Well → Mild	(0.0471, 0.189)	(0.0250, 0.228)	0.0812
Well → Death	(0.0456, 0.0485)	(0.0219, 0.428)	0.0445
Mild → Severe	(0.253, 0.562)	(0.269, 0.542)	0.330
Mild → Death	(0.00202, 1.497)	(0.00128, 2.667)	0.0635
Severe → Death	(0.246, 0.358)	(0.269, 0.356)	0.289

Table 6.10: Hazard for the models from Gamma process model alternative 1, Gamma process model alternative 2 and the time-homogeneous Markov models without covariates

In Table 6.10, the Markov model and the exponential distribution have equal hazard values, in line with our results in Chapter 5. For the Gamma process models, the hazard function are in some cases high and above 1. This is for example seen in **Mild → Death**. When the hazard function is above 1, the instantaneous risk of dying is very high, and it is very unlikely that an individual survives in that time period. Other than the state $1 \to 3$ transition, the Gamma process models and the Markov model seems to correspond well with each other.

In the next part, we compare the total survival probability. Therefore, we calculate the total survival probability in our proposed model. We start in state 0 and want to investigate the probability of not being dead before time t. There are three possibilities

1. Stay in state 0
2. Move from state 0 to state 1 and stay there
3. Move from state 0 to state 1 to state 2 and stay there

The probability of staying in state 0 is

$$Pr(T_0 > t, T_{03} > t) = S_0(t)S_{03}(t).$$

The probability of moving from state 0 to state 1 between 0 and t, and staying there is

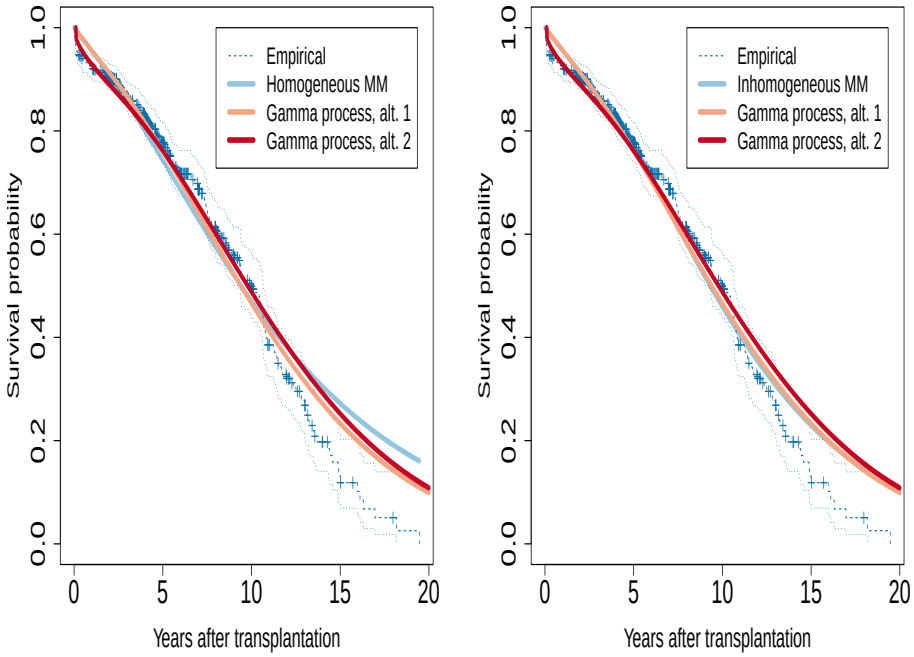
$$\begin{aligned} Pr(T_0 + T_1 > t) &= Pr(0 < T_0 < t, T_0 + T_1 > t, T_0 + T_{13} > t, T_{03} > T_0) \\ &= \int_0^t f_0(s)S_1(t-s)S_{03}(s)S_{13}(t-s)ds. \end{aligned}$$

The probability of moving from state 0 to state 1 to state 2 between 0 and t, and staying there is

$$Pr(T_0 + T_1 + T_2 > t) = Pr(0 < T_0 < t, 0 < T_0 + T_1 < t, T_0 + T_1 + T_2 > t, T_{03} > T_0, T_{13} > T_1)$$
$$= \int_0^t \int_0^{t-s} f_0(s) f_1(u) S_2(t-s-u) S_{03}(s) S_{13}(u) du ds.$$

Therefore, the total survival is

$$Pr(\text{total survival}) = Pr(\text{Staying in state 0}) + Pr(\text{Staying in state 1}) + Pr(\text{Staying in state 2})$$
$$= S_0(t) S_{03}(t) + \int_0^t f_0(s) S_1(t-s) S_{03}(s) S_{13}(t-s) ds$$
$$+ \int_0^t \int_0^{t-s} f_0(s) f_1(u) S_2(t-s-u) S_{03}(s) S_{13}(u) du ds.$$



(a) Comparison to homogeneous MM (b) Comparison to inhomogeneous MM

Figure 6.17: Total survival probability function without covariates from Kaplan-Meier estimates, Gamma process models alternative 1 and 2 and a homogeneous or inhomogeneous Markov model

Figure 6.17 (a) shows the total survival probability for the Gamma process models, alternative 1 and alternative 2, the homogeneous Markov model and the empirical Kaplan-Meier estimate. As we discussed in Chapter 2, the Kaplan-Meier estimate can be used as an informal way of estimating goodness-of-fit. If the survival probability curves are outside of the 95% confidence interval, then the model does not fit the data.

Both of the Gamma process model alternatives seem to follow the empirical survival probability closer than the homogeneous Markov model. The Gamma process model alternative 2 is closer to the empirical survival probability in the beginning. Then after around 10 years, the Gamma process model alternative 1 is closer to the empirical survival probability. Both of the Gamma process models are almost always inside the 95% confidence interval bands. They are barely outside after around 15 years. However, the Markov model is quite close to the the lower 95% confidence band in the beginning and upper confidence band in the enda.

Figure 6.17 (b) shows the total survival probability for the Gamma process models alternative 1 and alternative 2, the inhomogeneous Markov model and the empirical Kaplan-Meier estimate. The Gamma process model alternative 1 and the inhomogeneous Markov model are very close to each other, so close it is hard to separate the two lines in the plot. The inhomogeneous Markov model is therefore closer to the empirical Kaplan-Meier estimate than the homogeneous Markov model.

Model	$-2\log(\mathcal{L})$	k	**AIC**
MM, homogeneous/Exp.	2877.069	5	2887.07
MM, inhomogeneous	2853.011	10	2873.01
GP alternative 1	2849.462	10	2869.46
GP alternative 2	2812.062	15	2842.06

Table 6.11: AIC in a four-state illness death models, without covariates

We compare the models by calculating the AIC for the different models. Table 6.11 presents the AIC in the different models. The preferred model, which is the model with the lowest AIC, is the Gamma process model alternative 2.

There are at least two possible reasons to why the Gamma process models are preferred over the Markov models for the CAV data. One reason may be the Markov property. For example, the probability of transitioning from state 1 to state 2 may depend on when the individual entered state and left state 0. If the Markov property is clearly violated, we would also have detected a big change in the AIC between the inhomogeneous Markov model and the Gamma process models.

The second possibility is that it has something to do with the flexibility of the Gamma process models. The Gamma process model alternative 2 gives a much lower AIC than alternative 1. This indicates that including t^b and not just t gives a decrease in AIC. One reason is that the hazard functions in Gamma process model alternative 2 are more flexible, and therefore captures the true hazard functions better than the other model. For example, in Section 6.3.1.2, the hazard functions changed quite a bit compared to in Section 6.3.1.1. This explanation is therefore likely, because of this difference in AIC between Gamma process model alternative 1 and 2.

6.5.2 With Covariates

Model	$-2\log(\mathcal{L})$	k	AIC
MM, homogeneous	2821.21	15	2851.21
MM, inhomogeneous	2792.116	20	2832.116
GP alternative 1	2783.778	20	2823.778
GP alternative 2	2779.496	13	2805.496

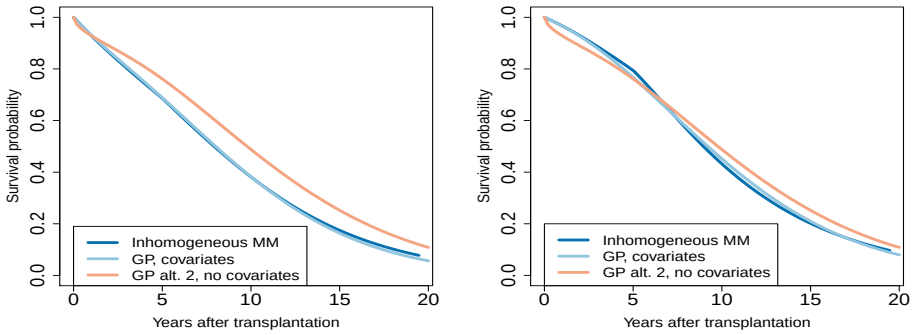
Table 6.12: AIC in a model, with covariates

In this part, we compare the different models with covariates. Table 6.12 shows the AIC-values when we include covariates for the Gamma process models, the homogeneous Markov model and the inhomogeneous Markov model. Also in this case, the Gamma process models have a lower AIC than both the homogeneous and inhomogeneous Markov model. The Gamma process model alternative 2 also has a lower AIC than alternative 1 and is therefore the preferred model when covariates are included. In this model, we only include covariates in the first transition and we only include the motor function t^b for the transition from state 0 to state 3.

Next, we want to compare how different values of the covariates affect the total survival probability for the different models. We therefore compare the total survival probability for the inhomogeneous Markov model and the Gamma process model alternative 1 with covariates, to the Gamma process model alternative 2 without covariates. We choose the inhomogeneous Markov model, because it is the Markov model with the lowest AIC. The reason we choose Gamma process model alternative 1 is that we want to compare the Markov model to a Gamma process model where covariates are included in all the transitions.

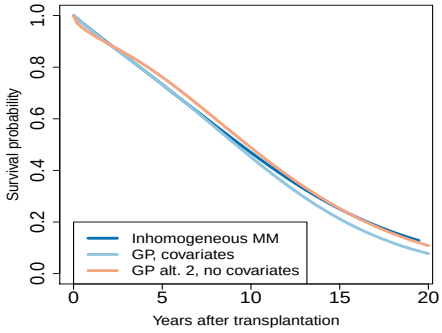
Figure 6.18 shows the the total survival probability functions for the inhomogeneous Markov model with covariates, the Gamma process model alternative 1 with covariates and the Gamma process model alternative 2 without covariates. The plots are of the total survival probability from state 0, for different values of the covariates. We include these plots in order to show whether certain values of the covariates raise or decrease the survival probabilities. Since the total survival probability for the Gamma process models and the inhomogeneous Markov model without covariates were quite equal, we compare the effect of the covariates to only Gamma process model alternative 2 without covariates. By including only one of these models without covariates, it is easier to see the effect in the plot of including covariates. For the rest of this section, Gamma process model without covariates means Gamma process model alternative 2 without covariates and Gamma process model with covariates, means Gamma process model alternative 1 with covariates.

In Figure 6.18 (a), we show a plot of the total survival probability function when $dage_{st} = 1$ and $ihd = 1$. When $dage_{st} = 1$, the age of the donor is around 43 years. That $ihd = 1$ means the initial diagnosis of the patient was *ihd*. Firstly, it seems like the Gamma process model with covariates and the inhomogeneous Markov model with covariates are close to each other. They are both a bit below the Gamma process model without covariates. The effect of having two

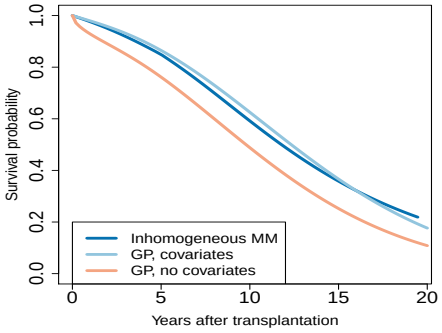


(a) $dage_{st} = 1, ihd = 1$

(b) $dage_{st} = -1, ihd = 1$



(c) $dage_{st} = 1, ihd = 0$



(d) $dage_{st} = -1, ihd = 0$

Figure 6.18: Total survival probability functions for the Gamma process model alternative 2 without covariates, thr inhomogeneous Markov model and the Gamma process model alternative 1 with covariates, for different values of *ihd* and $dage_{st}$

risk factors, lower the survival probability for both the inhomogeneous Markov model and the Gamma process model with covariates compared to no covariates.

The inhomogeneous Markov model and the Gamma process model with covariates are also close to each other in Figure 6.18 (b). In the beginning, the inhomogeneous Markov model and the Gamma process model with covariates are a bit above the model without covariates. After around 7 years, the inhomogeneous Markov model and the Gamma process model with covariates drop a bit below the model without covariates. Compared to Figure 6.18 (a), we have that when $dage_{st}$ changes from 1 to -1, meaning the donor age changes from 43 years to 18 years, the survival probability in both the Markov model and the Gamma process model with covariates increases a bit and are therefore close to the Gamma process model without covariates.

In Figure 6.18 (c), all of the models are very close to each other. For these covariate values, the survival probability for the inhomogeneous Markov model and Gamma process model with covariates are almost equal to the Gamma process model without covariates.

Lastly, in Figure 6.18 (d) both the inhomogeneous Markov model and the Gamma process model with covariates are higher than the survival probability function for the Gamma process model without covariates. The survival probability for the inhomogeneous Markov model and the Gamma process model with covariates are also close to each other for these covariate values.

In conclusion, it seems like the covariates have similar effects in the inhomogeneous Markov model and in the Gamma process model, at least when considering the total survival probability. If you are in one of the two risk groups, meaning either $dage_{st}$ is high or ihd= 1, the total survival probability for the models with covariates are quite equal to the model without covariates. If you are in both or neither of the risk groups, then the survival probability is a bit lower or higher respectively. In conclusion, when $dage_{st} = -1$ and $ihd = 0$ there is a small positive effect on the survival probability compared to when no covariates are included. When $dage_{st} = 1$ and $ihd = 1$ there is a small negative effect on the survival probability compared to when no covariates are included. If either $dage_{st} = 1$ or $ihd = 1$, the survival probability is almost equal to when no covariates are included.

CHAPTER 7

Conclusions and Future Work

In this thesis, we have introduced modeling of transition times as the threshold crossing times for Gamma processes in multi-state models for interval-censored data. We have constructed a general likelihood framework, investigated the MLEs through simulated data and investigated the Markov property. In addition, we have applied our model framework on a real dataset.

We started this thesis with presenting the theoretical background in Chapter 2. Then in Chapter 3, we created a general likelihood framework for a three-state progressive model, an illness-death model, a four-state progressive model and a four-state illness-death model. In addition, the different multi-state models are divided into when the transition to the absorbing state is observed exactly and when it is not observed exactly.

In Chapter 4, we simulated data from the Gamma process and computed the estimated maximum likelihood parameters. We checked the large-sample properties of the MLEs and found that these properties are satisfied.

Further, in Chapter 5, we discussed the Markov property in our likelihood construction. Since we wanted to investigate whether the Markov property is fulfilled in our model, we calculated the transition probabilities. In our calculations, we found that if the transition times are assumed to follow an exponential distribution, we have a homogeneous Markov model. However, the Markov property is not fulfilled when the transition times are modeled as the threshold crossing times for Gamma processes.

Finally, in Chapter 6, we applied our model framework on a dataset called CAV, where CAV is a post-transplant complication. The four-state illness-death model with two Gamma process model alternatives, a homogeneous and an inhomogeneous Markov model were considered. In addition, we considered models both with and without covariates. The included covariates are *ihd* as factor, which means whether the patient initially was diagnosed with *ihd* or not, and $dage_{st}$, which is the standardized donor age. We found that if the patient was originally not diagnosed with *ihd* and the age of the donor was low, then the survival probability was in general higher. In addition, we found that one of the Gamma process model alternatives gave the lowest AIC, both with and without covariates. For the models without covariates, we also checked if these estimated total survival probability curves followed the empirical Kaplan-Meier survival curve. We found that both of the Gamma processes model alternatives followed the empirical Kaplan-Meier survival curves closely.

There are several way of extending our framework. A possible and doable first step is to consider for example first-hitting time models based on an

underlying Wiener process instead of Gamma process. As we presented in Section 2.7.1, in a Wiener process the lifetime T follows an inverse Gaussian process if $\mu \leq 0$. If we in addition assume $\sigma = 1$, the parameters of interest are the threshold c and the mean μ. By switching the processes in the code, this is both easy and doable. It would then be possible to compare these two different first-hitting time models.

Another possible extension is to include dependencies between the different transitions. This means that in an illness-death model, one can assume that T_0, T_1 and T_{02} are not independent. In this case, one must include dependency in the log-likelihood and estimate this dependency in some way. However, some of these dependencies may also be captured with covariates.

By using Williams et al. (2020) as an inspiration, there is for example possible to include more states and different possible transitions between the states. We have extended the model to a multi-state model where the possible transitions and states are illustrated in Figure 7.1. This is the multi-state model discussed in Williams et al. (2020). With 7 states and the possible transitions as in Figure 7.1, there would be at least 86 different likelihood types. In addition we would have 13 transition times and for a Gamma process model, there would be at least 26 parameters to be found through optimization. This shows it is possible to build on the likelihood construction in this thesis and use it in a framework with more complex data. If both the data and the multi-state model can be more complex, then it is possible to use this framework in various settings. Having 86 different types may be a problem when it comes to computation time. Especially when we have to extend to formulas for the likelihood contributions with triple integrals. A possibility is then to program these likelihood contributions into a faster program than R, for example C and C++. This is also a limitation of our likelihood framework, that it may be quite complex when we include more states and transitions.

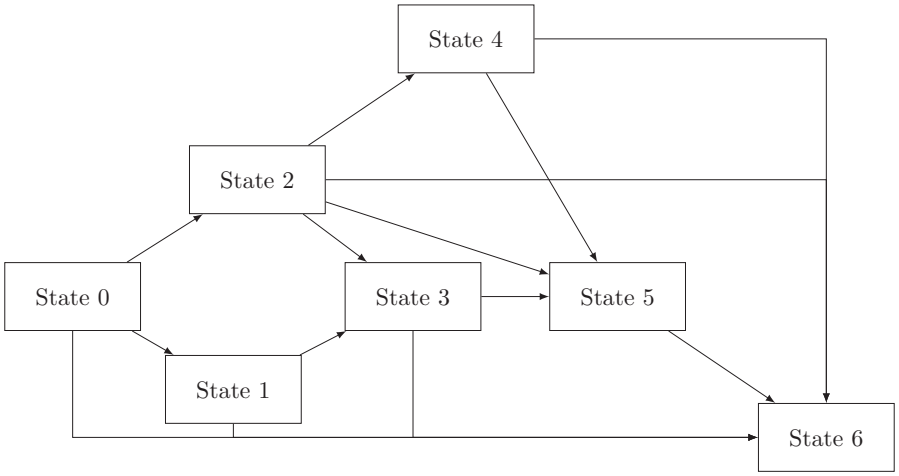


Figure 7.1: A seven-state illness-death model

As we stated above, with the states and possible transitions in Williams et al. (2020) there are at least 86 different types. When the model becomes more complex, the probability of making an error and forgetting a type gets higher. A possible expansion is then to make an algorithm which automatically finds

the likelihood contributions for each individual. The idea is that the algorithm can find which type the individual is and then likelihood contribution from this individual, for different multi-state models.

Inspired by the hidden Markov model in Williams et al. (2020), it would also be possible to make a hidden version of our model, based on the likelihood contributions in Chapter 3. Let us consider a three-state progressive model, as in Figure 7.2 . We start by only having one screening at time point t. In order

Figure 7.2: Three-state progressive model

to make the model hidden, we do not observe in which state an individual is. However, we observe a different response called y, which for example is normally distributed. We assume the transition to the absorbing state is also unknown. The distribution may for example be

$$y = \begin{cases} \sim N(\mu_0, 1) & \text{State 0} \\ \sim N(\mu_1, 1) & \text{State 1} \\ \sim N(\mu_2, 1) & \text{State 2} \end{cases}$$

When we construct the hidden model, we follow the approach and the formulas in Williams et al. (2020) with some modifications. The likelihood contribution from the ith subject at time t, where $X_{i,t}$ is the state at time t for individual i becomes

$$\begin{aligned} \mathcal{L}_{i,t} &= \sum_{X_{i,t}=0}^{2} f(y_{i,t}, X_{i,t}) \\ &= \sum_{X_{i,t}=0}^{2} P(X_{i,t}) f(y_{i,t}|X_{i,t}) \\ &= S_0(t) N(y_{i,t}|\mu_0, 1) + \int_0^t f_0(s) S_1(t-s) ds N(y_{i,t}|\mu_1, 1) \\ &+ \int_0^t f_0(s)(1 - S_1(t-s)) ds N(y_{i,t}|\mu_2, 1), \end{aligned}$$

where $N(y_{i,t}|\mu, 1)$ is the normal density with mean μ and variance 1.

Suppose we now have two screenings, t_1 and t_2. The likelihood contribution from the ith subject becomes

$$\mathcal{L}_{i,t_1,t_2} = \sum_{X_{i,t_1}=0}^{2} P(X_{i,t_1}) f(y_{i,t_1}|X_{i,t_1}) \cdot \sum_{X_{i,t_2}=0}^{2} P(X_{i,t_2}|X_{i,t_1}) f(y_{i,t_2}|X_{i,t_2})$$

$$
\begin{aligned}
&= S_0(t_1)N(y_{i,t_1}|\mu_0,1)\Big[S_0(t_2)N(y_{i,t_2}|\mu_0,1) + \int_{t_1}^{t_2} f_0(s)S_1(t_2 s)dsN(y_{i,t_2}|\mu_1,1) \\
&+ \int_{t_1}^{t_2} f_0(s)(1-S_1(t_2-s))dsN(y_{i,t_2}|\mu_2,1)\Big] \\
&+ \int_0^{t_1} f_0(s)S_1(t_1-s)dsN(y_{i,t_1}|\mu_1,1)\Big[\int_0^{t_1} f_0(s)S_1(t_2-s)dsN(y_{i,t_2}|\mu_1,1) \\
&+ \int_0^{t_1} f_0(s)(S_1(t_1-s)-S_1(t_2-s))dsN(y_{i,t_2}|\mu_2,1)\Big] \\
&+ \int_0^{t_1} f_0(s)(1-S_1(t_1-s))dsN(y_{i,t_1}|\mu_2,1).
\end{aligned}
$$

If we have three screenings, t_1, t_2 and t_3, the formula for the likelihood contribution is

$$
\begin{aligned}
\mathcal{L}_{i,t_1,t_2,t_3} = &\sum_{X_{i,t_1}=0}^{2} P(X_{i,t_1})f(y_{i,t_1}|X_{i,t_1}) \\
&\sum_{X_{i,t_2}=0}^{2} P(X_{i,t_2}|X_{i,t_1})f(y_{i,t_2}|X_{i,t_2}) \\
&\sum_{X_{i,t_3}=0}^{2} P(X_{i,t_3}|X_{i,t_1},X_{i,t_2})f(y_{i,t_3}|X_{i,t_3}).
\end{aligned} \tag{7.1}
$$

In a hidden Markov model, we have from Williams et al. (2020) that the likelihood contribution would be

$$
\begin{aligned}
\mathcal{L}_{i,t_1,t_2,t_3} = &\sum_{X_{i,t_1}=0}^{2} P(X_{i,t_1})f(y_{i,t_1}|X_{i,t_1}) \\
&\sum_{X_{i,t_2}=0}^{2} P(X_{i,t_2}|X_{i,t_1})f(y_{i,t_2}|X_{i,t_2}) \\
&\sum_{X_{i,t_3}=0}^{2} P(X_{i,t_3}|X_{i,t_2})f(y_{i,t_3}|X_{i,t_3}).
\end{aligned} \tag{7.2}
$$

Equation 7.1 is not in general equal to Equation 7.2, since in Equation 7.1 we condition on the whole state history. Our approach to the hidden model, would then be different and more complex than in a hidden Markov model. It will probably be an interesting and perhaps useful expansion of the models in this thesis.

An additional extension to the multi-state Gamma process models is to include a Bayesian approach. This is especially useful in situations with a priori knowledge about the disease of interest. Let us go back to the three-state progressive model. We assume a simple prior for all of the parameters and a random walk proposal. A possible MCMC-algorithm becomes

1. Define a starting point $\boldsymbol{\theta}^0 = (c_0^0, a_0^0, c_1^0, a_1^0)$

2. For $t = 2, \ldots, n$

   a) Sample a proposal $\theta^* = (c_0^*, a_0^*, c_1^*, a_1^*)$ from a proposal distribution

   $$J_t(\theta^*|\theta^{t-1}) = \theta^{t-1} + s,$$

   where for example $s \sim N(0, 0.1)$, which means we sample 4 random numbers from this distribution.

   b) Calculate

   $$r = \frac{Pr(c_0^*, a_0^*, c_1^*, a_1^*|t)}{Pr(c_0^{t-1}, a_0^{t-1}, c_1^{t-1}, a_1^{t-1}|t)}.$$

   c) Simulate a number $u \sim U[0, 1]$. If $u < r$

   $$\theta^t = (c_0^t, a_0^t, c_1^t, a_1^t) = \theta^* = (c_0^*, a_0^*, c_1^*, a_1^*),$$

   else:

   $$\theta^t = (c_0^t, a_0^t, c_1^t, a_1^t) = \theta^{t-1} = (c_0^{t-1}, a_0^{t-1}, c_1^{t-1}, a_1^{t-1}).$$

This MCMC proposal should be possible to do and expand to four-state cases as well. One can then include informative or uninformative priors on the parameters.

In relation to the Bayesian approach, it is probably also possible to include Bayesian nonparametrics in some way. For example, it may be possible to consider prior processes on each of the different transitions. Then one would probably need some assumptions about having enough people in all of the different transitions, the time interval must be long enough and so on. There probably exist a well-defined version of maximum-likelihood procedure. This may be quite complicated, but we think it is doable.

Lastly, it is also possible to have a greater focus on model selection and goodness-of-fit for multi-state models with interval-censored data. A possible extension is to create a powerful goodness-of-fit framework for these interval-censored data in this model framework. Then it would be possible to give a more accurate evaluation of the goodness-of-fit for this model. When it comes to model selection, it would also be interesting to compare the Gamma process models and the Markov model by using FIC from Claeskens and Hjort (2003).

Appendices

APPENDIX A

Likelihood when the Exact Time of Entry into the Absorbing State is not Known

A.1 Three-State Progressive Model

Suppose the individuals are screened $t_1, t_2, \ldots, t_n$ times. The different contributions to the likelihood are

1. Suppose an individual is only observed in state 0 at all the screening time points, where t_n is the last screening. Then

$$Pr(T_0 > t_n) = S_0(t_n)$$

2. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is still in state 1 at the last screening point, t_n. Then

$$\begin{aligned}
&Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 < t_n, T_0 + T_1 > t_{i+1}, T_0 + T_1 > t_n) \\
&= Pr(t_i < T_0 < t_{i+1}, T_0 + T_1 > t_n) = \int_{t_i}^{t_{i+1}} f_0(s) Pr(T_0 + T_1 > t_n | T_0 = s) ds \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(T_1 > t_n - s) ds = \int_{t_i}^{t_{i+1}} f_0(s) S_1(t_n - s) ds
\end{aligned}$$

.

3. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is observed in state 2 at t_{i+k}, where $k > 1$. Then

$$\begin{aligned}
&Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 + T_1 > t_{i+k-1}, T_0 + T_1 < t_{i+k}) \\
&= Pr(t_i < T_0 < t_{i+1}, t_{i+k-1} < T_0 + T_1 < t_{i+k}) \\
&= \int_{t_i}^{t_{i+1}} f_0(s)(F_1(t_{i+k} - s) - F_1(t_{i+k-1} - s)) ds \\
&= \int_{t_i}^{t_{i+1}} f_0(s)(S_1(t_{i+k-1} - s) - S_1(t_{i+k} - s)) ds.
\end{aligned}$$

4. Suppose an individual is only observed in state 1 at all the screening time points, where t_n is the last screening. Then

$$\begin{aligned} &Pr(T_0 < t_1, T_0 < t_n, T_0 + T_1 > t_1, T_0 + T_1 > t_n) \\ &= Pr(T_0 < t_1, T_0 + T_1 > t_n) = \int_0^{t_1} f_0(s) S_1(t_n - s) ds. \end{aligned}$$

5. Suppose an individual is observed in state 1 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. Then

$$\begin{aligned} &Pr(T_0 < t_1, T_0 + T_1 > t_i, T_0 + T_1 < t_{i+1}) \\ &= Pr(T_0 < t_1, t_i < T_0 + T_1 < t_{i+1}) \\ &= \int_0^{t_1} f_0(s) Pr(t_i < T_0 + T_1 < t_{i+1} | T_0 = s) ds \\ &= \int_0^{t_1} f_0(s)(S_1(t_i - s) - S_1(t_{i+1} - s)) ds. \end{aligned}$$

6. Suppose an individual is seen to be in state 0 for the time points t_1 to t_i and it is observed in state 2 at time point t_{i+1}. Then

$$\begin{aligned} &Pr(T_0 > t_i, T_0 + T_1 < t_{i+1}) \\ &= Pr(t_1 < T_0 < t_{i+1}, T_0 + T_1 < t_{i+1}) \\ &= \int_{t_i}^{t_{i+1}} f_0(s)(1 - S_1(t_{i+1} - s)) ds. \end{aligned}$$

7. Suppose an individual is observed in state 2 at the first screening point, t_1, without any intermittent screening.Then

$$\begin{aligned} Pr(T_0 + T_1 < t_1) &= \int_0^{t_1} f_0(s) F_1(t_1 - s) ds \\ &= \int_0^{t_1} f_0(s)(1 - S_1(t_1 - s)) ds. \end{aligned}$$

The full likelihood for the individuals $p = 1, \ldots, m$ becomes

$$\begin{aligned} \mathcal{L} = &\prod_{(I)} S_0(t_{n,p}, \boldsymbol{\theta} | x_p) \prod_{(II)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta} | x_p) S_1(t_{n,p} - s, \boldsymbol{\theta} | x_p) ds \\ &\prod_{(III)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta} | x_p)(S_1(t_{i+k-1,p} - s, \boldsymbol{\theta} | x_p) - S_1(t_{i+k,p} - s, \boldsymbol{\theta} | x_p)) ds \\ &\prod_{(IV)} \int_0^{t_{1,p}} f_0(s, \boldsymbol{\theta} | x_p) S_1(t_{n,p} - s, \boldsymbol{\theta} | x_p) ds \\ &\prod_{(V)} \int_0^{t_{1,p}} f_0(s, \boldsymbol{\theta} | x_p)(S_1(t_{i,p} - s, \boldsymbol{\theta} | x_p) - S_1(t_{i+1,p} - s, \boldsymbol{\theta} | x_p)) ds \\ &\prod_{(VI)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta} | x_p)(1 - S_1(t_{i+1,p} - s, \boldsymbol{\theta} | x_p)) ds \\ &\prod_{(VII)} \int_0^{t_{1,p}} f_0(s, \boldsymbol{\theta} | x_p)(1 - S_1(t_{1,p} - s, \boldsymbol{\theta} | x_p)) ds. \end{aligned}$$

A.2 Illness-Death Model

Suppose the individuals are screened $t_1, t_2, \ldots, t_n$ times. The different contributions to the likelihood are

1. Suppose an individual is seen to be in state 0 at all the different screening time points, where t_n is the last screening. Then
$$Pr(T_0 > t_n, T_{02} > t_n) = S_0(t_n)S_{02}(t_n)$$
with no further information about T_1.

2. Suppose an individual is seen to be in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is still in state 1 at the last screening point t_n. We also have that $T_{02} > T_0$. Then
$$\begin{aligned}
&Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 < t_n, T_0 + T_1 > t_{i+1}, T_0 + T_1 > t_n, T_{02} > T_0) \\
&= Pr(t_i < T_0 < t_{i+1}, T_0 + T_1 > t_n) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(T_0 + T_1 > t_n | T_0 = s) Pr(T_{02} > T_0 | T_0 = s) ds \\
&= \int_{t_i}^{t_{i+1}} f_0(s) S_1(t_n - s) S_{02}(s) ds.
\end{aligned}$$

3. Suppose an individual is seen to be in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is observed in state 2 at t_{i+k}, where $k > 1$. We also have that $T_{02} > T_0$. Then
$$\begin{aligned}
&Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 + T_1 > t_{i+k-1}, T_0 + T_1 < t_{i+k}, T_{02} > T_0) \\
&= Pr(t_i < T_0 < t_{i+1}, t_{i+k-1} < T_0 + T_1 < t_{i+k}, T_{02} > T_0) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k-1} < T_0 + T_1 < t_{i+k} | T_0 = s) Pr(T_{02} > T_0 | T_0 = s) ds \\
&= \int_{t_i}^{t_{i+1}} f_0(s)(F_1(t_{i+k} - s) - F_1(t_{i+k-s} - s)) S_{02}(s) ds \\
&= \int_{t_i}^{t_{i+1}} f_0(s)(S_1(t_{i+k-1} - s) - S_1(t_{i+k} - s)) S_{02}(s) ds.
\end{aligned}$$

4. Suppose an individual is only observed in state 1 at all the screening time points, where t_n is the last screening. We also have that $T_{02} > T_0$. Then
$$\begin{aligned}
&Pr(T_0 < t_1, T_0 < t_n, T_0 + T_1 > t_1, T_0 + T_1 > t_n, T_{02} > T_0) \\
&= Pr(T_0 < t_1, T_0 + T_1 > t_n, T_{02} > T_0) \\
&= \int_0^{t_1} f_0(s) S_1(t_n - s) S_{02}(s) ds.
\end{aligned}$$

5. Suppose an individual is observed in state 1 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. We also have that $T_{02} > T_0$. Then
$$\begin{aligned}
&Pr(T_0 < t_1, T_0 + T_1 > t_i, T_0 + T_1 < t_{i+1}, T_{02} > T_0) \\
&= Pr(T_0 < t_1, t_i < T_0 + T_1 < t_{i+1}, T_{02} > T_0) \\
&= \int_0^{t_1} f_0(s)(S_1(t_i - s) - S_1(t_{i+1} - s)) S_{02}(s) ds.
\end{aligned}$$

6. Suppose an individual is observed in state 1 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. We also have that $T_{02} > T_0$. Then

$$\begin{aligned} &Pr(T_0 > t_i, T_0 + T_1 < t_{i+1}, T_{02} > T_0) \\ &= Pr(t_1 < T_0 < t_{i+1}, T_0 + T_1 < t_{i+1}, T_{02} > T_0) \\ &= \int_{t_i}^{t_{i+1}} f_0(s)(1 - S_1(t_{i+1} - s))S_{02}(s)ds. \end{aligned}$$

7. Suppose an individual is observed in state 2 at the first screening point t_1, without any intermittent screening. We also have that $T_{02} > T_0$. Then

$$\begin{aligned} Pr(T_0 + T_1 < t_1, T_{02} > T_0) &= \int_0^{t_1} f_0(s)F_1(t_1 - s)S_{02}(s)ds \\ &= \int_0^{t_1} f_0(s)(1 - S_1(t_1 - s))S_{02}(s)ds. \end{aligned}$$

8. Suppose an individual is observed in state 2 at the first screening point t_1, without any intermittent screening. We also have that $T_{02} < T_0$. Then

$$Pr(T_{02} < t_1, T_{02} < T_0) = \int_0^{t_1} f_{02}(s)S_0(s)ds$$

9. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. We also have that $T_{02} < T_0$. Then

$$Pr(t_i < T_{02} < t_{i+1}, T_{02} < T_0) = \int_{t_i}^{t_{i+1}} f_{02}(s)S_0(s)ds$$

The full likelihood for a dataset with these type of screenings for the individuals $p = 1, \ldots, m$ then becomes

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) = &\prod_{(I)} S_0(t_{n,p}, \boldsymbol{\theta}|x_p)S_{02}(t_{n,p}, \boldsymbol{\theta}|x_p) \prod_{(II)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p)S_1(t_{n,p} - s, \boldsymbol{\theta}|x_p)S_{02}(s, \boldsymbol{\theta}|x_p)ds \\ &\prod_{(III)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p)(S_1(t_{i+k-1,p} - s, \boldsymbol{\theta}|x_p) - S_1(t_{i+k,p} - s, \boldsymbol{\theta}|x_p))S_{02}(s, \boldsymbol{\theta}|x_p)ds \\ &\prod_{(IV)} \int_0^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p)S_1(t_{n,p} - s, \boldsymbol{\theta}|x_p)S_{02}(s, \boldsymbol{\theta}|x_p)ds \\ &\prod_{(V)} \int_0^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p)(S_1(t_{i,p} - s, \boldsymbol{\theta}|x_p) - S_1(t_{i+1,p} - s, \boldsymbol{\theta}|x_p))S_{02}(s, \boldsymbol{\theta}|x_p)ds \\ &\prod_{(VI)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p)(1 - S_1(t_{i+1,p} - s, \boldsymbol{\theta}|x_p))S_{02}(s, \boldsymbol{\theta}|x_p)ds \\ &\prod_{(VII)} \int_0^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p)(1 - S_1(t_{1,p} - s, \boldsymbol{\theta}|x_p))S_{02}(s, \boldsymbol{\theta}|x_p)ds \\ &\prod_{(VIII)} \int_0^{t_{1,p}} f_{02}(s, \boldsymbol{\theta}|x_p)S_0(s, \boldsymbol{\theta}|x_p)ds \\ &\prod_{(IX)} \int_{t_{i,p}}^{t_{i+1,p}} f_{02}(s, \boldsymbol{\theta}|x_p)S_0(s, \boldsymbol{\theta}|x_p)ds. \end{aligned}$$

A.3 Four-State Progressive Model

Suppose we have screened an individual $t_1, t_2, \ldots, t_n$ times. The likelihood contributions are

1. Suppose an individual is only observed in state 0 at all the screening time points, where t_n is the last screening. Then

$$Pr(T_0 > t_n) = S_0(t_n)$$

2. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is still in state 1 at the last screening point t_n. Then

$$\begin{aligned}
&Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 < t_n, T_0 + T_1 > t_{i+1}, T_0 + T_1 > t_n) \\
&= Pr(t_i < T_0 < t_{i+1}, T_0 + T_1 > t_n) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(T_0 + T_1 > t_n | T_0 = s) ds \\
&= \int_{t_i}^{t_{i+1}} f_0(s) S_1(t_n - s) ds.
\end{aligned}$$

3. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is observed in state 1 until t_{i+k-1}, where $k > 1$. At t_{i+k}, the individual is observed in state 2. The individual is still in state 2 at the last screening point t_n. Then

$$\begin{aligned}
&Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 + T_1 > t_{i+1}, T_0 + T_1 < t_{i+k}, \\
&T_0 + T_1 + T_2 > t_{i+k}, T_0 + T_1 + T_2 > t_n) \\
&= Pr(t_i < T_0 < t_{i+1}, t_{i+k-1} < T_0 + T_1 < t_{i+k}, T_0 + T_1 + T_2 > t_n) \\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) Pr(T_2 > t_n - u - s) du ds \\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) S_2(t_n - u - s) du ds.
\end{aligned}$$

4. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is observed in state 1 until t_{i+k-1}, where $k > 1$. The individual is observed in state 2 at t_{i+k}. The individual is observed in state 2 until $t_{i+k+l-1}$, where $l > 1$. At t_{i+k+l}, the individual is observed in state 3. Then

$$\begin{aligned}
&Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 + T_1 > t_{i+1}, T_0 + T_1 < t_{i+k}, \\
&T_0 + T_1 + T_2 > t_{i+k}, T_0 + T_1 + T_2 < t_{i+k+l}) \\
&= Pr(t_i < T_0 < t_{i+1}, t_{i+k-1} < T_0 + T_1 < t_{i+k}, t_{i+k+l-1} < T_0 + T_1 + T_2 < t_{i+k+l}) \\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) Pr(t_{i+k+l-1} - s - u < T_2 < t_{i+k+l} - s - u) du ds \\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) (S_2(t_{i+k+l-1} - s - u) - S_2(t_{i+k+l} - s - u)) du ds.
\end{aligned}$$

5. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 3.

$$
\begin{aligned}
&Pr(T_0 > t_i, T_0 + T_1 + T_2 < t_{i+1}) = Pr(t_i < T_0 < t_{i+1}, T_0 + T_1 + T_2 < t_{i+1}) \\
&= Pr(t_i < T_0 < t_{i+1}, t_i < T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 < t_{i+1}) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_i < T_0 + T_1 < t_{i+1} | T_0 = s) Pr(T_0 + T_1 + T_2 < t_{i+1} | T_0 = s) ds \\
&= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) F_2(t_{i+1} - u - s) du ds \\
&= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) (1 - S_2(t_{i+1} - s - u)) du ds.
\end{aligned}
$$

6. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. The individual is still in state 2 at the last screening point t_n. Then

$$
\begin{aligned}
&Pr(T_0 > t_i, T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_n) \\
&= Pr(t_i < T_0 < t_{i+1}, t_i < T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_n) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_i < T_0 + T_1 < t_{i+1} | T_0 = s) Pr(T_0 + T_1 + T_2 > t_n | T_0 = s) \\
&= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) S_2(t_n - s - u) du ds.
\end{aligned}
$$

7. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. The individual is observed in state 2 until t_{i+k-1}, where $k > 1$. At t_{i+k}, the individual is observed in state 3. Then

$$
\begin{aligned}
&Pr(T_0 > t_i, T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_{i+k-1}, T_0 + T_1 + T_2 < t_{i+k}) \\
&= Pr(t_i < T_0 < t_{i+1}, t_i < T_0 + T_1 < t_{i+1}, t_{i+k-1} < T_0 + T_1 + T_2 < t_{i+k}) \\
&= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) (S_2(t_{i+k-1} - u - s) - S_2(t_{i+k} - u - s)) du ds.
\end{aligned}
$$

8. Suppose an individual is only observed in state 1 at all the screening time points., where t_n is the last screening. Then

$$
\begin{aligned}
&Pr(T_0 < t_1, T_0 < t_n, T_0 + T_1 > t_1, T_0 + T_1 > t_n) \\
&= Pr(T_0 < t_1, T_0 + T_1 > t_n) \\
&= \int_0^{t_1} f_0(s) S_1(t_n - s) ds.
\end{aligned}
$$

9. Suppose an individual is observed in state 1 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. The individual is still in state 2 at the last screening point t_n. Then

$$
\begin{aligned}
&Pr(T_0 < t_1, T_0 + T_1 > t_i, T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_n) \\
&= Pr(T_0 < t_1, t_i < T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_n) \\
&= \int_0^{t_1} f_0(s) Pr(t_i < T_0 + T_1 < t_{i+1} | T_0 = s) Pr(T_0 + T_1 + T_2 > t_n | T_0 = s) ds \\
&= \int_0^{t_1} \int_{t_i - s}^{t_{i+1}-s} f_0(s) f_1(u) S_2(t_n - s - u) du ds.
\end{aligned}
$$

10. Suppose an individual is observed in state 1 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. The individual is observed in state 2 until t_{i+k-1}, where $k > 1$. At t_{i+k}, the individual is observed in state 3. Then

$$\begin{aligned}
&Pr(T_0 < t_1, T_0 + T_1 > t_i, T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_{i+k-1}, T_0 + T_1 + T_2 < t_{i+k}) \\
&= Pr(T_0 < t_1, t_i < T_0 + T_1 < t_{i+1}, t_{i+k-1} < T_0 + T_1 + T_2 < t_{i+k}) \\
&= \int_0^{t_1} \int_{t_i - s}^{t_{i+1} - s} f_0(s) f_1(u) Pr(t_{i+k-1} < T_0 + T_1 + T_2 < t_{i+k} | T_0 = s, T_1 = u) du ds \\
&= \int_0^{t_1} \int_{t_i - s}^{t_{i+1} - s} f_0(s) f_1(u) (S_2(t_{i+k-1} - s - u) - S_2(t_{i+k} - s - u)) du ds.
\end{aligned}$$

11. Suppose an individual is observed in state 1 from t_1 to t_i. At t_{i+1}, the individual is observed in state 3. Then

$$\begin{aligned}
&Pr(T_0 < t_1, T_0 + T_1 > t_i, T_0 + T_1 + T_2 < t_{i+1}) \\
&= Pr(T_0 < t_1, t_i < T_0 + T_1 < t_{i+1}, t_i < T_0 + T_1 + T_2 < t_{i+1}) \\
&= \int_0^{t_1} \int_{t_i - s}^{t_{i+1} - s} f_0(s) f_1(u) (1 - S_2(t_{i+1} - s - u)) du ds.
\end{aligned}$$

12. Suppose an individual is only observed in state 2 at all the screening time points, where t_n is the last screening. Then

$$\begin{aligned}
&Pr(T_0 < t_1, T_0 < t_n, T_0 + T_1 < t_1, T_0 + T_1 < t_n, T_0 + T_1 + T_2 > t_n) \\
&= Pr(T_0 < t_1, T_0 + T_1 < t_1, T_0 + T_1 + T_2 > t_n) \\
&= \int_0^{t_1} \int_0^{t_1 - s} f_0(s) f_1(u) S_2(t_n - u - s) du ds.
\end{aligned}$$

13. Suppose an individual is observed in state 2 from t_1 to t_i. At t_{i+1}, the individual is observed in state 3. Then

$$\begin{aligned}
&Pr(T_0 < t_1, T_0 + T_1 < t_1, t_i < T_0 + T_1 + T_2 < t_{i+1}) \\
&= \int_0^{t_1} \int_0^{t_1 - s} f_0(s) f_1(u) (S_2(t_i - s - u) - S_2(t_{i+1} - s - u)) du ds.
\end{aligned}$$

14. Suppose an individual is observed in state 3 at the first screening point t_1, without any intermittent screening. Then

$$\begin{aligned}
&Pr(T_0 + T_1 + T_2 < t_1) \\
&= \int_0^{t_1} \int_0^{t_1 - s} f_0(s) f_1(u) (1 - S_2(t_1 - s - u)) du ds.
\end{aligned}$$

15. Suppose an individual is observed in state 0 from t_1 to t_i. The individual is observed in state 2 at t_{i+1}. The individual is observed in state 2 until t_{i+k-1}, where $k > 1$. At t_{i+k}, the individual is observed in state 3. Then

$$\begin{aligned}
&Pr(t_i < T_0 < t_{i+1}, T_0 + T_1 > t_{i+k-1}, T_0 + T_1 + T_2 < t_{i+k}) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k-1} < T_0 + T_1 < t_{i+k} | T_0 = s) Pr(T_0 + T_1 + T_2 < t_{i+k} | T_0 = s) ds \\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1} - s}^{t_{i+k} - s} f_0(s) f_1(u) (1 - S_2(t_{i+k} - s - u)) du ds.
\end{aligned}$$

A. Likelihood when the Exact Time of Entry into the Absorbing State is not Known

The full likelihood for the individuals $p = 1, \ldots, m$ then becomes

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) = & \prod_{(I)} S_0(t_{n,p}|x_p) \prod_{(II)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s,\boldsymbol{\theta}|x_p) S_1(t_n - s,\boldsymbol{\theta}|x_p) ds \\
& \prod_{(III)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p}-s}^{t_{i+k,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) S_2(t_{n,p} - u - s,\boldsymbol{\theta}|x_p) du ds \\
& \prod_{(IV)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p}-s}^{t_{i+k,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) (S_2(t_{i+k+l-1,p} - s - u,\boldsymbol{\theta}|x_p) \\
& - S_2(t_{i+k+l,p} - s - u,\boldsymbol{\theta}|x_p)) du ds \\
& \prod_{(V)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{0}^{t_{i+1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) (1 - S_2(t_{i+1,p} - s - u,\boldsymbol{\theta}|x_p)) du ds \\
& \prod_{(VI)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{0}^{t_{i+1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) S_2(t_{n,p} - s - u,\boldsymbol{\theta}|x_p) du ds \\
& \prod_{(VII)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{0}^{t_{i+1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) (S_2(t_{i+k-1,p} - u - s,\boldsymbol{\theta}|x_p) \\
& - S_2(t_{i+k,p} - u - s,\boldsymbol{\theta}|x_p)) du ds \\
& \prod_{(VIII)} \int_{0}^{t_{1,p}} f_0(s,\boldsymbol{\theta}|x_p) S_1(t_{n,p} - s,\boldsymbol{\theta}|x_p) ds \\
& \prod_{(IX)} \int_{0}^{t_{1,p}} \int_{t_{i,p}-s}^{t_{i+1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) S_2(t_{n,p} - s - u,\boldsymbol{\theta}|x_p) du ds \\
& \prod_{(X)} \int_{0}^{t_{1,p}} \int_{t_{i,p}-s}^{t_{i+1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) (S_2(t_{i+k-1,p} - s - u,\boldsymbol{\theta}|x_p) - \\
& S_2(t_{i+k,p} - s - u,\boldsymbol{\theta}|x_p)) du ds \\
& \prod_{(XI)} \int_{0}^{t_{1,p}} \int_{t_{i,p}-s}^{t_{i+1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) (1 - S_2(t_{i+1,p} - s - u,\boldsymbol{\theta}|x_p)) du ds \\
& \prod_{(XII)} \int_{0}^{t_{1,p}} \int_{0}^{t_{1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) S_2(t_{n,p} - u - s,\boldsymbol{\theta}|x_p) du ds \\
& \prod_{(XIII)} \int_{0}^{t_{1,p}} \int_{0}^{t_{1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) (S_2(t_{i,p} - s - u,\boldsymbol{\theta}|x_p) - \\
& S_2(t_{i+1,p} - s - u,\boldsymbol{\theta}|x_p)) du ds \\
& \prod_{(XIV)} \int_{0}^{t_{1}} \int_{0}^{t_{1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) (1 - S_2(t_{1,p} - s - u,\boldsymbol{\theta}|x_p)) du ds \\
& \prod_{(XV)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p}-s}^{t_{i+k,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) (1 - S_2(t_{i+k,p} - s - u,\boldsymbol{\theta}|x_p)) du ds.
\end{aligned}
$$

A.4 Four-State Illness-Death Model

Suppose we have screened an individual $t_1, t_2, \ldots, t_n$ times. The likelihood contributions are

1. Suppose an individual is only observed in state 0 at all the screening time points, where t_n is the last screening. Then

$$Pr(T_0 > t_n, T_{03} > t_n) = S_0(t_n)S_{03}(t_n)$$

2. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is still in state 1 at the last screening point t_n. We have that $T_{03} > T_0$ and $T_{13} + T_0 > t_n$. Then

$$\begin{aligned}
&Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 < t_n, T_0 + T_1 > t_{i+1}, T_0 + T_1 > t_n,\\
&T_{03} > T_0, T_{13} + T_0 > t_n)\\
&= Pr(t_i < T_0 < t_{i+1}, T_0 + T_1 > t_n, T_{03} > T_0, T_{13} + T_0 > t_n)\\
&= \int_{t_i}^{t_{i+1}} f_0(s)Pr(T_1 > t_n - s)Pr(T_{03} > s)Pr(T_{13} > t_n - s)ds\\
&= \int_{t_i}^{t_{i+1}} f_0(s)S_1(t_n - s)S_{03}(s)S_{13}(t_n - s)ds.
\end{aligned}$$

3. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1} the individual is observed in state 1. The individual is observed in state 1 until t_{i+k-1}, where $k > 1$. At t_{i+k}, the individual is observed in state 2. The individual is still in state 2 at the last screening point t_n. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$\begin{aligned}
&Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 + T_1 > t_{i+1}, T_0 + T_1 < t_{i+k},\\
&T_0 + T_1 + T_2 > t_{i+k}, T_0 + T_1 + T_2 > t_n, T_{03} > T_0, T_{13} > T_1)\\
&= Pr(t_i < T_0 < t_{i+1}, t_{i+k-1} < T_0 + T_1 < t_{i+k}, T_0 + T_1 + T_2 > t_n,\\
&T_{03} > T_0, T_{13} > T_1)\\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s)f_1(u)Pr(T_0 + T_1 + T_2 > t_n | T_0 = s, T_1 = u)\\
&Pr(T_{03} > s)Pr(T_{13} > T_1 | T_1 = u)duds\\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s)f_1(u)Pr(T_2 > t_n - u - s)S_{03}(s)S_{13}(u)duds\\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s)f_1(u)S_2(t_n - u - s)S_{03}(s)S_{13}(u)duds.
\end{aligned}$$

4. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is observed in state 1 until t_{i+k-1}, where $k > 1$. The individual is observed in state 2 at t_{i+k}. The individual is observed in state 2 until $t_{i+k+l-1}$, where $l > 1$. At t_{i+k+l}, the individual is observed in state 3. We also have that $T_{03} > T_0$

and $T_{13} > T_1$. Then

$$
\begin{aligned}
&Pr(T_0 > t_i, T_0 < t_{i+1}, T_0 + T_1 > t_{i+1}, T_0 + T_1 < t_{i+k},\\
&T_0 + T_1 + T_2 > t_{i+k}, T_0 + T_1 + T_2 < t_{i+k+l}, T_{03} > T_0, T_{13} > T_1)\\
&= Pr(t_i < T_0 < t_{i+1}, t_{i+k-1} < T_0 + T_1 < t_{i+k}, t_{i+k+l-1} < T_0 + T_1 + T_2 < t_{i+k+l},\\
&T_{03} > T_0, T_{13} > T_1)\\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) Pr(t_{i+k+l-1} < T_0 + T_1 + T_2 < t_{i+k+l} | T_0 = s, T_1 = u)\\
&Pr(T_{03} > s) Pr(T_{13} > T_1 | T_1 = u) du ds\\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) (F_2(t_{i+k+l} - s - u) - F_2(t_{i+k+l-1} - s - u))\\
&S_{03}(s) S_{13}(u) du ds\\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) (S_2(t_{i+k+l-1} - s - u) - S_2(t_{i+k+l} - s - u))\\
&S_{03}(s) S_{13}(u) du ds.
\end{aligned}
$$

5. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 3. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$
\begin{aligned}
&Pr(T_0 > t_i, T_0 + T_1 + T_2 < t_{i+1}, T_{03} > T_0, T_{13} > T_1)\\
&= Pr(t_i < T_0 < t_{i+1}, T_0 + T_1 + T_2 < t_{i+1}, T_{03} > T_0, T_{13} > T_1)\\
&= \int_{t_i}^{t_{i+1}} \int_{0}^{t_{i+1}-s} f_0(s) f_1(u) Pr(T_2 < t_{i+1} - s - u) S_{03}(s) S_{13}(u) du ds\\
&= \int_{t_i}^{t_{i+1}} \int_{0}^{t_{i+1}-s} f_0(s) f_1(u) F_2(t_{i+1} - u - s) S_{03}(s) S_{13}(u) du ds\\
&= \int_{t_i}^{t_{i+1}} \int_{0}^{t_{i+1}-s} f_0(s) f_1(u) (1 - S_2(t_{i+1} - s - u)) S_{13}(s) S_{13}(u) du ds.
\end{aligned}
$$

6. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. The individual is still in state 2 at the last screening point t_n. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$
\begin{aligned}
&Pr(T_0 > t_i, T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_n, T_{03} > T_0, T_{13} > T_1)\\
&= Pr(t_i < T_0 < t_{i+1}, t_i < T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_n, T_{03} > T_0, T_{13} > T_1)\\
&= \int_{t_i}^{t_{i+1}} \int_{0}^{t_{i+1}-s} f_0(s) f_1(u) Pr(T_0 + T_1 + T_2 > t_n | T_0 = s, T_1 = u)\\
&S_{03}(s) Pr(T_{13} > T_1 | T_1 = u) du ds\\
&= \int_{t_i}^{t_{i+1}} \int_{0}^{t_{i+1}-s} f_0(s) f_1(u) S_2(t_n - s - u) S_{03}(s) S_{13}(u) du ds.
\end{aligned}
$$

7. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. The individual is observed in state 2 until t_{i+k-1}, where $k > 1$. At t_{i+k}, the individual is observed in state 3.

We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$
\begin{aligned}
&Pr(T_0 > t_i, T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_{i+k-1}, T_0 + T_1 + T_2 < t_{i+k}, \\
&T_{03} > T_0, T_{13} > T_1) \\
&= Pr(t_i < T_0 < t_{i+1}, t_i < T_0 + T_1 < t_{i+1}, t_{i+k-1} < T_0 + T_1 + T_2 < t_{i+k}, \\
&T_{03} > T_0, T_{13} > T_1) \\
&= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) (F_2(t_{i+k} - u - s) - F_2(t_{i+k-1} - u - s)) \\
&S_{03}(s) S_{13}(u) du ds \\
&= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1}-s} f_0(s) f_1(u) (S_2(t_{i+k-1} - u - s) - S_2(t_{i+k} - u - s)) \\
&S_{03}(s) S_{13}(u) du ds.
\end{aligned}
$$

8. Suppose an individual is only observed in state 1 at all the screening time points, where t_n is the last screening. We also have that $T_{03} > T_0$ and $T_{13} + T_0 > t_n$. Then

$$
\begin{aligned}
&Pr(T_0 < t_1, T_0 < t_n, T_0 + T_1 > t_1, T_0 + T_1 > t_n, T_{03} > T_0, T_0 + T_{13} > t_n) \\
&= Pr(T_0 < t_1, T_0 + T_1 > t_n, T_{03} > T_0, T_0 + T_{13} > T_1) \\
&= \int_0^{t_1} f_0(s) S_1(t_n - s) S_{03}(s) S_{13}(t_n - s) ds.
\end{aligned}
$$

9. Suppose an individual is observed in state 1 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. The individual is still in state 2 at the last screening point t_n. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$
\begin{aligned}
&Pr(T_0 < t_1, T_0 + T_1 > t_i, T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_n, \\
&T_{03} > T_0, T_{13} > T_1) \\
&= Pr(T_0 < t_1, t_i < T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_n, \\
&T_{03} > T_0, T_{13} > T_1) \\
&= \int_0^{t_1} \int_{t_i - s}^{t_{i+1}-s} f_0(s) f_1(u) Pr(T_0 + T_1 + T_2 > t_n | T_0 = s, T_1 = u) \\
&S_{03}(s) S_{13}(u) du ds \\
&= \int_0^{t_1} \int_{t_i - s}^{t_{i+1}-s} f_0(s) f_1(u) S_2(t_n - s - u) S_{03}(s) S_{13}(u) du ds.
\end{aligned}
$$

10. Suppose an individual is observed in state 1 from t_1 to t_i. At t_{i+1}, the individual is observed in state 2. The individual is observed in state 2 until t_{i+k-1}, where $k > 1$. At t_{i+k}, the individual is observed in state 3.

We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$Pr(T_0 < t_1, T_0 + T_1 > t_i, T_0 + T_1 < t_{i+1}, T_0 + T_1 + T_2 > t_{i+k-1},$$
$$T_0 + T_1 + T_2 < t_{i+k}, T_{03} > T_0, T_{13} > T_1)$$
$$= Pr(T_0 < t_1, t_i < T_0 + T_1 < t_{i+1}, t_{i+k-1} < T_0 + T_1 + T_2 < t_{i+k},$$
$$T_{03} > T_0, T_{13} > T_1)$$
$$= \int_0^{t_1} \int_{t_i - s}^{t_{i+1}-s} f_0(s) f_1(u) Pr(t_{i+k-1} < T_0 + T_1 + T_2 < t_{i+k} | T_0 = s, T_1 = u)$$
$$S_{03}(s) S_{13}(u) du ds$$
$$= \int_0^{t_1} \int_{t_i - s}^{t_{i+1}-s} f_0(s) f_1(u) (S_2(t_{i+k-1} - s - u) - S_2(t_{i+k} - s - u)) S_{03}(s) S_{13}(u) du ds.$$

11. Suppose an individual is observed in state 1 from t_1 to t_i. At t_{i+1} the individual is observed in state 3. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$Pr(T_0 < t_1, T_0 + T_1 > t_i, T_0 + T_1 + T_2 < t_{i+1}, T_{03} > T_0, T_{13} > T_1)$$
$$= Pr(T_0 < t_1, t_i < T_0 + T_1 < t_{i+1}, t_i < T_0 + T_1 + T_2 < t_{i+1}, T_{03} > T_0, T_{13} > T_1)$$
$$= \int_0^{t_1} \int_{t_i - s}^{t_{i+1}-s} f_0(s) f_1(u) Pr(t_i < T_0 + T_1 + T_2 < t_{i+1} | T_0 = s, T_1 = u)$$
$$Pr(T_{03} > T_0 | T_0 = s) Pr(T_{13} > T_1) du ds$$
$$= \int_0^{t_1} \int_{t_i - s}^{t_{i+1}-s} f_0(s) f_1(u) (1 - S_2(t_{i+1} - s - u)) S_{03}(s) S_{13}(u) du ds.$$

12. Suppose an individual is only observed in state 2 at all the screening time points, where t_n is the last screening. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$Pr(T_0 < t_1, T_0 < t_n, T_0 + T_1 < t_1, T_0 + T_1 < t_n, T_0 + T_1 + T_2 > t_n, T_{03} > T_0, T_{13} > T_1)$$
$$= Pr(T_0 < t_1, T_0 + T_1 < t_1, T_0 + T_1 + T_2 > t_n, T_{03} > T_0, T_{13} > T_1)$$
$$= \int_0^{t_1} \int_0^{t_1 - s} f_0(s) f_1(u) S_2(t_n - u - s) S_{03}(s) S_{13}(u) du ds.$$

13. Suppose an individual is observed in state 2 from t_1 to t_i. At t_{i+1}, the individual is observed in state 3. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$Pr(T_0 < t_1, T_0 + T_1 < t_1, t_i < T_0 + T_1 + T_2 < t_{i+1}, T_{03} > T_0, T_{13} > T_1)$$
$$= \int_0^{t_1} f_0(s) Pr(T_0 + T_1 < t_1 | T_0 = s) Pr(t_i < T_0 + T_1 + T_2 < t_{i+1} | T_0 = s)$$
$$Pr(T_{03} > T_0 | T_0 = s) Pr(T_{13} > T_1) ds$$
$$= \int_0^{t_1} \int_0^{t_1 - s} f_0(s) f_1(u) (S_2(t_i - s - u) - S_2(t_{i+1} - s - u)) S_{03}(s) S_{13}(u) du ds.$$

14. Suppose an individual is observed in state 3 at the first screening point t_1, without any intermittent screening. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$Pr(T_0 + T_1 + T_2 < t_1, T_{03} > T_0, T_{13} > T_1)$$
$$= Pr(0 < T_0 < t_1, 0 < T_0 + T_1 < t_1, 0 < T_0 + T_1 + T_2 < t_1, T_{03} > T_0, T_{13} > T_1)$$
$$= \int_0^{t_1} \int_0^{t_1 - s} f_0(s) f_1(u) (1 - S_2(t_1 - s - u)) S_{03}(s) S_{13}(u) du ds.$$

15. Suppose an individual is observed in state 0 from t_1 to t_i. The individual is observed in state 2 at t_{i+1}. The individual is observed in state 2 until t_{i+k-1}, where $k > 1$. At t_{i+k}, the individual is observed in state 3. We also have that $T_{03} > T_0$ and $T_{13} > T_1$. Then

$$\begin{aligned}
&Pr(t_i < T_0 < t_{i+1}, T_0 + T_1 > t_{i+k-1}, T_0 + T_1 + T_2 < t_{i+k}, T_{03} > T_0, T_{13} > T_1) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) Pr(t_{i+k-1} < T_0 + T_1 < t_{i+k} | T_0 = s) Pr(T_0 + T_1 + T_2 < t_{i+k} | T_0 = s) \\
&Pr(T_{03} > T_0 | T_0 = s) Pr(T_{13} > T_1) ds \\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) f_1(u) (1 - S_2(t_{i+k} - s - u)) S_{03}(s) S_{13}(s) du ds.
\end{aligned}$$

16. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 3. We also have that $T_{03} < T_0$. Then

$$Pr(t_i < T_{03} < t_{i+1}, T_{03} < T_0) = \int_{t_i}^{t_{i+1}} f_{03}(s) S_0(s) ds$$

17. Suppose an individual is observed in state 3 at time point t_1, without any intermittent screening. We also have that $T_{03} < T_0$. Then

$$Pr(t_1 > T_{03}, T_{03} < T_0) = \int_{0}^{t_1} f_{03}(s) S_0(s) ds$$

18. Suppose an individual is observed in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 1. The individual is observed in state 1 until t_{i+k-1}, where $k > 1$. At t_{i+k}, the individual is observed in state 3. We also have that $T_{03} > T_0$ and $T_{13} < T_1$. Then

$$\begin{aligned}
&Pr(t_i < T_0 < t_{i+1}, t_{i+k-1} < T_0 + T_{13} < t_{i+k}, T_{03} > T_0, T_{13} < T_1) \\
&= Pr(t_i < T_0 < t_{i+1}, t_{i+k-1} - T_0 < T_{13} < t_{i+k}, T_{03} > T_0, T_{13} < T_1) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) S_{03}(s) Pr(t_{i+k-1} - s < T_{13} < t_{i+k} - s) Pr(T_{13} < T_1) ds \\
&= \int_{t_i}^{t_{i+1}} \int_{t_{i+k-1}-s}^{t_{i+k}-s} f_0(s) S_{03}(s) f_{13}(u) S_1(u) du ds.
\end{aligned}$$

19. Suppose an individual is observed in state 1 from t_1 to t_i. At t_{i+1}, the individual is observed in state 3. We also have that $T_{03} > T_0$ and $T_{13} < T_1$. Then

$$\begin{aligned}
&Pr(T_0 < t_1, t_i < T_0 + T_{13} < t_{i+1}, T_{03} > T_0, T_{13} < T_1) \\
&= Pr(0 < T_0 < t_1, t_i - T_0 < T_{13} < t_{i+1} - T_0, T_{03} > T_0, T_{13} > T_1) \\
&= \int_{0}^{t_1} f_0(s) S_{03}(s) Pr(t_i - s < T_{13} < t_{i+1} - s) Pr(T_{13} > T_1) ds \\
&= \int_{0}^{t_1} \int_{t_i-s}^{t_{i+1}-s} f_0(s) S_{03}(s) f_{13}(u) S_1(u) du ds.
\end{aligned}$$

20. Suppose an individual is observed in state 3 at the first screening point t_1 with no intermittent screening. We also have that $T_{03} > T_0$ and $T_{13} < T_1$. Then

$$\begin{aligned}
&Pr(t_1 > T_{13} + T_0, T_{13} < T_1, T_{03} > T_0) \\
&= Pr(0 < T_0 < t_1, 0 < T_0 + T_{13} < t_1, T_{13} < T_1, T_{03} > T_0) \\
&= \int_0^{t_1} \int_0^{t_1 - s} f_0(s) S_{03}(s) f_{13}(u) S_1(u) du ds.
\end{aligned}$$

21. Suppose an individual is seen in state 0 from t_1 to t_i. At t_{i+1}, the individual is observed in state 3. We also have that $T_{03} > T_0$ and $T_{13} < T_1$. Then

$$\begin{aligned}
&Pr(t_i < T_0 < t_{i+1}, T_{03} > T_0, T_{13} < T_1, T_0 + T_{13} < t_{i+1}) \\
&= \int_{t_i}^{t_{i+1}} f_0(s) S_{03}(s) Pr(T_{13} < t_{i+1} - T_0 | T_0 = s) Pr(T_{13} < T_1) ds \\
&= \int_{t_i}^{t_{i+1}} \int_0^{t_{i+1} - s} f_0(s) S_{03}(s) f_{13}(u) S_1(u) du ds.
\end{aligned}$$

The full likelihood for the individuals $p = 1, \ldots, m$ then becomes

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) = &\prod_{(I)} S_0(t_{n,p}, \boldsymbol{\theta}|x_p) S_{03}(t_{n,p}, \boldsymbol{\theta}|x_p) \\
&\prod_{(II)} \int_{t_{i,p}}^{t_{i+1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_1(t_{n,p} - s, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(t_{n,p} - s, \boldsymbol{\theta}|x_p) ds \\
&\prod_{(III)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p}-s}^{t_{i+k,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) S_2(t_{n,p} - u - s, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
&\prod_{(IV)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p}-s}^{t_{i+k,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) (S_2(t_{i+k+l-1,p} - s - u, \boldsymbol{\theta}|x_p) - S_2(t_{i+k+l} - s - u, \boldsymbol{\theta}|x_p)) \\
&S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
&\prod_{(V)} \int_{t_{i,p}}^{t_{i+1,p}} \int_0^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) (1 - S_2(t_{i+1,p} - s - u, \boldsymbol{\theta}|x_p)) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
&\prod_{(VI)} \int_{t_{i,p}}^{t_{i+1,p}} \int_0^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) S_2(t_{n,p} - s - u, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
&\prod_{(VII)} \int_{t_{i,p}}^{t_{i+1,p}} \int_0^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) (S_2(t_{i+k-1,p} - u - s, \boldsymbol{\theta}|x_p) - S_2(t_{i+k,p} - u - s, \boldsymbol{\theta}|x_p)) \\
&S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
&\prod_{(VIII)} \int_0^{t_{1,p}} f_0(s, \boldsymbol{\theta}|x_p) S_1(t_{n,p} - s, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(t_{n,p} - s, \boldsymbol{\theta}|x_p) ds \\
&\prod_{(IX)} \int_0^{t_{1,p}} \int_{t_{i,p}-s}^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) S_2(t_{n,p} - s - u, \boldsymbol{\theta}|x_p) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
&\prod_{(X)} \int_0^{t_{1,p}} \int_{t_{i,p}-s}^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) (S_2(t_{i+k-1,p} - s - u, \boldsymbol{\theta}|x_p) - S_2(t_{i+k,p} - s - u, \boldsymbol{\theta}|x_p)) \\
&S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds \\
&\prod_{(XI)} \int_0^{t_{1,p}} \int_{t_{i,p}-s}^{t_{i+1,p}-s} f_0(s, \boldsymbol{\theta}|x_p) f_1(u, \boldsymbol{\theta}|x_p) (1 - S_2(t_{i+1,p} - s - u, \boldsymbol{\theta}|x_p)) S_{03}(s, \boldsymbol{\theta}|x_p) S_{13}(u, \boldsymbol{\theta}|x_p) du ds
\end{aligned}$$

$$\prod_{(XII)} \int_0^{t_{1,p}} \int_0^{t_{1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) S_2(t_{n,p}-u-s,\boldsymbol{\theta}|x_p) S_{03}(s,\boldsymbol{\theta}|x_p) S_{13}(u,\boldsymbol{\theta}|x_p) duds$$

$$\prod_{(XIII)} \int_0^{t_{1,p}} \int_0^{t_{1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) (S_2(t_{i,p}-s-u,\boldsymbol{\theta}|x_p) - S_2(t_{i+1,p}-s-u,\boldsymbol{\theta}|x_p))$$
$$S_{03}(s,\boldsymbol{\theta}|x_p) S_{13}(u,\boldsymbol{\theta}|x_p) duds$$

$$\prod_{(XIV)} \int_0^{t_{1,p}} \int_0^{t_{1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) (1 - S_2(t_{1,p}-s-u,\boldsymbol{\theta}|x_p))$$
$$S_{03}(s,\boldsymbol{\theta}|x_p) S_{13}(u,\boldsymbol{\theta}|x_p) duds$$

$$\prod_{(XV)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1}}^{t_{i+k,p}-s} f_0(s,\boldsymbol{\theta}|x_p) f_1(u,\boldsymbol{\theta}|x_p) (1 - S_2(t_{i+k,p}-s-u,\boldsymbol{\theta}|x_p))$$
$$S_{03}(s,\boldsymbol{\theta}|x_p) S_{13}(u,\boldsymbol{\theta}|x_p) duds$$

$$\prod_{(XVI)} \int_{t_{i,p}}^{t_{i+1,p}} f_{03}(s,\boldsymbol{\theta}|x_p) S_0(s,\boldsymbol{\theta}|x_p) ds$$

$$\prod_{(XVII)} \int_0^{t_{1,p}} f_{03}(s,\boldsymbol{\theta}|x_p) S_0(s,\boldsymbol{\theta}|x_p) ds$$

$$\prod_{(XVIII)} \int_{t_{i,p}}^{t_{i+1,p}} \int_{t_{i+k-1,p}-s}^{t_{i+k,p}-s} f_0(s,\boldsymbol{\theta}|x_p) S_{03}(s,\boldsymbol{\theta}|x_p) f_{13}(u,\boldsymbol{\theta}|x_p) S_1(u,\boldsymbol{\theta}|x_p) duds$$

$$\prod_{(XIX)} \int_0^{t_{1,p}} \int_{t_{i,p}-s}^{t_{i+1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) S_{03}(s,\boldsymbol{\theta}|x_p) f_{13}(u,\boldsymbol{\theta}|x_p) S_1(u,\boldsymbol{\theta}|x_p) duds$$

$$\prod_{(XX)} \int_0^{t_{1,p}} \int_0^{t_{1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) S_{03}(s,\boldsymbol{\theta}|x_p) f_{13}(u,\boldsymbol{\theta}|x_p) S_1(u,\boldsymbol{\theta}|x_p) duds$$

$$\prod_{(XXI)} \int_{t_{i,p}}^{t_{i+1,p}} \int_0^{t_{i+1,p}-s} f_0(s,\boldsymbol{\theta}|x_p) S_{03}(s,\boldsymbol{\theta}|x_p) f_{13}(u,\boldsymbol{\theta}|x_p) S_1(u,\boldsymbol{\theta}|x_p) duds.$$

APPENDIX B

Recipe for Simulations for Multiple Screenings

B.1 Three-State Progressive Model

We consider the following simulation scheme in a three-state progressive model for multiple screenings

1. Simulate $T_{0,p}$ and $T_{1,p}$ as described in Section 4.1
2. Draw the time points for the screenings from a uniform distribution

$$t_1 \sim U[0.5, 30],\ t_2 \sim U[t_1, 40],\ t_3 \sim U[t_1, 50],,$$

$$t_4 \sim U[t_1, 60],\ t_5 \sim U[t_1, 70],\ t_6 \sim U[t_1, 80],,$$

$$t_7 \sim U[t1, 90],\ t_8 \sim U[t_1, 100],\ t_9 \sim U[t_1, 110],,$$

$$t_{10} \sim U[t_1, 120],\ t_{11} \sim U[t_1, 130],\ t_{12} \sim U[t_1, 140],$$

$$t_{13} \sim U[t_1, 150],\ t_{14} \sim U[t_1, 160],\ t_{15} \sim U[t_1, 170],$$

   Then $\boldsymbol{t}$ will consist of the data which has increasing order. This means if $t_1 < t_2 < t_3 < t_4 > t_5$, then $\boldsymbol{t} = [t_1, t_2, t_3, t_4]$. For n screenings

   a) if $T_{0,p} > t_n$, then the patient is type 1.
   b) if any $\boldsymbol{t} < T_{0,p}$, any $\boldsymbol{t} > T_{0,p}$ and in addition if $t_n < T_{0,p} + T_{1,p}$, then the patient is type 2.
   c) if any $\boldsymbol{t} > T_{0,p} + T_{1,p}$, any $\boldsymbol{t} < T_{0,p}$ and if the maximum value of $\boldsymbol{t} < T_{0,p} + T_{1,p}$ is larger or equal to the minimum value where $\boldsymbol{t} > T_{0,p}$, then the patient is type 3.
   d) if $t_1 > T_{0,p}$ and $t_n < T_{0,p} + T_{1,p}$, then the patient is type 4.
   e) if any $\boldsymbol{t} > T_{0,p} + T_{1,p}$, $t_1 > T_{0,p}$ and if any $\boldsymbol{t} < T_{0,p} + T_{1,p}$, then the patient is type 5.
   f) if any $\boldsymbol{t} > T_{0,p} + T_{1,p}$ and if any $\boldsymbol{t} < T_{0,p}$ and it is not type 3, then the patient is type 6.
   g) if $t_1 > T_{0,p} + T_{1,p}$, then the patient is type 7.

3. Use these data to optimize the log-likelihood function and find the maximum likelihood estimates.

If we observe the exact time of death, we have that the time of death is exactly $T_{0,p} + T_{1,p}$. However, if we do not observe the exact time of death, the time of death is the first $\boldsymbol{t} > T_{0,p} + T_{1,p}$.

B.2 Illness-Death Model

We consider the following simulation scheme in an illness-death model for multiple screenings

1. Simulate $T_{0,p}$, $T_{1,p}$ and $T_{02,p}$ as described in Section 4.1
2. Draw the time points for the screenings from a uniform distribution

$$t_1 \sim U[0.5, 20],\ t_2 \sim U[t_1, 30],\ t_3 \sim U[t_1, 40],$$
$$t_4 \sim U[t_1, 50],\ t_5 \sim U[t_1, 60],\ t_6 \sim U[t_1, 70],$$
$$t_7 \sim U[t1, 80],\ t_8 \sim U[t_1, 90],\ t_9 \sim U[t_1, 100],$$
$$t_{10} \sim U[t_1, 110],\ t_{11} \sim U[t_1, 120],\ t_{12} \sim U[t_1, 130],$$
$$t_{13} \sim U[t_1, 140],\ t_{14} \sim U[t_1, 150],\ t_{15} \sim U[t_1, 160],$$

   Then $\boldsymbol{t}$ will consist of the data which has increasing order. This means if $t_1 < t_2 < t_3 < t_4 > t_5$, then $\boldsymbol{t} = [t_1, t_2, t_3, t_4]$. When we have n screenings

   a) if $T_{0,p} > t_n$ and $T_{02,p} > t_n$, then the patient is type 1.
   b) if $T_{02,p} > T_{0,p}$, any $\boldsymbol{t} < T_{0,p}$, any $\boldsymbol{t} > T_{0,p}$ and in addition if $t_n < T_{0,p} + T_{1,p}$, then the patient is type 2.
   c) if $T_{02,p} > T_{0,p}$, any $\boldsymbol{t} > T_{0,p} + T_{1,p}$ and any $\boldsymbol{t} < T_{0,p}$, the patient may be of type 3. However, we must also have that the maximum value where $\boldsymbol{t} < T_{0,p} + T_{1,p}$ is larger than the minimum value where $\boldsymbol{t} > T_{0,p}$ for the patient to be type 3.
   d) if $T_{02,p} > T_{0,p}$, $t_1 > T_{0,p}$ and $t_n < T_{0,p} + T_{1,p}$, then the patient is type 4.
   e) if $T_{02,p} > T_{0,p}$, $t_1 > T_{0,p}$, any $\boldsymbol{t} > T_{0,p}+T_{1,p}$ and if any $\boldsymbol{t} < T_{0,p}+T_{1,p}$, then the patient is type 5.
   f) if $T_{02,p} > T_{0,p}$, any $\boldsymbol{t} < T_{0,p}$, any $\boldsymbol{t} > T_{0,p} + T_{1,p}$ the patient may be of type 6. However, we must also have that if $t_{s,p} < T_{0,p}$, then $t_{s+1,p} > T_{0,p} + T_{1,p}$ for the patient to be type 6.
   g) if $t_1 > T_{0,p} + T_{1,p}$ and $T_{02,p} > T_{0,p}$, then the patient is type 7.
   h) if $T_{02,p} < T_{0,p}$ and $t_1 > T_{02,p}$, then the patient is type 8.
   i) if $T_{02,p} < T_{0,p}$, any $T_{02,p} > \boldsymbol{t}$ and any $\boldsymbol{t} > T_{02,p}$, then the patient is type 9.

3. Use these data to optimize the log-likelihood function and find the maximum likelihood estimates.

If we observe the exact time of death, the time of death is $T_{0,p} + T_{1,p}$ or $T_{02,p}$ depending on the type the individual is. However, if we do not observe the exact time of death, the time of death is the first $\boldsymbol{t} > T_{0,p} + T_{1,p}$ or $\boldsymbol{t} > T_{02,p}$, depending on the type the individual is.

B.3 Four-State Progressive Model

We consider the following simulation scheme in a four-state progressive model for multiple screenings

1. Simulate $T_{0,p}$, $T_{1,p}$ and $T_{2,p}$ as described in Section 4.1
2. Draw the time points for the screenings from a uniform distribution

$$t_1 \sim U[0.5, 50],\ t_2 \sim U[t_1, 60],\ t_3 \sim U[t_1, 70],$$
$$t_4 \sim U[t_1, 80],\ t_5 \sim U[t_1, 90],\ t_6 \sim U[t_1, 100],$$
$$t_7 \sim U[t1, 110],\ t_8 \sim U[t_1, 120],\ t_9 \sim U[t_1, 130],$$
$$t_{10} \sim U[t_1, 140],\ t_{11} \sim U[t_1, 150],\ t_{12} \sim U[t_1, 160],$$
$$t_{13} \sim U[t_1, 157],\ t_{14} \sim U[t_1, 180],\ t_{15} \sim U[t_1, 190],$$

   Then $\boldsymbol{t}$ will consist of the data which has increasing order. This means if $t_1 < t_2 < t_3 < t_4 > t_5$, then $\boldsymbol{t} = [t_1, t_2, t_3, t_4]$ When we have n screenings

   a) if $T_{0,p} > t_n$, then the patient is type 1.
   b) if any $\boldsymbol{t} < T_{0,p}$ and any $\boldsymbol{t} > T_{0,p}$ and in addition if $t_n < T_{0,p} + T_{1,p}$, then the patient is type 2.
   c) if any $\boldsymbol{t} < T_{0,p}$ and any $\boldsymbol{t} > T_{0,p} + T_{1,p}$ and $T_{0,p} + T_{1,p} + T_{2,p} > t_n$, then the patient is type 3.
   d) if any $\boldsymbol{t} < T_{0,p}$, any $\boldsymbol{t} > T_{0,p} + T_{1,p}$ and any $\boldsymbol{t} > T_{0,p} + T_{1,p} + T_{2,p}$, the patient may be type 4. However, we must also have that the minimum value where $\boldsymbol{t} > T_{0,p} + T_{1,p}$ is smaller than the minimum value where $\boldsymbol{t} > T_{0,p} + T_{1,p} + T_{2,p}$ for the patient to be type 4.
   e) if any $\boldsymbol{t} < T_{0,p}$ and any $\boldsymbol{t} > T_{0,p} + T_{1,p} + T_{2,p}$, then the patient may be type 5. However, we must also have that for the $\boldsymbol{t} < T_{0,p}$, for example $t_{s,p} < T_{0,p}$, then $t_{s+1,p} > T_{0,p} + T_{1,p} + T_{2,p}$, for the patient to be type 5.
   f) if any $\boldsymbol{t} < T_{0,p}$, any $\boldsymbol{t} > T_{0,p} + T_{1,p}$ and $t_n < T_{0,p} + T_{1,p} + T_{2,p}$, then the patient may be type 6. However, we must also have for any i, if $t_{s,p} < T_{0,p}$, then $t_{s+1,p} > T_{0,p} + T_{1,p}$, for the patient to be type 6.
   g) if any $\boldsymbol{t} < T_{0,p}$, any $\boldsymbol{t} > T_{0,p} + T_{1,p}$ and any $\boldsymbol{t} > T_{0,p} + T_{1,p} + T_{2,p}$, then the patient may be of type 7. However, we must also have for any i, if for example $t_{s,p} < T_{0,p}$, then $t_{s+1,p} > T_{0,p} + T_{1,p}$ for the the patient to be type 7.
   h) if $t_1 > T_{0,p}$, $t_1 < T_{0,p} + T_{1,p}$ and $t_n < T_{0,p} + T_{1,p}$, then the patient is of type 8.
   i) if $t_1 > T_{0,p}$, $t_1 < T_{0,p} + T_{1,p}$, any $\boldsymbol{t} > T_{0,p} + T_{1,p}$ and $t_n < T_{0,p} + T_{1,p} + T_{2,p}$, then the patient is type 9.
   j) if $t_1 > T_{0,p}$, $t_1 < T_{0,p} + T_{1,p}$, any $\boldsymbol{t} > T_{0,p} + T_{1,p}$ and any $\boldsymbol{t} > T_{0,p} + T_{1,p} + T_{2,p}$, then the patient may be of type 10. However, we must also have that the minimum value where $\boldsymbol{t} > T_{0,p} + T_{1,p}$ is smaller than the minimum value where $\boldsymbol{t} > T_{0,p} + T_{1,p} + T_{2,p}$ for the patient to be type 10.

k) if $t_1 > T_{0,p}$, $t_1 < T_{0,p} + T_{1,p}$ and any $\boldsymbol{t} > T_{0,p} + T_{1,p} + T_{2,p}$, then the patient may be type 11. However, we must also have that if $t_{s,p} < T_{0,p} + T_{1,p}$ then $t_{s+1,p} > T_{0,p} + T_{1,p} + T_{2,p}$, for the patient to be type 11.

l) if $t_1 > T_{0,p} + T_{1,p}$ and $T_{0,p} + T_{1,p} + T_{2,p} > t_n$, then the patient is type 12.

m) if $t_1 > T_{0,p}+T_{1,p}$, $t_1 < T_{0,p}+T_{1,p}+T_{2,p}$ and any $\boldsymbol{t} > T_{0,p}+T_{1,p}+T_{2,p}$, then the patient is type 13.

n) if $t_1 > T_{0,p} + T_{1,p} + T_{2,p}$, then the patient is type 14.

o) if any $\boldsymbol{t} < T_{0,p}$, any $\boldsymbol{t} > T_{0,p}$ and any $\boldsymbol{t} > T_{0,p} + T_{1,p} + T_{2,p}$, then the patient may be type 15. However we must also have that for the $\boldsymbol{t} > T_{0,p}$, $t_{s,p} > T_{0,p}$, then $t_{s,p} < T_{0,p} + T_{1,p}$ and $t_{s+1,p} > T_{0,p} + T_{1,p} + T_{2,p}$, for the patient to be type 15.

3. Use these data to optimize the log-likelihood function and find the maximum likelihood estimates.

B.4 Four-State Illness-death Model

We consider the following simulation scheme in a four-state illness-death model for multiple screenings

1. Simulate $T_{0,p}$, $T_{1,p}$, $T_{2,p}$, $T_{03,p}$ and $T_{13,p}$ as described in Section 4.1

2. Draw the time points for the screenings from a uniform distribution

$$t_1 \sim U[0.5, 50],\ t_2 \sim U[t_1, 60],\ t_3 \sim U[t_1, 70],$$

$$t_4 \sim U[t_1, 80],\ t_5 \sim U[t_1, 90],\ t_6 \sim U[t_1, 100],$$

$$t_7 \sim U[t1, 110],\ t_8 \sim U[t_1, 120],\ t_9 \sim U[t_1, 130],$$

$$t_{10} \sim U[t_1, 140],\ t_{11} \sim U[t_1, 150],\ t_{12} \sim U[t_1, 160],$$

$$t_{13} \sim U[t_1, 170],\ t_{14} \sim U[t_1, 180],\ t_{15} \sim U[t_1, 190],$$

Then $\boldsymbol{t}$ will consist of the data which has increasing order. This means if $t_1 < t_2 < t_3 < t_4 > t_5$, then $\boldsymbol{t} = [t_1, t_2, t_3, t_4]$ When we have n screenings

a) if $T_{0,p} > t_n$ and $T_{03,p} > t_n$, then the patient is type 1.

b) if any $\boldsymbol{t} < T_{0,p}$ and any $\boldsymbol{t} > T_{0,p}$, in addition if $T_{0,p} + T_{1,p} > t_n$, $T_{03,p} > T_{0,p}$ and $T_{13,p} + T_{0,p} > t_n$ then the patient is type 2.

c) if $T_{03,p} > T_{0,p}$, $T_{13,p} > T_{1,p}$, any $\boldsymbol{t} < T_{0,p}$, any $\boldsymbol{t} > T_{0,p} + T_{1,p}$ and $T_{0,p} + T_{1,p} + T_{2,p} > t_n$, then the patient is type 3.

d) if $T_{03,p} > T_{0,p}$, $T_{13,p} > T_{1,p}$, any $\boldsymbol{t} < T_{0,p}$ and any $\boldsymbol{t} > T_{0,p} + T_{1,p}$, then the patient may be type 4. However, we must also have that the minimum value where $\boldsymbol{t} > T_{0,p} + T_{1,p}$, is smaller than the minimum value where $\boldsymbol{t} > T_{0,p} + T_{1,p} + T_{2,p}$, for the patient to be type 4.

e) if $T_{03,p} > T_{0,p}$, $T_{13,p} > T_{1,p}$, any $\boldsymbol{t} < T_{0,p}$ and any $\boldsymbol{t} > T_{0,p} + T_{1,p} + T_{2,p}$, the patient may be type 5. However, we must also have that for any i, if $t_{s,p} < T_{0,p}$, then $t_{s+1,p} > T_{0,p} + T_{1,p} + T_{2,p}$, for the patient to be type 5.

f) if $T_{03,p} > T_{0,p}$, $T_{13,p} > T_{1,p}$, any $\boldsymbol{t} < T_{0,p}$, any $\boldsymbol{t} > T_{0,p} + T_{1,p}$ and $T_{0,p} + T_{1,p} + T_{2,p} > t_n$, then the patient may be type 6. However, we must also have that for any i, if $t_{s,p} < T_{0,p}$, then $t_{s+1,p} > T_{0,p} + T_{1,p}$, for the patient to be type 6.

g) if $T_{03,p} > T_{0,p}$, $T_{13,p} > T_{1,p}$, any $\boldsymbol{t} < T_{0,p}$, any $\boldsymbol{t} > T_{0,p} + T_{1,p}$ and any $\boldsymbol{t} > T_{0,p} + T_{1,p} + T_{2,p}$, then the patient may be type 7. However, we must also have that for any i, if $t_{s,p} < T_{0,p}$, then $t_{s+1,p} > T_{0,p} + T_{1,p}$, for the patient to be type 7.

h) if $T_{03,p} > T_{0,p}$, $T_{13,p} + T_{0,p} > t_n$, $t_1 > T_{0,p}$, $t_1 < T_{0,p} + T_{1,p}$ and $t_n < T_{0,p} + T_{1,p}$, then the patient is type 8.

i) if $T_{03,p} > T_{0,p}$, $T_{13,p} > T_{1,p}$, $t_1 > T_{0,p}$, $t_1 < T_{0,p} + T_{1,p}$, any $\boldsymbol{t} > T_{0,p} + T_{1,p}$ and $t_n < T_{0,p} + T_{1,p} + T_{2,p}$, then the patient i type 9.

j) if $T_{03,p} > T_{0,p}$, $T_{13,p} > T_{1,p}$, $t_1 > T_{0,p}$, $t_1 < T_{0,p} + T_{1,p}$, any $\boldsymbol{t} > T_{0,p} + T_{1,p}$ and any $\boldsymbol{t} > T_{0,p} + T_{1,p} + T_{2,p}$, then the patient may be type 10. However, we must also have that the minimum value where $\boldsymbol{t} > T_{0,p} + T_{1,p}$ is smaller than the minimum value where $\boldsymbol{t} > T_{0,p} + T_{1,p} + T_{2,p}$ for the patient to be type 10.

k) if $T_{03,p} > T_{0,p}$, $T_{13,p} > T_{1,p}$, $t_1 > T_{0,p}$, $t_1 < T_{0,p} + T_{1,p}$ and any $\boldsymbol{t} > T_{0,p} + T_{1,p} + T_{2,p}$, then the patient may be type 11. However, we must also have that if $t_{s,p} < T_{0,p} + T_{1,p}$,then $t_{s+1,p} > T_{0,p} + T_{1,p} + T_{2,p}$ for the patient to be type 11.

l) if $T_{03,p} > T_{0,p}$, $T_{13,p} > T_{1,p}$, $t_1 > T_{0,p} + T_{1,p}$ and $T_{0,p} + T_{1,p} + T_{2,p} > t_n$, then the patient is type 12.

m) if $T_{03,p} > T_{0,p}$, $T_{13,p} > T_{1,p}$, $t_1 > T_{0,p} + T_{1,p}$, $t_1 < T_{0,p} + T_{1,p} + T_{2,p}$ and any $\boldsymbol{t} > T_{0,p} + T_{1,p} + T_{2,p}$, then the patient is type 13.

n) if $T_{03,p} > T_{0,p}$, $T_{13,p} > T_{1,p}$, $t_1 > T_{0,p} + T_{1,p} + T_{2,p}$, then the patient is type 14.

o) if $T_{03,p} > T_{0,p}$, $T_{13,p} > T_{1,p}$, any $\boldsymbol{t} < T_{0,p}$, any $\boldsymbol{t} > T_{0,p} + T_{1,p} + T_{2,p}$, then the patient may be type 15. However, we must also have that for any i, if $t_{s,p} > T_{0,p}$, then $t_{s,p} < T_{0,p} + T_{1,p}$ and $t_{s+1,p} > T_{0,p} + T_{1,p} + T_{2,p}$, for the patient to be type 15.

p) if $T_{03,p} < T_{0,p}$, any $\boldsymbol{t} < T_{03,p}$ and any $\boldsymbol{t} > T_{03,p}$, then the patient is of type 16.

q) if $T_{03,p} < T_{0,p}$ and $t_1 > T_{03,p}$, then the patient is type 17.

r) if $T_{03,p} > T_{0,p}$, $T_{13,p} < T_{1,p}$, any $\boldsymbol{t} < T_{0,p}$, any $\boldsymbol{t} > T_{0,p}$, any $\boldsymbol{t} > T_{0,p} + T_{13,p}$ then the patient may be of type 18. However, we must also have that the maximum value where $\boldsymbol{t} < T_{0,p}$ is smaller than the maximum value where $\boldsymbol{t} < T_{0,p} + T_{13,p}$, for the patient to be type 18.

s) if $T_{03,p} > T_{0,p}$, $T_{13,p} < T_{1,p}$, $t_1 > T_{0,p}$ and any $\boldsymbol{t} > T_{0,p} + T_{13,p}$, then the patient is type 19.

t) if $T_{03,p} > T_{0,p}$, $T_{13,p} < T_{1,p}$ and $T_{13,p} + T_{0,p} < T_{1,p}$, then the patient is type 20.

u) if $T_{03,p} > T_{0,p}$, $T_{13,p} < T_{1,p}$, any $\boldsymbol{t} < T_{0,p}$, any $\boldsymbol{t} > T_{0,p}$ and any $\boldsymbol{t} > T_{0,p} + T_{13,p}$, then the patient may be type 21. However, we must also have that the maximum value where $\boldsymbol{t} < T_{0,p}$, for example $t_{s,p} < T_{0,p}$, then $t_{s+1,p} > T_{0,p} + T_{13,p}$ for the patient to be type 21.

3. Use these data to optimize the log-likelihood function and find the maximum likelihood estimates.

APPENDIX C

Analysis of the CAV-Data in the Illness-Death Model

In Chapter 6, we presented the analysis for a four-state illness-death model using the CAV data. In this appendix, we consider the illness-death model illustrated in Figure C.1.

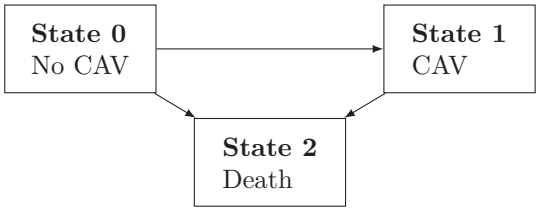


Figure C.1: Illness-death model

Firstly, if we observe an individual only in state 0 and state 2, the individual can either transfer directly from state 0 to state 2 or go through state 1 on the way to state 2. We are then considering type 1, 2, 3 and the combination 6 or 9. The log-likelihood becomes

$$\begin{aligned}\ell = &\sum_{(I)} \log\Big(S_0(t_{n,p},\boldsymbol{\theta}|x_p)S_{02}(t_{n,p},\boldsymbol{\theta}|x_p)\Big)\\ &+\sum_{(II)} \log\Big(\int_{t_{i,p}}^{t_{i+1,p}} f_0(s,\boldsymbol{\theta}|x_p)S_1(t_{n,p}-s,\boldsymbol{\theta}|x_p)S_{02}(s|x_p)ds\Big)\\ &+\sum_{(III)} \log\Big(\int_{t_{i,p}}^{t_{i+1,p}} f_0(s,\boldsymbol{\theta}|x_p)f_1(t_{i+k,p}-s,\boldsymbol{\theta}|x_p)S_{02}(s,\boldsymbol{\theta}|x_p)ds\Big)\\ &+\sum_{(IV)} \log\Big(\int_{t_{i,p}}^{t_{i+1,p}} f_0(s,\boldsymbol{\theta}|x_p)f_1(t_{i+1,p}-s,\boldsymbol{\theta}|x_p)S_{02}(s,\boldsymbol{\theta}|x_p)ds\\ &+f_{02}(t_{i+1,p},\boldsymbol{\theta}|x_p)S_0(t_{i+1,p},\boldsymbol{\theta}|x_p)\Big).\end{aligned}$$

C.1 Gamma Process Model, Alternative 1

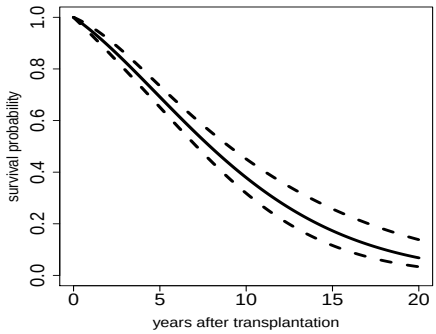
We consider a Gamma process model without covariates. Let the survival function from state 0 to state 1 be $S_0(t) = Gam(c_0, a_0 t, 1)$, and likewise for S_1

and S_{02}.

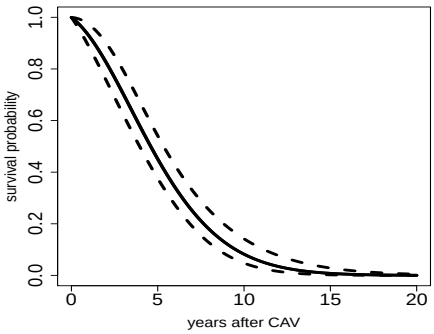
Parameter	**Estimate**	**Standard error**
$\hat{c}_0$	0.730	0.223
$\hat{a}_0$	0.131	0.0315
$\hat{c}_1$	1.229	0.443
$\hat{a}_1$	0.337	0.0849
$\hat{c}_{02}$	0.0183	0.0514
$\hat{a}_{02}$	0.0127	0.00985

Table C.1: Estimates and standard errors in an illness-death Gamma process model without covariates, alternative 1

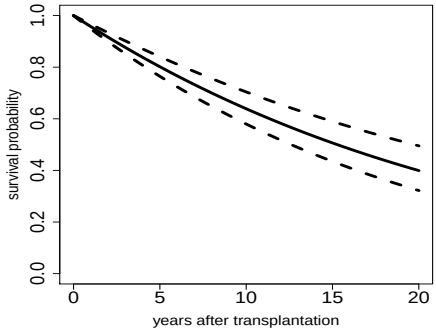
The maximum likelihood estimates of the parameters and their standard errors are presented in Table C.1. They are calculated in the same way as in Chapter 6. The standard error is the square root of the diagonal of the inverse Hessian matrix. The estimated threshold, $\hat{c}_{02}$, is quite close to 0, while $\hat{c}_0$ is a bit below 1 and $\hat{c}_1$ is a bit above 1. at is the shape parameter, and $\hat{a}_0$, $\hat{a}_1$ and $\hat{a}_{02}$ are very close to 0. The thresholds and the shape parameters decides how fast an individual transfers to the next state. For example, after 10 years, the probability of not transitioning from healthy to CAV is 0.380. The probability of not transitioning from CAV to death 10 year after the individual was diagnosed with CAV is 0.0820. Finally, the probability of not transitioning from healthy to death after 10 years i 0.638.



(a) State 0 to state 1



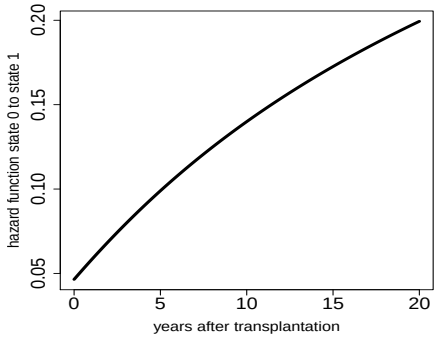
(b) State 1 to state 2



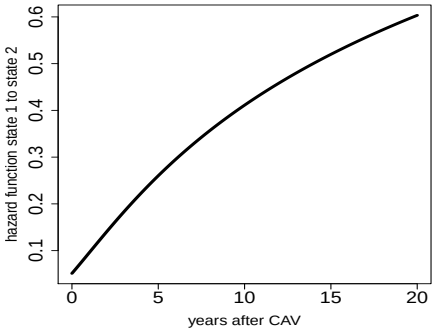
(c) State 0 to state 2

Figure C.2: Survival functions in an illness-death Gamma process model without covariates, alternative 1

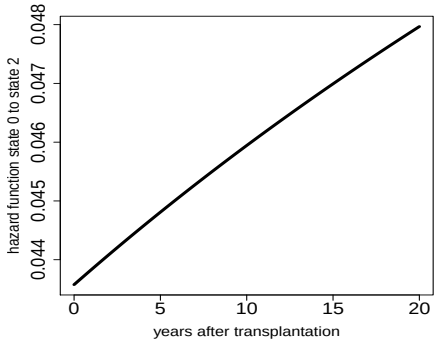
Figure C.2 shows the survival functions S_0, S_1 and S_{02} with a 95% pointwise confidence interval. Survival for S_0 means not entering state 1, survival for S_1 and S_{02} means not entering the absorbing state 2. The probability of not entering state 1 from state 0 is decreasing and around 0.1 after 20 years. The probability of not entering state 2 from state 0 is very slowly decreasing, and after 20 years, it is around 0.4. However, if the individual get the diagnosis CAV, then the probability of surviving 10 years is almost 0. The confidence intervals follows the shape of the survival functions quite closely. Both in Figure C.2 (a) and (c), the confidence intervals becomes wider as time goes. This follows from fewer people transitioning after 15 years, compared to after 5 years.



(a) State 0 to state 1



(b) State 1 to state 2



(c) State 0 to state 2

Figure C.3: Hazard functions in an illness-death Gamma process model without covariates, alternative 1

Figure C.3 shows the hazard functions. All of the hazard functions are increasing, meaning the instantaneous risk of transitioning becomes larger with time.

C.2 Gamma Process Model, Alternative 2

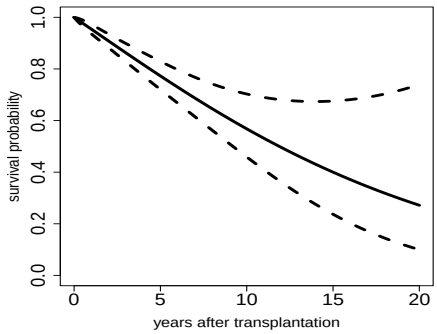
We also consider Gamma process model alternative 2, where the survival function is $S_0 = Gam(c_0, a_0 t^{b_0}, 1)$, with similar shape for S_1 and S_{02}. Table C.2 presents the maximum likelihood estimates. For the transition from state 0 to state 1, we have slightly different parameters compared to alternative 1. $\hat{b}_0$ is larger than 1, while both $\hat{a}_0$ and $\hat{c}_0$ are smaller in alternative 1 compared to alternative 2. $\hat{b}_1$ is a bit smaller than 1. $\hat{c}_1$ and $\hat{a}_1$ are also smaller in alternative 2 compared to alternative 1. However, for the transition from state 0 to state 2, we have that both $\hat{c}_{02}$ and $\hat{a}_{02}$ are larger compared to alternative 1. In addition we have that $\hat{b}_{02}$ smaller than 1. It also makes it possible for the hazard function to have a different shape. In Table C.2, we also calculate the p-value for the null hypothesis $H_0 : b = 1$. We find that we can reject the null hypothesis at a 1%-level for b_{02}. However, we can not reject the null hypothesis for b_0 and b_1.

Parameter	**Estimate**	**Standard error**	**p-value** ($H_0 : b = 1$)
$\hat{c}_0$	0.262	0.478	
$\hat{b}_0$	1.251	0.243	0.301
$\hat{a}_0$	0.0454	0.0663	
$\hat{c}_1$	0.491	0.692	
$\hat{b}_1$	0.976	0.240	0.973
$\hat{a}_1$	0.214	0.259	
$\hat{c}_{02}$	0.556	0.723	
$\hat{b}_{02}$	0.477	0.0801	6.846×10^{-11}
$\hat{a}_{02}$	0.134	0.171	

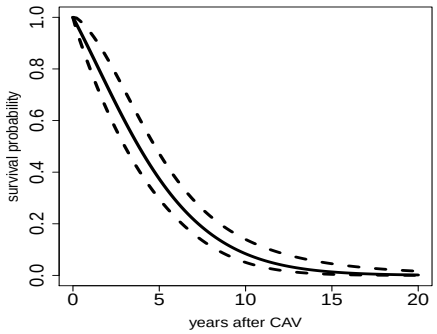
Table C.2: Estimates, standard errors and Wald-test for the b-parameters in a illness-death Gamma process model without covariates, alternative 2

Figure C.4 shows the survival functions S_0, S_1 and S_{02} with a pointwise 95% confidence interval. Survival for S_0 means not entering state 1, survival for S_1 and S_{02} means not entering the absorbing sate 2. A 10-year survival probability is around 0.3 for S_0 and around 0.8 for S_{02}, while it is around 0.1 for S_1. This means that if you get CAV, the probability of surviving 10 years after you get CAV is around 0.1. If you do not get CAV, the probability of surviving 10 years without going through state 1 is 0.8. The survival probability is decreasing faster from state 0 to state 1, quite equal from state 1 to state 2, but slower for state 0 to state 2.

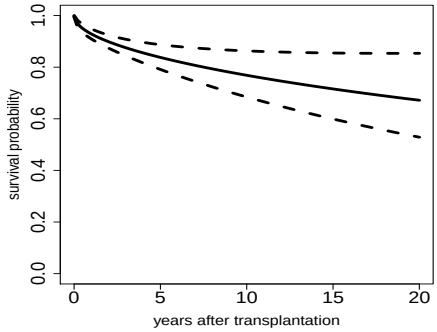
We present the plots of the hazard functions in Figure C.5. The hazard functions are increasing and concave for the transitions $0 \to 1$ and $1 \to 2$. The hazard function for the transition $0 \to 2$ decreases fast in the beginning, but then it is quite constant and decreasing toward 0. The shape of the hazard function for Figure C.5 (c) says that the instantaneous risk of dying in a small time interval is much higher in the beginning, before it becomes very low. In alternative 1, the hazard function for the transition from $0 \to 2$, was increasing and concave in the complete time period. Since $\hat{b}_{02}$ in alternative 2 is much smaller than 1, we are able to capture this effect. In addition, we have from Table C.2 that we can reject the null hypothesis of b being equal to 1 even at a 1%-level. From a medical point of view, the hazard function for the transition $0 \to 2$ also makes sense. Since the body has undergone a massive change if you get a transplant, the probability of dying is very high in the beginning. Then the probability of dying decreases, even though you can get complications later



(a) State 0 to state 1



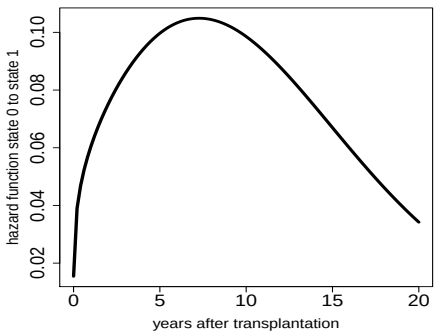
(b) State 1 to state 2



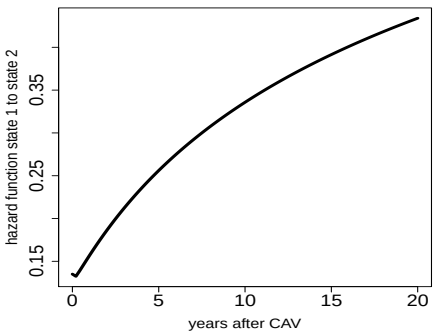
(c) State 0 to state 2

Figure C.4: Survival functions in an illness-death Gamma process model without covariates, alternative 2

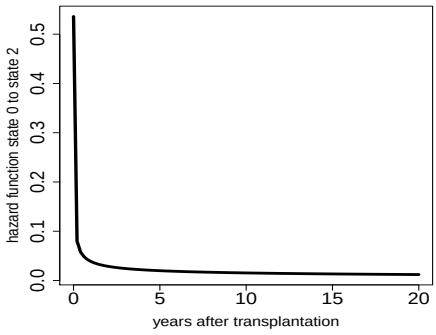
on.



(a) State 0 to state 1



(b) State 1 to state 2



(c) State 0 to state 2

Figure C.5: Hazard functions in an illness-death Gamma process model without covariates, alternative 2

APPENDIX D

Code

The statistical programming language used in this thesis is R (R Core Team, 2019). In this appendix, we include two examples of the code. In the first example, we show how we have written the code for the simulations in the four-state illness-death model from Chapter 4. In the second example, we provide the code for the CAV-analysis using the Gamma process model alternative 1 without covariates. We only provide these two examples, because including all of the scrips become too comprehensive. The rest of the code is available upon request. It should be possible to reproduce the results for the rest of the simulations and the CAV-data from these examples and the explanations provided in Chapter 4, Chapter 6 and Appendix B.

D.1 Simulations in a Four-State Illness-Death Model

In this part, we present the code from Section 4.5. We start with defining the true values for the parameters. In the next step, we calculate T_0, T_1, T_2, T_{03} and T_{13} in the same way as we explained in Section 4.1. Then we divide the individuals and place them in the correct likelihood type with the relevant timepoints. We also define the survival functions and their corresponding densities. Further, we use the timepoints when we create the functions for the different likelihood types. In the end, we define the function for the complete log-likelihood, which we optimize. From the maximum likelihood estimates, we also calculate the inverse of the Hessian-matrix. Note that we define the negative log-likelihood, since *optim* minimizes the function.

```
library(Rlab)
library(numDeriv)
nn = 100

theta_hat_par = matrix(NA, nrow = nn, ncol = 10)
theta_hat_solve_hessian = matrix(NA, nrow = nn, ncol = 10)

for (mn in 1:nn)
{
  set.seed(mn)
  print(mn)
  eps = 0.01
  c_0 = 5
  a_0 = 0.2
  c_1 = 3
  a_1 = 0.2
  c_2 = 4
```

```
a_2 = 0.1
c_03 = 6
a_03 = 0.15
c_13 = 4
a_13 = 0.15
n = 500

T_0 = rep(NA, n)
T_1 = rep(NA, n)
T_2 = rep(NA, n)
T_03 = rep(NA, n)
T_13 = rep(NA, n)
t_i1 = c()
t_i2 = matrix( , nrow = 0, ncol = 3)
t_i3 = matrix( , nrow = 0, ncol = 5)
t_i4 = matrix( , nrow = 0, ncol = 5)
t_i5 = matrix( , nrow = 0, ncol = 2)
t_i6 = matrix( , nrow = 0, ncol = 3)
t_i7 = matrix( , nrow = 0, ncol = 3)
t_i8 = matrix( , nrow = 0, ncol = 2)
t_i9 = matrix( , nrow = 0, ncol = 4)
t_i10 = matrix( , nrow = 0, ncol = 4)
t_i11 = matrix( , nrow = 0, ncol = 3)
t_i12 = matrix( , nrow = 0, ncol = 2)
t_i13 = matrix( , nrow = 0, ncol = 2)
t_i14 = c()
t_i15 = matrix(, nrow = 0, ncol = 4)
t_i16 = c()
t_i17 = c()
t_i18 = matrix(, nrow = 0, ncol = 3)
t_i19 = matrix(, nrow = 0, ncol = 2)
t_i20 = c()
t_i21 = matrix(, nrow = 0, ncol = 2)

for (i in 1:n)
{
  ## Solve equation for T_0, T_1, T_2, T_03, T_13
  unif1 = runif(1, min = 0, max = 1)
  S0 = function(T)
  {
    pgamma(c_0, a_0*T, 1) - unif1
  }
  T_0[i] = uniroot(S0, interval = c(1.e-14, 1e04), tol = 1e-9)$root

  unif2 = runif(1, min = 0, max = 1)
  S1 = function(T)
  {
    pgamma(c_1, a_1*T, 1) - unif2
  }
  T_1[i] = uniroot(S1, interval = c(1.e-14, 1e04), tol = 1e-9)$root

  unif3 = runif(1, min = 0, max = 1)
  S2 = function(T)
  {
    pgamma(c_2, a_2*T, 1) - unif3
  }
  T_2[i] = uniroot(S2, interval = c(1.e-14, 1e04), tol = 1e-9)$root

  unif4 = runif(1, min = 0, max = 1)
  S03 = function(T)
  {
    pgamma(c_03, a_03*T, 1) - unif4
```

```
  }
  T_03[i] = uniroot(S03, interval = c(1.e-14, 1e04), tol = 1e-9)$root
  unif5 = runif(1, min = 0, max = 1)
  S13 = function(T)
  {
    pgamma(c_13, a_13*T, 1) - unif5
  }
  T_13[i] = uniroot(S13, interval = c(1.e-14, 1e04), tol = 1e-9)$root

  ## Simulating time points
  t1 = runif(1, 0.5, 50)
  t2 = runif(1, t1, 60)
  t3 = runif(1, t1, 70)
  t4 = runif(1, t1, 80)
  t5 = runif(1, t1, 90)
  t6 = runif(1, t1, 100)
  t7 = runif(1, t1, 110)
  t8 = runif(1, t1, 120)
  t9 = runif(1, t1, 120)
  t10 = runif(1, t1, 140)
  t11 = runif(1, t1, 150)
  t12 = runif(1, t1, 160)
  t13 = runif(1, t1, 170)
  t14 = runif(1, t1, 180)
  t15 = runif(1, t1, 190)
  t = c(t1, t2, t3, t4, t5, t6, t7, t8, t9, t10, t11, t12, t13, t14, t15)
  if(all(t[1:14] - t[2:15]<0))
  {
    t = t
  } else if(all(t[1:13] - t[2:14]<0))
  {
    t = t[1:14]
  } else if(all(t[1:12] - t[2:13]<0))
  {
    t = t[1:13]
  } else if(all(t[1:11] - t[2:12]<0))
  {
    t = t[1:12]
  } else if(all(t[1:10] - t[2:11]<0))
  {
    t = t[1:11]
  } else if(all(t[1:9] - t[2:10]<0))
  {
    t = t[1:10]
  } else if (all(t[1:8] - t[2:9]<0))
  {
    t = t[1:9]
  } else if (all(t[1:7] - t[2:8]<0))
  {
    t = t[1:8]
  } else if (all(t[1:6] - t[2:7]<0))
  {
    t = t[1:7]
  } else if (all(t[1:5] - t[2:6]<0))
  {
    t = t[1:6]
  } else if (all(t[1:4] - t[2:5]<0))
  {
    t = t[1:5]
  } else if (all(t[1:3] - t[2:4]<0))
  {
    t = t[1:4]
```

```
142     } else if (all(t[1:2] - t[2:3]<0))
143     {
144       t = t[1:3]
145     } else if (all(t[1:1] - t[2:2]<0))
146     {
147       t = t[1:2]
148     }
149
150     ## Splitting the individuals into types
151     t_n = length(t)
152     if(t[t_n] < T_0[i] & t[t_n] < T_03[i]){
153       t_i1 = c(t_i1, t[t_n])
154     } else if(any(t < T_0[i]) & any(t > T_0[i]) & T_0[i] + T_1[i] > t[t_n] &
155               T_03[i] > T_0[i] & T_13[i] + T_0[i] > t[t_n]){
156       check_12 = max(which(t < T_0[i]))
157       check_22 = min(which(t > T_0[i]))
158       t_i2 = rbind(t_i2, c(t[check_12], t[check_22], t[t_n]))
159     } else if(t[1] > T_0[i] & t[t_n] < T_0[i] + T_1[i] &
160               T_03[i] > T_0[i] & T_13[i] + T_0[i] > t[t_n]){
161       t_i8 = rbind(t_i8, c(t[1], t[t_n]))
162     } else if(any(t > T_0[i] + T_1[i] + T_2[i]) & T_03[i] > T_0[i] & T_13[i] >
        T_1[i]){
163       if(any(t < T_0[i])){
164         if (max(which(t < T_0[i] + T_1[i])) >= min(which(t > T_0[i])) &
165             min(which(t > T_0[i] + T_1[i])) <= max(which(t < T_0[i] + T_1[i] +
        T_2[i]))){
166           check_14 = max(which(t < T_0[i]))
167           check_24 = min(which(t > T_0[i]))
168           check_34 = max(which(t < T_0[i] + T_1[i]))
169           check_44 = min(which(t > T_0[i] + T_1[i]))
170           t_i4 = rbind(t_i4, c(t[check_14], t[check_24], t[check_34],
171                                t[check_44], T_0[i] + T_1[i] + T_2[i]))
172         } else if(max(which(t < T_0[i])) + 1 == min(which(t > T_0[i]+T_1[i]+
173                                                           T_2[i]))){
174           check_15 = max(which(t < T_0[i]))
175           t_i5 = rbind(t_i5, c(t[check_15], T_0[i] + T_1[i] + T_2[i]))
176         } else if(max(which(t < T_0[i] + T_1[i])) >= min(which(t > T_0[i])) &
177                   max(which(t < T_0[i] + T_1[i])) + 1 == min(which(t > T_0[i]+
        T_1[i]+
178                                                                    T_2[i]))){
179           check_115 = max(which(t < T_0[i]))
180           check_215 = min(which(t > T_0[i]))
181           check_315 = max(which(t < T_0[i] + T_1[i]))
182           t_i15 = rbind(t_i15, c(t[check_115], t[check_215], t[check_315],
183                                  T_0[i] + T_1[i] + T_2[i]))
184         } else if(max(which(t < T_0[i])) + 1 == min(which(t > T_0[i]+T_1[i]))
        &
185                   min(which(t > T_0[i]+T_1[i])) <= max(which(t > T_0[i] +
186                                                              T_1[i] + T_2[i]))
        ){
187           check_17 = max(which(t < T_0[i]))
188           check_27 = min(which(t > T_0[i] + T_1[i]))
189           t_i7 = rbind(t_i7, c(t[check_17], t[check_27], T_0[i] + T_1[i] + T_
        2[i]))
190         }
191       } else if(T_0[i] < t[1] & any(t < T_0[i] + T_1[i])){
192         if(max(which(t < T_0[i] + T_1[i] + T_2[i])) >= min(which(t > T_0[i] +
        T_1[i]))){
193           check_110 = max(which(t < T_0[i] + T_1[i]))
194           check_210 = min(which(t > T_0[i] + T_1[i]))
195           t_i10 = rbind(t_i10, c(t[1], t[check_110], t[check_210],T_0[i] + T_
        1[i] + T_2[i]))
```

```
      }
      else if(max(which(t < T_0[i] + T_1[i])) + 1 == min(which(t > T_0[i]+T_1[i]+
                                                                T_2[i]))){
        check_111 = max(which(t < T_0[i] + T_1[i]))
        t_i11 = rbind(t_i11, c(t[1], t[check_111], T_0[i] + T_1[i] + T_2[i]))
      }
    } else if(t[1] > T_0[i] + T_1[i] & t[1] < T_0[i] + T_1[i] + T_2[i]){
      if(max(which(t < T_0[i] + T_1[i] + T_2[i])) <= min(which(t > T_0[i]+T_1[i]+T_2[i]))){
        t_i13 = rbind(t_i13, c(t[1], T_0[i] + T_1[i] + T_2[i]))
      }
    } else if(t[1] > T_0[i] + T_1[i] + T_2[i]){
      t_i14 = c(t_i14, T_0[i] + T_1[i] + T_2[i])
    }
  } else if(any(t < T_0[i]) & T_03[i] > T_0[i] & T_13[i] > T_1[i]){
    if(max(which(t < T_0[i] + T_1[i])) >= min(which(t > T_0[i])) &
       min(which(t > T_0[i] + T_1[i])) <= max(which(t < T_0[i] + T_1[i] + T_2[i]))){
      check_13 = max(which(t < T_0[i]))
      check_23 = min(which(t > T_0[i]))
      check_33 = max(which(t < T_0[i] + T_1[i]))
      check_43 = min(which(t > T_0[i] + T_1[i]))
      t_i3 = rbind(t_i3, c(t[check_13], t[check_23], t[check_33], t[check_43], t[t_n]))
    } else if(max(which(t < T_0[i])) + 1 == min(which(t > T_0[i]+T_1[i]))){
      check_16 = max(which(t < T_0[i]))
      check_26 = min(which(t > T_0[i] + T_1[i]))
      t_i6 = rbind(t_i6, c(t[check_16], t[check_26], t[t_n]))
    }
  } else if(t[1] > T_0[i] & any(t < T_0[i] + T_1[i]) & any(t > T_0[i] + T_1[i]) &
            T_03[i] > T_0[i] & T_13[i] > T_1[i]){
    check_19 = max(which(t < T_0[i] + T_1[i]))
    check_29 = min(which(t > T_0[i] + T_1[i]))
    t_i9 = rbind(t_i9, c(t[1], t[check_19], t[check_29], t[t_n]))
  } else if(t[1] > T_0[i] + T_1[i] & T_03[i] > T_0[i] & T_13[i] > T_1[i]){
    t_i12 = rbind(t_i12, c(t[1], t[t_n]))
  } else if(T_03[i] < T_0[i] & any(t > T_03[i]) & any(t < T_03[i])){
    t_i16 = c(t_i16, T_03[i])
  }
  else if(T_03[i] < T_0[i] & t[1] > T_03[i]){
    t_i17 = c(t_i17, T_03[i])
  }
  else if(T_03[i] > T_0[i] & T_13[i] < T_1[i] & any(t < T_0[i]) &
          any(t > T_0[i]) & any(t > T_0[i] + T_13[i])){
    if(max(which(t < T_0[i])) < max(which(t < T_0[i] + T_13[i]))){
      check_118 = max(which(t < T_0[i]))
      check_218 = min(which(t > T_0[i]))
      t_i18 = rbind(t_i18, c(t[check_118], t[check_218], T_0[i] + T_13[i]))
    } else if(max(which(t < T_0[i])+1) == min(which(t > T_0[i] + T_13[i]))){
      check_121 = max(which(t < T_0[i]))
      t_i21 = rbind(t_i21, c(t[check_121], T_0[i] + T_13[i]))
    }
  }
  else if(T_03[i] > T_0[i] & T_13[i] < T_1[i] & t[1] > T_0[i] &
          t[1] < T_0[i] + T_13[i] & any(t > T_0[i] + T_13[i])){
    t_i19 = rbind(t_i19, c(t[1], T_0[i] + T_13[i]))
  }
  else if(T_03[i] > T_0[i] & T_13[i] + T_0[i] < t[1] & T_13[i] < T_1[i]){
    t_i20 = c(t_i20, T_0[i] + T_13[i])
```

```
  }
}

## Initial survival functions
S_0 = function(a, t){pgamma(a[1], shape = t*a[2], rate = 1)}
S_1 = function(a, t){pgamma(a[3], shape = t*a[4], rate = 1)}
S_2 = function(a, t){pgamma(a[5], shape = t*a[6], rate = 1)}
S_03 = function(a, t){pgamma(a[7], shape = t*a[8], rate = 1)}
S_13 = function(a, t){pgamma(a[9], shape = t*a[10], rate = 1)}

## Initial density functions - derivative of -survival function
f_0 = function(a, t){(S_0(a, t) - S_0(a, t + eps))/(eps)}
f_1 = function(a, t){(S_1(a, t) - S_1(a, t + eps))/(eps)}
f_2 = function(a, t){(S_2(a, t) - S_2(a, t + eps))/(eps)}
f_03 = function(a, t){(S_03(a, t) - S_03(a, t + eps))/(eps)}
f_13 = function(a, t){(S_13(a, t) - S_13(a, t + eps))/(eps)}

## Other functions
f0_S1_S03_S13 = function(a, w, t){f_0(a, t)*S_03(a, t)*S_1(a, w - t)*S_13(a, w - t)}
f0_f1_S2_S03_S13 = function(a, w, t, u){f_0(a, t)*f_1(a, u)*S_2(a, w - u - t)*S_03(a, t)*S_13(a, u)}
f0_f1_f2_S03_S13 = function(a, w, t, u){f_0(a, t)*f_1(a, u)*f_2(a, w - t - u)*S_03(a, t)*S_13(a, u)}
f0_S1_S03_f13 = function(a, w, t){f_0(a, t)*S_03(a, t)*S_1(a, w - t)*f_13(a, w - t)}

## Type 1
type_1 = function(a){-sum(sapply(t_i1, function(t) (log(S_0(a, t)) + log(S_03(a, t)))))}

## Type 2
type_2 = function(a){-sum(mapply(function(t1, t2, t3)(log(as.numeric(
  integrate(f0_S1_S03_S13, lower = t1, upper = t2, w = t3, a = a)[1]))), t_i2[,1], t_i2[,2], t_i2[,3]))}

## Type 3
type_3 = function(a){-sum(mapply(function(t1, t2, t3, t4, t5){
  log(as.numeric(integrate(function(t){sapply(t, function(z){integrate(f0_f1_S2_S03_S13,
                                                                       lower = t3-z, upper = t4-z, w = t5, t = z, a = a)$value})},
                           lower = t1, upper = t2)[1]))}, t_i3[,1], t_i3[,2], t_i3[,3], t_i3[,4],
  t_i3[,5]))}

## Type 4
type_4 = function(a){ifelse(nrow(t_i4) > 0, -sum(mapply(function(t1, t2, t3, t4, t5){
  log(as.numeric(integrate(function(t){
    sapply(t, function(z){integrate(f0_f1_f2_S03_S13,
                                    lower = t3-z, upper = t4-z, w = t5, t = z, a = a)$value})},
    lower = t1, upper = t2)[1]))}, t_i4[,1], t_i4[,2], t_i4[,3], t_i4[,4], t_i4[,5])), 0)}

## Type 5
type_5 = function(a){ifelse(nrow(t_i5)>0,-sum(mapply(function(t1, t2){
  log(as.numeric(integrate(function(t){sapply(t, function(z){integrate(f0_f1_f2_S03_S13,
                                                                       lower = 0, upper = t2 - z, w = t2, t = z, a = a)$value})},
```

```
                              lower = t1, upper = t2)[1]))
}, t_i5[,1], t_i5[,2])), 0)}

## Type 6
type_6 = function(a){-sum(mapply(function(t1, t2, t3){
  log(as.numeric(integrate(function(t){sapply(t, function(z){integrate(f0_f1
    _S2_S03_S13,
                           lower = 0, upper = t2 - z, w = t3, t = z, a = a)$
    value})},
                              lower = t1, upper = t2)[1]))
}, t_i6[,1], t_i6[,2], t_i6[,3]))}

## Type 7
type_7 = function(a){ifelse(nrow(t_i7) > 0, -sum(mapply(function(t1, t2, t3)
    {
  log(as.numeric(integrate(function(t){sapply(t, function(z){integrate(f0_f1
    _f2_S03_S13,
                            lower = 0, upper = t2 - z, w = t3, t = z, a = a)
    $value})},
                              lower = t1, upper = t2)[1]))
}, t_i7[,1], t_i7[,2], t_i7[,3])),0)}

## Type 8
type_8 = function(a){-sum(mapply(function(t1, t2){
  log(as.numeric(integrate(f0_S1_S03_S13,
                              lower = 0, upper = t1, w = t2, a = a)[1]))
}, t_i8[,1], t_i8[,2]))}


## Type 9
type_9 = function(a){-sum(mapply(function(t1, t2, t3, t4){
  log(as.numeric(integrate(function(t){sapply(t, function(z){integrate(f0_f1
    _S2_S03_S13,
                           lower = t2-z, upper = t3-z, w = t4, t = z, a = a)$
    value})},
                              lower = 0, upper = t1)[1]))
}, t_i9[,1], t_i9[,2], t_i9[,3], t_i9[,4]))}

## Type 10
type_10 = function(a){-sum(mapply(function(t1, t2, t3, t4){
  log(as.numeric(integrate(function(t){sapply(t, function(z){integrate(f0_f1
    _f2_S03_S13,
                           lower = t2-z, upper = t3-z, w = t4, t = z, a = a)$
    value})},
                              lower = 0, upper = t1)[1]))
}, t_i10[,1], t_i10[,2], t_i10[,3], t_i10[,4]))}

## Type 11
type_11 = function(a){ifelse(nrow(t_i11) > 0, -sum(mapply(function(t1, t2,
    t3){
  log(as.numeric(integrate(function(t){sapply(t, function(z){integrate(f0_f1
    _f2_S03_S13,
                           lower = t2-z, upper = t3-z, w = t3, t = z, a = a)$
    value})},
                              lower = 0, upper = t1)[1]))
}, t_i11[, 1], t_i11[,2], t_i11[,3])), 0)}

## Type 12
type_12 = function(a){-sum(mapply(function(t1, t2){
  log(as.numeric(integrate(function(t){sapply(t, function(z){integrate(f0_f1
    _S2_S03_S13,
                           lower = 0, upper = t1 - z, w = t2, t = z, a = a)$
```

```
   value})},
                              lower = 0, upper = t1)[1]))
}, t_i12[, 1], t_i12[, 2]))}

## Type 13
type_13 = function(a){-sum(mapply(function(t1, t2){
  log(as.numeric(integrate(function(t){sapply(t, function(z){integrate(f0_f1
   _f2_S03_S13,
                             lower = 0, upper = t1-z, w = t2, t = z, a = a)$
   value})},
                              lower = 0, upper = t1)[1]))
}, t_i13[, 1], t_i13[, 2]))}

## Type 14
type_14 = function(a){ifelse(length(t_i14) > 0,
                               -sum(sapply(t_i14, function(t1) (log(as.numeric
   (integrate(function(t){
                                 sapply(t, function(z){integrate(f0_f1_f2_S03_
   S13,
                               lower = 0, upper = t1-z, w = t1, t = z, a = a)$
   value})},
                                 lower = 0, upper = t1)[1])))))), 0)}

## Type 15
type_15 = function(a){ifelse(nrow(t_i15) > 0, -sum(mapply(function(t1, t2,
   t3, t4){
  log(as.numeric(integrate(function(t){sapply(t, function(z){integrate(f0_f1
   _f2_S03_S13,
                             lower = t3 - z, upper = t4-z, w = t4, t = z, a = a
   )$value})},
                              lower = t1, upper = t2)[1]))
}, t_i15[, 1], t_i15[, 2], t_i15[, 3], t_i15[, 4])), 0)}

## Type 16
type_16 = function(a){-sum(sapply(t_i16, function(t) (log(S_0(a, t)) + log(f
   _03(a, t)))))}

## Type 17
type_17 = function(a){-sum(sapply(t_i17, function(t) (log(S_0(a, t)) + log(f
   _03(a, t)))))}

## Type 18
type_18 = function(a){-sum(mapply(function(t1, t2, t3)(log(as.numeric(
  integrate(f0_S1_S03_f13, lower = t1, upper = t2, w = t3, a = a)[1]))), t_
   i18[,1], t_i18[,2], t_i18[,3]))}

## Type 19
type_19 = function(a){-sum(mapply(function(t1, t2)(log(as.numeric(
  integrate(f0_S1_S03_f13, lower = 0, upper = t1, w = t2, a = a)[1]))), t_
   i19[,1], t_i19[,2]))}

## Type 20
type_20 = function(a){ifelse(length(t_i20) > 0,
                               -sum(sapply(t_i20, function(t1) (log(as.numeric
   (integrate(f0_S1_S03_f13,
                             lower = 0, upper = t1, w = t1,  a = a)$value))))),
    0)}

## Type 21
type_21 = function(a){-sum(mapply(function(t1, t2)(log(as.numeric(
  integrate(f0_S1_S03_f13, lower = t1, upper = t2, w = t2, a = a)[1]))), t_
   i21[,1], t_i21[,2]))}
```

```
395
396   ## Sum of all
397   sum_all = function(a){
398     type_1(a) + type_2(a) + type_3(a) + type_4(a) +
399       type_5(a) + type_6(a) + type_7(a) + type_8(a) +
400       type_9(a) + type_10(a) + type_11(a) + type_12(a) +
401       type_13(a) + type_14(a) + type_15(a) + type_16(a) +
402       type_17(a) + type_18(a) + type_19(a) + type_20(a) + type_21(a)
403   }
404   ## Constraint above 0 for all parameters
405   theta_hat_new = optim(c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1), sum_all, method = "L-
        BFGS-B",
406                         lower = c(eps, eps, eps, eps, eps, eps, eps, eps, eps,
        eps), hessian = TRUE)
407   theta_hat_hessian_solved = solve(theta_hat_new$hessian)
408 }
```

D.2 Application of CAV for Gamma Process Alternative 1 without Covariates

In this example, we have present the code we have written for the analysis in Section 6.3.1.1. We start by loading the CAV-data and excluding the individuals which transitions the wrong way. Then we divide the individuals into the correct types and store the relevant timepoints. Again, we create functions for the likelihood contributions for the different types and collect them to a final function which is the complete log-likelihood. Then we use *optim* for optimization.

```
1  library(data.table)
2  library("msm")
3  library(Rlab)
4  require(plyr)
5  library(numDeriv)
6
7  eps = 10^-5
8
9  cav = cav[!is.na(cav$pdiag),]
10 cav
11 ## 1: no CAV, 2: mild/moderate CAV, 3: severe CAV,
12 ## 4: recorded at the date of death
13
14 cav = as.data.table(cav)
15
16 cav_check = list()
17 cav_check2 = list()
18 for (i in 2:nrow(cav))
19 {
20   if (cav[i, state] < cav[i-1,state] && cav[i,PTNUM] == cav[i-1, PTNUM])
21   {
22     cav_check2 = rbind(cav[i-1,], cav_check2)
23     cav_check = rbind(cav[i,], cav_check)
24   }
25 }
26 cav_check
27 cav_check2
28
29 ## Remove observations from people going wrong way
30 cav = cav[!cav_check, on=.(PTNUM)]
31 ## The parameters can not be 0 in a Gamma distribution
32 cav$years = ifelse(cav$years == 0, 0.0000001, cav$years)
```

```
33
34 ## Time points
35 t_i1 = c()
36 t_i2 = matrix( , nrow = 0, ncol = 3)
37 t_i3 = matrix( , nrow = 0, ncol = 5)
38 t_i4 = matrix( , nrow = 0, ncol = 5)
39 t_i6 = matrix( , nrow = 0, ncol = 3)
40 t_i7 = matrix( , nrow = 0, ncol = 3)
41 t_i16 = matrix(, nrow = 0, ncol = 2)
42 t_i18 = matrix(, nrow = 0, ncol = 4)
43
44 nn = length(unique(cav[, PTNUM]))
45
46 unique_PTNUM = unique(cav[,PTNUM])
47
48 ## Placing the individuals into different types
49 for (i in 1:nn){
50   cav_PTNUM = cav[PTNUM == unique_PTNUM[i], ]
51   if(all(cav_PTNUM$state == 1)){
52     t_i1 = c(t_i1, cav_PTNUM$years[nrow(cav_PTNUM)])
53   } else if(any(cav_PTNUM$state == 1) & any(cav_PTNUM$state == 2) & !any(cav_
       PTNUM$state == 3) &
54             !any(cav_PTNUM$state == 4)){
55     check_12 = max(which(cav_PTNUM$state == 1))
56     check_22 = min(which(cav_PTNUM$state == 2))
57     t_i2 = rbind(t_i2, c(cav_PTNUM$years[check_12], cav_PTNUM$years[check_22],
       cav_PTNUM$years[nrow(cav_PTNUM)]))
58   } else if(any(cav_PTNUM$state == 1) & any(cav_PTNUM$state == 2) & any(cav_
       PTNUM$state == 3) & !any(cav_PTNUM$state == 4)){
59     check_13 = max(which(cav_PTNUM$state == 1))
60     check_23 = min(which(cav_PTNUM$state == 2))
61     check_33 = max(which(cav_PTNUM$state == 2))
62     check_43 = min(which(cav_PTNUM$state == 3))
63     t_i3 = rbind(t_i3, c(cav_PTNUM$years[check_13], cav_PTNUM$years[check_23],
         cav_PTNUM$years[check_33], cav_PTNUM$years[check_43],
64                          cav_PTNUM$years[nrow(cav_PTNUM)]))
65   } else if (any(cav_PTNUM$state == 1) & any(cav_PTNUM$state == 2) & any(cav_
       PTNUM$state == 3) & any(cav_PTNUM$state == 4)){
66     check_14 = max(which(cav_PTNUM$state == 1))
67     check_24 = min(which(cav_PTNUM$state == 2))
68     check_34 = max(which(cav_PTNUM$state == 2))
69     check_44 = min(which(cav_PTNUM$state == 3))
70     check_54 = which(cav_PTNUM$state == 4)
71     t_i4 = rbind(t_i4, c(cav_PTNUM$years[check_14], cav_PTNUM$years[check_24],
         cav_PTNUM$years[check_34], cav_PTNUM$years[check_44],
72                          cav_PTNUM$years[check_54]))
73   }  else if(any(cav_PTNUM$state == 1) & any(cav_PTNUM$state == 3) & !any(cav_
       PTNUM$state == 2) & !any(cav_PTNUM$state == 4)){
74     check_16 = max(which(cav_PTNUM$state == 1))
75     check_26 = min(which(cav_PTNUM$state == 3))
76     t_i6 = rbind(t_i6, c(cav_PTNUM$years[check_16], cav_PTNUM$years[check_26],
         cav_PTNUM$years[nrow(cav_PTNUM)]))
77   } else if(any(cav_PTNUM$state == 1) & any(cav_PTNUM$state == 3) & !any(cav_
       PTNUM$state == 2) & any(cav_PTNUM$state == 4)){
78     check_17 = max(which(cav_PTNUM$state == 1))
79     check_27 = min(which(cav_PTNUM$state == 3))
80     check_37 = which(cav_PTNUM$state == 4)
81     t_i7 = rbind(t_i7, c(cav_PTNUM$years[check_17], cav_PTNUM$years[check_27],
         cav_PTNUM$years[check_37]))
82   } else if(any(cav_PTNUM$state == 1) & !any(cav_PTNUM$state == 2) & !any(cav_
       PTNUM$state == 3) &
83             any(cav_PTNUM$state == 4)){
```

```
    check_116 = max(which(cav_PTNUM$state == 1))
    check_216 = which(cav_PTNUM$state == 4)
    t_i16 = rbind(t_i16, c(cav_PTNUM$years[check_116], cav_PTNUM$years[check_216]))
  } else if(any(cav_PTNUM$state == 1) & any(cav_PTNUM$state == 2) & !any(cav_PTNUM$state == 3) &
            any(cav_PTNUM$state == 4)){
    check_118 = max(which(cav_PTNUM$state == 1))
    check_218 = min(which(cav_PTNUM$state == 2))
    check_318 = max(which(cav_PTNUM$state == 2))
    check_418 = which(cav_PTNUM$state == 4)
    t_i18 = rbind(t_i18, c(cav_PTNUM$years[check_118], cav_PTNUM$years[check_218], cav_PTNUM$years[check_318], cav_PTNUM$years[check_418]))
  }
}

print(length(t_i1) + nrow(t_i2) + nrow(t_i3) + nrow(t_i4) + nrow(t_i6) + nrow(t_i7) + nrow(t_i16) + nrow(t_i18))
nn


## Initial survival functions
S_0 = function(a, t){pgamma(exp(a[1]), shape = t*exp(a[2]), rate = 1)}
S_1 = function(a, t){pgamma(exp(a[3]), shape = t*exp(a[4]), rate = 1)}
S_2 = function(a, t){pgamma(exp(a[5]), shape = t*exp(a[6]), rate = 1)}
S_03 = function(a, t){pgamma(exp(a[7]), shape = t*exp(a[8]), rate = 1)}
S_13 = function(a, t){pgamma(exp(a[9]), shape = t*exp(a[10]), rate = 1)}

## Initial density functions - derivative of -survival function
f_0 = function(a, t){(S_0(a, t) - S_0(a, t + eps))/(eps)}
f_1 = function(a, t){(S_1(a, t) - S_1(a, t + eps))/(eps)}
f_2 = function(a, t){(S_2(a, t) - S_2(a, t + eps))/(eps)}
f_03 = function(a, t){(S_03(a, t) - S_03(a, t + eps))/(eps)}
f_13 = function(a, t){(S_13(a, t) - S_13(a, t + eps))/(eps)}

## Other functions
f0_S1_S03_S13 = function(a, w, t){f_0(a, t)*S_03(a, t)*S_1(a, w - t)*S_13(a, w - t)}
f0_f1_S2_S03_S13 = function(a, w, t, u){f_0(a, t)*f_1(a, u)*S_2(a, w - t - u)*S_03(a, t)*S_13(a, u)}
f0_f1_f2_S03_S13 = function(a, w, t, u){f_0(a, t)*f_1(a, u)*f_2(a, w - t - u)*S_03(a, t)*S_13(a, u)}
f0_S1_S03_f13 = function(a, w, t){f_0(a, t)*S_03(a, t)*S_1(a, w - t)*f_13(a, w - t)}

## Type 1
type_1 = function(a){-sum(sapply(t_i1, function(t) (log(S_0(a, t)) + log(S_03(a, t)))))}

## Type 2
type_2 = function(a){-sum(mapply(function(t1, t2, t3)(log(as.numeric(
  integrate(f0_S1_S03_S13, lower = t1, upper = t2, w = t3, a = a)$value))), t_i2[,1], t_i2[,2], t_i2[,3]))}

## Type 3
type_3 = function(a){-sum(mapply(function(t1, t2, t3, t4, t5){
  log(as.numeric(integrate(function(t){sapply(t, function(z){integrate(f0_f1_S2_S03_S13,
                                                                       lower = t3-z, upper = t4-z, w = t5, t = z, a = a)$value})},
                           lower = t1, upper = t2)[1]))}, t_i3[,1], t_i3[,2], t_i3[,3], t_i3[,4],
```

```
  t_i3[,5]))}


## Type 4
type_4 = function(a){-sum(mapply(function(t1, t2, t3, t4, t5){
  log(as.numeric(integrate(function(t){
    sapply(t, function(z){integrate(f0_f1_f2_S03_S13,
                                    lower = t3-z, upper = t4-z, w = t5, t = z,
     a = a)$value})},
    lower = t1, upper = t2)[1]))}, t_i4[,1], t_i4[,2], t_i4[,3], t_i4[,4], t_
    i4[,5]))}

## Type 6
type_6 = function(a){-sum(mapply(function(t1, t2, t3){
  log(as.numeric(integrate(function(t){sapply(t, function(z){integrate(f0_f1_
    S2_S03_S13,
                                                                lower =
     0, upper = t2 - z, w = t3, t = z, a = a)$value})},
                           lower = t1, upper = t2)[1]))}, t_i6[,1], t_i6[,2],
    t_i6[,3]))}

## Type 7
type_7 = function(a){-sum(mapply(function(t1, t2, t3){
  log(as.numeric(integrate(function(t){sapply(t, function(z){integrate(f0_f1_
    f2_S03_S13,
                                                                lower =
     0, upper = t2 - z, w = t3, t = z, a = a)$value})},
                           lower = t1, upper = t2)[1]))}, t_i7[,1], t_i7[,2],
    t_i7[,3]))}


## Combination of type 5, 16 and 21
type_16 = function(a){-sum(mapply(function(t1, t2){
  log(as.numeric(integrate(function(t){sapply(t, function(z){integrate(f0_f1_
    f2_S03_S13,
                                                                lower =
     0, upper = t2 - z, w = t2, t = z, a = a)$value})},
                           lower = t1, upper = t2)[1]) + S_0(a, t2)*f_03(a,t2)
     + as.numeric(
                             integrate(f0_S1_S03_f13, lower = t1, upper = t2,
    w = t2, a = a)[1]))}, t_i16[,1], t_i16[,2]))}


## Combination of type 15 and 18
type_18 = function(a){-sum(mapply(function(t1, t2, t3, t4)(log(as.numeric(
  integrate(f0_S1_S03_f13, lower = t1, upper = t2, w = t4, a = a)[1]) +
    as.numeric(integrate(function(t){sapply(t, function(z){integrate(f0_f1_f2_
    S03_S13,
                                                                  lower =
    t3-z, upper = t4 - z, w = t4, t = z, a = a)$value})},
                         lower = t1, upper = t2)[1]))), t_i18[,1], t_i18[,2],
    t_i18[,3], t_i18[,4]))}


## Sum of all
sum_all = function(a){
  type_1(a) + type_2(a) + type_3(a) + type_4(a) +
    type_6(a) + type_7(a)  + type_16(a) + type_18(a)
}
theta_hat_new = optim(c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1), sum_all, method = "L-
    BFGS-B", hessian = FALSE)
```

Bibliography

Aalen, O. O., Borgan, Ø., & Gjessing, H. K. (2008). *Survival and event history analysis: A process point of view*. Springer Science & Business Media. https://doi.org/10.1007/978-0-387-68560-1

Aalen, O. O., & Gjessing, H. K. (2001). Understanding the shape of the hazard rate: a process point of view (With comments and a rejoinder by the authors). *Statistical Science*, *16*(1), 1–22. https://doi.org/10.1214/ss/998929473

Aalen, O. O., & Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*, *5*(3), 141–150. http://www.jstor.org/stable/4615704

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, *60*(2), 255–265. https://doi.org/10.1093/biomet/60.2.255

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Andersen, P. K., & Keiding, N. (2002). Multi-state models for event history analysis. *Statistical Methods in Medical Research*, *11*(2), 91–115. https://doi.org/10.1191/0962280202SM276ra

Armitage, P., & Doll, R. (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. *British Journal of Cancer*, *8*(1), 1–12. https://doi.org/10.1038/bjc.1954.1

Berman, S. M. (1963). Note on extreme values, competing risks and semi-Markov processes. *The Annals of Mathematical Statistics*, *34*(3), 1104–1106. http://www.jstor.org/stable/2238495

Caroni, C. (2017). *First hitting time regression models: Lifetime data analysis based on underlying stochastic processes*. ISTE Ltd/John Wiley & Sons Inc.

Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Brooks/Cole Cengage Learning.

Claeskens, G., & Hjort, N. L. (2003). The focused information criterion. *Journal of the American Statistical Association*, *98*(464), 900–916. https://doi.org/10.1198/016214503000000819

Claeskens, G., & Hjort, N. L. (2008). *Model selection and model averaging*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790485

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003). Survival analysis part I: Basic concepts and first analyses. *British Journal of Cancer*, *89*(2), 232–238. https://doi.org/10.1038/sj.bjc.6601118

Commenges, D. (2002). Inference for multi-state models from interval-censored data. *Statistical Methods in Medical Research*, *11*(2), 167–182. https://doi.org/10.1191/0962280202sm279ra

Cox, D. R. (1959). The analysis of exponentially distributed life-times with two types of failure. *Journal of the Royal Statistical Society: Series B (Methodological)*, *21*(2), 411–421. https://doi.org/10.1111/j.2517-6161.1959.tb00349.x

de Uña-Álvarez, J., & Meira-Machado, L. (2015). Nonparametric estimation of transition probabilities in the non-Markov illness-death model: A comparative study. *Biometrics*, *71*(2), 364–375. https://doi.org/10.1111/biom.12288

Fix, E., & Neyman, J. (1951). A simple stochastic model of recovery, relapse, death and loss of patients. *Human Biology*, *23*(3), 205–241. http://www.jstor.org/stable/41448000

Foucher, Y., Giral, M., Soulillou, J.-P., & Daures, J.-P. (2007). A semi-markov model for multistate and interval-censored data with multiple terminal events. application in renal transplantation. *Statistics in Medicine*, *26*(30), 5381–5393. https://doi.org/10.1002/sim.3100

Frank, S. A. (2004). Commentary: Mathematical models of cancer progression and epidemiology in the age of high throughput genomics. *International Journal of Epidemiology*, *33*(6), 1179–1181. https://doi.org/10.1093/ije/dyh222

Frydman, H. (1995). Nonparametric estimation of a Markov 'illness-death' process from interval-censored observations, with application to diabetes survival data. *Biometrika*, *82*(4), 773–789. https://doi.org/10.1093/biomet/82.4.773

Gentleman, R. C., & Geyer, C. J. (1994). Maximum likelihood for interval censored data: Consistency and computation. *Biometrika*, *81*(3), 618–623. https://doi.org/10.1093/biomet/81.3.618

Gentleman, R. C., Lawless, J. F., Lindsey, J. C., & Yan, P. (1994). Multi-state Markov models for analysing incomplete disease history data with illustrations for hiv disease. *Statistics in Medicine*, *13*(8), 805–821. https://doi.org/10.1002/sim.4780130803

Gil, J., Preux, P.-M., Alioum, A., Ketzoian, C., Desport, J.-C., Druet-Cabanac, M., & Couratier, P. (2007). Disease progression and survival in ALS: First multi-state model approach. *Amyotrophic Lateral Sclerosis*, *8*(4), 224–229. https://doi.org/10.1080/17482960701278562

Hougaard, P. (1999). Multi-state models: A review. *Lifetime Data Analysis*, *5*(3), 239–264. https://doi.org/10.1023/A:1009672031531

Hougaard, P. (2000). *Analysis of multivariate survival data*. Springer-Verlag.

Huang, J., & Wellner, J. A. (1997). Interval censored survival data: A review of recent progress. In D. Y. Lin & T. R. Fleming (Eds.), *Lecture notes in statistics: Vol. 123. Proceedings of the first seattle symposium in biostatistics* (pp. 123–169). Springer-Verlag.

Jack Jr., C. R., Therneau, T. M., Wiste, H. J., Weigand, S. D., Knopman, D. S., Lowe, V. J., Mielke, M. M., Vemuri, P., Roberts, R. O., Machulda, M. M., Senjem, M. L., Gunter, J. L., Rocca, W. A., & Petersen, R. C. (2016).

Transition rates between amyloid and neurodegeneration biomarker states and to dementia: A population-based, longitudinal cohort study. *The Lancet Neurology*, *15*(1), 56–64. https://doi.org/10.1016/S1474-4422(15)00323-3

Jackson, C. (2011). Multi-state models for panel data: The msm package for R. *Journal of Statistical Software, Articles*, *38*(8), 1–28. https://doi.org/10.18637/jss.v038.i08

Jackson, C. (2016). Flexsurv: A platform for parametric survival modeling in R. *Journal of Statistical Software*, *70*(8), 1–33. https://doi.org/10.18637/jss.v070.i08

Jackson, C. (2019). *Package msm*. https://cran.r-project.org/web/packages/msm/vignettes/msm-manual.pdf

Kalbfleisch, J. D., & Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*, *80*(392), 863–871. http://www.jstor.org/stable/2288545

Kay, R. (1986). A Markov model for analysing cancer markers and disease states in survival studies. *Biometrics*, *42*(4), 855–865. https://doi.org/10.2307/2530699

Klotz, J. H., & Sharples, L. D. (1994). Estimation for a Markov heart transplant model. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *43*(3), 431–438. https://doi.org/10.2307/2348579

Lawless, J. F. (2003). *Statistical models and methods for lifetime data* (2nd ed.). Wiley-Interscience.

Le-Rademacher, J. G., Peterson, R. A., Therneau, T. M., Sanford, B. L., Stone, R. M., & Mandrekar, S. J. (2018). Application of multi-state models in cancer clinical trials. *Clinical Trials*, *15*(5), 489–498. https://doi.org/10.1177/1740774518789098

Leung, K.-M., & Elashoff, R. M. (1996). A three-state disease model with interval-censored data: Estimation and applications to AIDS and cancer. *Lifetime Data Analysis*, *2*(2), 175–194. https://doi.org/10.1007/BF00128574

Lindsey, J. C., & Ryan, L. M. (1998). Methods for interval-censored data. *Statistics in Medicine*, *17*(2), 219–238. https://doi.org/10.1002/(SICI)1097-0258(19980130)17:2<219::AID-SIM735>3.0.CO;2-O

Meira-Machado, L., de Uña-Álvarez, J., & Cadarso-Suárez, C. (2006). Nonparametric estimation of transition probabilities in a non-Markov illness–death model. *Lifetime Data Analysis*, *12*(3), 325–344. https://doi.org/10.1007/s10985-006-9009-x

Meira-Machado, L., de Uña-Álvarez, J., Cadarso-Suárez, C., & Andersen, P. K. (2009). Multi-state models for the analysis of time-to-event data. *Statistical Methods in Medical Research*, *18*(2), 195–222. https://doi.org/10.1177/0962280208092301

Nelson, W. (1970). Hazard plotting methods for analysis of life data with different failure modes. *Journal of Quality Technology*, *2*(3), 126–149. https://doi.org/10.1080/00224065.1970.11980427

Peto, R. (1973). Experimental survival curves for interval-censored data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *22*(1), 86–91. https://doi.org/10.2307/2346307

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T., & Breslow, N. E. (1978). The analysis of failure times in the presence

of competing risks. *Biometrics*, *34*(4), 541–554. https://doi.org/10.2307/2530374

Putter, H., Fiocco, M., & Geskus, R. B. (2007). Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine*, *26*(11), 2389–2430. https://doi.org/10.1002/sim.2712

Putter, H., van der Hage, J., de Bock, G. H., Elgalta, R., & van de Velde, C. J. H. (2006). Estimation and prediction in a multi-state model for breast cancer. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *48*(3), 366–380. https://doi.org/10.1002/bimj.200510218

Qiu, Q., & Cui, L. (2019). Gamma process based optimal mission abort policy. *Reliability Engineering & System Safety*, *190*, 106496. https://doi.org/10.1016/j.ress.2019.106496

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rabiner, L., & Juang, B. (1986). An introduction to hidden Markov models. *IEEE ASSP Magazine*, *3*(1), 4–16. https://doi.org/10.1109/MASSP.1986.1165342

Sildnes, B., & Lindqvist, B. H. (2018). Modeling of semi-competing risks by means of first passage times of a stochastic process. *Lifetime Data Analysis*, *24*(1), 153–175. https://doi.org/10.1007/s10985-017-9399-y

Silverman, B. W., Jones, M. C., Wilson, J. D., & Nychka, D. W. (1990). A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography. *Journal of the Royal Statistical Society: Series B (Methodological)*, *52*(2), 271–303. https://doi.org/10.1111/j.2517-6161.1990.tb01788.x

Sverdrup, E. (1965). Estimates and test procedures in connection with stochastic models for deaths, recoveries and transfers between different states of health. *Scandinavian Actuarial Journal*, *1965*(3-4), 184–211. https://doi.org/10.1080/03461238.1965.10405687

Titman, A. C., & Sharples, L. D. (2010). Model diagnostics for multi-state models. *Statistical Methods in Medical Research*, *19*(6), 621–651. https://doi.org/10.1177/0962280209105541

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *38*(3), 290–295. https://doi.org/10.1111/j.2517-6161.1976.tb01597.x

Van Den Hout, A. (2017). *Multi-state survival models for interval-censored data*. Chapman & Hall/CRC. https://doi.org/10.1201/9781315374321

van Noortwijk, J. M. (2009). A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety*, *94*(1), 2–21. https://doi.org/10.1016/j.ress.2007.03.019

Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*(2), 228–243. https://doi.org/10.1037/a0027127

Williams, J. P., Storlie, C. B., Therneau, T. M., Jack Jr, C. R., & Hannig, J. (2020). A Bayesian approach to multistate hidden Markov models: Application to dementia progression. *Journal of the American Statistical Association*, *115*(529), 16–31. https://doi.org/10.1080/01621459.2019.1594831