Estimating transition probabilities for the illness-death model
The Aalen-Johansen estimator under violation of the Markov assumption
Torunn Heggland
Master’s Thesis for the degree Modelling and Data Analysis (MOD5960), November 2015
Acknowledgements

When I now complete my master’s degree in statistics, it’s been 18 years since I first got the idea: “I want to become a statistician!” I went in the 4th grade, and had an interest in numbers, tables and it was fun to look at name statistics from Statistics Norway. Through my school years, I have had good and inspiring teachers in mathematics. I appreciate their efforts!

My years at the university have contained a lot of work, challenges, eye-openers, good friendships and x number of dinners at Frederikke. The last years I have had the pleasure being part of study hall B800 ”Parameterrommet”. You guys are cool, and I thank you for a good working environment, conversations and gatherings.

A great thank you to my great supervisor Ørnulf Borgan for all talks and constructive feedback. I also want to thank the Faculty of Mathematics and Natural Sciences for letting programming be such an integrated part of the studies right from the start. I went from having no clue, to really love it.

My parents, Lillian and Olav, thank you for all your support through my studies. You are amazing! And the rest of my nearest family, Marit, Håvard and Sunniva, Aron Olai, Iben Emanuel: You bring such a joy to my life.

Blindern, November 2015
Torunn Heggland
Abstract

The Aalen-Johansen estimator for calculation of transition probabilities in a multi-state model, builds on the assumption that the data are Markovian. For real data, the Markov property may not be fulfilled, and it is then of interest to study how the estimator behaves.

In this thesis, the study is restricted to the three-state illness-death model, and in addition to the Aalen-Johansen estimator, two other methods for calculation of transition probabilities are considered. The first one is a method building on the assumption that the data are semi-Markovian, while the other is a general method not building on any assumptions.

Simulated data with known properties are used to study the performance of the methods for different situations. As known in advance, we see that the Aalen-Johansen estimator works well for state occupation probabilities, although the Markov assumption is not fulfilled. Further, it turns out that the semi-Markov method performs well only when the data are semi-Markovian. The Aalen-Johansen estimator is more robust to small deviations from the Markov assumption, than the semi-Markov method is to small deviations from the semi-Markov assumption. For the general method, it is seen that approximately unbiased estimates are produced in all the situations considered, but that the variance here is larger than for the two other methods. For state occupation probabilities, the simulations show that the available variance estimator for the Aalen-Johansen estimator works well also in non-Markovian cases.

The Brier score is investigated as a mean to find the best method to use on real data. The Brier score calculated for one method, is compared to the Brier score for another one. Based on this comparison, one gets an idea of the over- or underestimation of the methods.

KEY WORDS: Aalen-Johansen estimator; Brier score; Illness-death model; Kaplan-Meier estimator; Markov condition; Multi-state models; Survival analysis; Transition probabilities.
# Contents

Acknowledgements

Abstract

1 Introduction

2 Survival and event history analysis
   2.1 Data example
   2.2 Survival models
      2.2.1 Estimation in the survival model
   2.3 Multi-state models
   2.4 Markov models
      2.4.1 Estimation in Markov models
   2.5 The progressive illness-death model

3 Alternative estimators
   3.1 Semi-Markov illness-death model
      3.1.1 Estimation in the semi-Markov model
   3.2 A non-Markov/general illness-death model
      3.2.1 Estimation in the general model

4 Comparison of the methods on real data
   4.1 The Brier score for our setting
   4.2 Another data example

5 Comparison of the methods
   5.1 Setup for data generation and comparison
   5.2 Comparison of the methods on some data sets
      5.2.1 Setting 1: Markov data with constant hazards
      5.2.2 Setting 2: Markov data with nonconstant hazards
      5.2.3 Setting 3: Semi-Markov data
      5.2.4 Setting 4: Frailty non-Markov data
      5.2.5 Setting 5: Non-Markov Meira data
   5.3 Summary of the simulation results
6 Concluding remarks 85
  6.1 Discussion and conclusion ............................................. 85
  6.2 Further work and challenges .......................................... 86

References 89

A Some calculations 91
  A.1 The expression for $P_{12}(s,t)$ in Section 2.5 ................. 91
  A.2 Estimated variance of $\hat{P}_{12}(s,t)$ in Section 2.5 .......... 92
  A.3 The expectation of the Brier score in Section 4.1 .............. 92
  A.4 A check of the simulation procedure in Section 5.1 ............ 93
  A.5 Setting 4: Frailty non-Markov data in Section 5.2 ............. 95
  A.6 Setting 5: Non-Markov data in Section 5.2 ...................... 96

B R-scripts 99
  B.1 Estimation of transition probabilities and calculation of Brier scores 99
  B.2 Simulation study: Markovian data (setting 1) ..................... 105

C Plots 111
Chapter 1

Introduction

In medicine we are interested in the prognosis for a patient. For a cancer patient who undergoes an operation of some kind, it is of interest to know the probability to have a relapse of the cancer within a year, within two years, or some other period. If no such event has happened within, for instance, a two year period after the operation, the probability to have a relapse during the following year may have changed, compared to the probability right after the operation. This setting may be extended by including the possibility of death. The patient may die after a relapse; right afterwards or years later, or he/she could die without having had a relapse first.

The situation above may be described by a model with three states. The states will here be ‘operated’, ‘relapsed’ and ‘dead’. When an event happens to an individual, a transition from one state to another occurs. The transition probabilities in cases like this are unknown, but we could estimate them based on what we observe. For this purpose we need collected data. In Norway, there are 17 central health registers (Sekretariat for Nasjonalt helseregisterprosjekter, 2015); the Cancer Registry of Norway and the Norwegian Patient Register are probably the most well-known, and data from these registers are used for research. In this thesis we will consider two data sets from the European Registry for Blood and Marrow Transplantation.

Event history analysis gives a methodology to analyze data in settings like the one above. This methodology extends survival analysis, which is considering situations where only one event could happen to an individual. Events happen over time, and could well happen for individuals after a study is completed. Observations for such individuals are then said to be right-censored. The survival and event history analysis may handle data where censoring is present.

There have been developed various methods to calculate estimates of the transition probabilities. The methods build on models, where the model is trying to describe reality. We will mainly restrict ourselves to irreversible three-state models; such a model is called a progressive illness-death model. The Aalen-Johansen estimator is the method that traditionally has been most used. This method builds on a Markov model, but in real situations we do not know if the Markov property is fulfilled, and the method is then not guaranteed to work well. It is of interest to study how wrong the estimates may turn out to be.
under violation of the Markov assumption. We will also consider two alternative methods; one building on a semi-Markov model, another on a model without any assumptions. It is then of interest to check how these methods behave compared to the Aalen-Johansen estimator in various settings.

The methods for multi-state models are not so widely used yet. A main reason for this could be the lack of software. However, the last years some packages have been developed for the statistical software \textit{R} (\cite{R_2011}). Two examples are the mstate package (\cite{deWrede_2011}) and the newly developed TPmsm package (\cite{Araujo_2014}). These packages will be used in the data examples and for the simulations.

The outline of the thesis is as follows. In Chapter 2 we will first go into the theory of survival analysis. Key elements here are the Nelson-Aalen and the Kaplan-Meier estimators, and these will be studied in detail. These estimators are also important building blocks for the event history analysis. Further we introduce multi-state models, and study Markov models and estimation in such models. We end the chapter by focusing on the three-state illness-death model under the Markov condition. In Chapter 3 we will consider the alternative estimators. We look at the illness-death model in the two cases, and how estimation may be performed. For our two real data sets we do not know if the Markov property, or the semi-Markov property, is fulfilled. When we have various estimation procedures available, as is the case here, it is natural to choose the procedure that fits the data best. In Chapter 4 we will present and use the Brier score for this purpose. In Chapter 5 we study the behavior of our three methods through simulations. Data with chosen properties; Markovian data, semi-Markovian data and data where none of these assumptions are fulfilled, are generated, and the three methods are applied to the data. Here we will also study the behavior of the variance estimator for the Aalen-Johansen estimator. In Chapter 6 we will give concluding remarks, and also mention possible extensions to what has been done.
Chapter 2

Survival and event history analysis

Lifetimes differ from lots of other measurements. Looking at a time period and a group of people, there is a chance that some of the individuals will not die in that period. All we know about them, is that they will live longer than the given endpoint of the study. If this is the case we will have problems to compute even such a simple quantity as the mean lifetime, because of the incomplete data. Hence we understand that classical statistics is not satisfactory for this kind of data.

A model for lifetimes is called a survival model, and the statistical methodology we need to study lifetimes is called survival analysis. Often we are interested in more complex situations than lifetimes. There are for instance different causes of death; cancer, heart disease, etc, and taking this into account makes the modeling more complex. There could also be more than one event happening to each individual as the time goes by. For these purposes we use multi-state models, and the methodology for such models is called event history analysis.

At any time in the time period we are considering, each individual is said to be in a state. The individuals move among different states, and we would be interested in the probabilities of transitions between states, and the probabilities of being in the different states at specified times, called state occupation probabilities.

In this chapter we will look at the theory of survival and event history analysis. In Section 2.4 we will meet the already mentioned Markov assumption. This assumption says that the history of an individual is irrelevant for the probabilities to make future transitions. Much of the theory in this chapter is taken from Aalen et al. (2008, Chap 3 and Appendix A.1, A.2).

2.1 Data example

Example 1.1

In this, and the two next chapters, we will for illustration consider a data set from the European Registry for Blood and Marrow Transplantation, consisting of 1977 patients transplanted for chronic myelogenous leukemia (CML). This
CHAPTER 2. SURVIVAL AND EVENT HISTORY ANALYSIS

Figure 2.1: Number of individuals in each state at different times after transplant for the CML data. The black curve shows the number who have not had an event, the red curve the number of patients who are in the relapsed state, while the green curve is the number of patients who are dead.

data set is available from the mstate package in R, under the name ebmt, see Appendix B.1. CML is a cancer where the bone marrow makes too many white blood cells. The condition may be attempted cured by a bone marrow transplant, where the damaged bone marrow is replaced with healthy bone marrow stem cells from a donor (U.S. National Library of Medicine, 2015).

Each individual in the data set is followed from the transplant onwards, and it is recorded if and when the patient had a relapse of CML, or died. In Section 2.2 we will think of the time when one of these events happened, as a survival time. The two states relapsed and dead, are then merged to one. Some of the individuals did not have an event during the observation period. They are said to be right-censored. In Section 2.5 we will analyze the data with a multi-state model. We will then look at relapse and death as two separate states, and hence consider a three-state model. Again censoring will be present.

To get an overview of the data, Figure 2.1 shows the number of individuals in each state at different times. The red curve shows how many who are in the relapsed state at each time point. Individuals enter this state, but they may also leave it. The green curve shows how many who are dead, while the black one shows the number that have not had any events after the transplant. A year after the transplant, 185 individuals are in the relapsed state, 698 are dead, while
1012 individuals have had no events after the transplant. This means that the observation of 82 individuals have been censored during the first year after the transplant. Five years after the transplant, 77 individuals are in the relapsed state, 863 are dead, while 198 individuals have had no event.

### 2.2 Survival models

Starting out with lifetimes, we let $T \geq 0$ be a random survival time with the well-known survival function

\[ S(t) = P(T > t). \]  

In words, this is the probability that the lifetime is greater than the time $t$, where $t$ is time since an initial timepoint.

The hazard rate $\alpha(t)$ is the instantaneous risk of dying at time $t$

\[ \alpha(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}. \]  

This is the probability of dying shortly, given that the individual has survived up to time $t$. When $\alpha(t)$ exists, that is when $S(t)$ is absolutely continuous, (2.2) gives us the relation

\[ \alpha(t) = \lim_{\Delta t \to 0} \frac{-(S(t + \Delta t) - S(t))}{S(t)\Delta t} = -\frac{dS(t)}{S(t)}. \]

This motivates that the cumulative hazard $A(t)$ may be defined by the Stieltjes integral

\[ A(t) = \int_0^t \alpha(u)du = -\int_0^t \frac{dS(u)}{S(u)}. \]  

for all kind of distributions. (2.3) is a general expression for the cumulative hazard, and it leads to the differential equation

\[ dS(t) = -S(t-)dA(t), \]  

which we will come back to in (2.7). We can also write (2.4) as an integral equation

\[ S(t) = 1 - \int_0^t S(u-)dA(u). \]

To study the survival function (2.1), it is useful to express it as a product of conditional survival functions. For times $u > v$, we define the conditional survival function as

\[ S(u \mid v) = P(T > u \mid T > v) = \frac{S(u)}{S(v)}. \]
We make a partition of the time interval \((0, t]\) into \(K\) subintervals \((0, t_1], (t_1, t_2], \ldots, (t_{K-1}, t]\), and can then write
\[
S(t) = P(T > t_1)P(T > t_2 | T > t_1) \cdots P(T > t | T > t_{K-1})
\]
\[
= \prod_{k=1}^{K} S(t_k | t_{k-1}),
\]
with \(t_0 = 0\) and \(t_K = t\). Further, from (2.4) we have the approximation
\[
S(t_k) - S(t_{k-1}) \approx -S(t_{k-1})(A(t_k) - A(t_{k-1}))
\]
or, by dividing both sides by \(S(t_{k-1})\),
\[
S(t_k | t_{k-1}) \approx 1 - (A(t_k) - A(t_{k-1})).
\]
Now, by using (2.8) in (2.6), we get the approximation
\[
S(t) \approx \prod_{k=1}^{K} (1 - (A(t_k) - A(t_{k-1}))).
\]
Letting \(K\) increase, while the lengths of the intervals go to zero uniformly, the product on the right-hand side in (2.9) will approach the product-integral
\[
S(t) = \prod_{u \leq t} (1 - dA(u)).
\]
A product-integral has the same relation to a product, as the well-known integral has to a sum. (2.10) shows the general relation between the survival function and the cumulative hazard.

When the cumulative hazard \(A(t)\) is absolutely continuous, we have that \(dA(u) = \alpha(u)du\). Using the approximation \(\exp(-\alpha(u)du) \approx 1 - \alpha(u)du\), valid for small \(du\), we end up with
\[
S(t) = \prod_{u \leq t} (1 - dA(u)) = \prod_{u \leq t} (1 - \alpha(u)du)
\]
\[
= \exp(-\int_{u \leq t} \alpha(u)du) = \exp(-A(t)).
\]
For the discrete case \(S(t) = \prod_{u \leq t} (1 - \alpha_u)\), where \(\alpha_u = P(T = u | T \geq u)\) is the discrete hazard.

### 2.2.1 Estimation in the survival model

To estimate the hazard rate \(\alpha(t)\) and the survival function \(S(t)\) we need to consider a group of \(n\) individuals. Each of them will have the possibly censored survival time \(\tilde{T}_i\), and the associated indicator function \(D_i = I(\tilde{T}_i = T_i)\) for this time to be a survival time. Instead of estimating the hazard rate directly, which is hard, we will estimate the cumulative hazard \(A(t)\). This turns out to be easy.
2.2. SURVIVAL MODELS

When we have an estimator of $A(t)$, we see from (2.10) that it will be easy to estimate $S(t)$.

A common approach to survival estimation is to consider counting processes. A counting process is counting the number of events observed during a time period. In survival analysis we are looking at the occurrences of events, and it is hence natural to count them as they occur and use this information for estimation purposes.

The Nelson-Aalen estimator

For each of the individuals, we define the counting process

$$N_i(t) = I\{\tilde{T}_i \leq t, D_i = 1\},$$

which is counting one when an uncensored event happens. The intensity process of $N_i(t)$ takes the form

$$\lambda_i(t) = \alpha(t)Y_i(t),$$

where $Y_i(t) = I(\tilde{T}_i \geq t)$ is an indicator for being at risk 'just before' time $t$.

For all the individuals, the aggregated counting process is now

$$N(t) = \sum_{i=1}^{n} N_i(t).$$

$N(t)$ tells us for how many of the individuals an uncensored event has happened up to time $t$. Likewise, we have that $Y(t)$, the total number of individuals at risk at time $t$, is given by $Y(t) = \sum_{i=1}^{n} Y_i(t)$, while the intensity process of $N(t)$ is

$$\lambda(t) = \sum_{i=1}^{n} \lambda_i(t) = \alpha(t)Y(t).$$

The relation between the counting process $N(t)$ and its intensity process $\lambda(t)$ is given by the zero-mean martingale

$$M(t) = N(t) - \int_{0}^{t} \lambda(u)du = N(t) - \int_{0}^{t} \alpha(u)Y(u)du.$$ (2.11)

This equation, (2.11), makes us express the increment $dN(t)$ as

$$dN(t) = \alpha(t)Y(t)dt + dM(t).$$

By multiplying with the indicator function $J(t) = I(Y(t) > 0)$, dividing by $Y(t)$ and then integrating, we get

$$\int_{0}^{t} \frac{J(u)}{Y(u)}dN(u) = \int_{0}^{t} J(u)\alpha(u)du + \int_{0}^{t} \frac{J(u)}{Y(u)}dM(u).$$ (2.12)

The last term in (2.12) is a stochastic integral and hence it has expectation zero. When $P(Y(t) = 0)$ is small we have that

$$\int_{0}^{t} J(u)\alpha(u)du \approx A(t),$$
hence we may estimate the cumulative hazard by the Nelson-Aalen estimator

\[
\hat{A}(t) = \int_0^t \frac{J(u)}{Y(u)} dN(u) = \sum_{\{j: \tilde{T}_j \leq t, D_j = 1\}} \frac{1}{Y(\tilde{T}_j)},
\]

where the second equality follows since the counting process makes jumps only at event times. The estimated hazard will now be the slope of this cumulative function (2.13).

The variance of the Nelson-Aalen estimator

We will also be interested in the variance of the Nelson-Aalen estimator. In that way we will be able to construct confidence intervals for the cumulative hazard. Introducing the notation

\[
A^*(t) = \int_0^t J(u) \alpha(u) du,
\]

(2.12) may be written as

\[
\hat{A}(t) - A^*(t) = \int_0^t \frac{J(u)}{Y(s)} dM(u).
\]

(2.14)

Because \(M\) is a zero-mean martingale, this stochastic integral, (2.14), is a zero-mean martingale as well. Then \(E\{\hat{A}(t) - A^*(t)\} = 0\), which means that \(\hat{A}(t)\) is an unbiased estimator of \(A^*(t)\). We would want an unbiased estimator of \(A(t)\), but this is impossible since \(\alpha(t)\) can not be estimated when \(Y(t) = 0\).

From the theory of martingales we have that

\[
\text{Var}(\hat{A}(t) - A^*(t)) = E\left\{[\hat{A} - A^*](t) \right\},
\]

where \([\hat{A} - A^*]\) means the optional variation process. In words, the variance of (2.14) is the expectation of its optional variation process. Thus

\[
\hat{\sigma}^2(t) = \text{Var}(\hat{A}(t)) = [\hat{A} - A^*](t) = \int_0^t \frac{J(u)}{Y(u)^2} dN(u) = \sum_{\{j: \tilde{T}_j \leq t, D_j = 1\}} \frac{1}{Y(\tilde{T}_j)^2},
\]

which is an unbiased estimator.

The Nelson-Aalen estimator, evaluated at a given time \(t\), is approximately normally distributed in large samples. A standard 100\((1 - \alpha)\)% confidence interval (CI) is then given by

\[
\hat{A}(t) \pm z_{\alpha/2} \hat{\sigma}(t),
\]

(2.15)

where \(z_{\alpha/2}\) is the 100\((1 - \alpha/2)\)th percentile of the standard normal distribution. An alternative, and better interval is achived by using a log transformation resulting in the interval

\[
\hat{A}(t) \exp\{\pm z_{\alpha/2} \hat{\sigma}(t)/\hat{A}(t)\}.
\]

(2.16)
The Kaplan-Meier estimator and its variance

From (2.10) a natural estimator for $S(t)$, called the Kaplan-Meier estimator, is

$$
\hat{S}(t) = \prod_{u \leq t} \{1 - d\hat{A}(u)\} = \prod_{\{j:\tilde{T}_j \leq t, D_j = 1\}} \left\{1 - \frac{1}{Y(\tilde{T}_j)}\right\}, \quad (2.17)
$$

where the last equality follows since $\hat{A}(t)$, (2.13), is a step function with increment $\frac{1}{Y(\tilde{T}_j)}$ when $\tilde{T}_j$ is a survival time. In large samples, $\hat{S}(t)$ is approximately normally distributed, when evaluated at a given time $t$. To estimate the variance of (2.17), and hence be able to construct confidence intervals, we introduce

$$
S^*(t) = \prod_{u \leq t} \{1 - dA^*(u)\},
$$

which is nearly $S(t)$. It can be shown that

$$
\frac{\hat{S}(t) - S^*(t)}{S^*(t)} = -\int_0^t \frac{\hat{S}(u-)}{S^*(u)} d(\hat{A} - A^*)(u). \quad (2.18)
$$

Thus we have the approximation

$$
\frac{\hat{S}(t)}{S(t)} - 1 \approx -\int_0^t d(\hat{A} - A)(u),
$$
or

$$
\hat{S}(t) - S(t) \approx -S(t) \left( \hat{A}(t) - A(t) \right). \quad (2.19)
$$

From (2.19) we get that

$$
\text{Var}(\hat{S}(t)) \approx S(t)^2 \text{Var}(\hat{A}(t)).
$$

The variance of the Kaplan-Meier estimator may now be estimated by

$$
\hat{\tau}^2(t) = \text{Var}(\hat{S}(t)) = \hat{S}(t)^2 \sum_{\{j:\tilde{T}_j \leq t, D_j = 1\}} \frac{1}{Y(\tilde{T}_j)^2}. \quad (2.20)
$$

Another alternative is to estimate the variance by Greenwood’s formula

$$
\hat{\tau}^2(t) = \hat{S}(t)^2 \sum_{\{j:\tilde{T}_j \leq t, D_j = 1\}} \frac{1}{Y(\tilde{T}_j)\{Y(\tilde{T}_j) - 1\}}.
$$

A standard $100(1 - \alpha)%$ confidence interval (CI) for $S(t)$ is now given by

$$
\hat{S}(t) \pm z_{\alpha/2}\hat{\tau}(t), \quad (2.21)
$$

or we could use a log-minus-log transformation to get a better interval

$$
\hat{S}(t)^{\exp\{\pm z_{\alpha/2}\hat{\tau}(t)/(\hat{S}(t)\log\hat{S}(t))\}}. \quad (2.22)
$$
Figure 2.2: Estimated cumulative hazard (left) and survival function (right) with standard 95\% confidence intervals for the 1977 patients transplanted for CML. To survive means to stay event-free.

Example 1.2: Estimation in the survival case

We are continuing Example 1.1 in Section 2.1, and are now looking at the case where relapsed and dead are considered as one common state. The left-hand plot in Figure 2.2 shows the Nelson-Aalen estimated cumulative hazard (full line) with a 95\% standard confidence interval (dashed lines). It is the slope of this curve we will be interested in. The first year after the transplant, the slope is steeper than later on. This means that the instantaneous risk of relapse or death is decreasing as time goes by.

The other plot gives the Kaplan-Meier estimated survival function, with a 95\% standard confidence interval. The probability not to have had an event during the first year after the transplant is 53.8\%, CI: (51.6\%,56.0\%), hence the probability of relapse or death in the same period is 46.2\%. The probability to stay event-free the first five years after the transplant is 37.8\%, CI: (35.4\%,40.3\%), while the probability to stay event-free up to eight years after the transplant is 33.8\%, with the confidence interval (30.4\%,37.3\%). When \( t \) is close to the maximal survival time which here is 8.45, the data are scarce. When we follow this example further, we will cut off at \( t = 7 \) years. □

2.3 Multi-state models

We will now start focusing on multi-state models. A multi-state model is modeling a stochastic process \( X(t) \) with a set of discrete states (at least two) called the state space \( S \). The value of the process at time \( t \) denotes the state being occupied at that time. Our interests are the probabilities of transitions between the states,
2.4. MARKOV MODELS

the intensities for the transitions and the probabilities to occupy the different states.

For an individual who is in state \( g \) at time \( s \), we will be interested in the probability that he/she is in state \( h \) at time \( t \) \((s < t)\). This is the transition probability, and it is written as

\[
P_{gh}(s, t) = P(X(t) = h \mid X(s) = g, \mathcal{F}_{s-}),
\]

where \( \mathcal{F}_{s-} \) is the history of the process up to time \( s \), i.e. information about the earlier transitions of the process. \( P_{gh}(s, t) \) is the \( gh \)-element of the transition probability matrix \( P(s, t) \); which is showing the probabilities for transitions between all the states in \( S \).

The instantaneous risk of making a transition from \( g \) to \( h \) in a small time interval at time \( t \) (assuming that \( P_{gh} \) is absolutely continuous) is given by the transition intensity

\[
\alpha_{gh}(t) = \lim_{\Delta t \to 0} \frac{P_{gh}(t, t + \Delta t)}{\Delta t}, \quad g \neq h.
\]

The transition intensity matrix \( \alpha(t) \) contains all the transition intensities, where \( \alpha_{gg}(t) \) is defined to be \( \alpha_{gg}(t) = -\sum_{h \neq g} \alpha_{gh}(t) \).

The probability to be in state \( h \in S \) at time \( t \) is denoted \( p_h(t) = P(X(t) = h) \). This state occupation probability is given by the linear combination

\[
p_h(t) = \sum_{j \in S} p_j(0) P_{jh}(0, t).
\]

Expression (2.24) simplifies to \( p_h(t) = P_{1h}(0, t) \), when all the individuals start out in state 1.

The simplest multi-state model is the survival model discussed in Section 2.2 with the two states ‘alive’ and ‘dead’. Two other quite simple multi-state models are the competing risks model and the illness-death model. In a competing risks model we are considering different causes of death. The ‘dead’ state in the survival model is divided into two or more states. The illness-death model will be considered closely in Section 2.5.

2.4 Markov models

Multi-state models are often assumed to be Markov models. This means that the present state of the process is all that matters for future transitions. The past and the future are independent given the present.

Formally we say that a process \( X(t) \) is Markov if

\[
P(X(t) = h \mid X(s) = g, \mathcal{F}_{s-}) = P(X(t) = h \mid X(s) = g).
\]

We will study the matrix versions \( P \) and \( \alpha \), defined in Section 2.3, when the process is Markovian.
For Markov processes, we have the Chapman-Kolmogorov equations
\[ P_{gh}(s, t) = \sum_{l \in S} P_{gl}(s, u) P_{lh}(u, t). \] (2.26)

Using the Chapman-Kolmogorov equations in the case of absolutely continuous transition probabilities, we have that
\[ P(s, t + \Delta t) - P(s, t) = P(s, t)P(t, t + \Delta t) - P(s, t) \]
\[ = P(s, t)(P(t, t + \Delta t) - I) \]
\[ \approx P(s, t)\alpha(t)\Delta t, \]
where
\[ \alpha(t) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} (P(t, t + \Delta t) - I). \]

Hence the Kolmogorov forward equation holds
\[ \frac{\partial}{\partial t} P(s, t) = P(s, t)\alpha(t). \] (2.27)

In the general case, the forward equation may be expressed as
\[ P(s, t) = I + \int_s^t P(s, u-)dA(u). \] (2.28)

This is the multi-state equivalent of (2.5). \( A(t) \) is the matrix of cumulative transition intensities; it is the elementwise integral of \( \alpha(t) \) in the absolutely continuous case. We will now find a solution of (2.28). As in the survival case in Section 2.2, we make a partition of the time interval \((0, t]\) into \( K \) subintervals \((s, t_1], (t_1, t_2], \ldots, (t_{K-1}, t]\). By using the Chapman-Kolmogorov equation, we have that
\[ P(s, t) = P(t_0, t_1)P(t_1, t_2)\ldots P(t_{K-1}, t_k), \]
and by using (2.28), we can write
\[ P(s, t) \approx \prod_{k=1}^K \{I + (A(t_k) - A(t_{k-1}))\}. \]

This matrix product needs to be taken in the increasing order from left to right. Letting the lengths of the subintervals go to zero, the solution of (2.28) is the matrix product-integral
\[ P(s, t) = \prod_{u \in (s, t]} \{I + dA(u)\}. \] (2.29)

This expression is not restricted to the situation where transition intensities exist. In the continuous case, (2.29) will be
\[ P(s, t) = \prod_{u \in (s, t]} \{I + \alpha(u)du\}. \]
2.4. MARKOV MODELS

2.4.1 Estimation in Markov models
As for the survival function, we will use counting processes to estimate the transition probabilities. We will estimate the transition probability matrix by

$$\hat{P}(s, t) = \prod_{u \in (s, t]} \{ I + d\hat{A}(u) \}. \quad (2.30)$$

Thus we need to estimate the matrix of cumulative transition intensities $\hat{A}(t)$. For this purpose, we define $N_{gh}(t)$ to be the number of individuals who are observed to go from state $g$ to state $h$ in the interval $[0, t]$, and $Y_g(t)$ to be the number observed in state $g$ right before time $t$. Then $\hat{A}(t)$ is a matrix of Nelson-Aalen estimators, where the $gh$th element is given by

$$\hat{A}_{gh}(t) = \int_0^t dN_{gh}(u) \frac{Y_g(u)}{Y_g(T_j)}, \quad (2.31)$$

for $h \neq g$ and $\hat{A}_{gg}(t) = -\sum_{h \neq g} \hat{A}_{gh}(t)$. Here the $T_j$’s are the observed transition times between all states. The variance of (2.31) may be estimated by

$$\hat{\sigma}_{gh}^2(t) = \text{Var} \left( \hat{A}_{gh}(t) \right) = \sum_{T_j \leq t} \frac{\Delta N_{gh}(T_j)}{Y_g(T_j)^2} \quad (2.32)$$

Since (2.31) is a step function, the product-integral (2.30) is the finite matrix product

$$\hat{P}(s, t) = \prod_{s < T_j \leq t} \left( I + \Delta \hat{A}(T_j) \right). \quad (2.33)$$

The matrix product needs to be taken in the order of increasing transition times $T_j$. The estimator (2.33) is the Aalen-Johansen estimator. In Section 2.5 we will see that we get nice expressions for the elements of $\hat{P}(s, t)$ when we have a three-state model without recovery.

We will consider some steps leading to the large sample distribution of $\hat{P}(s, t)$. We introduce the indicator function $J_g(t) = I(Y_g(t) > 0)$ for each $g \in S$, and for all $g, h \in S$ we define

$$A_{gh}^*(t) = \int_0^t J_g(u)dA_{gh}(u). \quad (2.34)$$

We let $A^*(t)$ be the matrix with these elements, and introduce $P^*(s, t) = \prod_{u \in (s, t]} \{ I + dA^*(u) \}$, which is almost the same as $P(s, t)$ when $P(Y_g(u) = 0)$ is small for $u \in (s, t]$. By Duhamel’s equation, we may now write

$$\hat{P}(s, t)P^*(s, t)^{-1} - I = \int_{(s, t]} \hat{P}(s, u-)d(\hat{A} - A^*)(u)P^*(s, u)^{-1}. \quad (2.35)$$

This is the multi-state version of (2.18). The matrix $\hat{A} - A^*$ is a matrix of martingales, cf. (2.14). Then we have that the right-hand side of (2.35) is a matrix-valued
stochastic integral, which means that \( \{ \tilde{P}(s, t)P^*(s, t)^{-1} - I \} \) is a matrix of mean zero martingales. Thus
\[
E\{ \tilde{P}(s, t)P^*(s, t)^{-1} \} = I,
\]
which shows that the Aalen-Johansen estimator is almost unbiased. For large sample purposes, \( P^* \) may be replaced by \( P \). The Aalen-Johansen estimator is uniformly consistent, thus we may replace \( \tilde{P} \) by \( P \) on the right-hand side of (2.35). Then
\[
\tilde{P}(s, t)P(s, t)^{-1} - I \approx \int_{(s,t]} P(s, u) d(\hat{A} - A^*)(u)P(s, u)^{-1}.
\]
By multiplying by \( P(s, t) = P(s, u)P(u, t) \), we get
\[
\hat{P}(s, t) - P(s, t) \approx \int_{(s,t]} P(s, u) d(\hat{A} - A^*)(u)P(u, t). \tag{2.36}
\]
From (2.36) one may derive the large sample distribution of \( \hat{P}(s, t) \), see Aalen et al. (2008, Section 3.4.5) for details. In large samples, the elements of \( \hat{P}(s, t) \) are approximately normally distributed, and for any \( g, h, m, r \in S \), one may estimate the covariance between \( \hat{P}_{gh}(s, t) \) and \( \hat{P}_{mr}(s, t) \) by
\[
\hat{cov}(\hat{P}_{gh}(s, t), \hat{P}_{mr}(s, t)) = \sum_{l=1}^{k} \sum_{q \neq l}^{k} \sum_{s < t_{j} \leq t} \{ \hat{P}_{gq}(s, T_j) \hat{P}_{mq}(s, T_j) \}
\times [\hat{P}_{lh}(T_j, t) - \hat{P}_{qh}(T_j, t)][\hat{P}_{lr}(T_j, t) - \hat{P}_{qr}(T_j, t)] \Delta \hat{\sigma}_{ql}^2(T_j), \tag{2.37}
\]
where \( \Delta \hat{\sigma}_{ql}^2(T_j) \) is the increment of (2.32) at time \( T_j \).

2.5 The progressive illness-death model

In the previous section we considered Markov multi-state models. Now we will restrict ourselves to three states, and consider the progressive illness-death model. The state space is then \( S = \{1, 2, 3\} \), and the model is depicted in Figure 2.3. For this model there are three possible transitions, \( 1 \rightarrow 2 \), \( 1 \rightarrow 3 \) and \( 2 \rightarrow 3 \).

State 1: Healthy is the initial state. From state 1 it is possible to go to state 2: Diseased, or to state 3: Dead (i.e. to die of another cause than the illness we are looking at here). State 2 is a transient state; it is possible to leave it, and move on to state 3. Once an individual have reached state 3, it is not possible to move on. The state is absorbing. The path for a patient will be \( 1 \rightarrow 2 \rightarrow 3 \) or \( 1 \rightarrow 3 \), but due to censoring we will not necessarily observe the whole path.

For this model the transition intensity matrix is given by
\[
\alpha(t) = \begin{pmatrix}
-\alpha_{12}(t) - \alpha_{13}(t) & \alpha_{12}(t) & \alpha_{13}(t) \\
0 & -\alpha_{23}(t) & \alpha_{23}(t) \\
0 & 0 & 0
\end{pmatrix}.
\]
2.5. THE PROGRESSIVE ILLNESS-DEATH MODEL

![Illness-death model diagram]

**Figure 2.3:** Illness-death model. It is only possible to move in the directions of the arrows.

When the model is Markovian, we find the transition probabilities as the solution of (2.27). Since state 3 is absorbing we know that \( P_{33}(s, t) = 1 \), and since the model is irreversible, we get that \( P_{ij}(s, t) = 0 \) when \( i > j \). By (2.27), we have that

\[
\frac{\partial}{\partial t} P_{11}(s, t) = - (\alpha_{12}(t) + \alpha_{13}(t)) P_{11}(s, t),
\]

thus we get the solution

\[
P_{11}(s, t) = \exp\left( - \int_{s}^{t} \alpha_{12}(u) + \alpha_{13}(u) du \right). \tag{2.38}
\]

Similarly,

\[
P_{22}(s, t) = \exp\left( - \int_{s}^{t} \alpha_{23}(u) du \right). \tag{2.39}
\]

Further, we have \( P_{23}(s, t) = 1 - P_{22}(s, t) \) and \( P_{13}(s, t) = 1 - P_{11}(s, t) - P_{12}(s, t) \). The last probability we need, to have all the solutions to the Kolmogorov equations, is \( P_{12}(s, t) \). For this probability we have the differential equation

\[
\frac{\partial}{\partial t} P_{12}(s, t) = \alpha_{12}(t) P_{11}(s, t) - \alpha_{23}(t) P_{12}(s, t), \tag{2.40}
\]

which is solved by

\[
P_{12}(s, t) = \exp\left( - \int_{s}^{t} \alpha_{23}(u) du \right) \int_{s}^{t} \alpha_{12}(u) P_{11}(s, u) \exp\left( \int_{s}^{u} \alpha_{23}(v) dv \right) du
\]

\[
= \int_{s}^{t} P_{11}(s, u) \alpha_{12}(u) P_{22}(u, t) du. \tag{2.41}
\]

The steps leading to (2.41) are given in Appendix A.1.
CHAPTER 2. SURVIVAL AND EVENT HISTORY ANALYSIS

Estimation

In Section 2.4.1 we defined $N_{gh}(t)$ to be the number of individuals observed to go from state $g$ to state $h$ in the interval $[0,t]$. For the illness-death model we then have $N_{12}(t)$, $N_{13}(t)$ and $N_{23}(t)$, while $N_{21}(t)$, $N_{32}(t)$ and $N_{31}(t)$ are zero for all $t$ values. We record the exact times of the observed events; when individuals get the disease, or die, $T_1 < T_2 < \ldots$. For the total number of individuals leaving state 1 in $[0,t]$, we use the notation $N_1(t) = N_{12}(t) + N_{13}(t)$, while $Y_1(t)$ and $Y_2(t)$ are the number of healthy and diseased individuals, respectively, right before time $t$.

Then estimators for the transition probabilities, the elements of $\hat{P}(s,t)$ in (2.33), are given by

$$
\hat{P}_{11}(s,t) = \prod_{s<T_j \leq t} \left( 1 - \frac{\Delta N_{12}(T_j)}{Y_1(T_j)} \right),
$$

$$
\hat{P}_{22}(s,t) = \prod_{s<T_j \leq t} \left( 1 - \frac{\Delta N_{23}(T_j)}{Y_2(T_j)} \right),
$$

which both are Kaplan-Meier estimators, and

$$
\hat{P}_{12}(s,t) = \sum_{s<T_j \leq t} \hat{P}_{11}(s,T_{j-1}) \Delta \hat{A}_{12}(T_j) \hat{P}_{22}(T_j,t).
$$

Here

$$
\Delta \hat{A}_{12}(T_j) = \frac{\Delta N_{12}(T_j)}{Y_1(T_j)},
$$

which is the increment of the Nelson-Aalen estimator (2.31) at time $T_j$. The state occupation probability (2.24) may be estimated by

$$
\hat{p}_h(t) = \sum_{g=1}^{3} \hat{p}_g(0) \hat{P}_{gh}(0,t),
$$

where $p_g(0)$ is estimated by the proportion of the individuals who start out in state $g$ at time zero. When all individuals start out in state 1, (2.45) is just

$$
\hat{p}_h(t) = \hat{P}_{1h}(0,t).
$$

Variances of the transition probability estimators

Since the estimators (2.42) and (2.43) are Kaplan-Meier estimators, their variances can be estimated as in the survival case, cf. Section 2.2.1.
We will use (2.37) to estimate the variance of \( \hat{P}_{12}(s, t) \). The estimated variance of a \( g \) to \( h \) transition probability is given by

\[
\hat{\text{Var}}(\hat{P}_{gh}(s, t)) = \hat{\text{Cov}}(\hat{P}_{gh}(s, t), \hat{P}_{gh}(s, t))
\]

\[
= \sum_{l=1}^{3} \sum_{q \neq l} \sum_{s<T_j \leq t} \{ \hat{P}_{gq}(s, T_j)^2[\hat{P}_{lh}(T_j, t) - \hat{P}_{qh}(T_j, t)]^2 \Delta \hat{\sigma}_{ql}^2(T_j) \},
\]

where \( \Delta \hat{\sigma}_{ql}^2(T_j) \) is the increment of (2.32) at \( T_j \). Since the model is without recovery, we obtain

\[
\hat{\text{Var}}(\hat{P}_{12}(s, t)) = \sum_{s<T_j \leq t} \hat{P}_{11}(s, T_j)^2[\hat{P}_{22}(T_j, t) - \hat{P}_{12}(T_j, t)]^2 \Delta \hat{\sigma}_{12}^2(T_j)
\]

\[
+ \sum_{s<T_j \leq t} [\hat{P}_{12}(s, T_j)\hat{P}_{22}(T_j, t)]^2 \Delta \hat{\sigma}_{13}^2(T_j)
\]

\[
+ \sum_{s<T_j \leq t} [\hat{P}_{12}(s, T_j)\hat{P}_{22}(T_j, t)]^2 \Delta \hat{\sigma}_{23}^2(T_j).
\]

(2.46)

The full expression of (2.46), where also those terms which are zero, are included, is given in Appendix A.2. The standard deviation of \( \hat{P}_{12}(s, t) \) may then be estimated by

\[
\hat{\sigma}_{\hat{P}_{12}(s,t)} = (\hat{\text{Var}}(\hat{P}_{12}(s, t)))^{1/2}.
\]

(2.47)

Since \( \hat{P}_{12}(s, t) \) is approximately normally distributed in large samples, a standard 100\( (1 - \alpha) \)% confidence interval for \( P_{12}(s, t) \) is given by

\[
\hat{P}_{12}(s, t) \pm z_{\alpha/2} \cdot \hat{\sigma}_{\hat{P}_{12}(s,t)}.
\]

(2.48)

Alternatively, one may use the log-transformed confidence interval

\[
\hat{P}_{12}(s, t) \exp(\pm z_{\alpha/2} \hat{\sigma}_{\hat{P}_{12}(s,t)}/\hat{P}_{12}(s, t)),
\]

(2.49)

or the log-minus-log transformed confidence interval

\[
\hat{P}_{12}(s, t) \exp(\pm z_{\alpha/2} \hat{\sigma}_{\hat{P}_{12}(s,t)}/\hat{P}_{12}(s, t) \log(\hat{P}_{12}(s, t))).
\]

(2.50)

**Example 1.3: Markov illness-death model**

We continue our example presented in Section 2.1, and we will now consider three different states. After the bone marrow transplant, a patient may have a relapse of CML. Relapse will be state 2 in our model. Later on, the patient may die, and hence enter state 3. State 3 may also be reached without a foregoing relapse. To be in state 1 means to stay event-free. In all states, the patient may be censored.
Figure 2.4: Estimated cumulative transition intensities for the CML data. The left plot shows the cumulative transition intensities the first 7 years after transplant. The right plot shows the cumulative transition intensities the first year after transplant. The black curve is for transition 1 → 2, the red for transition 1 → 3, and the green for transition 2 → 3.

Figure 2.4 shows the estimated cumulative transition intensities for the CML data we are considering. The left-hand plot shows the estimates for all times up to 7 years, while the right-hand plot is restricted to the first year to get a clearer picture of what is happening just after the transplant. The black curve is for transition 1 → 2 (to have a relapse), the red for transition 1 → 3 (to die without a foregoing relapse) and the green for transition 2 → 3 (to die after a relapse). We see that, right after the transplant, the instantaneous risk of dying without a foregoing relapse is higher than the risk of relapse. After a month, the risk of dying after a relapse, is much higher than the two other risks. The risk of a 1 → 2 transition is approximately constant the first year, before it is leveling off. The risk of a 1 → 3 transition is higher than the risk of a 1 → 2 transition the first half a year. The next six months they are approximately equal. Further on, the risk of a 1 → 2 transition is larger than the risk of a 1 → 3 transition, but the risks are smaller than during the first year after transplant. The risk of dying after a relapse is decreasing as time goes by.

Figure 2.5 shows the estimated transition probabilities from state 1, starting from three different time points after the transplant. We are interested in looking at how the transition probabilities changes for patients, who still have had no event at given times after the transplant. We choose two time points s, in addition to the initial point s = 0. These have been chosen such that the probability to be in state 1 is approximately 80% and 50%. For our data we then get s = 0.25 and s = 1.35, which corresponds to approximately 90 and 500 days post transplant. The black curve in each plot shows an estimate of
2.5. THE PROGRESSIVE ILLNESS-DEATH MODEL

Figure 2.5: Transition probabilities for the CML data. The first plot shows the state occupation probabilities with standard 95% CI. The black curve is the probability to be in state transplanted; no other events have happened. The green one is the probability to have had a relapse, but still be alive, while the red is the probability of being dead. The two other plots show the transition probabilities from 0.25 and 1.35 years after the transplant.

\[ P_{11}(s,t) \]. Note that \( \hat{P}_{11}(0,t) \) is the same as the estimated survival curve that we considered in Figure 2.2. The green and red curves are estimates of \( P_{12}(s,t) \) and \( P_{13}(s,t) \), respectively. The first plot, where \( s = 0 \), shows the estimated state occupation probabilities; since all the individuals start out as transplanted at known times. Right after the transplant, the estimated probability to have a relapse and then no other event during the next year is 10.2% with a standard 95% CI: (8.8%,11.5%). The probability to die in that period, with or without a foregoing relapse, is 36.1%, CI: (33.9%, 38.2%), while the probability of no event is 53.8%, CI: (51.6%, 56.0%). If we have a larger horizon, the probability to have a relapse and not leave state 2 during the next five years, is 14.5%, (12.5%,16.5%), while the probability of death during that time period is 47.7%, CI: (45.2%,50.1%). From \( s = 0.25 \), the estimated probability to have a relapse and no other event during the following year has increased a bit, compared
to for $s = 0$, and is now 11.3%, CI: (9.7%, 12.8%), while the probability of death has decreased to 25.3%, CI: (23.1%, 27.4%). For a five years period, the probability of relapse and then no other event is 17.0%, CI: (14.5%, 19.4%), while the probability to die during that period is 37.3%, CI: (34.5%, 40.1%). We see that the survival prognosis for a patient who has stayed event-free the first 90 days post transplant, is better than right after the transplant. The probability of death during the following year has decreased from 36.1% to 25.3%, and for the five years period it has decreased from 47.7% to 37.3%.

For a patient who is still event-free 500 days post transplant ($s = 1.35$), the probability of death during the following year is only 4.2%, CI: (2.9%, 5.5%), while it is 13.2%, CI: (10.1%, 16.3%) for the following five years period. The probabilities of relapse and no other events are 6.5%, CI: (4.9%, 8.1%) and 16.3%, CI: (12.5%, 20.1%), for these two time periods, respectively. R-script for estimation of the transition probabilities is given in Appendix B.1. 

When we are analyzing real data we do not know whether the Markov assumption is fulfilled or not. In Chapter 3 we will consider two other models and methods for the transition probabilities in an illness-death model. Further on, the Aalen-Johansen estimator (2.33), will be denoted the Markov method, while the standard deviation estimator (2.47) will be denoted the Markov standard deviation estimator.
Chapter 3

Alternative estimators

In the previous chapter, we studied the illness-death model under the Markov condition. The transition intensities $\alpha_{12}(t)$, $\alpha_{13}(t)$ and $\alpha_{23}(t)$, are then functions of time $t$ since the initial event, called the global time. Now we will consider the illness-death model under a semi-Markov assumption, and for the case without any of these assumptions. In Section 3.1, we will consider the semi-Markov model and a method for estimation. In this model the time is reset to zero when a new state is reached, hence the time scale is called ’clock reset’. What then matters regarding transition intensities, is for how long the individual has been in the current state, called the duration time $d$, while the global time $t$ is irrelevant. The transition intensities are then $\alpha_{12}(t)$, $\alpha_{13}(t)$ and $\alpha_{23}(d)$, where it is used that the duration time and the global time are the same for state 1. In Section 3.2 we will consider a general model, and one way to do estimation here. The transition intensities are then functions of both the duration time $d$ and the global time $t$. Below, we will not focus on transition intensities, but rather go directly to the transition probabilities.

3.1 Semi-Markov illness-death model

We say that a process $X(t)$ is semi-Markov if the only interesting part of the history $\mathcal{F}_{s-}$ in

$$P(X(t) = h \mid X(s) = g, \mathcal{F}_{s-}),$$

is the time since state $g$ was reached. A semi-Markov model is also called a Markov renewal model. In the Markov model the time runs from the initial time point, and only the current state and the time since the initial time matters for future transitions. In the semi-Markov model the time is reset to zero when a new state is entered, and the current state and the time since this state was reached is all that matters for future transitions.

We will only be interested in the illness-death model, and the time is then set to zero when an individual is entering state 2. We consider transition probabilities from state 1, and the expression for $P_{11}(s,t)$ will be as for the Markov model. Hence we will consider the expression for $P_{12}(s,t)$. Compared to (2.41), we now
need to give another expression for \( P_{22}(u, t) \). The probability to stay in state 2 for a time period longer than \( u \), after first reaching the state, is

\[
P_{22}^*(0, u) = P(T_{23} > u),
\]

where \( T_{23} \) is the potential time an individual is in state 2. Hence \( P_{12}(s, t) \) is now given by

\[
P_{12}(s, t) = \int_s^t P_{11}(s, v)\alpha_{12}(v)P_{22}^*(0, t - v)dv.
\]

### 3.1.1 Estimation in the semi-Markov model

We estimate \( P_{22}^*(0, v) \) by the Kaplan-Meier estimator

\[
\hat{P}_{22}^*(0, v) = \prod_{v_j \leq v} \left( 1 - \frac{\Delta N_{23}^*(v_j)}{Y_2^*(v_j)} \right).
\]  

(3.1)

Here \( v_1 < v_2 < \ldots \) are observed sojourn times in state 2 (observed values of \( T_{23} \)). \( N_{23}^*(v_j) \) is the number of individuals who go to state 3 within a time period of length \( v_j \) after state 2 was reached, while \( Y_2^*(v_j) \) is the number of individuals with sojourn time in state 2 at least \( v_j \). Now we estimate \( P_{12}(s, t) \) by

\[
\hat{P}_{12}(s, t) = \sum_{s < T_j \leq t} \hat{P}_{11}(s, T_{j-1})\hat{A}_{12}(T_j)\hat{P}_{22}^*(0, t - T_j).
\]  

(3.2)

One may prove that \( \frac{1}{n} \text{Var}(\hat{P}_{12}(s, t) - P_{12}(s, t)) \) converges in distribution to \( Z(s, t) \), where the expression of the Gaussian process \( Z \) is given in Voelkel and Crowley (1984). Hence, the variance of \( \hat{P}_{12}(s, t) \) is approximately given by \( \frac{1}{n} \text{Var}(Z(s, t)) \), where \( n \) is the number of individuals. We will not go any further into this material in this thesis. Because of the lack of software, we haven’t calculated the variance by this method neither in our examples nor in the simulations in Chapter 5. We can however use bootstrapping to estimate variances.

### Example 1.4: Semi-Markov probabilities

We continue to consider the CML data from Chapter 2. Now we want to see how the semi-Markov method performs, compared to the Markov method on these data. \( P_{11}(s, t) \) is estimated in the same way for the two methods, hence \( P_{12}(s, t) \) is the only interesting probability to consider. Figure 3.1 shows the transition probability \( P_{12}(s, t) \) estimated by the semi-Markov method (red curve) together with the transition probability estimated by the Markov method (green curve) for the three \( s \) values considered in Example 1.3. For \( s = 0 \), the probabilities estimated by the semi-Markov method are higher than those estimated by the Markov method, up to approximately \( t = 3 \) years. For larger \( t \) values, it is the
other way around. For $s = 0.25$, the semi-Markov estimates are a bit higher than the Markov estimates up to $t = 1.5$ years, while for larger $t$ values, the Markov estimates are higher. For $s = 1.35$, the transition probabilities estimated by the semi-Markov method are lower than those estimated by the Markov method. We will consider the same transition probabilities as in Example 1.3. For $s = 0$, the estimated probability to have a relapse and then no other event during the following year is 11.8% (10.2% for the Markov method), while it for the following five years period is 13.8% (14.5% for the Markov method). Since $P_{11}(s, t)$ is estimated in the same way for the two methods, we now have that the estimated probability to die during the following year after transplant is 34.4%, while it is 36.1% for the Markov method. Hence, this method gives a better one year survival prognosis for a patient, than the Markov method does, but a less
good five years prognosis. The estimated probability to die during the first five years after transplant is 48.4% for the semi-Markov method, and 47.7% for the Markov method. For $s = 0.25$, the estimated probability to have a relapse and then no other event during the following year is 11.7% (11.3% for the Markov method), and during the following five years period it is 15.0% (17.0% for the Markov method). For $s = 1.35$, the estimated probability to have a relapse and then no other event during the following year is 5.3% (6.5% for the Markov method), and during the following five years period it is 11.3% (16.3% for the Markov method). We see that the differences between the methods become quite large as $s$ increases. R-script for estimation of the semi-Markov transition probabilities is given in Appendix B.1.

### 3.2 A non-Markov/general illness-death model

We will now consider a general illness-death model, and follow the work by Meira-Machado et al. (2006). In this section we derive expressions for the transition probabilities (1): $P_{11}(s, t)$, (2): $P_{12}(s, t)$ and (3): $P_{22}(s, t)$.

We introduce the random vector $(T_{12}, T_{13}, T_{23})$, where $T_{ij}$ is the potential time spent in state $i$ before transition to state $j$. Starting from the initial state, state 1, there are two courses for an individual:

- **a)** $1 \rightarrow 2 \rightarrow 3$ : At time $T_{12}$ there is a transition from state 1 to 2, and then at time $T_{12} + T_{23}$ there is a transition further to state 3,

- **b)** $1 \rightarrow 3$ : At time $T_{13}$ there is a direct transition from state 1 to state 3.

If $T_{12} \leq T_{13}$, the individual follows course a). The value of $T_{13}$ is then censored at time $T_{12}$ for that individual. The potential time spent in state 1, called the potential sojourn time in that state, we denote by $Z = \min(T_{12}, T_{13})$, while the potential total time to state 3 is reached is denoted $T$, and can be expressed by $T = Z + T_{23}I(Z = T_{12})$.

The events involved in (1) are that the process $X(t)$ was in state 1 at time $s$, and that it is still there at time $t$. That means that both $T_{12}$ and $T_{13}$ are larger than both $s$ and $t$, giving that $Z > s, t$, and the probability expression is hence

$$P_{11}(s, t) = P(Z > t \mid Z > s) = \frac{P(Z > t)}{P(Z > s)} = \frac{1 - H(t)}{1 - H(s)}.$$  \hspace{1cm} (3.3)

Here $H(z)$ is the cumulative distribution function of $Z$, and $1 - H(z)$ is hence the survival function of $Z$. For (2), the process $X(t)$ was in state 1 at time $s$, hence $Z > s$. Further, it follows course a), hence $T_{12} \leq T_{13}$. The process leaves state 1 before, or at, time $t$ and has not yet arrived in state 3 at time $t$, hence $T_{12} \leq t$ and
3.2. A NON-MARKOV/GENERAL ILLNESS-DEATH MODEL

\[T_{12} + T_{23} > t.\] The probability expression is then

\[P_{12}(s, t) = P(T_{12} \leq t, T_{12} \leq T_{13}, T_{12} + T_{23} > t \mid Z > s)\]

\[= \frac{P(s < T_{12} \leq t, T_{12} \leq T_{13}, T_{12} + T_{23} > t)}{P(Z > s)}\]

\[= \frac{E \{ I(s < T_{12} \leq t, T_{12} + T_{23} > t, T_{12} \leq T_{13}) \}}{E \{ I(Z > s) \}}. \quad (3.4)\]

For (3), the process was in state 2 at time \(s\). That means that the process left state 1 before, or at time \(s\), followed course a) and didn’t leave state 2 before, or at, \(s\). Hence \(T_{12} \leq s, T_{12} \leq T_{13}\) and \(T_{12} + T_{13} > s\). The process is in state 2 at time \(t\), meaning that \(T_{12} + T_{23} > t\). Thus the probability expression is

\[P_{22}(s, t) = P(T_{12} + T_{23} > t \mid T_{12} \leq s, T_{12} \leq T_{13}, T_{12} + T_{23} > s)\]

\[= \frac{P(T_{12} + T_{23} > t, T_{12} \leq s, T_{12} \leq T_{13})}{P(T_{12} \leq s, T_{12} + T_{23} > s, T_{12} \leq T_{13})}\]

\[= \frac{E \{ I(T_{12} \leq s, T_{12} + T_{23} > t, T_{12} \leq T_{13}) \}}{E \{ I(T_{12} \leq s, T_{12} + T_{23} > s, T_{12} \leq T_{13}) \}}. \quad (3.5)\]

For estimation purposes it is convenient to have simpler notations for the expressions (3.4) and (3.5). We introduce the function

\[S(\phi) = E \{ \phi(T_{12}, T_{12} + T_{23})I(T_{12} \leq T_{13}) \}. \quad (3.6)\]

Using (3.6) we can then write

\[P_{12}(s, t) = \frac{S(I(s < T_{12} \leq t, T_{12} + T_{23} > t))}{1 - H(s)} = \frac{S(\phi_{s,t})}{1 - H(s)} \]

and

\[P_{22}(s, t) = \frac{S(I(T_{12} \leq s, T_{12} + T_{23} > t))}{S(I(T_{12} \leq s, T_{12} + T_{23} > s))} = \frac{S(\tilde{\phi}_{s,t})}{S(\phi_{s,s})}, \]

where

\[\phi_{s,t}(u, v) = I(s < u \leq t, v > t) \quad \text{and} \quad \tilde{\phi}_{s,t}(u, v) = I(u \leq s, v > t). \quad (3.7)\]

We see that \(S(\phi)\) only covers course a) above. To include course b), we consider \(\phi(Z, T)\), a function of \(Z\); the potential time in state 1, and the potential survival time \(T\). The survival time \(T\) can be expressed as

\[T = I(T_{12} \leq T_{13})(T_{12} + T_{23}) + I(T_{12} > T_{13})(T_{13}).\]

Now we have that

\[\phi(Z, T) = \begin{cases} 
\phi(T_{12}, T_{12} + T_{23}) & \text{if } T_{12} \leq T_{13}, \\
\phi(T_{13}, T_{13}) & \text{if } T_{12} > T_{13}.
\end{cases}\]
and the expectation of $\phi(Z, T)$ becomes

$$E \{ \phi(Z, T) \} = E \left\{ \phi(T_{12}, T_{12} + T_{23}) I(T_{12} \leq T_{13}) \right\} + E \left\{ \phi(T_{13}, T_{13}) I(T_{12} > T_{13}) \right\}$$

$$= S(\phi) + E \left\{ \phi(T_{13}, T_{13}) I(T_{12} > T_{13}) \right\}.$$

For our two $\phi$ functions, given by $\phi_{s,t}$ and $\tilde{\phi}_{s,t}$ in (3.7), we see that

$$\phi_{s,t}(T_{13}, T_{13}) = \tilde{\phi}_{s,t}(T_{13}, T_{13}) = 0 \quad \text{for} \quad 0 \leq s < t.$$

Hence, in the cases interesting for us,

$$S(\phi) = E \left\{ \phi(Z, T) \right\}. \quad (3.8)$$

Exact calculation of $S(\phi)$ assumes that the joint distribution of $Z$ and $T$ is known. When we in the next subsection give an estimator for $S(\phi)$, we look at how this distribution may be estimated.

### 3.2.1 Estimation in the general model

The data we use for estimation may be censored. Therefore we introduce a censoring variable $C$, which is assumed to be independent of $(T_{12}, T_{13}, T_{23})$. We let $U$ be the first transition or censoring time, $U = \min(T_{12}, T_{13}, C) = \min(Z, C)$, and $\delta$ an indicator for this time to be a $1 \to 2$ transition time, $\delta = I(T_{12} < T_{13}, T_{12} \leq C)$. We also introduce an indicator $\eta$ for $U$ to be a $1 \to 3$ transition time, $\eta = I(T_{13} < T_{12}, T_{13} \leq C)$. Further, we let $V$ be the time from a $1 \to 2$ transition takes place, to a new transition or censoring occurs, $V = \min(T_{23}, C - T_{12})$, while $\rho$ is an indicator for this time to be a transition time, $\rho = I(T_{23} \leq C - T_{12})$. Then $U$ and $V$ are the observed sojourn times in state 1 and state 2, respectively.

We also introduce the uncensoring indicator of $Z$, $\gamma = I(Z \leq C) = \delta + (1 - \delta) \eta$. For estimation of $S(\phi)$ we introduce $Y = \min(T, C) = U + \delta V$, where $T$ is the potential survival time, and $\xi = I(T \leq C) = (1 - \delta) \eta + \delta \rho$. $\xi$ is now an indicator for the observed $Y$ to be a survival time.

For each individual the total sample information is given by

$$(U_i, \delta_i, \delta_i V_i, \delta_i \rho_i, (1 - \delta_i) \eta_i), \quad 1 \leq i \leq n,$$

and this will be used for estimation of the transition probabilities (3.3) - (3.5). For estimation of the numerator and denominator of (3.3), and the denominator of (3.4), we may, since $C_i$ and $Z_i$ are independent, use the Kaplan-Meier estimator. We have that

$$U_i = \begin{cases} Z_i & \text{if } \gamma_i = 1, \\ C_i & \text{if } \gamma_i = 0, \end{cases}$$

The survival function of $Z$ may then be estimated by

$$\hat{P}(Z > z) = \prod_{i: U_i \leq z, \gamma_i = 1} \left\{ 1 - \frac{1}{K(i)} \right\} \equiv 1 - \hat{H}(z),$$
where $K(i)$ is the number in state 1 just before time $U_i$. As an estimator for (3.3), we define

$$\hat{P}_{11}(s,t) = \frac{1 - \hat{H}(t)}{1 - \hat{H}(s)}.$$  \hspace{1cm} (3.9)

This is equivalent to the Aalen-Johansen estimator of $P_{11}(s,t)$, (2.42).

How should $S(\phi)$ in (3.8) be estimated for our $\phi$ functions $\phi_{s,t}$ and $\tilde{\phi}_{s,t}$ given in (3.7)? Because of right-censoring, we will not necessarily observe $Z$ and $T$, but we know that $Z$ is uncensored whenever $T$ is. This fact makes us use the multivariate Kaplan-Meier estimators considered by Stute (1993).

For each individual we have $Y_i = U_i + \delta_i V_i$ and $\xi_i = (1 - \delta_i) \eta_i + \delta_i \rho_i$, $1 \leq i \leq n$. We are interested in the ordered sample $Y(1) \leq \ldots \leq Y(n)$, and the pair $(U_{[i]}, \xi_{[i]})$ attached to the $Y(i)$ value, $1 \leq i \leq n$. Consider the estimator

$$\hat{S}(\phi) = \sum_{i=1}^{n} W_i \phi(U_{[i]}, Y(i)),$$  \hspace{1cm} (3.10)

which is an empirical integral, where

$$W_i = \frac{\xi_{[i]} - 1}{L(i) \prod_{j=1}^{i-1} \left(1 - \frac{\xi_{[j]} L(j)}{L(i)}\right)}.$$  \hspace{1cm} (3.10)

Here $L(i) = n - i + 1$ is the number of individuals who are in state 1 or 2 right before time $Y(i)$. The weight $W_i$ for individual $i$, is the change in the Kaplan-Meier estimator at time $Y(i)$:

$$W_i = - \left( \prod_{j=1}^{i} \left[1 - \frac{\xi_{[j]} L(j)}{L(i)}\right] \right) - \prod_{j=1}^{i-1} \left[1 - \frac{\xi_{[j]} L(j)}{L(i)}\right]$$

$$= - \left( \left(1 - \frac{\xi_{[i]} L(i)}{L(i)}\right) - 1 \right) \prod_{j=1}^{i-1} \left[1 - \frac{\xi_{[j]} L(j)}{L(i)}\right]$$

$$= \frac{\xi_{[i]} - 1}{L(i) \prod_{j=1}^{i-1} \left(1 - \frac{\xi_{[j]} L(j)}{L(i)}\right)}.$$  \hspace{1cm} (3.10)

Meira-Machado et al. (2006) show that the estimator (3.10) is consistent for our choices of $\phi$. Hence estimators of (3.4) and (3.5) are

$$\hat{P}_{12}(s,t) = \frac{\hat{S}(\phi_{s,t})}{1 - \hat{H}(s)},$$  \hspace{1cm} (3.11)

and

$$\hat{P}_{22}(s,t) = \frac{\hat{S}(\tilde{\phi}_{s,t})}{\hat{S}(\tilde{\phi}_{s,s})},$$  \hspace{1cm} (3.12)
respectively. In their paper it is also shown that, $n^{1/2} [\hat{P}_{12}(s,t) - P_{12}(s,t)]$ converges in distribution to $N(0, B(s,t))$, where $B(s,t)$ is some limit variance function. This is valid for $s < \tau_0$, where $\tau_0$ is the upper bound of the support of $U$. We get an estimator for $B(s,t)$ by plugging in estimators of its parameters, but this estimator is not necessarily consistent when $\phi$ has noncompact support as in our case. For $\hat{P}_{22}$ there is a similar large sample result. Also here, as for the semi-Markov method, one can estimate the variances with bootstrapping.

**Example 1.5: Non-Markov transition probabilities**

For the CML data, we will now consider estimation of $P_{12}(s,t)$ by the non-Markov method. We use the TPmsm package in R, which also produces bootstrap
confidence intervals, see Appendix B.1. We do not plot these, but give confidence intervals for the specific transition probabilities considered. Figure 3.2 shows $\hat{P}_{12}(s, t)$ for the three already chosen $s$ values, for the non-Markov method (blue curve), and the Markov method (green curve). For $s = 0$, the curves follow each other quite closely, meaning that the methods are giving approximately the same estimates. For both $s = 0.25$ and $s = 1.35$, the non-Markov estimates are a bit lower than the Markov estimates from $t = 3$ years onwards. The differences are larger for the $s = 1.35$ case, than for $s = 0.25$.

Again we consider some of the transition probabilities from Example 1.3. For $s = 0$, the estimated probability to have a relapse and then no other event during the following year is 10.2%, as for the Markov method. The bootstrap confidence interval is (8.9%, 11.6%), which is approximately as the CI for the Markov method. For the five years period, this estimated transition probability is 14.6% (14.5% for the Markov method), with the confidence interval (12.5%, 16.4%). For $s = 0.25$, the estimated probability to have a relapse and then no other event during the following year is 11.2% (11.3% for the Markov method), CI: (9.6%, 12.8%), and during the following five years period it is 16.5% (17.0% for the Markov method), CI: (14.0%, 19.0%). For $s = 1.35$, the estimated probabilities are 6.4% (6.5% for the Markov method), CI: (4.9%, 8.1%), and 14.7% (16.3% for Markov), CI: (10.8%, 18.8%), for the one year and five years periods, respectively. Some of the confidence intervals here are a bit wider than those for the Markov method. The confidence intervals for the Markov method are based on the standard deviations estimated by the standard deviation estimator (2.47).

In the example above we saw that the Markov method and the non-Markov method gave estimates closer together, than was the case for the Markov and the semi-Markov methods. The true probabilities are unknown, hence, this information does not tell us which method is the "correct" one. In the next chapter, we will consider the Brier score as a mean to determine which method performs best on real data.
Chapter 4

Comparison of the methods on real data

In the two previous chapters we have considered three methods for estimation of transition probabilities in an illness-death model. When we have real data, we do not know if a process is Markov or semi-Markov. Which of the three methods should we use on our data? It is natural to choose the method which fits the data best. The Brier score is a measure of inaccuracy. The score was formulated back in 1950, and it was then used to verify weather forecasts (Brier 1950). Glenn Brier looked at probabilities of rain and no-rain on a particular day $i$, $p_i$ and $1 - p_i$, respectively, and considered the sum of the squared discrepancies between the actual weather that day $e_i$ ($e_i = 1$ for rain, $e_i = 0$ for no-rain) and the probability of rain, and between the actual weather and the probability of no-rain. The mean of these results from $n$ days, was then the Brier score for this particular example. This score could then be compared to the score calculated for other probabilities.

We see that the idea is to compare probabilities with what is actually being observed in the data. The Brier score was formulated for survival times, where the censoring is assumed to be random, by Graf et al. (1999). They used an inverse probability of censoring weighting, and showed that the loss of information due to censoring can be accounted for by this. In Section 4.1, we study the Brier score for our setting, and apply this on the CML data, before we in Section 4.2 give one more data example.

4.1 The Brier score for our setting

We will be interested in how well an estimated transition probability $\hat{P}_{12}(s, t)$ is able to predict the event status $I(X(t) = 2)$ at the specific time point $t > s \geq 0$, where $s$ is a given time. We will start by looking at the case where $s = 0$, since this case gives a simpler derivation, than for $s \in (0, t)$. The Brier score may also be calculated for the other states. The illness-death model contains three states, and we could be interested in how well $\hat{P}_{11}(s, t)$ and $\hat{P}_{13}(s, t)$ are able to predict $I(X(t) = 1)$ and $I(X(t) = 3)$, respectively. Hence, we will consider the Brier score for state $h$, where $h = 1, 2$ or 3. In our derivations, we will use $\Pi_{1h}(s, t^*)$ as
a model for the transition probability, and assume that $\Pi_{1h}(s, t)$ is non-random, to avoid too hard calculations. In practice different estimates of $P_{1h}(s, t)$ will be used. The Brier score in the case without censoring, is now given by

$$BS_{1h}(t) = \frac{1}{n} \sum_{i=1}^{n} \left( I(X_i(t) = h) - \Pi_{1h}(0, t) \right)^2.$$ \hspace{1cm} (4.1)

This is an estimator for the mean squared error measure

$$E\{BS_{1h}(t)\} = E\left\{ \left( I(X(t) = h) - \Pi_{1h}(0, t) \right)^2 \right\},$$ \hspace{1cm} (4.2)

where iid variables are required. We introduce the survival time for individual $i$; $T_i = \min\{t : X_i(t) = 3\}$. This is the time when individual $i$ arrives at state 3, and we can then write (4.2) as

$$E\{BS_{1h}(t)\} = E\left\{ \left( I(X(t) = h) - \Pi_{1h}(0, t) \right)^2 I(T_i \leq t) + I(T_i > t) \right\}$$

$$= E\left\{ \left( I(X(t) = h) - \Pi_{1h}(0, t) \right)^2 I(T_i \leq t) \right\}$$

$$+ E\left\{ \left( I(X(t) = h) - \Pi_{1h}(0, t) \right)^2 I(T_i > t) \right\}.$$ 

For the case with random censoring we introduce

$\tilde{T}_i = \min(T_i, C_i),$

where $C_i$ is a censoring time, independent of $T_i$, and also

$$\tilde{X}_i(t) = \begin{cases} X_i(t) & t < \tilde{T}_i \\ X_i(t)I(\tilde{T}_i = T_i) & t \geq \tilde{T}_i \end{cases}. \hspace{1cm} (4.3)$$

In words, (4.3) says that, if individual $i$ is in state 1 or 2 at time $t$, then $\tilde{X}_i(t)$ equals the state the individual is in at time $t$. If the survival or censoring time is smaller than or equal $t$, $\tilde{X}_i(t) = 3$ if $\tilde{T}_i$ was a survival time, and $\tilde{X}_i(t) = 0$ if $\tilde{T}_i$ was a censoring time. When censoring is present we need to weight the Brier score, (4.1), to compensate for the loss in information.

As suggested by Graf et al. (1999), the Brier score in the case of random censoring, for $h = 1, 2$ or 3, can now be given by

$$BS_{1h}^c(t) = \frac{1}{n} \sum_{i=1}^{n} \left( I(\tilde{X}_i(t) = h) - \Pi_{1h}(0, t) \right)^2 w_i(t),$$ \hspace{1cm} (4.4)

where the weights $w_i(t)$ are

$$w_i(t) = \frac{I(\tilde{T}_i \leq t, \tilde{X}_i(t) \neq 0)}{G(\tilde{T}_i)} + \frac{I(\tilde{T}_i > t)}{G(t)}.$$ \hspace{1cm} (4.5)
4.1. THE BRIER SCORE FOR OUR SETTING

Here $G(t) = P(C_i > t)$. From (4.5) we see that $w_i(t) = 0$ for observations censored before or at time $t$. We show that this score is reasonable by considering its expectation. The expectation of (4.4) is now

$$E \{ BS_{1h}^c(t) \} = E \left\{ (I(\tilde{X}_i(t) = h) - \Pi_{1h}(0,t))^2 w_i(t) \right\}$$

$$= E \left\{ (I(\tilde{X}_i(t) = h) - \Pi_{1h}(0,t))^2 \frac{I(\tilde{T}_i \leq t, \tilde{X}_i(t) \neq 0)}{G(\tilde{T}_i)} \right\}$$

$$+ E \left\{ (I(\tilde{X}_i(t) = h) - \Pi_{1h}(0,t))^2 \frac{I(\tilde{T}_i > t)}{G(t)} \right\}$$

$$= \text{Part I} + \text{Part II}.$$

For a better overview we will look at the two parts separately. For $\tilde{T}_i \leq t$ and $\tilde{X}_i(t) \neq 0$, it is the case that $\tilde{X}_i(t) = X_i(t)$, $\tilde{T}_i = T_i$, and hence that $T_i \leq C_i$. Now we have

**Part I**

$$E \left\{ (I(\tilde{X}_i(t) = h) - \Pi_{1h}(0,t))^2 \frac{I(\tilde{T}_i \leq t, \tilde{X}_i(t) \neq 0)}{G(\tilde{T}_i)} \right\}$$

$$= E \left\{ (I(X_i(t) = h) - \Pi_{1h}(0,t))^2 \frac{I(T_i \leq t)I(C_i \geq T_i)}{G(T_i)} \right\}$$

$$= E \left\{ E \left\{ (I(X_i(t) = h) - \Pi_{1h}(0,t))^2 \frac{I(T_i \leq t)I(C_i > T_i)}{G(T_i)} \right| T_i \right\}$$

$$= E \left\{ (I(X_i(t) = h) - \Pi_{1h}(0,t))^2 I(T_i \leq t) \right\} ,$$

where we are using that $E \{ I(C_i > T_i) \mid T_i \} = G(T_i)$. For the second part we have, since $\tilde{X}_i(t) = X_i(t)$ for $\tilde{T}_i > t$, and $T_i$ and $C_i$ are independent

**Part II**

$$E \left\{ (I(\tilde{X}_i(t) = h) - \Pi_{1h}(0,t))^2 \frac{I(\tilde{T}_i > t)}{G(t)} \right\}$$

$$= E \left\{ (I(X_i(t) = h) - \Pi_{1h}(0,t))^2 \frac{I(T_i > t)I(C_i > t)}{G(t)} \right\}$$

$$= E \left\{ (I(X_i(t) = h) - \Pi_{1h}(0,t))^2 I(T_i > t) \right\} E\{ I(C_i > t) \}$$

$$= E \left\{ (I(X_i(t) = h) - \Pi_{1h}(0,t))^2 I(T_i > t) \right\} .$$
Hence

\[ E \{ BS_{1h}(t) \} = E \left\{ \left( I(\bar{X}_i(t) = h) - \Pi_{1h}(0, t) \right)^2 w_i(t) \right\} \]

\[ = E \left\{ \left( I(\bar{X}_i(t) = h) - \Pi_{1h}(0, t) \right)^2 I(T_i \leq t) \right\} \]

\[ + E \left\{ \left( I(\bar{X}_i(t) = h) - \Pi_{1h}(0, t) \right)^2 I(T_i > t) \right\} \]

\[ = E \{ BS_h(t) \}, \]

Our weighted Brier score for censored data, (4.4), has the same expected value as the Brier score for a complete dataset, (4.1), and is thus succeeding.

For practical use, the censoring distribution \( G(t) = P(C > t) \) needs to be estimated, and we will use the Kaplan-Meier estimator for this purpose. We introduce \( \xi = I(T \leq C) \), and let \( \hat{G}(\cdot) \) denote the Kaplan-Meier estimate of the censoring distribution based on \((\bar{T}_i, 1 - \xi_i)\), for all \( n \) individuals. Additionally, we replace \( \Pi_{1h}(0, t) \) by \( \hat{\Pi}_{1h}(0, t) \), for a chosen method.

We are not only interested in the Brier score for estimated state occupation probabilities, but also for estimated transition probabilities from time \( s > 0 \). The Brier score, for the case without censoring, is now given by

\[ BS_{1h}(s, t) = \frac{1}{\sum_{i=1}^n I(X_i(s) = 1)} \sum_{i: X_i(s) = 1} (I(X_i(t) = h) - \Pi_{1h}(s, t))^2, \quad (4.6) \]

for \( t \geq s \). In (4.6) we are only interested in those individuals who are in state 1 at time \( s \). This score is an estimator of

\[ E \{ BS_{1h}(s, t) \} = E \left\{ (I(X_i(t) = h) - \Pi_{1h}(s, t))^2 \mid X_i(s) = 1 \right\} \]

\[ = E \left\{ (I(X_i(t) = h) - \Pi_{1h}(s, t))^2 I(s < T_i \leq t) \mid X_i(s) = 1 \right\} \]

\[ + E \left\{ (I(X_i(t) = h) - \Pi_{1h}(s, t))^2 I(T_i > t) \mid X_i(s) = 1 \right\}. \]

For the case with random censoring we introduce the weights

\[ w_i(t \mid s) = \frac{I(s < \bar{T}_i \leq t, \bar{X}_i(t) \neq 0)}{G(\bar{T}_i \mid s)} + \frac{I(\bar{T}_i > t)}{G(t \mid s)}, \]

where

\[ G(t \mid s) = P(C_i > t \mid C_i > s) = \frac{G(t)}{G(s)}. \]

The Brier score is then

\[ BS_{1h}^w(s, t) = \frac{1}{\sum_{i=1}^n I(\bar{X}_i(s) = 1)} \sum_{i: X_i(s) = 1} (I(\bar{X}_i(t) = h) - \Pi_{1h}(s, t))^2 w_i(t \mid s). \quad (4.7) \]

The expectation of the Brier score (4.7), is equal to the expectation of (4.6). The derivation is given in Appendix A.3.
4.1. THE BRIER SCORE FOR OUR SETTING

Integrated Brier score

Above, we considered the Brier score for fixed time points. By looking at different \( t \) values, we may plot the Brier score as a function of time. This curve is then a prediction error curve, and such a curve could be made for each method we are considering. If a curve produced by one of the methods, is smaller than the others for all time points, we know that this method fits the data best. But the curves may cross each other, so that it is difficult to give an overall picture of which method is the best one, by just looking at the graphs. The integrated Brier score, with respect to some weight function \( W(t) \) over the interval \([s, t^*]\),

\[
IBS_{1h}^c(s, t^*) = \sum_{i=1}^{n} I(\tilde{X}_i(s) = 1) \int_s^{t^*} \left( I(\tilde{X}_i(t) = h) - \Pi_{1h}(s, t) \right)^2 w_i(t \mid s) dW(t)
\]

could then be used for comparison. A natural choice for the weight function is \( W(t) = t / (t^* - s) \). In our cases the Brier score only changes when transitions or censorings happen. Hence we may, when we insert estimates for the transition probabilities, write (4.8) as

\[
IBS_{1h}^c(s, t^*) = \sum_{i=1}^{n} I(\tilde{X}_i(s) = 1) \int_s^{t^*} \frac{1}{\sum_{i=1}^{n} I(\tilde{X}_i(s) = 1)} \left( I(\tilde{X}_i(t) = h) - \Pi_{1h}(s, t) \right)^2 w_i(t \mid s) dW(t)
\]

\[
= \int_s^{t^*} BS_{1h}^c(s, t) dW(t),
\]

(4.8)

Example 1.6: Brier score

For the CML data we have been considering in Example 1.1-1.5, we will now calculate the Brier score \( BS_{12}^c(s, t) \), for the three \( s \) values and for all \( t \) values where there are changes in the transition probabilities. In Figure 4.1, the prediction error curve for \( s = 0 \) is plotted. The green prediction error curve in the top plot shows the Brier scores when the transition probabilities are estimated by the Markov method. The red and the blue curve in the two bottom plots show the changes in Brier scores compared to the Markov Brier scores, when the semi-Markov method and the non-Markov method, respectively, have been used to estimate the transition probabilities. For the non-Markov Brier scores there are only small deviations from the Markov case. They are approximately equal. The semi-Markov Brier scores differ more, and they are larger than the Markov scores.
CHAPTER 4. COMPARISON OF THE METHODS ON REAL DATA

Figure 4.1: For the CML data: The top plot shows the Brier score for the probability to stay in state 2, for the Markov method when $s = 0$. The second plot shows the change in Brier score compared to for the Markov method, for the semi-Markov method. While the bottom plot shows this change for the non-Markov method compared to the Markov method. Note the different scales for the y-axes.

Brier scores for nearly all $t$ values. From this we learn that the Markov and the non-Markov methods, for $s = 0$, estimate transition probabilities that fit the data better, than the semi-Markov method does. We do not know the true transition probabilities, but on the basis of our available data there is reason to believe that the Markov and the non-Markov methods estimate transition probabilities closer to the true probabilities, than is the case for the semi-Markov method.

Going back to the first plot in Figure 3.1, where $\hat{P}_{12}(0, t)$ is calculated with the Markov method (green) and the semi-Markov method (red), we now have got a tool to conclude that the semi-Markov method overestimates the transition probability for $t \leq 3$, and underestimates for larger $t$ values.

In Table 4.1 we have given the integrated Brier scores. For $s = 0$, the difference between the Markov method and the non-Markov method is minimal, while the semi-Markov method gives a slightly higher value. In this case it was
Figure 4.2: For the CML data: The first and the third plot show the Brier scores for the Markov method when $s = 0.25$ and $s = 1.35$, respectively. The second and the fourth plot show the change in Brier scores compared to for the Markov method, for the semi-Markov method (red) and the non-Markov method (blue), for the two $s$ values considered. Note the different scales for the y-axes.
clear from the prediction error curves which of the methods that was the best, in this case it was two of them. In other cases the integrated Brier scores may be more useful than here.

In Figure 4.2, the Brier scores for $s = 0.25$ and $s = 1.35$ are plotted in a similar way as for $s = 0$, but now the changes for the semi-Markov method, and the non-Markov method, compared to the Markov method, are given in the same plots. For $s = 0.25$ we see that the red curve is a little higher than the two other around 1 year post transplant. Looking at the middle plot in Figure 3.1, we see that the red curve is a bit higher than the green one in this area, and we can conclude that the green curve fits the data better here. From 2.75 years onwards, the prediction error curve for the Markov method is clearly lower than the curve for the semi-Markov method, and by looking at Figure 3.1 it is hence reason to believe that the semi-Markov method underestimates the transition probabilities. For the non-Markov prediction error curve, we see that the Brier scores are a little bit higher than the Markov Brier scores from around 3 years post transplant. From Figure 3.2 we now conclude that the non-Markov method slightly underestimates the transition probabilities in this area. For $s = 1.35$, the Brier scores are lower, or equal, for the Markov method compared to the two other methods up to 6 years post transplant. For the period 6 to 7 years post transplant, the non-Markov prediction error curve is lower than the Markov curve for large parts. In Figure 3.2 we now know that in this area, the non-Markov method produces transition probabilities that fits the data better, for most of the $t$ values. Also the semi-Markov method produces transition probabilities that fit the data better, than is the case for the Markov method, for some of the $t$ values in this period. The integrated Brier score is smallest for the Markov method, and largest for the semi-Markov method.

From this analysis it is clear that the semi-Markov method is a bad method choice for the CML data. The two other methods are approximately equal, and it seems to be okey to use the Markov method. R-script for calculation of the Brier score is given in Appendix B.1. 

### 4.2 Another data example

In Example 1, which we have followed through Chapter 2-4, it was the case that transitions happened after some time. We will now give an example where most of the $1 \rightarrow 2$ transitions happen early.
4.2. ANOTHER DATA EXAMPLE

Figure 4.3: Number of individuals in each state at different times after transplant for the platelet data in Example 2. The black curve shows the number still in state 1, the red curve the number of patients who are in state 2, while the green curve is the number who have reached state 3.

Example 2

We will consider another data set from the European Registry for Blood and Marrow Transplantation, concerning platelet recovery and relapse after a transplant. The data includes information about 2204 patients, who had a marrow transplant in the period 1995-98, and is available from the mstate package in R, under the name ebmt3.

A transplant destroys the body’s ability to make platelets. Platelets are blood cells that help blood to clot, and a low platelet count may therefore lead to bleeding problems. According to the American Cancer Society (2015), the platelet counts are low for at least 3 weeks after transplant. For each patient, it was recorded if and when the platelets returned to a normal level. There could also be a relapse of the pre-transplant condition, or the patient could die. These two events are in the data considered as a joint, absorbing state.

Hence, we are still in the case where we are considering a progressive illness-death model. As long as no interesting event have happened to a patient after a transplant (no platelet recovery, no relapse, and he/she is still alive), the patient is in state 1. A transition to state 2 happens if and when the platelets are recovered. If the patient has a relapse or dies after the recovery, a transition from 2 to 3 happens. There could also be a direct transition from 1 to 3, meaning that
Figure 4.4: Estimated cumulative transition intensities for the platelet data. The left-hand plot shows the cumulative transition intensities the first five years after transplant. The right-hand plot shows the cumulative transition intensities the first year after transplant. The black curve is for transition $1 \rightarrow 2$, the red for transition $1 \rightarrow 3$, and the green for transition $2 \rightarrow 3$.

the patient has a relapse or dies without a foregoing platelet recovery. Some of the patients did not reach the absorbing state during the observation period, or disappeared from the study for some reason. They are censored.

Figure 4.3 shows how many who were in the different states at different times. We see that the number of individuals in state 2 (red curve) is increasing rapidly during the first 3-4 months. At later times the number is decreasing. More patients are then leaving state 2; they are having a relapse or die, or are being censored, than those who are entering the state. After 7 years, approximately all the patients are in state 3 (green curve), or they are censored. We will not consider the data for such a long period of time, because there will be few patients left in the non-absorbing states as time goes by. As we will see in Figure 4.5, the probability to be in state 1 is approximately 24% after five years, and we cut off our analysis here.

Figure 4.4 shows the cumulative transition intensities, estimated as in Section 2.4.1. We consider the curves for five years, and have also included a plot of the cumulative transition intensities during the first 0.2 year (73 days), to get a better idea of how the transition intensities are developing in this short time period. The black curve shows the cumulative transition intensity from state 1 to 2, the red curve from 1 to 3, and the green from 2 to 3. The first 10 days post transplant, the black curve, and hence also the green one, is approximately horizontal. The transition intensity is close to zero during this time. This is in accordance with the assumption saying that it will normally take at least 3 weeks before the platelets are at a normal level. When no one has entered state 2, it is
impossible to go from state 2 to 3. The transition intensity from 1 to 3 is not zero, but it is small. From 10 to 26 days post transplant, the 1 to 2 transition intensity is increasing. The two other transition intensities are increasing a bit; the 1 to 3 intensity more than that from 2 to 3. It seems natural that the risk of relapse or death without a foregoing platelet recovery, is larger than the risk of relapse or death after the platelets are back at a normal level. Further, up to around 10 months post transplant, the 1 to 2 intensity is decreasing, and from 10 months onwards, it is close to zero. From one year onwards, the transition intensities from 1 to 3, and from 2 to 3, are approximately equal, and they are much smaller than during the first year post transplant.

Figure 4.5 shows all the transition probabilities from state 1 for three $s$ values, estimated by the Markov method. We consider the state occupation probabilities ($s = 0$), but we are also interested in the probabilities of transition for individuals who are still in the initial state at later times. We choose $s = 0.06$ (22 days) and
Figure 4.6: A section of $\hat{P}_{12}(s,t)$ calculated by the Markov method (green), the semi-Markov method (red) and the non-Markov method (blue) for three $s$ values, for the platelet data. For $s = 0$, the green and the blue curves are coinciding. Note that the scales on the y-axes are different in the plots.

$s = 0.14$ (51 days), where the probability to be in state 1 is approximately 20\% and 50\%, respectively. Right after the transplant ($s = 0$), the probability to have had the platelets recovered and still be in that state, during the following 6 months is 46.1\%, CI: (44.0\%, 48.1\%). The probability to have a relapse, or to die, with or without a foregoing platelet recovery, is 18.9\%, CI: (17.2\%,20.5\%), during this 6 month period. The probability of no event is 35.1\%, CI: (33.1\%,37.1\%). If an individual has had no events during the first 3 weeks ($s = 0.06$) after the transplant, the probability to have had the platelets recovered and still be in that state, during the following 6 months, has decreased to 39.7\%, CI: (37.5\%,41.8\%). The probability of relapse or to die during the period is 20.0\%, CI: (18.3\%,21.8\%). The desired condition for a patient, is to enter state 2 and not leave it. Right after the transplant, the probability for this, during the following half a year, is 46.1\%, while for a patient who has had no event during the first three weeks, we see that this probability has decreased to 39.7\%. For $s = 0.14$, the probability to
4.2. ANOTHER DATA EXAMPLE

Table 4.2: For the platelet data: Estimated transition probabilities for being in state 2 in half a year, when the individual is in state 1 at time s, for our three methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\hat{P}_{12}(s, s+0.5)$</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>s = 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Markov</td>
<td>46.1%</td>
<td>(44.0%, 48.1%)</td>
</tr>
<tr>
<td>Non-Markov</td>
<td>46.1%</td>
<td>(44.0%, 48.2%)</td>
</tr>
<tr>
<td>Semi-Markov</td>
<td>44.1%</td>
<td></td>
</tr>
<tr>
<td>s = 0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Markov</td>
<td>39.7%</td>
<td>(37.5%, 41.8%)</td>
</tr>
<tr>
<td>Non-Markov</td>
<td>39.4%</td>
<td>(37.1%, 41.5%)</td>
</tr>
<tr>
<td>Semi-Markov</td>
<td>38.2%</td>
<td></td>
</tr>
<tr>
<td>s = 0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Markov</td>
<td>11.8%</td>
<td>(10.0%, 13.6%)</td>
</tr>
<tr>
<td>Non-Markov</td>
<td>12.5%</td>
<td>(10.6%, 14.4%)</td>
</tr>
<tr>
<td>Semi-Markov</td>
<td>11.5%</td>
<td></td>
</tr>
</tbody>
</table>

have had the platelets recovered and still be in that state during the following 6 months has fallen down to 11.8%, CI: (10.0%, 13.6%). The probability of relapse or to die has further increased a bit, and is now 21.6%, CI: (19.3%, 24.0%).

Figure 4.6 shows $\hat{P}_{12}(s, t)$ calculated by our three methods for the three $s$ values we are considering. The green curve is for the Markov method, the red for the semi-Markov method, and the blue for the non-Markov method. To capture the interesting parts of the curves, different scales on the y-axes are used. For $s = 0$, the red curve is a bit lower than the two other from around two months post transplant. Before this, the curves are approximately equal, but by zooming in it is possible to see that the red curve is a bit higher than the two other. The green and the blue curves are approximately coinciding. For $s = 0.06$, the curves are approximately coinciding the two following months, but again the red curve is a bit higher than the others. Further on, the green curve is higher than the two other, while the red and blue curves are crossing each other some places. For $s = 0.14$, the non-Markov method estimates the $1 \rightarrow 2$ transition probabilities highest, while the semi-Markov method estimates them lowest, for $t$ values from approximately half a year post the transplant. In Table 4.2 we have given Markov, non-Markov and semi-Markov estimates for those $1 \rightarrow 2$ transition probabilities that were given above for the Markov method. Here we see that the semi-Markov transition probabilities are a bit lower than the two other for all the $s$ values. For $s = 0$, the Markov and non-Markov estimates are the same. For $s = 0.06$, the estimate is largest for the Markov method, while it for $s = 0.14$ is largest for the non-Markov method. The bootstrap confidence intervals for the non-Markov method are a bit wider than the Markov confidence intervals where the Markov standard deviation estimator (2.47) is used. Bootstrap confidence intervals are not given for the semi-Markov method, due to the lack of software.

We will now use the Brier score to check which of the methods that produce transition probabilities closest to the observed state 2 occupation for the patients.
The upper plot in Figure 4.7 shows the prediction error curve for $s = 0$ for the Markov method, while the two other show the changes in Brier scores compared to the Markov Brier scores for the semi-Markov (red) and the non-Markov (blue) method. The Brier scores for the semi-Markov method are larger than those for the Markov method, hence we can say that the semi-Markov estimated probabilities in Figure 4.6 are too low relative to what we believe about the probabilities on the basis of our data. For the non-Markov Brier scores there are only small deviations, note the scale for the y-axis here. From about four years post transplant, the non-Markov Brier score is the smallest. For $t = 4.5$ for instance, we have that $\hat{P}_{12}(0, t) \approx 33.7\%$ for the semi-Markov method, $\hat{P}_{12}(0, t) \approx 34.4\%$ for the Markov method, and $\hat{P}_{12}(0, t) \approx 34.5\%$ for the non-Markov method. The non-Markov method gives the smallest Brier score here, but as we see, the Markov method and the non-Markov method
4.2. ANOTHER DATA EXAMPLE

Figure 4.8: For the platelet data: The first and the third plot show the Brier scores for the Markov method when $s = 0.06$ and $s = 0.14$, respectively. The second and the fourth plot show the change in Brier scores compared to for the Markov method, for the semi-Markov method (red) and the non-Markov method (blue), for the two $s$ values considered. Note the different scales for the y-axes.
give approximately the same transition probability. Looking at the integrated Brier scores in Table 4.3, we see that the scores for the Markov and the non-Markov method for \( s = 0 \) are the same. The differences are of order \( 10^{-6} \). The two first plots in Figure 4.8 show the Brier scores for \( s = 0.06 \) for the Markov method (green curve), and the changes in Brier scores for the two other methods compared to for the Markov method (red curve for the semi-Markov method, and blue curve for the non-Markov method). The probabilities estimated by the Markov method are closest to the data at approximately all times, while the two other methods give too low estimates. The integrated Brier score is smallest for the Markov method, and largest for the semi-Markov method. The two last plots in Figure 4.8 show the Brier scores for \( s = 0.14 \) for the Markov method, and the changes in Brier scores for the two other methods compared to for the Markov method. Now the Brier scores are smallest for the non-Markov method, and largest for the semi-Markov method.

Also for these data, it is clear that the semi-Markov method is not preferable. The differences for the Markov method and the non-Markov method for \( s = 0.14 \), may indicate that the Markov assumption is not fulfilled.

#### Cross-validation for the Brier score

In the Brier score (4.7), we use all individuals to estimate the state occupation probability for individual \( i \) at time \( t \). It is generally not a good idea to include individual \( i \) in such an estimation. Instead of (4.7), an alternative is to use leave-one-out cross-validation, and hence, for \( s = 0 \), consider

\[
\text{cvBS}^c_{12}(t) = \frac{1}{n} \sum_{i=1}^{n} \left( I(X_i(t) = 2) - \hat{P}^{(-i)}_{12}(0, t) \right)^2 w_i(t),
\tag{4.9}
\]

where \( \hat{P}^{(-i)}_{12}(0, t) \) is the transition probability estimated without individual \( i \). For each \( t \) value, \( n \) transition probabilities will have to be estimated, while the Brier score (4.7) just require one estimated transition probability. This will be a time consuming process, especially for those methods where fast software is not available, as is the case for the semi-Markov method. Since we are not including any covariates in our analysis, we do not expect each of the \( n \) individuals to have a large impact on the estimated transition probabilities, and hence not on the Brier score. We calculated the cross-validated Brier score (4.9) for some \( t \) values for the CML data considered in Example 1, by using the Markov method. The results, together with the Brier score (4.7) for these \( t \) values, are given in

---

**Table 4.3:** Integrated Brier scores for the platelet data where \( t^* = 5 \) years.

<table>
<thead>
<tr>
<th></th>
<th>Markov</th>
<th>Semi-Markov</th>
<th>Non-Markov</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s = 0 )</td>
<td>0.23396</td>
<td>0.23420</td>
<td>0.23396</td>
</tr>
<tr>
<td>( s = 0.06 )</td>
<td>0.22298</td>
<td>0.22315</td>
<td>0.22308</td>
</tr>
<tr>
<td>( s = 0.14 )</td>
<td>0.10282</td>
<td>0.10304</td>
<td>0.10267</td>
</tr>
</tbody>
</table>
4.2. ANOTHER DATA EXAMPLE

Table 4.4: Brier score and cross-validated Brier score for some chosen $t$ values for the Markov method. The data used are the CML data from Example 1.

<table>
<thead>
<tr>
<th>$t$</th>
<th>1</th>
<th>3</th>
<th>5</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$BS_{12}(0,t)$</td>
<td>0.089852</td>
<td>0.114759</td>
<td>0.123301</td>
<td>0.134425</td>
</tr>
<tr>
<td>$cvBS_{12}(0,t)$</td>
<td>0.089852</td>
<td>0.114753</td>
<td>0.123290</td>
<td>0.134416</td>
</tr>
<tr>
<td>abs. diff</td>
<td>$&lt; 10^{-6}$</td>
<td>$6 \cdot 10^{-6}$</td>
<td>$1.1 \cdot 10^{-5}$</td>
<td>$9 \cdot 10^{-6}$</td>
</tr>
</tbody>
</table>

Table 4.4. The differences for these $t$ values are of size $10^{-5}$ or less, and this is not much, relative to the Brier scores for the other methods. Hence it seems to be ok to use the Brier score (4.7).
CHAPTER 4. COMPARISON OF THE METHODS ON REAL DATA
Chapter 5

Comparison of the methods

We want to use simulation to compare the three transition probability methods presented in chapters 2 and 3. For this purpose we will generate data with known properties: Markov data, semi-Markov data and data that do not fulfill any of these requirements. We will then use our three methods on each of the data sets. We are considering transition probabilities from state 1, and because $P_{11}(s,t)$ is estimated in the same way for the methods, we will only consider estimation of $P_{12}(s,t)$ in this chapter.

When we have data from a known model we can compute the true transition probabilities, and we will choose models that make these computations not too hard. We may then compare the three different estimates with the true transition probability, for some chosen $t$ values. In Section 5.1 we will explain how to generate data, and how we will compare our three methods. In Section 5.2 we consider five different settings.

5.1 Setup for data generation and comparison

We may simulate a process by simulating potential times. For the illness-death model we will hence simulate the potential times $T_{12}$, $T_{13}$ and $T_{23}$. Further, we will let the shortest time of $T_{12}$ and $T_{13}$ be the potential sojourn time in state 1, and we have then decided if the process goes to state 2 or to state 3. If state 2 is reached, $T_{23}$ is the potential sojourn time in state 2. So far we haven’t included any censoring. We will do this by introducing a censoring process independent of the potential sojourn times. In this section we will show how data where the Markov assumption is fulfilled, may be generated. Necessary changes for the other settings will be mentioned when needed in Section 5.2.

Generating a Markov illness-death model

One method of generating data is to draw uniformly distributed numbers on the interval $[0,1]$, and then form the data we want. We let $T_{ij}$ be a potential time spent in state $i$ before transition to state $j$, with hazard $\alpha_{ij}(t)$. The survival
function is then expressed by
\[ S(t) = P(T_{ij} > t) = e^{-\int_0^t \alpha_{ij}(u)du} = e^{-A_{ij}(t)}. \]
We will set
\[ S(T_{ij}) = e^{-A_{ij}(T_{ij})} = U \sim \text{uniform}[0,1]. \]

\[ T_{ij} \text{ is then obtained by the transformation } T_{ij} = S^{-1}(U): \]
\[ e^{-A_{ij}(T_{ij})} = U \]
\[ A_{ij}(T_{ij}) = -\log U \]
\[ \rightarrow T_{ij} = A_{ij}^{-1}(-\log U). \quad (5.1) \]

To summarize, we can generate \( T_{ij} \) by first generating \( U \sim \text{uniform}[0,1] \), and then use the relation (5.1). We choose hazards for the three times of the Weibull form
\[ \alpha_{12}(u) = a_{12} u^{b_{12}}, \quad \alpha_{13}(u) = a_{13} u^{b_{13}} \quad \text{and} \quad \alpha_{23}(u) = a_{23} u^{b_{23}}. \quad (5.2) \]

For \( T_{12} \), the potential time spent in state 1 before transition to state 2, the calculation will now be
\[ A_{12}(T_{12}) = \int_0^{T_{12}} \alpha_{12}(u)du = \int_0^{T_{12}} a_{12} u^{b_{12}}du = \frac{a_{12}}{b_{12} + 1} T_{12}^{b_{12}+1} = -\log U \]
\[ \rightarrow T_{12} = \left(-\left(\frac{b_{12} + 1}{a_{12}} \log U\right)\right)^{1/(b_{12}+1)}. \quad (5.3) \]

Potential \( T_{13} \) is drawn in a similar way. For \( T_{23} \) we need to go the way through the global time \( T_3 \) when state 3 is reached, after first visiting state 2. That is
\[ \int_{T_{12}}^{T_3} \alpha_{23}(u)du = A_{23}(T_3) - A_{23}(T_{12}) = -\log U \]
\[ A_{23}(T_3) = -\log U + A_{23}(T_{12}) \]
\[ \frac{a_{23}}{b_{23} + 1} T_3^{b_{23}+1} = -\log U + \frac{a_{23}}{b_{23} + 1} T_{12}^{b_{23}+1} \]
\[ T_3^{b_{23}+1} = \frac{b_{23} + 1}{a_{23}} \left(-\log U + \frac{a_{23}}{b_{23} + 1} T_{12}^{b_{23}+1}\right) \]
\[ T_3 = \left(-\frac{b_{23} + 1}{a_{23}} \log U + T_{12}^{b_{23}+1}\right)^{1/(b_{23}+1)}. \quad (5.4) \]

Potential time spent in state 2 is then \( T_{23} = T_3 - T_{12} \). When the data are generated in this way, the process
\[
X(t) = \begin{cases} 
1 & \text{if } T_{12} > t \text{ and } T_{13} > t, \\
2 & \text{if } T_{12} \leq T_{13}, T_{12} \leq t \text{ and } T > t, \\
3 & \text{if } T_{13} \leq T_{12} \text{ and } T_{13} \leq t, \quad \text{or if } T_{12} \leq T_{13} \text{ and } T \leq t,
\end{cases} \quad (5.5)
\]
is the Markov illness-death process defined in Section 2.5. See Appendix A.4 for details.

**Observed sojourn times**

To simulate a simple censoring mechanism we will choose a uniform distribution, meaning that individuals have the same probability to be censored at any time. We will also choose a threshold value \( \tau \) when the observation will be stopped. This threshold value will be chosen such that about 20% of the individuals are left in state 1 at time \( \tau \). We let \( U_{\text{cens}} \sim \text{Unif}[0,a] \), where \( a \geq \tau \), and the censoring variable is then \( C = \min(U_{\text{cens}}, \tau) \).

The observed sojourn time in state 1 will now be \( T_1 = \min(T_{12}, T_{13}, C) \), and we introduce the indicator \( \delta = I(T_1 = T_{12}) \), for transition to state 2. If \( \delta = 1 \), state 2 is reached, and then the observed sojourn time in state 2 will be \( T_2 = \min(T_{23}, C - T_1) \). We let \( T \) be the total time before the process is reaching state 3 or is being censored. As an indicator for the time \( T \) to be a noncensored lifetime we introduce \( \eta = I(T \leq C) \). Table 5.1 shows an example of data simulated in this way. For the first line, \( T_1 = T_{12} \), since \( \delta = 1 \). We also have that \( \eta = 1 \), hence the total time \( T \) is a survival time. For the second line, \( T_1 = T_{12} \), while \( \eta = 0 \), and total time is 60. This is the threshold value \( \tau \), hence the uncensored life time was originally greater than 60, but was censored at \( \tau = 60 \).

For line 3, the process never leaves state 1, but is censored at 59.73. For line 4, \( T_1 = T_{13} \) and we have a direct transition to state 3. Since \( \eta = 1 \) we know that \( T_1 \) is not a censoring time.

**Comparison with the true transition probability**

We will use our three methods to estimate \( P_{12}(s,t) \). For comparison of the three methods we will be interested in:

- bias = \( E\{ \hat{P}_{12}(s,t) \} - P_{12}(s,t) \),
- relative bias (%) = \( \frac{\text{bias}}{P_{12}(s,t)} \times 100 \),
- variance = \( V(\hat{P}_{12}(s,t)) \),

**Table 5.1: Example of simulated data.**

<table>
<thead>
<tr>
<th></th>
<th>( T_1 )</th>
<th>( \delta )</th>
<th>( T_2 )</th>
<th>( T )</th>
<th>( \eta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13.57</td>
<td>1</td>
<td>31.50</td>
<td>45.07</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>20.27</td>
<td>1</td>
<td>39.73</td>
<td>60.00</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>59.73</td>
<td>0</td>
<td>0.00</td>
<td>59.73</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>12.46</td>
<td>0</td>
<td>0.00</td>
<td>12.46</td>
<td>1</td>
</tr>
</tbody>
</table>
mean squared error (MSE)  
\[ \text{MSE} = \mathbb{E}\left\{ (\hat{P}_{12}(s,t) - P_{12}(s,t))^2 \right\} = \text{bias}^2 + \text{variance}. \]

The bias is giving us the expected error produced by the estimator. A small bias is preferable, but it is also important to take the variance into account. It is ok with some bias, if the variance is not too large. MSE incorporates bias and variance. For comparison of estimators through MSE, one consider the estimator with the smallest MSE value as the best one. To estimate the quantities we will generate \( K \) data sets, each containing \( N \) ‘individuals’. We then have the estimators:

- \( \hat{\text{bias}} = \frac{1}{K} \sum_{k=1}^{K} \hat{P}_{12,k}(s,t) - P_{12}(s,t) \),
- \( \hat{\text{variance}} = \frac{1}{K-1} \sum_{k=1}^{K} (\hat{P}_{12,k}(s,t) - \frac{1}{K} \sum_{k=1}^{K} \hat{P}_{12,k}(s,t))^2 \),
- \( \hat{\text{MSE}} = \frac{1}{K} \sum_{k=1}^{K} (\hat{P}_{12,k}(s,t) - P_{12}(s,t))^2 \),

where \( \hat{P}_{12,k} \) is the estimated transition probability for the \( k \)th generated data set.

For each data set we use the three transition probability methods to compute, for a given \( s \), estimates of \( P_{12}(s,t) \) for a number of \( t \) values. We then estimate the bias, standard deviation (\( \sqrt{\text{variance}} \)) and MSE for each \( t \) value, for each of the methods.

To integrate the absolute bias, the variance and the MSE over \( t \) for each method, gives summaries of how good the methods are compared to the others. We are interested in estimates of

\[ \int_{s}^{T} |\text{bias}(t)| dt, \quad \int_{s}^{T} \text{var}(\hat{P}_{12}(s,t)) dt \quad \text{and} \quad \int_{s}^{T} \text{MSE}(\hat{P}_{12}(s,t)) dt. \] (5.6)

We will estimate the integrals by sums, where \( \Delta \) is the constant distance between successive \( t \) values. Hence, we are evaluating the bias, MSE and variance for each method at the \( t \) values: \( s, s + \Delta, s + 2\Delta, \ldots, \tau \). We are then adding the contributions times \( \Delta \), which gives us the estimates:

\[ \sum_{t=s}^{T} |\text{bias}(t)| \Delta, \sum_{t=s}^{T} \text{variance}(\hat{P}_{12}(s,t)) \Delta \text{ and } \sum_{t=s}^{T} \text{MSE}(\hat{P}_{12}(s,t)) \Delta. \]

For the Markov method we have the standard deviation estimator (2.47). We are interested in the expectation of this estimator:

\[ E\left\{ \hat{\sigma}_{\hat{P}_{12}(s,t)} \right\}, \] (5.7)

for comparison with the empirical standard deviation. To estimate (5.7), we use

\[ \frac{1}{K} \sum_{k=1}^{K} \hat{\sigma}_{\hat{P}_{12,k}(s,t)}', \]

where \( \hat{\sigma}_{\hat{P}_{12,k}(s,t)}' \) is the estimated standard deviation of \( \hat{P}_{12}(s,t) \) for the \( k \)th generated data set. We will also study the coverage probability of the standard
5.2. COMPARISON OF THE METHODS ON SOME DATA SETS

confidence interval (2.48) for $P_{12}(s, t)$). The coverage probability will be estimated by the proportion of the $K$ confidence intervals that contain the true transition probability $P_{12}(s, t)$. While mean in tables like Table 5.2 gives the mean of the $K$ transition probabilities estimated by the Markov method, and m.std in tables like Table 5.4 gives the mean of the $K$ standard deviations estimated by the standard deviation estimator, we are now considering the proportion of these $K$ confidence intervals that contain $P_{12}(s, t)$.

5.2 Comparison of the methods on some data sets

In the previous section we described a way of simulating data, and presented bias, variance/standard deviation and MSE as tools to compare our methods. An overall picture of how good a method is compared to another one, is achieved by the integrated measures (5.6). We will now look at some different scenarios, and will start by simulating Markovian data with constant hazards. Since the hazards then are time independent, the data will also turn out to be semi-Markovian. In the second setting we will let the hazards be functions of global time $t$, and hence simulate Markov data. Further on, we will simulate semi-Markov data by letting the hazards be functions of duration time $d$, before we generate some general data (not fulfilling the Markov or semi-Markov properties) by introducing frailty in the hazards (the term frailty will be explained in Section 5.2.4), and by looking at one other case.

In all the simulations in this chapter we have chosen $K = 1000$ and $N = 400$. The proportion censored is about 50%. For all the settings, we have chosen the constants so that at $\tau = 60$, we have $P_{11}(0, \tau) \approx 0.2$. Further we have chosen $a_{12} = 2 \cdot a_{13}$ and $b_{12} = b_{13} = b_{23}$, such that $2/3$ of the individuals will have the potential transitions $1 \rightarrow 2 \rightarrow 3$, while $1/3$ potentially will go $1 \rightarrow 3$. For each setting we are considering three $s$ values. In addition to consider the state occupation probabilities; $s = 0$, we are also interested in the transition probabilities for individuals who are still in state 1 at later time points. We choose to consider time points where the probability to be in state 1 is approximately 80% and 50%, that is $P_{11}(0, s) \approx 0.8$ and $P_{11}(0, s) \approx 0.5$. In some of our settings, the individuals leave state 1 early, and we get two low $s$ values in addition to $s = 0$, while in other settings it is the other way around. For each setting, a table, such as Table 5.2, is given. This table is giving us the relative bias, empirical standard deviation and MSE for $t = \lceil s \rceil_{10}, \lceil s \rceil_{10} + 10, \cdots, 60$, where $\lceil s \rceil_{10}$ is $s$ rounded up to the nearest tenth. The bias with empirical standard deviation for each method is showed in a figure, see Figure 5.1 as an example. This figure only gives the bias for up to six $t$ values. We have chosen fairly few $t$ values to be able to compare the three methods’ bias and standard deviation in one figure, in a good way. As a supplement we have also included a figure showing bias for every integer from $s$ to $\tau$, see Figure C.1 in Appendix C.
5.2.1 Setting 1: Markov data with constant hazards

We will start by simulating Markovian data with constant hazards. We use the Weibull hazards given in (5.2), and set the $b_{ij}$s equal to zero. The expression for the true transition probability $P_{12}(s,t)$ is given by (2.41), and with our chosen hazards we get

$$P_{12}(s,t) = \int_s^t P_{11}(s,u)\alpha_{12}(u)P_{22}(u,t)du$$

$$= \int_s^t \exp \left( -\frac{a_{12}}{b_{12}+1} u^{b_{12}+1} - \frac{a_{13}}{b_{13}+1} u^{b_{13}+1} + \frac{a_{12}}{b_{12}+1} s^{b_{12}+1} + \frac{a_{13}}{b_{13}+1} s^{b_{13}+1} \right) \times a_{12} u^{b_{12}} \exp \left( -\frac{a_{23}}{b_{23}+1} t^{b_{23}+1} + \frac{a_{23}}{b_{23}+1} u^{b_{23}+1} \right) du. \quad (5.8)$$

For this setting we will consider the $s$ values $s = 0$, $s = 8$, and $s = 23$. The results from using our three methods on the simulated data are given in Table 5.2, while a visualization of bias and standard deviation is given in Figure 5.1. In Table 5.3 the integrated measures are given. R-script for this setting is given in Appendix B.2.

For $s = 0$, all the methods are approximately unbiased, with relative biases ranging from $-0.58\%$ to $0.35\%$ among the $t$ values we are considering. The integrated absolute bias is smallest for the semi-Markov method and largest for the Markov method, but the differences are small. The pointwise standard deviations are smallest for the semi-Markov method. They are a little bit smaller for the Markov method, than for the non-Markov method for some of the $t$ values, but the differences are of size $10^{-4}$. For all the $t$ values considered, the pointwise MSE is smallest for the semi-Markov method, which is natural because the method gives the smallest standard deviations, and all the methods are approximately unbiased. For the two other methods, the MSE values are equal, or a bit lower for the Markov method. The integrated MSE values follow the same ranking.

For $s = 8$, all the methods are still performing well. By looking at Figure C.1, we see that the absolute bias for the semi-Markov method, is smaller than, or approximately equal to, the absolute biases for the two other methods for large parts of the interval $[0, 60]$. We find the same in the integrated absolute bias; it is clearly smallest for the semi-Markov method, while it is largest for the non-Markov method. The non-Markov method seems to be better in the tails than the two other methods. Here we get relative bias of $0.19\%$ for $t = 10$, and $0.15\%$ for $t = 60$ for the non-Markov method, and larger relative biases for the two other methods. The pointwise standard deviations are smallest for the semi-Markov method, but now we see that the standard deviations for the Markov method, differ more from those for the non-Markov method, than for $s = 0$. The standard deviations are largest for the non-Markov method. The pointwise MSE values are smallest for the semi-Markov method, and largest for the non-Markov method, and the integrated MSE values follow the same ranking.
Table 5.2: Results of the simulations with Markovian data with constant hazards. True is the true transition probability $P_{12}(s, t)$, mean is the mean of the $K = 1000$ estimated transition probabilities, rb % means relative bias given in percent. Std is the empirical standard deviation for the estimated transition probabilities and MSE is the mean squared error. Here we have used the constants $\tau = 60$, $a_{12} = 0.02$, $a_{13} = 0.01$, $a_{23} = 0.025$, $b_{12} = b_{13} = b_{23} = 0$, $a = 160$.

<table>
<thead>
<tr>
<th>s</th>
<th>t</th>
<th>true</th>
<th>Markov mean (rb %)</th>
<th>std</th>
<th>MSE-10^4</th>
<th>Semi-Markov mean (rb %)</th>
<th>std</th>
<th>MSE-10^4</th>
<th>Non-Markov mean (rb %)</th>
<th>std</th>
<th>MSE-10^4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10</td>
<td>0.1519</td>
<td>0.1525 (0.35)</td>
<td>0.0186</td>
<td>0.35</td>
<td>0.1523 (0.27)</td>
<td>0.0174</td>
<td>0.30</td>
<td>0.1525 (0.35)</td>
<td>0.0186</td>
<td>0.35</td>
</tr>
<tr>
<td>0</td>
<td>20</td>
<td>0.2309</td>
<td>0.2295 (-0.58)</td>
<td>0.0219</td>
<td>0.48</td>
<td>0.2298 (-0.45)</td>
<td>0.0199</td>
<td>0.40</td>
<td>0.2296 (-0.57)</td>
<td>0.0219</td>
<td>0.48</td>
</tr>
<tr>
<td>0</td>
<td>30</td>
<td>0.2632</td>
<td>0.2629 (-0.09)</td>
<td>0.0229</td>
<td>0.52</td>
<td>0.2629 (-0.11)</td>
<td>0.0210</td>
<td>0.44</td>
<td>0.2629 (-0.10)</td>
<td>0.0230</td>
<td>0.53</td>
</tr>
<tr>
<td>0</td>
<td>40</td>
<td>0.2667</td>
<td>0.2667 (-0.03)</td>
<td>0.0244</td>
<td>0.59</td>
<td>0.2666 (-0.04)</td>
<td>0.0219</td>
<td>0.48</td>
<td>0.2666 (-0.06)</td>
<td>0.0245</td>
<td>0.60</td>
</tr>
<tr>
<td>0</td>
<td>50</td>
<td>0.2535</td>
<td>0.2537 (0.07)</td>
<td>0.0248</td>
<td>0.61</td>
<td>0.2538 (0.12)</td>
<td>0.0233</td>
<td>0.54</td>
<td>0.2536 (0.04)</td>
<td>0.0249</td>
<td>0.62</td>
</tr>
<tr>
<td>0</td>
<td>60</td>
<td>0.2313</td>
<td>0.2320 (0.29)</td>
<td>0.0254</td>
<td>0.64</td>
<td>0.2321 (0.33)</td>
<td>0.0242</td>
<td>0.59</td>
<td>0.2319 (0.26)</td>
<td>0.0255</td>
<td>0.65</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>0.0379</td>
<td>0.0380 (0.46)</td>
<td>0.0109</td>
<td>0.12</td>
<td>0.0380 (0.43)</td>
<td>0.0109</td>
<td>0.12</td>
<td>0.0379 (0.19)</td>
<td>0.0109</td>
<td>0.12</td>
</tr>
<tr>
<td>8</td>
<td>20</td>
<td>0.1726</td>
<td>0.1708 (-1.03)</td>
<td>0.0211</td>
<td>0.45</td>
<td>0.1711 (-0.84)</td>
<td>0.0206</td>
<td>0.42</td>
<td>0.1707 (-1.08)</td>
<td>0.0220</td>
<td>0.49</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>0.2404</td>
<td>0.2399 (-0.19)</td>
<td>0.0237</td>
<td>0.56</td>
<td>0.2400 (-0.16)</td>
<td>0.0229</td>
<td>0.52</td>
<td>0.2398 (-0.23)</td>
<td>0.0253</td>
<td>0.64</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
<td>0.2657</td>
<td>0.2655 (-0.11)</td>
<td>0.0259</td>
<td>0.67</td>
<td>0.2653 (-0.17)</td>
<td>0.0241</td>
<td>0.58</td>
<td>0.2652 (-0.21)</td>
<td>0.0281</td>
<td>0.79</td>
</tr>
<tr>
<td>8</td>
<td>50</td>
<td>0.2651</td>
<td>0.2652 (0.02)</td>
<td>0.0264</td>
<td>0.70</td>
<td>0.2652 (0.02)</td>
<td>0.0247</td>
<td>0.61</td>
<td>0.2646 (-0.18)</td>
<td>0.0287</td>
<td>0.82</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>0.2496</td>
<td>0.2503 (0.28)</td>
<td>0.0276</td>
<td>0.76</td>
<td>0.2502 (0.26)</td>
<td>0.0250</td>
<td>0.63</td>
<td>0.2500 (0.15)</td>
<td>0.0302</td>
<td>0.91</td>
</tr>
<tr>
<td>23</td>
<td>30</td>
<td>0.1155</td>
<td>0.1160 (0.43)</td>
<td>0.0232</td>
<td>0.54</td>
<td>0.1160 (0.46)</td>
<td>0.0231</td>
<td>0.53</td>
<td>0.1160 (0.41)</td>
<td>0.0243</td>
<td>0.59</td>
</tr>
<tr>
<td>23</td>
<td>40</td>
<td>0.2131</td>
<td>0.2134 (0.15)</td>
<td>0.0307</td>
<td>0.94</td>
<td>0.2133 (0.09)</td>
<td>0.0300</td>
<td>0.90</td>
<td>0.2136 (0.24)</td>
<td>0.0338</td>
<td>1.14</td>
</tr>
<tr>
<td>23</td>
<td>50</td>
<td>0.2572</td>
<td>0.2577 (0.20)</td>
<td>0.0313</td>
<td>0.98</td>
<td>0.2575 (0.10)</td>
<td>0.0304</td>
<td>0.92</td>
<td>0.2575 (0.10)</td>
<td>0.0357</td>
<td>1.28</td>
</tr>
<tr>
<td>23</td>
<td>60</td>
<td>0.2679</td>
<td>0.2692 (0.50)</td>
<td>0.0329</td>
<td>1.08</td>
<td>0.2688 (0.35)</td>
<td>0.0296</td>
<td>0.87</td>
<td>0.2691 (0.45)</td>
<td>0.0382</td>
<td>1.46</td>
</tr>
</tbody>
</table>
For \( s = 23 \), there is a clear tendency of some overestimation, while it was underestimation for parts of the interval for the two other \( s \) values, but all the methods are still giving approximately unbiased estimates. For the three largest \( t \) values in Table 5.2, the semi-Markov method gives the smallest relative bias, while the non-Markov method gives the smallest for \( t = 30 \). We see this same picture for the integrated bias; it is smallest for the semi-Markov method, and largest for the Markov method. The pointwise standard deviations are smallest for the semi-Markov method, and largest for the non-Markov method. We see that the standard deviations have increased compared to for \( s = 0 \) and \( s = 8 \). Considering the integrated variance for the semi-Markov method, we see that this measure has increased more compared to for \( s = 8 \), than is the case for the Markov method. The pointwise MSE values are smallest for the semi-Markov method, and largest for the non-Markov method.
5.2. **COMPARISON OF THE METHODS ON SOME DATA SETS**

Table 5.3: *Estimated integrated measures for the Markovian data with constant hazards.*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0222</td>
<td>0.0215</td>
<td>0.0221</td>
<td>0.0290</td>
<td>0.0250</td>
<td>0.0292</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.0286</td>
<td>0.0267</td>
<td>0.0354</td>
<td>0.0286</td>
<td>0.0257</td>
<td>0.0331</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>0.0222</td>
<td>0.0186</td>
<td>0.0201</td>
<td>0.0298</td>
<td>0.0280</td>
<td>0.0369</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In this setting we have seen that the semi-Markov method is a little bit better than the Markov method. The non-Markov method makes it a little worse than the two other mainly because of larger standard deviations. However, the differences are small.

We continue by considering the standard deviations estimated by the Markov standard deviation estimator (2.47). Table 5.4 gives the mean of the K estimated standard deviations, which will be denoted method standard deviation (m.std), in addition to the empirical standard deviation (e.std) and the relative difference (r.diff), for our six t values. For s = 0, there is a mixture of over- and underestimation. For t = 20 the method standard deviation gives 0.0218, while the empirical standard deviation is 0.0219. This is a relative difference of only −0.33%. For t = 30 the method standard deviation gives 0.0234, while the empirical standard deviation is 0.0229. The relative difference is here 2.17%. And for t = 60, the method standard deviation gives 0.0246, while the empirical standard deviation is 0.0254. This is a relative difference of −2.97%. As we see, the differences are small, and we can say that the estimator performs well. Also

Table 5.4: *Markov data with constant hazards: m.std gives the mean of the K standard deviations estimated by the Markov standard deviation estimator. e.std gives the empirical standard deviation for the data. r.diff(%) = 100 · (m.std - e.std)/e.std.*

<table>
<thead>
<tr>
<th>t</th>
<th>s = 0 m.std</th>
<th>e.std</th>
<th>r.diff(%)</th>
<th>s = 8 m.std</th>
<th>e.std</th>
<th>r.diff(%)</th>
<th>s = 23 m.std</th>
<th>e.std</th>
<th>r.diff(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.0182</td>
<td>0.0186</td>
<td>-2.09</td>
<td>0.0108</td>
<td>0.0109</td>
<td>-0.69</td>
<td>0.0236</td>
<td>0.0232</td>
<td>1.53</td>
</tr>
<tr>
<td>20</td>
<td>0.0218</td>
<td>0.0219</td>
<td>-0.33</td>
<td>0.0212</td>
<td>0.0211</td>
<td>0.33</td>
<td>0.0295</td>
<td>0.0307</td>
<td>-3.88</td>
</tr>
<tr>
<td>30</td>
<td>0.0234</td>
<td>0.0229</td>
<td>2.17</td>
<td>0.0243</td>
<td>0.0237</td>
<td>2.54</td>
<td>0.0236</td>
<td>0.0232</td>
<td>1.53</td>
</tr>
<tr>
<td>40</td>
<td>0.0242</td>
<td>0.0244</td>
<td>-0.78</td>
<td>0.0256</td>
<td>0.0259</td>
<td>-0.94</td>
<td>0.0295</td>
<td>0.0307</td>
<td>-3.88</td>
</tr>
<tr>
<td>50</td>
<td>0.0245</td>
<td>0.0248</td>
<td>-0.85</td>
<td>0.0264</td>
<td>0.0264</td>
<td>-0.02</td>
<td>0.0314</td>
<td>0.0313</td>
<td>0.41</td>
</tr>
<tr>
<td>60</td>
<td>0.0246</td>
<td>0.0254</td>
<td>-2.97</td>
<td>0.0268</td>
<td>0.0276</td>
<td>-2.79</td>
<td>0.0323</td>
<td>0.0329</td>
<td>-1.67</td>
</tr>
</tbody>
</table>
Table 5.5: Estimated coverage probabilities for the Markov method. The data are the Markovian data with constant hazards.

<table>
<thead>
<tr>
<th>$t$</th>
<th>$s = 0$</th>
<th>$s = 8$</th>
<th>$s = 23$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.939</td>
<td>0.927</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.951</td>
<td>0.943</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.951</td>
<td>0.950</td>
<td>0.945</td>
</tr>
<tr>
<td>40</td>
<td>0.943</td>
<td>0.955</td>
<td>0.943</td>
</tr>
<tr>
<td>50</td>
<td>0.944</td>
<td>0.949</td>
<td>0.945</td>
</tr>
<tr>
<td>60</td>
<td>0.930</td>
<td>0.935</td>
<td>0.934</td>
</tr>
</tbody>
</table>

for $s = 8$ and $s = 23$, the estimator performs well. The relative differences vary from $-3.88\%$ to $2.54\%$.

The coverage probabilities in Table 5.5 are all close to 0.95. For all the $s, t$ values we are considering, the coverage probabilities are between 0.927 and 0.955. The small deviations are mainly due to uncertainty in the simulations.

### 5.2.2 Setting 2: Markov data with nonconstant hazards

We will now consider Markovian data with nonconstant hazards. We use the same setup as in setting 1, but now the $b_{ij}$s are not equal to zero. The expression for the true transition probability $P_{12}(s, t)$ is hence already given in (5.8). The data are no longer semi-Markovian. For this setting we are considering the $s$ values $s = 0$, $s = 30$ and $s = 46$. The results from using our three methods on the simulated data are given in Table 5.6, while the visualization of bias and standard deviation is given in Figure 5.2. Here we see that both the Markov and the non-Markov method give biases around zero for all the $s$ values. In Table 5.7 the integrated measures are given.

For $s = 0$, the results in Table 5.6 are virtually equal for the Markov and the non-Markov method. Looking at the integrated measures, we see though, that the Markov method gives a bit smaller biases and variances. The semi-Markov method gives smaller pointwise standard deviations than for the two other methods, but the biases are large. For $t = 30$, the relative bias is $-16.16\%$, while it, in contrast, is only 0.24\% for the two other methods. Looking at Figure C.2, we see that the method underestimates up to around $t = 55$, and overestimates for larger $t$ values. Although the standard deviations are small, the biases are too large, for the method to compete with the two others.

For $s = 30$, it is alternating which of the Markov and the non-Markov method that gives the smallest relative bias. Again, the integrated absolute bias is smallest for the Markov method. Now the pointwise standard deviations are smaller for the Markov method, than for the non-Markov method, but they are still smallest for the semi-Markov method. The semi-Markov method underestimates a bit up to $t = 48$. For larger $t$ values, the bias is positive and increases rapidly for increasing $t$ values. We see this clearly in Figure C.2. For $t = 50$, we see that, because of the small pointwise standard deviation, and the
Table 5.6: Results of the simulations with Markovian data with nonconstant hazards. True is the true transition probability $P_{12}(s, t)$, mean is the mean of the $K = 1000$ estimated transition probabilities, rb % means relative bias given in percent. Std is the empirical standard deviation for the estimated transition probabilities and MSE is the mean squared error. Here we have used the constants $\tau = 60, a_{12} = 1.4 \cdot 10^{-5}, a_{13} = 7 \cdot 10^{-6}, a_{23} = 2 \cdot 10^{-5}, b_{12} = b_{13} = b_{23} = 2, a = 560$.

<table>
<thead>
<tr>
<th>s</th>
<th>t</th>
<th>true</th>
<th>Markov</th>
<th>Semi-Markov</th>
<th>Non-Markov</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10</td>
<td>0.0046</td>
<td>0.0048 (2.50)</td>
<td>0.0044 (-5.77)</td>
<td>0.0048 (2.50)</td>
</tr>
<tr>
<td>0</td>
<td>20</td>
<td>0.0353</td>
<td>0.0356 (0.62)</td>
<td>0.0307 (-13.16)</td>
<td>0.0356 (0.62)</td>
</tr>
<tr>
<td>0</td>
<td>30</td>
<td>0.1048</td>
<td>0.1050 (0.24)</td>
<td>0.0878 (-16.16)</td>
<td>0.1050 (0.24)</td>
</tr>
<tr>
<td>0</td>
<td>40</td>
<td>0.1929</td>
<td>0.1926 (-0.15)</td>
<td>0.1646 (-14.64)</td>
<td>0.1926 (-0.15)</td>
</tr>
<tr>
<td>0</td>
<td>50</td>
<td>0.2483</td>
<td>0.2482 (-0.04)</td>
<td>0.2325 (-6.35)</td>
<td>0.2482 (-0.04)</td>
</tr>
<tr>
<td>0</td>
<td>60</td>
<td>0.2304</td>
<td>0.2302 (-0.11)</td>
<td>0.2595 (12.61)</td>
<td>0.2302 (-0.12)</td>
</tr>
<tr>
<td>30</td>
<td>40</td>
<td>0.1341</td>
<td>0.1335 (0.47)</td>
<td>0.1282 (-4.38)</td>
<td>0.1337 (-0.31)</td>
</tr>
<tr>
<td>30</td>
<td>50</td>
<td>0.2341</td>
<td>0.2339 (-0.11)</td>
<td>0.2365 (1.02)</td>
<td>0.2338 (-0.13)</td>
</tr>
<tr>
<td>30</td>
<td>60</td>
<td>0.2425</td>
<td>0.2421 (-0.15)</td>
<td>0.2873 (18.47)</td>
<td>0.2419 (-0.23)</td>
</tr>
<tr>
<td>46</td>
<td>50</td>
<td>0.1069</td>
<td>0.1073 (0.39)</td>
<td>0.1098 (2.73)</td>
<td>0.1072 (0.35)</td>
</tr>
<tr>
<td>46</td>
<td>60</td>
<td>0.2461</td>
<td>0.2460 (-0.04)</td>
<td>0.2898 (17.73)</td>
<td>0.2456 (-0.20)</td>
</tr>
</tbody>
</table>
Figure 5.2: The data are the Markovian data with nonconstant hazards. The points are showing the biases for the given $t$ values, while the line segments are showing one empirical standard deviation in each direction. The green points and lines are for the Markov method, the red for the semi-Markov method and the blue for the non-Markov method.

relatively small bias here, the pointwise MSE for the semi-Markov method is smaller than for the two other methods. For $t = 60$, the picture is completely different, hence we realize the importance of considering the estimator at lots of points. The integrated MSE value is smallest for the Markov method, and largest for the semi-Markov method. The standard deviations are a bit higher than for $s = 0$.

For $s = 46$, the integrated bias is a bit smaller for the Markov method than for the non-Markov method. Now, we have a case where the pointwise standard deviations in Table 5.6, are smaller for the Markov method, than for the semi-Markov method. Also the integrated variance is smaller for the Markov method, than for the semi-Markov method. The non-Markov method gives the largest standard deviations. The semi-Markov method gives a large bias. The method is overestimating the true transition probabilities for all $t$ values; for $t = 60$, the
5.2. COMPARISON OF THE METHODS ON SOME DATA SETS

Table 5.7: Estimated integrated measures for the Markovian data with nonconstant hazards.

<table>
<thead>
<tr>
<th>s = 0</th>
<th>Markov</th>
<th>Semi-Markov</th>
<th>Non-Markov</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrated abs. bias</td>
<td>0.0180</td>
<td>0.7228</td>
<td>0.0180</td>
</tr>
<tr>
<td>Integrated variance</td>
<td>0.0149</td>
<td>0.0115</td>
<td>0.0149</td>
</tr>
<tr>
<td>Integrated MSE</td>
<td>0.0148</td>
<td>0.0268</td>
<td>0.0149</td>
</tr>
<tr>
<td>s = 30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated abs. bias</td>
<td>0.0169</td>
<td>0.2864</td>
<td>0.0172</td>
</tr>
<tr>
<td>Integrated variance</td>
<td>0.0114</td>
<td>0.0104</td>
<td>0.0125</td>
</tr>
<tr>
<td>Integrated MSE</td>
<td>0.0114</td>
<td>0.0169</td>
<td>0.0125</td>
</tr>
<tr>
<td>s = 46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated abs. bias</td>
<td>0.0054</td>
<td>0.1965</td>
<td>0.0058</td>
</tr>
<tr>
<td>Integrated variance</td>
<td>0.0081</td>
<td>0.0086</td>
<td>0.0097</td>
</tr>
<tr>
<td>Integrated MSE</td>
<td>0.0081</td>
<td>0.0136</td>
<td>0.0097</td>
</tr>
</tbody>
</table>

relative bias is as large as 17.73%. This fact makes the method the worst, when integrated MSE is considered as a measure of goodness. The Markov method has the smallest integrated MSE value.

In this setting we have seen that the Markov method behaves good, and better than the two other methods. The non-Markov method also performs well, but it gives larger standard deviations than the Markov method. The semi-Markov method has problems with over- and underestimation, depending on the s and t value. For all the s values, the bias is rapidly increasing as t comes close to τ.

Table 5.8 shows that the standard deviation estimator performs well. There is a mixture of some over- and underestimation, but not much. The relative difference is as high as −6.28% for the combination s = 0, t = 10, but the true transition probability is here as small as 0.0046. We will then have few observations for calculation of the standard deviation, and this is a clear source of uncertainty.

Table 5.8: Markov data with nonconstant hazards: m.std gives the mean of the K standard deviations estimated by the Markov standard deviation estimator. e.std gives the empirical standard deviation for the data. r.diff(%) = 100 · (m.std - e.std)/e.std.

<table>
<thead>
<tr>
<th>t</th>
<th>s = 0</th>
<th>s = 30</th>
<th>s = 46</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m.std</td>
<td>e.std</td>
<td>r.diff(%)</td>
</tr>
<tr>
<td>10</td>
<td>0.0031</td>
<td>0.0033</td>
<td>-6.28</td>
</tr>
<tr>
<td>20</td>
<td>0.0093</td>
<td>0.0092</td>
<td>1.33</td>
</tr>
<tr>
<td>30</td>
<td>0.0156</td>
<td>0.0160</td>
<td>-2.90</td>
</tr>
<tr>
<td>40</td>
<td>0.0202</td>
<td>0.0200</td>
<td>1.06</td>
</tr>
<tr>
<td>50</td>
<td>0.0223</td>
<td>0.0219</td>
<td>1.94</td>
</tr>
<tr>
<td>60</td>
<td>0.0220</td>
<td>0.0229</td>
<td>-3.96</td>
</tr>
</tbody>
</table>
Table 5.9: Estimated coverage probabilities for the Markov method. The data are the Markovian data with nonconstant hazards.

<table>
<thead>
<tr>
<th>t</th>
<th>s = 0</th>
<th>s = 30</th>
<th>s = 46</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.869</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.930</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.937</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.949</td>
<td>0.955</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.954</td>
<td>0.949</td>
<td>0.944</td>
</tr>
<tr>
<td>60</td>
<td>0.938</td>
<td>0.939</td>
<td>0.940</td>
</tr>
</tbody>
</table>

The coverage probabilities in Table 5.9 are, except for the combination $s = 0$ and $t = 10$, close to 0.95. For $s = 0$, $t = 10$ the coverage probability is 0.869. Going back to Table 5.6, we see that the relative bias for the mean of the transition probabilities was 2.50%, and in Table 5.8 we see that the relative difference of the mean of the estimated standard deviations was $-6.28\%$. The intervals are a bit narrow, and they are located a bit too high, and this is resulting in this relatively low coverage probability.

5.2.3 Setting 3: Semi-Markov data

For the Markov case in setting 2, the hazards are functions of global time $t$. Now, we turn to a semi-Markov setting and let the hazards be functions of duration time $d$. For the three state illness-death model, this change in time scale has only an impact on the generation of $T_{23}$, the potential sojourn time in state 2, if state 2 is reached. $T_{12}$ and $T_{13}$ can be found as before, see (5.3), and since the global time when state 2 is reached, has no impact on $\alpha_{23}(d)$, $T_{23}$ can be found in the same way as $T_{12}$ and $T_{13}$. We use the Weibull hazards given in (5.2). The true transition probability is then given by

$$P_{12}(s,t) = \int_s^t P_{11}(s,u)\alpha_{12}(u)P_{22}(0,t-u)du$$

$$= \int_s^t \exp \left( - \frac{a_{12}}{b_{12} + 1} u^{b_{12}+1} - \frac{a_{13}}{b_{13} + 1} u^{b_{13}+1} + \frac{a_{12}}{b_{12} + 1} u^{b_{12}+1} + \frac{a_{13}}{b_{13} + 1} u^{b_{13}+1} \right)$$

$$\times a_{12}u^{b_{12}} \exp \left( - \frac{a_{23}}{b_{23} + 1} (t-u)^{b_{23}+1} \right) du.$$ 

For this setting we will consider the $s$ values $s = 0$, $s = 30$ and $s = 44$. The results from using our three methods on the simulated data are given in Table 5.10, while the visualization of bias and standard deviation is given in Figure 5.3. In Table 5.11 the integrated measures are given.

For $s = 0$, we see that all the methods give approximately unbiased estimates. For each of the $t$ values we are considering in Table 5.10, except for $t = 60$, the relative biases for the different methods have the same sign. Hence, the methods are following each other. The largest relative bias we observe among our 18 values, is 1.54%. The integrated absolute bias is smallest for the semi-Markov
Table 5.10: The results of the simulations with Semi-Markovian data. True is the true transition probability $P_{12}(s,t)$, mean is the mean of the $K=1000$ estimated transition probabilities, rb % means relative bias given in percent. Std is the empirical standard deviation for the estimated transition probabilities and MSE is the mean squared error. Here we have used the constants $\tau = 60$, $a_{12} = 1.6 \cdot 10^{-5}$, $a_{13} = 8 \cdot 10^{-6}$, $1.7 \cdot 10^{-3}$, $b_{12} = b_{13} = b_{23} = 2$, $a = 160$.

<table>
<thead>
<tr>
<th>s</th>
<th>t</th>
<th>true</th>
<th>Markov</th>
<th>Semi-Markov</th>
<th>Non-Markov</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>mean (rb %)</td>
<td>std</td>
<td>MSE-10^4</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>0.0052</td>
<td>0.0052 (0.94)</td>
<td>0.0036</td>
<td>0.01</td>
</tr>
<tr>
<td>0</td>
<td>20</td>
<td>0.0350</td>
<td>0.0351 (0.35)</td>
<td>0.0099</td>
<td>0.10</td>
</tr>
<tr>
<td>0</td>
<td>30</td>
<td>0.0888</td>
<td>0.0883 (-0.57)</td>
<td>0.0157</td>
<td>0.25</td>
</tr>
<tr>
<td>0</td>
<td>40</td>
<td>0.1426</td>
<td>0.1420 (-0.41)</td>
<td>0.0196</td>
<td>0.39</td>
</tr>
<tr>
<td>0</td>
<td>50</td>
<td>0.1647</td>
<td>0.1658 (0.68)</td>
<td>0.0218</td>
<td>0.48</td>
</tr>
<tr>
<td>0</td>
<td>60</td>
<td>0.1404</td>
<td>0.1404 (-0.02)</td>
<td>0.0215</td>
<td>0.46</td>
</tr>
<tr>
<td>30</td>
<td>40</td>
<td>0.1518</td>
<td>0.1229 (-19.02)</td>
<td>0.0210</td>
<td>1.28</td>
</tr>
<tr>
<td>30</td>
<td>50</td>
<td>0.2042</td>
<td>0.1828 (-10.45)</td>
<td>0.0253</td>
<td>1.10</td>
</tr>
<tr>
<td>30</td>
<td>60</td>
<td>0.1743</td>
<td>0.1656 (-4.96)</td>
<td>0.0254</td>
<td>0.72</td>
</tr>
<tr>
<td>44</td>
<td>50</td>
<td>0.1764</td>
<td>0.1428 (-19.06)</td>
<td>0.0289</td>
<td>1.97</td>
</tr>
<tr>
<td>44</td>
<td>60</td>
<td>0.2719</td>
<td>0.2085 (-23.32)</td>
<td>0.0338</td>
<td>5.16</td>
</tr>
</tbody>
</table>
method, and largest for the non-Markov method. The semi-Markov method gives the smallest pointwise standard deviations, while the non-Markov method gives the largest, but the differences are small. Because of this, and the fact that the biases are small, the pointwise MSE values are smallest for the semi-Markov method, and largest for the non-Markov method. We see the same ranking for the integrated measures. The integrated variance and the integrated MSE is smallest for the semi-Markov method and largest for the non-Markov method.

For \( s = 30 \), the Markov method fails, by underestimating the true transition probabilities. The relative bias comes up to \(-19.02\%\) among our three \( t \) values. We see that this refers approximately to the largest absolute bias for all \( t \leq \tau \), by looking at Figure C.3. Both the semi-Markov method and the non-Markov method perform well. We see that it is alternating which of the two methods that give the smallest relative biases. Looking at the integrated measures, we
see that the semi-Markov method has the smallest integrated absolute bias. The semi-Markov method gives smaller pointwise standard deviations than the non-Markov method. The pointwise standard deviations for the Markov method are between those for the two other methods, except for \( t = 40 \), where the standard deviation is smallest for the Markov method. The large biases for the Markov method make the pointwise MSE values for this method larger than for the two other methods. We find an exception for \( t = 60 \). Here the bias is not so large, and the pointwise MSE value turns out to be lower than for the non-Markov method. Considering the integrated MSE value we find that it is 0.0278 for the Markov method, and 0.0168 for the non-Markov method.

For \( s = 44 \), the Markov method gives more biased estimates than for \( s = 30 \). The relative bias is \(-23.32\%\) for \( t = 60 \), and from Figure C.3 we see that the bias is even larger for other \( t \) values, such as for \( t = 55 \). Both the semi-Markov and the non-Markov method are still approximately unbiased, and there are small differences; the integrated absolute biases are 0.0097 and 0.0102, respectively. For the Markov method this value is as large as 0.6918. The standard deviations are smaller for the semi-Markov method, than for the non-Markov method. But the standard deviations are even smaller for the Markov method. The integrated variances show the same ranking of the methods. Despite the small standard deviations for the Markov method, the large biases make the method bad. Both the pointwise MSE values and the integrated MSE are largest for the Markov method, while they are smallest for the semi-Markov method.

In this setting, the semi-Markov method works very well. For \( s = 0 \), we are looking at the state occupation probability, and then it may be shown that the Markov method works well (Datta and Satten 2001). Through our simulation in this setting, we see this to be true. For \( s > 0 \) the Markov method performs badly. The non-Markov method is approximately unbiased, but it gives larger standard deviations than the other methods.

In Table 5.12 we see that the standard deviation estimator (2.47) seems to
Table 5.12: Semi-Markov data: m.std gives the mean of the K standard deviations estimated by the Markov variance estimator. e.std gives the empirical standard deviation for the data. r.diff(%) = 100 \cdot (m.std - e.std.)/e.std.

<table>
<thead>
<tr>
<th>t</th>
<th>m.std</th>
<th>e.std</th>
<th>r.diff(%)</th>
<th>m.std</th>
<th>e.std</th>
<th>r.diff(%)</th>
<th>m.std</th>
<th>e.std</th>
<th>r.diff(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.0033</td>
<td>0.0036</td>
<td>-8.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.0095</td>
<td>0.0099</td>
<td>-3.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.0153</td>
<td>0.0157</td>
<td>-3.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.0196</td>
<td>0.0196</td>
<td>-0.23</td>
<td>0.0194</td>
<td>0.0210</td>
<td>-7.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.0217</td>
<td>0.0218</td>
<td>-0.41</td>
<td>0.0241</td>
<td>0.0253</td>
<td>-4.60</td>
<td>0.0267</td>
<td>0.0289</td>
<td>-7.77</td>
</tr>
<tr>
<td>60</td>
<td>0.0211</td>
<td>0.0215</td>
<td>-1.84</td>
<td>0.0246</td>
<td>0.0254</td>
<td>-2.99</td>
<td>0.0316</td>
<td>0.0338</td>
<td>-6.36</td>
</tr>
</tbody>
</table>

work well for \( s = 0 \). The relative difference here, for \( t = 10 \), is \(-8.43\%\), but as in setting 2, we have that the true transition probability is small, which is leading to few observations for calculation of the standard deviations. For \( s = 30 \) the method standard deviation seems to underestimate more than when the data are Markovian as in the two previous settings. For \( t = 40 \) the relative difference is \(-7.83\%\). As we see in Table 5.10, also the Markov method underestimates the true transition probability quite heavily here. For \( s = 44 \) we get relative differences of \(-7.77\%\) and \(-6.36\%\) for \( t = 50 \) and \( t = 60 \), respectively. From our results it is reason to believe that the standard deviation estimator does not work that well for \( s > 0 \).

For \( s = 0 \), the coverage probabilities in Table 5.13 are quite close to 0.95, except for \( t = 10 \). For this \( t \) value we have a similar case as for \( s = 0, t = 10 \) in setting 2. The confidence intervals are generally too narrow (m.std = 0.0033, while e.std = 0.0036), but now they are, in contrast to for setting 2, located in the right place (rb = 0.94\%). For \( s = 30 \) the coverage probabilities are far from 0.95 (not that much for \( t = 60 \)). For \( t = 40 \), the coverage probability is only 0.620. For this case, we have in Table 5.10 that the Markov method is clearly underestimating the true transition probability, and in Table 5.12 we see that the standard deviation estimator underestimates the standard deviation. The main reason for the low coverage probability is the biasedness of the Markov method. For \( s = 44, t = 60 \) the large underestimation of the Markov method (relative

Table 5.13: Estimated coverage probabilities for the Markov method. The data are the semi-Markovian data.

<table>
<thead>
<tr>
<th>t</th>
<th>s = 0</th>
<th>s = 30</th>
<th>s = 44</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.878</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.924</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.938</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.944</td>
<td>0.620</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.947</td>
<td>0.820</td>
<td>0.698</td>
</tr>
<tr>
<td>60</td>
<td>0.935</td>
<td>0.913</td>
<td>0.471</td>
</tr>
</tbody>
</table>
bias of \(-23.2\%\) results in a coverage probability of only 0.471.

### 5.2.4 Setting 4: Frailty non-Markov data

Individuals differ, and some are for instance more likely to get a specific disease than others. The reason for this is not always clear, and we denote them as more frail. Frailty is an unobserved heterogeneity, and each individual has its own frailty. To generate data with this property we will use the simple proportional frailty model:

\[
\alpha_{ij}(t \mid \Theta) = \Theta \cdot \alpha_{ij}(t). \tag{5.9}
\]

This model, (5.9), says that the individual hazard rate is the product of an individual quantity \(\Theta\) describing the frailty, and a basic rate \(\alpha_{ij}(t)\). For each individual, we use the same frailty for all the states. Data generated from this model will be non-Markovian. Now not only the current state, but also the individual frailty \(\Theta\) matter for future transitions.

We let the basic hazard rates be as in (5.2). Before we generate \(T_{12}, T_{13}\) and \(T_{23}\), we generate values of \(\Theta\); one for each individual. We choose \(\Theta \sim \text{Gam}(\frac{1}{\delta_f}, \frac{1}{\delta_f})\) since this choice makes the computations of expectations below, quite easy. For a given \(\theta\) we then have

\[
A_{12}(T_{12} \mid \theta) = \int_0^{T_{12}} \alpha_{12}(u \mid \theta) du = \int_0^{T_{12}} a_{12}\theta u^{b_{12}} du = \frac{a_{12}}{b_{12} + 1} \theta T_{12}^{b_{12} + 1} = -\log U
\]

\[
\rightarrow T_{12} = \left(-\left(\frac{b_{12} + 1}{a_{12}\theta} \cdot \log U\right)\right)^{1/(b_{12} + 1)},
\]

and similar for \(T_{13}\), while we use (5.4) for \(T_{23}\), but now the hazard is as in (5.9).

We set the \(b_{ij}\)'s to zero, and the transition probability will then be

\[
P_{12}(s,t) = \frac{P(X(t) = 2 \mid X(s) = 1)}{P(X(s) = 1)}
\]

\[
= \frac{E\{I(X(s) = 1, X(t) = 2)\}}{E\{I(X(s) = 1)\}}
\]

\[
= \frac{EE\{I(X(s) = 1, X(t) = 2) \mid \Theta\}}{EE\{I(X(s) = 1) \mid \Theta\}}
\]

\[
= \frac{E\{P_{11}(0,s \mid \Theta)P_{12}(s,t \mid \Theta)\}}{E\{P_{11}(0,s \mid \Theta)\}} \tag{5.10}
\]

\[
= \frac{E\left\{\frac{a_{12}}{a_{12} + a_{13} - a_{23}} \left(e^{-((a_{12} + a_{13} - a_{23})s + a_{23}t)\Theta} - e^{-(a_{12} + a_{13})t\Theta}\right)\right\}}{E\left\{e^{-(a_{12} + a_{13})s\Theta}\right\}}. \tag{5.11}
\]
See Appendix A.5 for the omitted calculations between (5.10) and (5.11). To calculate the expectations, we make use of the Laplace transform

\[ L(c) = E(e^{-c\Theta}) \]  

(5.12)

When \( \Theta \sim \text{Gam}(\frac{1}{\delta_f}, \frac{1}{\delta_f}) \), giving us that \( E(\Theta) = 1 \) and \( \text{Var}(\Theta) = \delta_f \), the Laplace transform (5.12) takes the simple form

\[
\begin{align*}
L(c) &= E \left\{ \exp(-c\Theta) \right\} = \int_0^\infty \exp(-c\theta) \frac{1/\delta_f}{\Gamma(1/\delta_f)} \theta^{1/\delta_f - 1} \exp \left( -\frac{\theta}{\delta_f} \right) d\theta \\
&= \frac{(1/\delta_f)^{1/\delta_f}}{\Gamma(1/\delta_f)} \int_0^\infty \theta^{1/\delta_f - 1} \exp \left( -\theta \left( \frac{\delta_c + 1}{\delta_f} \right) \right) d\theta \\
&= \left( \frac{1}{\delta_f} \right)^{1/\delta_f} \left( \frac{\delta_f}{\delta_f c + 1} \right)^{1/\delta_f} = (1 + \delta_f c)^{-1/\delta_f}.
\end{align*}
\]

The fourth equality follows since we are integrating a gamma density over its entire domain. The numerator of (5.10) is now

\[
E \left\{ P_{11}(0, s | \Theta) P_{12}(s, t | \Theta) \right\}
\]

\[
= \frac{a_{12}}{a_{12} + a_{13} - a_{23}} \left( E \left\{ e^{-((a_{12} + a_{13} - a_{23})s + a_{23}t)\Theta} \right\} - E \left\{ e^{-(a_{12} + a_{13})\Theta} \right\} \right)
\]

\[
= \frac{a_{12}}{a_{12} + a_{13} - a_{23}} (L(((a_{12} + a_{13} - a_{23})s + a_{23}t) - L((a_{12} + a_{13})t))
\]

\[
= \frac{a_{12}}{a_{12} + a_{13} - a_{23}} \left( \{1 + \delta_f((a_{12} + a_{13} - a_{23})s + a_{23}t)\}^{-1/\delta_f} - \{1 + \delta_f(a_{12} + a_{13}t)\}^{-1/\delta_f} \right),
\]

while the denominator is

\[
E \left\{ P_{11}(0, s | \Theta) \right\} = E \left\{ e^{-((a_{12} + a_{13})s)\Theta} \right\} = L((a_{12} + a_{13})s)
\]

\[
= \{1 + \delta_f(a_{12} + a_{13})s\}^{-1/\delta_f}.
\]

We will consider two cases. One where the frailty variance is \( \delta_f = 0.5 \), and one where \( \delta_f = 1.5 \). In setting 1, we have the case where the frailty is one for every individual.

For the \( \delta_f = 0.5 \) case, we consider the \( s \) values \( s = 0, s = 6 \) and \( s = 20 \). The results from using our three methods on the simulated data are given in Table 5.14, while the visualization of bias and standard deviation is given in Figure 5.4. In Table 5.15 the integrated measures are given.

For \( s = 0 \), both the Markov method and the non-Markov method are good. The differences in biases are small, but we see in Table 5.14, that for the four \( t \) values from \( t = 30 \) and upward, the relative bias is a bit lower for the non-Markov
Table 5.14: Results of the simulations with frailty data with $\delta_f = 0.5$. True is the true transition probability $P_{12}(s,t)$, mean is the mean of the $K = 1000$ estimated transition probabilities, rb % means relative bias given in percent. Std is the empirical standard deviation for the estimated transition probabilities and MSE is the mean squared error. Here we have used the constants $a_{12} = 0.028, a_{13} = 0.014, a_{23} = 0.033, b_{12} = b_{13} = b_{23} = 0, \tau = 60, a = 125, \delta_f = 0.5$.

<table>
<thead>
<tr>
<th>$s$</th>
<th>$t$</th>
<th>true</th>
<th>Markov mean (rb %)</th>
<th>Markov std</th>
<th>Markov MSE $10^{-3}$</th>
<th>Semi-Markov mean (rb %)</th>
<th>Semi-Markov std</th>
<th>Semi-Markov MSE $10^{-3}$</th>
<th>Non-Markov mean (rb %)</th>
<th>Non-Markov std</th>
<th>Non-Markov MSE $10^{-3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10</td>
<td>0.1673</td>
<td>0.1670 (-0.22)</td>
<td>0.0188</td>
<td>0.35</td>
<td>0.1737 (3.83)</td>
<td>0.0175</td>
<td>0.35</td>
<td>0.1669 (-0.23)</td>
<td>0.0188</td>
<td>0.35</td>
</tr>
<tr>
<td>0</td>
<td>20</td>
<td>0.2159</td>
<td>0.2158 (-0.02)</td>
<td>0.0220</td>
<td>0.48</td>
<td>0.2261 (4.73)</td>
<td>0.0197</td>
<td>0.49</td>
<td>0.2158 (-0.04)</td>
<td>0.0220</td>
<td>0.48</td>
</tr>
<tr>
<td>0</td>
<td>30</td>
<td>0.2210</td>
<td>0.2217 (0.29)</td>
<td>0.0238</td>
<td>0.57</td>
<td>0.2296 (3.90)</td>
<td>0.0214</td>
<td>0.53</td>
<td>0.2216 (0.26)</td>
<td>0.0238</td>
<td>0.57</td>
</tr>
<tr>
<td>0</td>
<td>40</td>
<td>0.2101</td>
<td>0.2107 (0.30)</td>
<td>0.0238</td>
<td>0.56</td>
<td>0.2129 (1.36)</td>
<td>0.0215</td>
<td>0.47</td>
<td>0.2106 (0.26)</td>
<td>0.0238</td>
<td>0.57</td>
</tr>
<tr>
<td>0</td>
<td>50</td>
<td>0.1938</td>
<td>0.1948 (0.50)</td>
<td>0.0237</td>
<td>0.56</td>
<td>0.1897 (-2.09)</td>
<td>0.0222</td>
<td>0.51</td>
<td>0.1947 (0.46)</td>
<td>0.0238</td>
<td>0.57</td>
</tr>
<tr>
<td>0</td>
<td>60</td>
<td>0.1765</td>
<td>0.1768 (0.18)</td>
<td>0.0243</td>
<td>0.59</td>
<td>0.1655 (-6.25)</td>
<td>0.0231</td>
<td>0.66</td>
<td>0.1767 (0.13)</td>
<td>0.0247</td>
<td>0.61</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>0.0820</td>
<td>0.0818 (-0.26)</td>
<td>0.0154</td>
<td>0.24</td>
<td>0.0828 (0.95)</td>
<td>0.0154</td>
<td>0.24</td>
<td>0.0819 (-0.05)</td>
<td>0.0160</td>
<td>0.25</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>0.1859</td>
<td>0.1852 (-0.38)</td>
<td>0.0219</td>
<td>0.48</td>
<td>0.1883 (1.30)</td>
<td>0.0207</td>
<td>0.44</td>
<td>0.1862 (0.17)</td>
<td>0.0238</td>
<td>0.57</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>0.2182</td>
<td>0.2177 (-0.20)</td>
<td>0.0249</td>
<td>0.62</td>
<td>0.2185 (0.15)</td>
<td>0.0227</td>
<td>0.51</td>
<td>0.2191 (0.42)</td>
<td>0.0267</td>
<td>0.71</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>0.2209</td>
<td>0.2202 (-0.32)</td>
<td>0.0256</td>
<td>0.65</td>
<td>0.2155 (-2.43)</td>
<td>0.0228</td>
<td>0.55</td>
<td>0.2222 (0.59)</td>
<td>0.0280</td>
<td>0.78</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>0.2114</td>
<td>0.2110 (-0.20)</td>
<td>0.0261</td>
<td>0.68</td>
<td>0.1989 (-5.92)</td>
<td>0.0233</td>
<td>0.70</td>
<td>0.2129 (0.69)</td>
<td>0.0287</td>
<td>0.83</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>0.1973</td>
<td>0.1961 (-0.59)</td>
<td>0.0273</td>
<td>0.75</td>
<td>0.1775 (-10.05)</td>
<td>0.0236</td>
<td>0.95</td>
<td>0.1979 (0.33)</td>
<td>0.0301</td>
<td>0.90</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td>0.1360</td>
<td>0.1354 (-0.41)</td>
<td>0.0259</td>
<td>0.67</td>
<td>0.1330 (-2.20)</td>
<td>0.0251</td>
<td>0.64</td>
<td>0.1365 (0.38)</td>
<td>0.0283</td>
<td>0.80</td>
</tr>
<tr>
<td>20</td>
<td>40</td>
<td>0.1955</td>
<td>0.1937 (-0.90)</td>
<td>0.0288</td>
<td>0.83</td>
<td>0.1845 (-5.63)</td>
<td>0.0268</td>
<td>0.84</td>
<td>0.1961 (0.32)</td>
<td>0.0326</td>
<td>1.07</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>0.2179</td>
<td>0.2158 (-0.96)</td>
<td>0.0313</td>
<td>0.98</td>
<td>0.1974 (-9.42)</td>
<td>0.0278</td>
<td>1.19</td>
<td>0.2187 (0.36)</td>
<td>0.0365</td>
<td>1.33</td>
</tr>
<tr>
<td>20</td>
<td>60</td>
<td>0.2218</td>
<td>0.2184 (-1.54)</td>
<td>0.0336</td>
<td>1.14</td>
<td>0.1913 (-13.76)</td>
<td>0.0275</td>
<td>1.69</td>
<td>0.2224 (0.27)</td>
<td>0.0391</td>
<td>1.53</td>
</tr>
</tbody>
</table>
method. We find this for the integrated absolute bias as well. This integrated value is a bit lower for the non-Markov method, than for the Markov method. The semi-Markov method overestimates up to \( t = 40 \), and underestimates for larger \( t \) values. The integrated absolute bias is more than 10 times as large as for the two other methods. The pointwise standard deviations are smallest for the semi-Markov method, and largest for the non-Markov method, but the standard deviations for the Markov method are approximately as for the non-Markov method. The integrated variances follow the same ranking. Because of the small standard deviations, the pointwise MSE values for the semi-Markov method compete with the MSE values for the two other methods, and the integrated MSE is actually smallest for the semi-Markov method.

For \( s = 6 \), the Markov method underestimates the true probabilities a bit, for all the \( t \) values we are considering in Table 5.14, while the non-Markov method overestimates them. The semi-Markov method seems to work well for low \( t \)
values, but for \( t > 30 \) the method underestimates the true probabilities. For these \( t \) values, the absolute bias increases linearly for increasing \( t \). The integrated bias is smallest for the Markov method, and largest for the semi-Markov method. The pointwise standard deviations are smallest for the semi-Markov method, and largest for the non-Markov method. The integrated variances follow the same ranking. For the two largest \( t \) values, the biases for the semi-Markov method are large enough for the pointwise MSE values to be larger for the semi-Markov method, than for the Markov method. But for the integrated MSE, this value is, again, smallest for the semi-Markov method, and largest for the non-Markov method.

For \( s = 20 \), the Markov and the non-Markov methods behave approximately as for \( s = 6 \). But now the Markov method underestimates a bit more, while the relative biases have decreased for the non-Markov method. Also the semi-Markov method is underestimating for all \( t \) values. For \( t = 60 \), the relative bias for this method is \(-13.76\%\), while it is \(-1.54\%\) and \(0.27\%\) for the Markov method and the the non-Markov method, respectively, at this \( t \) value. Inte-

Table 5.16: Frailty data with \( \delta_f = 0.5 \): \( m.\text{std} \) gives the mean of the \( K \) standard deviations estimated by the Markov variance estimator. \( e.\text{std} \) gives the empirical standard deviation for the data. \( r.\text{diff}(\%) = 100 \cdot (m.\text{std} - e.\text{std.})/e.\text{std.} \).

<table>
<thead>
<tr>
<th>( s = 0 )</th>
<th>Markov</th>
<th>Semi-Markov</th>
<th>Non-Markov</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrated abs. bias</td>
<td>0.0298</td>
<td>0.3638</td>
<td>0.0269</td>
</tr>
<tr>
<td>Integrated variance</td>
<td>0.0286</td>
<td>0.0244</td>
<td>0.0287</td>
</tr>
<tr>
<td>Integrated MSE</td>
<td>0.0286</td>
<td>0.0273</td>
<td>0.0287</td>
</tr>
<tr>
<td>( s = 6 )</td>
<td>Integrated abs. bias</td>
<td>0.0307</td>
<td>0.3120</td>
</tr>
<tr>
<td>Integrated variance</td>
<td>0.0300</td>
<td>0.0251</td>
<td>0.0355</td>
</tr>
<tr>
<td>Integrated MSE</td>
<td>0.0300</td>
<td>0.0287</td>
<td>0.0355</td>
</tr>
<tr>
<td>( s = 20 )</td>
<td>Integrated abs. bias</td>
<td>0.0659</td>
<td>0.4988</td>
</tr>
<tr>
<td>Integrated variance</td>
<td>0.0309</td>
<td>0.0255</td>
<td>0.0397</td>
</tr>
<tr>
<td>Integrated MSE</td>
<td>0.0310</td>
<td>0.0352</td>
<td>0.0397</td>
</tr>
</tbody>
</table>

\[ \text{Integrated abs. bias} = \frac{\text{Mean of absolute differences}}{\text{Mean of standard deviations}} \]
Table 5.17: Estimated coverage probabilities for the Markov method. The data are the frailty data with $\delta_f = 0.5$.

<table>
<thead>
<tr>
<th>t</th>
<th>s = 0</th>
<th>s = 6</th>
<th>s = 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.946</td>
<td>0.944</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.945</td>
<td>0.949</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.930</td>
<td>0.923</td>
<td>0.928</td>
</tr>
<tr>
<td>40</td>
<td>0.943</td>
<td>0.940</td>
<td>0.939</td>
</tr>
<tr>
<td>50</td>
<td>0.946</td>
<td>0.948</td>
<td>0.943</td>
</tr>
<tr>
<td>60</td>
<td>0.936</td>
<td>0.933</td>
<td>0.926</td>
</tr>
</tbody>
</table>

Grated absolute bias is smallest for the non-Markov method, and largest for the semi-Markov method. The semi-Markov method gives the smallest pointwise standard deviations, and the non-Markov method the largest. The integrated variances follow the same ranking. Integrated MSE is now smallest for the Markov method, and largest for the non-Markov method. Hence we see that the bias for the semi-Markov method is now too large, for this method to be the best one.

For $\delta_f = 0.5$ we have seen that both the Markov method, and the semi-Markov method, is better than the non-Markov method for all the $s$ values, when integrated MSE is considered for comparison. Thus the two methods perform quite well. Further, we have seen that the clearly biased semi-Markov method gives larger pointwise MSE values than the Markov method when the biases are large enough compared to the standard deviations.

In Table 5.16 we consider the goodness of the Markov standard deviation estimator. For $s = 0$ the relative difference for $t = 30$ is $-5.21\%$, but there does not seem to be a systematic underestimation of this magnitude for the other $t$ values. Hence this could be a result of uncertainty in the simulations. Also for $s = 6$ and $s = 20$, the relative differences are at an acceptable level. Hence it seems like the standard deviation estimator works quite well for all the $s$ values.

The coverage probabilities in Table 5.17 are all quite close to 0.95.

For the $\delta_f = 1.5$ case, we consider the $s$ values $s = 0$, $s = 2$ and $s = 10$. The results from using our three methods on the simulated data are given in Table 5.18, while the visualization of bias and standard deviation is given in Figure 5.5. In Table 5.19 the integrated measures are given. For $s = 0$, we see that the Markov and the non-Markov methods are approximately unbiased. The semi-Markov method is overestimating up to about $t = 40$, and underestimates for larger $t$ values. The integrated absolute bias is 0.0263 for the Markov method, and 0.0266 for the non-Markov method, while it is as large as 0.7294 for the semi-Markov method. The pointwise standard deviations in Table 5.18 are smallest for the semi-Markov method, except for $t = 60$, where it is smallest for the Markov method. The non-Markov method gives the largest pointwise standard deviations. The integrated variance is smallest for the semi-Markov method, and largest for the non-Markov method. Regarding the pointwise MSE values, they are approximately equal for the Markov method and the non-Markov method,
Table 5.18: Results of the simulations with Frailty data with $\delta_f = 1.5$. True is the true transition probability $P_{12}(s,t)$, mean is the mean of the $K = 1000$ estimated transition probabilities, rb % means relative bias given in percent. Std is the empirical standard deviation for the estimated transition probabilities and MSE is the mean squared error. Here we have used the constants $a_{12} = 0.08$, $a_{13} = 0.04$, $a_{23} = 0.025$, $b_{12} = b_{13} = b_{23} = 0$, $\tau = 60$, $a = 125$.

<table>
<thead>
<tr>
<th>s</th>
<th>t</th>
<th>true</th>
<th>Markov</th>
<th>Semi-Markov</th>
<th>Non-Markov</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>mean (rb %)</td>
<td>std</td>
<td>MSE-10^3</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>0.257</td>
<td>0.2570 (-0.04)</td>
<td>0.0222</td>
<td>0.51</td>
</tr>
<tr>
<td>0</td>
<td>20</td>
<td>0.2754</td>
<td>0.2752 (-0.07)</td>
<td>0.0233</td>
<td>0.58</td>
</tr>
<tr>
<td>0</td>
<td>30</td>
<td>0.2652</td>
<td>0.2654 (0.08)</td>
<td>0.0237</td>
<td>0.60</td>
</tr>
<tr>
<td>0</td>
<td>40</td>
<td>0.2501</td>
<td>0.2502 (0.04)</td>
<td>0.0239</td>
<td>0.57</td>
</tr>
<tr>
<td>0</td>
<td>50</td>
<td>0.2351</td>
<td>0.2360 (0.38)</td>
<td>0.0242</td>
<td>0.60</td>
</tr>
<tr>
<td>0</td>
<td>60</td>
<td>0.2213</td>
<td>0.2223 (0.42)</td>
<td>0.0245</td>
<td>0.64</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>0.2170</td>
<td>0.2120 (-2.28)</td>
<td>0.0229</td>
<td>0.55</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>0.2700</td>
<td>0.2598 (-3.77)</td>
<td>0.0250</td>
<td>0.73</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>0.2752</td>
<td>0.2627 (-4.52)</td>
<td>0.0257</td>
<td>0.81</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>0.2680</td>
<td>0.2540 (-5.22)</td>
<td>0.0252</td>
<td>0.83</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>0.2573</td>
<td>0.2436 (-5.32)</td>
<td>0.0260</td>
<td>0.86</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>0.2460</td>
<td>0.2321 (-5.65)</td>
<td>0.0270</td>
<td>0.92</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>0.1696</td>
<td>0.1619 (-4.54)</td>
<td>0.0273</td>
<td>0.80</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>0.2336</td>
<td>0.2167 (-7.21)</td>
<td>0.0296</td>
<td>1.16</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>0.2609</td>
<td>0.2356 (-9.72)</td>
<td>0.0300</td>
<td>1.54</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>0.2722</td>
<td>0.2423 (-11.00)</td>
<td>0.0316</td>
<td>1.89</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
<td>0.2756</td>
<td>0.2412 (-12.47)</td>
<td>0.0332</td>
<td>2.28</td>
</tr>
</tbody>
</table>
and these are smaller than for the semi-Markov method. For $t = 40$ we find an exception. Here the bias for the semi-Markov method is small, compared to for the other $t$ values, and the small standard deviation contributes to a small pointwise MSE value. The integrated MSE value is smallest for the Markov method and largest for the semi-Markov method. The value for the non-Markov method, is close to the value for the Markov method.

For $s = 2$, the Markov method clearly underestimates the true transition probabilities, but the semi-Markov method is more biased. The semi-Markov method overestimates a bit for small $t$ values, before it underestimates heavily. For $t = 60$ the relative bias is $-5.65\%$ for the Markov method, while it for the same $t$ value is $-17.48\%$ for the semi-Markov method. The non-Markov method is approximately unbiased. The pointwise standard deviations are smallest for the semi-Markov method, and largest for the non-Markov method. The integrated variances follow the same ranking. For the three first $t$ values in Table
5.2. COMPARISON OF THE METHODS ON SOME DATA SETS

Table 5.19: Estimated integrated measures for the frailty data with $\delta_f = 1.5$.

<table>
<thead>
<tr>
<th></th>
<th>Markov</th>
<th>Semi-Markov</th>
<th>Non-Markov</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s = 0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated abs. bias</td>
<td>0.0263</td>
<td>0.7294</td>
<td>0.0266</td>
</tr>
<tr>
<td>Integrated variance</td>
<td>0.0328</td>
<td>0.0301</td>
<td>0.0332</td>
</tr>
<tr>
<td>Integrated MSE</td>
<td>0.0328</td>
<td>0.0417</td>
<td>0.0332</td>
</tr>
<tr>
<td>$s = 2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated abs. bias</td>
<td>0.6203</td>
<td>0.8726</td>
<td>0.0248</td>
</tr>
<tr>
<td>Integrated variance</td>
<td>0.0352</td>
<td>0.0318</td>
<td>0.0395</td>
</tr>
<tr>
<td>Integrated MSE</td>
<td>0.0428</td>
<td>0.0555</td>
<td>0.0395</td>
</tr>
<tr>
<td>$s = 10$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated abs. bias</td>
<td>0.9710</td>
<td>1.8448</td>
<td>0.0462</td>
</tr>
<tr>
<td>Integrated variance</td>
<td>0.0418</td>
<td>0.0356</td>
<td>0.0539</td>
</tr>
<tr>
<td>Integrated MSE</td>
<td>0.0665</td>
<td>0.1342</td>
<td>0.0539</td>
</tr>
</tbody>
</table>

5.18, the pointwise MSE values are smallest for the semi-Markov method, due to the small standard deviations. For the three last $t$ values, the non-Markov method gives the smallest MSE values. The integrated MSE value is smallest for the non-Markov method, and largest for the semi-Markov method. Compared to how the semi-Markov method behaves overall, the Markov method is not that bad.

For $s = 10$, we see in Figure 5.5 that the Markov method and the semi-Markov methods perform worse than for $s = 2$. For the Markov method, the relative bias is now $-12.47\%$ for $t = 60$, while it is $-28.51\%$ for the semi-Markov method. The non-Markov method still performs well. The integrated absolute bias is only 0.0462 for the non-Markov method, while it is 0.9710 for the Markov method, and nearly twice that value for the semi-Markov method. The standard deviations are smallest for the semi-Markov method, and largest for the non-Markov method for the $t$ values we are considering. The integrated variances follow the same ranking. The pointwise MSE values for the Markov method remain quite low, but they are larger than for the non-Markov method, except

Table 5.20: Frailty data with $\delta_f = 1.5$: $m.\text{std}$ gives the mean of the $K$ standard deviations estimated by the Markov variance estimator. $e.\text{std}$ gives the empirical standard deviation for the data. $r.\text{diff}(\%) = 100 \cdot (m.\text{std} - e.\text{std})/e.\text{std}.$

<table>
<thead>
<tr>
<th>$t$</th>
<th>$s = 0$</th>
<th></th>
<th></th>
<th></th>
<th>$s = 2$</th>
<th></th>
<th></th>
<th></th>
<th>$s = 10$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m.\text{std}</td>
<td>e.\text{std}</td>
<td>r.\text{diff}(%)</td>
<td>m.\text{std}</td>
<td>e.\text{std}</td>
<td>r.\text{diff}(%)</td>
<td>m.\text{std}</td>
<td>e.\text{std}</td>
<td>r.\text{diff}(%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.0222</td>
<td>0.0225</td>
<td>-1.32</td>
<td>0.0221</td>
<td>0.0229</td>
<td>-3.42</td>
<td>0.0258</td>
<td>0.0273</td>
<td>-5.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.0233</td>
<td>0.0240</td>
<td>-3.05</td>
<td>0.0239</td>
<td>0.0250</td>
<td>-4.41</td>
<td>0.0285</td>
<td>0.0296</td>
<td>-3.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.0237</td>
<td>0.0244</td>
<td>-3.18</td>
<td>0.0246</td>
<td>0.0257</td>
<td>-4.21</td>
<td>0.0285</td>
<td>0.0296</td>
<td>-3.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.0239</td>
<td>0.0240</td>
<td>-0.27</td>
<td>0.0250</td>
<td>0.0252</td>
<td>-0.85</td>
<td>0.0295</td>
<td>0.0300</td>
<td>-1.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.0242</td>
<td>0.0244</td>
<td>-0.86</td>
<td>0.0255</td>
<td>0.0260</td>
<td>-1.68</td>
<td>0.0305</td>
<td>0.0316</td>
<td>-3.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>0.0245</td>
<td>0.0252</td>
<td>-2.79</td>
<td>0.0260</td>
<td>0.0270</td>
<td>-3.62</td>
<td>0.0313</td>
<td>0.0332</td>
<td>-5.70</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 5. COMPARISON OF THE METHODS

Table 5.21: Estimated coverage probabilities for the Markov method. The data are the frailty data with $\delta_f = 1.5$.

<table>
<thead>
<tr>
<th>t</th>
<th>s = 0</th>
<th>s = 2</th>
<th>s = 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.942</td>
<td>0.929</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.932</td>
<td>0.894</td>
<td>0.908</td>
</tr>
<tr>
<td>30</td>
<td>0.940</td>
<td>0.906</td>
<td>0.873</td>
</tr>
<tr>
<td>40</td>
<td>0.946</td>
<td>0.898</td>
<td>0.838</td>
</tr>
<tr>
<td>50</td>
<td>0.938</td>
<td>0.900</td>
<td>0.797</td>
</tr>
<tr>
<td>60</td>
<td>0.945</td>
<td>0.890</td>
<td>0.762</td>
</tr>
</tbody>
</table>

for small $t$ values. The integrated MSE is smallest for the non-Markov method. This measure has the value $0.0539$, while it is $0.0664$ for the Markov method. We see that the small standard deviations for the Markov method makes the method not that bad compared to the non-Markov method. The integrated MSE value for the semi-Markov method is about twice that for the Markov method.

For $\delta_f = 1.5$ we have seen that both the Markov method, and the semi-Markov method, is clearly biased. The exception is the Markov method for state occupation probabilities. The biases for these two methods are large enough for the non-Markov method to be the best method, when integrated MSE is considered for comparison.

In Table 5.20 we see that the Markov standard deviation estimator seems to work well for $s = 0$. There is some underestimation, but the relative differences are not larger than $-3.18\%$ for the $t$ values we are considering. For $s = 2$, the underestimation seems to be larger than for $s = 0$, but it is still not that large. For $s = 10$, we get relative differences of $-5.48\%$ and $-5.70\%$, for $t = 20$ and $t = 60$, respectively. Also for the other $t$ values the standard deviation estimator underestimates. We see that for $s > 0$, the standard deviation estimator seems to work less good for $\delta_f = 1.5$, than for $\delta_f = 0.5$.

For $s = 0$ the coverage probabilities in Table 5.21 are a bit lower than $0.95$, but still quite close; they vary from $0.932$ to $0.946$. For $s = 2$, we see that the biasedness of the Markov method makes the coverage probabilities stay around $0.9$. Also for $s = 10$ the coverage probabilities are small due to biasedness of the Markov method. For $t = 60$, only 762 of the 1000 standard confidence intervals include $P_{12}(10, 60)$.

5.2.5 Setting 5: Non-Markov Meira data

We can also draw the potential times directly. We follow Meira-Machado et al. (2006), hence the data will be denoted the Meira data, and say that $T_{12} \sim \exp(\lambda_{12})$, $T_{13} \sim \exp(\lambda_{13})$ and $T_{23} = 1.7 \times T_{12}$. The data do not fulfill the Markov or the semi-Markov assumptions. The true transition probability is now given
by
\[ P_{12}(s,t) = P(T_{12} \leq t, T_{12} \leq T_{13}, T_{12} + T_{23} > t \mid Z > s) = P(T_{12} \leq t, T_{12} \leq T_{13}, T_{12} > t/2.7 \mid \min(T_{12}, T_{13}) > s) \]
\[ = \frac{P(T_{12} \leq t, T_{12} \leq T_{13}, T_{12} > t/2.7, T_{12} > s, T_{13} > s)}{P(\min(T_{12}, T_{13}))} \]
\[ = \frac{P(s < T_{12} \leq t, T_{12} \leq T_{13}, T_{12} > t/2.7)}{P(\min(T_{12}, T_{13}))} \]
\[ = \frac{\lambda_{12}}{\lambda_{12} + \lambda_{13}} \left\{ \left( 1 - e^{-\left(\lambda_{12} + \lambda_{13}\right)(t-s)} \right) \right. \]
\[ + \left( e^{\left(\lambda_{12} + \lambda_{13}\right)s} \left( e^{-\left(\lambda_{12} + \lambda_{13}\right)t/2.7} - e^{-\left(\lambda_{12} + \lambda_{13}\right)t} \right) \right) \]
\[ \text{if } \frac{t}{2.70} < s \]
\[ \frac{e^{\left(\lambda_{12} + \lambda_{13}\right)s} \left( e^{-\left(\lambda_{12} + \lambda_{13}\right)t/2.7} - e^{-\left(\lambda_{12} + \lambda_{13}\right)t} \right)}{} \]
\[ \text{if } \frac{t}{2.70} \geq s, \]

where the omitted calculations are given in Appendix A.6.

For this setting we will consider the \( s \) values \( s = 0 \), \( s = 8 \) and \( s = 25 \). The results from using our three methods on the simulated data are given in Table 5.22, while the visualization of bias and standard deviation is given in Figure 5.6. In Table 5.23 the integrated measures are given.

For \( s = 0 \), both the Markov method and the non-Markov method perform well, and the results in Table 5.22 are approximately equal. Considering just these two methods, the largest relative bias among our six \( t \) values, is 1.07%. For the two first \( t \) values, the relative biases are a bit larger for the Markov method, than for the non-Markov method, while it is the other way around for the other four \( t \) values. Overall, the integrated bias is smallest for the Markov method. The pointwise standard deviations are a bit smaller for the Markov method, than for the non-Markov method, hence the pointwise MSE values are approximately equal, or a bit lower for the Markov method. The semi-Markov method is very biased, and we see clearly in Figure C.6 how the bias evolves for increasing \( t \) values. The method overestimates for \( t \) values up to 40, and underestimates for \( t > 40 \). For \( t = 10 \), the relative bias is as large as 49.16%. For the three first \( t \) values in Table 5.22, the pointwise standard deviations for the semi-Markov method are approximately as for the two other methods, or a bit smaller, while they are larger for the three other \( t \) values. The integrated variance is largest for the semi-Markov method. The Markov method gives the smallest integrated MSE value, and the semi-Markov method, definitively, the largest.

For \( s = 8 \), also the Markov method fails. We see in Figure C.6 that the method underestimates the transition probabilities for all \( t \) values. For \( t = 20 \), the absolute bias is approximately at it largest, and here the relative bias is \(-22.80\%\). In terms of bias, the semi-Markov method is, compared to the Markov method, better for small \( t \) values (up to \( t = 40 \)), and worse for large \( t \) values. The non-Markov method is almost unbiased. The integrated bias is smallest for the non-Markov method, and largest for the semi-Markov method. The pointwise standard deviations are largest for the non-Markov method. For the two other methods, it is changing with the value of \( t \), which method that gives the smallest standard deviation. Overall, the semi-Markov method gives
Table 5.22: Results of the simulations with Non-Markovian Meira data. True is the true transition probability $P_{12}(s,t)$, mean is the mean of the $K = 1000$ estimated transition probabilities, rb % means relative bias given in percent. Std is the empirical standard deviation for the estimated transition probabilities and MSE is the mean squared error. Here we have used the constants $\lambda_{12} = 0.02$, $\lambda_{13} = 0.01$, $\tau = 60$, $a = 135$.

<table>
<thead>
<tr>
<th>$s$</th>
<th>$t$</th>
<th>true</th>
<th>Markov</th>
<th>Semi-Markov</th>
<th>Non-Markov</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>mean (rb %)</td>
<td>std</td>
<td>MSE-10^4</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>0.0943</td>
<td>0.0953 (1.07)</td>
<td>0.0148</td>
<td>0.22</td>
</tr>
<tr>
<td>0</td>
<td>20</td>
<td>0.1573</td>
<td>0.1574 (0.02)</td>
<td>0.0186</td>
<td>0.35</td>
</tr>
<tr>
<td>0</td>
<td>30</td>
<td>0.1973</td>
<td>0.1964 (-0.48)</td>
<td>0.0209</td>
<td>0.44</td>
</tr>
<tr>
<td>0</td>
<td>40</td>
<td>0.2205</td>
<td>0.2196 (-0.39)</td>
<td>0.0229</td>
<td>0.53</td>
</tr>
<tr>
<td>0</td>
<td>50</td>
<td>0.2315</td>
<td>0.2309 (-0.29)</td>
<td>0.0237</td>
<td>0.56</td>
</tr>
<tr>
<td>0</td>
<td>60</td>
<td>0.2339</td>
<td>0.2332 (-0.33)</td>
<td>0.0245</td>
<td>0.60</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>0.0350</td>
<td>0.0334 (-4.80)</td>
<td>0.0100</td>
<td>0.10</td>
</tr>
<tr>
<td>8</td>
<td>20</td>
<td>0.1845</td>
<td>0.1424 (-22.80)</td>
<td>0.0211</td>
<td>2.21</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>0.2449</td>
<td>0.2047 (-16.42)</td>
<td>0.0243</td>
<td>0.21</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
<td>0.2736</td>
<td>0.2414 (-11.77)</td>
<td>0.0265</td>
<td>1.74</td>
</tr>
<tr>
<td>8</td>
<td>50</td>
<td>0.2873</td>
<td>0.2608 (-9.24)</td>
<td>0.0274</td>
<td>1.45</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>0.2903</td>
<td>0.2676 (-7.84)</td>
<td>0.0285</td>
<td>1.33</td>
</tr>
<tr>
<td>25</td>
<td>30</td>
<td>0.0842</td>
<td>0.0788 (-6.46)</td>
<td>0.0204</td>
<td>0.44</td>
</tr>
<tr>
<td>25</td>
<td>40</td>
<td>0.2220</td>
<td>0.1863 (-16.07)</td>
<td>0.0302</td>
<td>2.18</td>
</tr>
<tr>
<td>25</td>
<td>50</td>
<td>0.3272</td>
<td>0.2506 (-23.42)</td>
<td>0.0337</td>
<td>7.01</td>
</tr>
<tr>
<td>25</td>
<td>60</td>
<td>0.4075</td>
<td>0.2860 (-29.82)</td>
<td>0.0368</td>
<td>16.12</td>
</tr>
</tbody>
</table>
5.2. COMPARISON OF THE METHODS ON SOME DATA SETS

Figure 5.6: The data are the non-Markovian Meira data. The points are showing the biases for the given $t$ values, while the line segments are showing one empirical standard deviation in each direction. The green points and lines are for the Markov method, the red for the semi-Markov method and the blue for the non-Markov method.

The smallest integrated variance, and the non-Markov method the largest. The integrated MSE value is smallest for the non-Markov method, and largest for the semi-Markov method.

For $s = 25$, both the Markov and semi-Markov method are clearly underestimating the transition probabilities. The relative bias at $t = 60$, is $-29.82\%$ for the Markov method, and $-41.41\%$ for the semi-Markov method. For the non-Markov method, the relative bias is only $-0.43\%$ at this $t$ value. This method is approximately unbiased for all the $t$ values. Regarding the pointwise standard deviations, the semi-Markov method gives the smallest, while the non-Markov method gives the largest. Because of the large biases, the pointwise MSE values are much smaller for the non-Markov method than for the two other methods, except for $t = 30$. The Markov method performs better than the semi-Markov method from about $t > 40$. The integrated MSE value is smallest for the non-Markov method, and largest for the semi-Markov method.
Table 5.23: Estimated integrated measures for the non-Markovian Meira data.

<table>
<thead>
<tr>
<th></th>
<th>Markov</th>
<th>Semi-Markov</th>
<th>Non-Markov</th>
</tr>
</thead>
<tbody>
<tr>
<td>s = 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated abs. bias</td>
<td>0.0373</td>
<td>2.5949</td>
<td>0.0392</td>
</tr>
<tr>
<td>Integrated variance</td>
<td>0.0245</td>
<td>0.0270</td>
<td>0.0249</td>
</tr>
<tr>
<td>Integrated MSE</td>
<td>0.0245</td>
<td>0.1640</td>
<td>0.0249</td>
</tr>
<tr>
<td>s = 8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated abs. bias</td>
<td>1.5478</td>
<td>2.0978</td>
<td>0.0378</td>
</tr>
<tr>
<td>Integrated variance</td>
<td>0.0302</td>
<td>0.0288</td>
<td>0.0330</td>
</tr>
<tr>
<td>Integrated MSE</td>
<td>0.0836</td>
<td>0.1706</td>
<td>0.0330</td>
</tr>
<tr>
<td>s = 25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated abs. bias</td>
<td>1.7602</td>
<td>2.1378</td>
<td>0.0210</td>
</tr>
<tr>
<td>Integrated variance</td>
<td>0.0312</td>
<td>0.0267</td>
<td>0.0361</td>
</tr>
<tr>
<td>Integrated MSE</td>
<td>0.1686</td>
<td>0.2471</td>
<td>0.0361</td>
</tr>
</tbody>
</table>

In this setting we have seen that the Markov method, and the non-Markov method, works well for $s = 0$. For the two other $s$ values, only the non-Markov method works well. Because of smaller bias, the Markov method works better than the semi-Markov method.

In Table 5.24, we see that the standard deviation estimator works well for the three lowest $t$ values for $s = 0$. Here we have relative differences between $-0.43\%$ and $0.46\%$. Also for the other $t$ values the relative differences are at acceptable levels. For $s = 8$, the relative differences from $t = 20$ onwards, are between $-5.85\%$ and $-9.38\%$ for the $t$ values we are considering. For $s = 25$, the relative differences from $t = 40$ onwards, are between $-8.35\%$ and $-12.16\%$. Hence, the standard deviation estimator does not seem to work that well for $s > 0$.

For $s = 0$, the coverage probabilities in Table 5.25 are quite close to 0.95. For $s = 8$ and $s = 25$ they are much lower, mainly due to heavily biasedness of the Markov method. For $s = 25$, $t = 60$ the estimated coverage probability is only 0.072.

Table 5.24: Non-Markovian Meira data: $m.std$ gives the mean of the $K$ standard deviations estimated by the Markov variance estimator. $e.std$ gives the empirical standard deviation for the data. $r.diff(\%) = 100 \cdot (m.std - e.std)/e.std$. 

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>s = 0</th>
<th></th>
<th></th>
<th>s = 8</th>
<th></th>
<th></th>
<th>s = 25</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>$m.std$</td>
<td>$e.std$</td>
<td>$r.diff(%)$</td>
<td>$m.std$</td>
<td>$e.std$</td>
<td>$r.diff(%)$</td>
<td>$m.std$</td>
<td>$e.std$</td>
<td>$r.diff(%)$</td>
</tr>
<tr>
<td>10</td>
<td>0.0148</td>
<td>0.0148</td>
<td>-0.43</td>
<td>0.0098</td>
<td>0.0100</td>
<td>-1.40</td>
<td>0.0194</td>
<td>0.0204</td>
<td>-4.66</td>
</tr>
<tr>
<td>20</td>
<td>0.0187</td>
<td>0.0186</td>
<td>0.46</td>
<td>0.0191</td>
<td>0.0211</td>
<td>-9.38</td>
<td>0.0277</td>
<td>0.0302</td>
<td>-8.35</td>
</tr>
<tr>
<td>30</td>
<td>0.0208</td>
<td>0.0209</td>
<td>-0.32</td>
<td>0.0226</td>
<td>0.0243</td>
<td>-6.97</td>
<td>0.0307</td>
<td>0.0337</td>
<td>-8.91</td>
</tr>
<tr>
<td>40</td>
<td>0.0221</td>
<td>0.0229</td>
<td>-3.82</td>
<td>0.0245</td>
<td>0.0265</td>
<td>-7.29</td>
<td>0.0327</td>
<td>0.0368</td>
<td>-12.16</td>
</tr>
<tr>
<td>50</td>
<td>0.0229</td>
<td>0.0237</td>
<td>-3.32</td>
<td>0.0258</td>
<td>0.0274</td>
<td>-5.85</td>
<td>0.0307</td>
<td>0.0337</td>
<td>-8.91</td>
</tr>
<tr>
<td>60</td>
<td>0.0234</td>
<td>0.0245</td>
<td>-4.47</td>
<td>0.0266</td>
<td>0.0285</td>
<td>-6.60</td>
<td>0.0323</td>
<td>0.0368</td>
<td>-12.16</td>
</tr>
</tbody>
</table>
5.3. SUMMARY OF THE SIMULATION RESULTS

Table 5.25: Estimated coverage probabilities for the Markov method. The data are the non-Markovian Meira data.

<table>
<thead>
<tr>
<th>t</th>
<th>s = 0</th>
<th>s = 8</th>
<th>s = 25</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.940</td>
<td>0.914</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.950</td>
<td>0.403</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.943</td>
<td>0.542</td>
<td>0.899</td>
</tr>
<tr>
<td>40</td>
<td>0.938</td>
<td>0.702</td>
<td>0.717</td>
</tr>
<tr>
<td>50</td>
<td>0.940</td>
<td>0.791</td>
<td>0.337</td>
</tr>
<tr>
<td>60</td>
<td>0.935</td>
<td>0.826</td>
<td>0.072</td>
</tr>
</tbody>
</table>

5.3 Summary of the simulation results

The Markov method

For all the cases we have been considering, the Markov method performs well for $s = 0$. Hence the Markov method is good on estimating state occupation probabilities. This is in agreement with the result of Datta and Satten (2001).

The method is quite robust. In setting 4, we saw that the method tolerate small deviations from the Markov assumption. When the frailty variance $\delta_f$ was 0.5, all the pointwise MSE values were smaller (or equal) for the Markov method, than for the non-Markov method, for all the $s$ values considered. When the deviations are larger, such as when $\delta_f = 1.5$, we see that the Markov method performs worse than the non-Markov method, but the difference is not that large.

For $s > 0$, our results show that the method is bad in handling non-Markovian data, but not that bad for setting 4 mentioned above. In all the cases where the Markov assumption was not fulfilled, that is, setting 3-5, the method underestimates the true $1 \rightarrow 2$ transition probabilities to a greater or lesser extent.

The semi-Markov method

For all the settings where the semi-Markov assumption is not fulfilled, the method gives clearly biased estimates. For our settings, it was the case that when transitions to state 2 happened quite early, the semi-Markov method overestimated the true transition probabilities up to some $t$ value, for then to underestimate for larger $t$ values. We see this in Figure C.4, Figure C.5 and Figure C.6 in Appendix C. When transitions happened later, such as in setting 2, we had the opposite case, visualized in Figure C.2.

For most of the cases we have been considering, the semi-Markov method gives smaller standard deviations than the two other methods. But when the bias is large, this does not help much. For the Markovian data with constant hazards (setting 1), this method performs better than the others, mainly because of smaller standard deviations, but also the biases are generally lower for this method. In these data, also the semi-Markov assumption is fulfilled. The semi-
Markov method gives smaller variation than the Markov method here. We find the reason for this in the calculation of \( \hat{P}_{22}^*(0, v) \) (3.1) and \( \hat{P}_{22}(u, t) \) (2.43). \( \hat{P}_{22}^*(0, v) \) uses all the individuals that enter state 2, while \( \hat{P}_{22}(u, t) \) only uses those who are in state 2 in the time interval \((u, t]\).

The method does not seem to be that robust to small changes/deviations from the semi-Markov assumption. In setting 4 we use the same basis as in setting 1, but now an individual frailty is included. The frailty has expectation one, and we considered the two cases with frailty variance \( \delta_f = 0.5 \) and \( \delta_f = 1.5 \). Already for \( \delta_f = 0.5 \), the method has problems with biasedness. The integrated MSE value is smallest for this method for \( s = 0 \) and \( s = 6 \), because of small standard deviations, but for \( s = 20 \), the integrated MSE value is larger for the semi-Markov method than for the Markov method. For this \( s \) value, the integrated MSE value for the semi-Markov method is smaller than for the non-Markov method, hence we see that the bias is not that large. For \( \delta_f = 1.5 \), the semi-Markov method clearly performs worse than the two other methods.

When the data are semi-Markovian, and the Markov assumption is not fulfilled (setting 3), this method beats the two other in MSE value. The semi-Markov method is only preferable in cases where there is a great reason to believe that the semi-Markov assumption is fulfilled.

The non-Markov method

In all our cases, the non-Markov method produces approximately unbiased estimates. The disadvantage is that the method gives larger standard deviations than the two other methods. For \( s = 0 \), the standard deviations are quite close to those for the Markov method, but for larger \( s \) values, the standard deviations for this method are clearly larger than for the two other methods. The MSE values are then useful to check the goodness of the non-Markov method compared to the two other methods. For the frailty data with \( \delta_f = 0.5 \), the MSE values are smaller for the Markov method than for the non-Markov method. For the other non-Markovian cases, that is for setting 3 and 5, and the frailty data with \( \delta_f = 1.5 \), the MSE values are generally smaller for the non-Markov method than for the two other methods, for \( s > 0 \). To summarize, we see that the non-Markov method is preferable, over the Markov method, in cases where the Markov assumption is strongly violated.

The Markov standard deviation estimator

When the Markov assumption is fulfilled, the standard deviation estimator (2.47) works well. It overestimates a bit in some cases, and underestimates in other, due to uncertainty in the simulations. For the non-Markovian settings (setting 3-5), we have seen that the estimator works quite well for \( s = 0 \). In setting 4, where we considered the frailty data, it seems that the estimator is following the same degree of robustness as the Markov method for the estimated transition probabilities. The estimator works quite well for all \( s \) values when the frailty
variance $\delta_f = 0.5$ is used, while there is some more underestimation when $\delta_f = 1.5$ is used. For setting 5, the estimator really fails for $s > 0$. Also for the semi-Markovian data the estimator does not work that well for $s > 0$.

The coverage probabilities

When the Markov assumption is fulfilled, both the Markov method (the Aalen-Johansen estimator) and the Markov standard deviation estimator perform well. As a consequence of this, the coverage probabilities of the confidence intervals should be close to 0.95, when we are using a confidence level of 95%. (We remember that $\hat{P}_{12}(s, t)$ is approximately normally distributed in large samples.) We see this to be true in our settings where the Markov assumption is fulfilled.

The coverage probabilities are also close to 0.95 in the non-Markovian settings, when $s = 0$. When $s > 0$ and the data are non-Markovian, the coverage probabilities are smaller than 0.95. Then, both the Markov method and the Markov standard deviation estimator underestimate the true quantities. The extremely low coverage probabilities for some of the cases is a result of the fact that the Markov method produces heavily biased estimates.
Chapter 6

Concluding remarks

The main aim of the thesis has been to study how the Markov method performs in cases where the Markov assumption is not fulfilled. At the same time, we were also interested in how the two other methods, and especially the semi-Markov method, behave in various cases. Since the non-Markov method does not build on any assumptions, we assumed this method to work quite well for all the cases. Chapter 2 and 3 presented the methods, and the methods were here applied on the CML data. Further on, we studied the Brier score, and gave one more data example, where the platelet data were used. In Chapter 5 we finally did a comparison of the methods through simulations.

6.1 Discussion and conclusion

We have seen that the Aalen-Johansen estimator (the Markov method) works well for state occupation probabilities, even in situations where the Markov condition is strongly violated, as was the case in setting 5, Chapter 5. This is in accordance with the results of Datta and Satten (2001). Our results in Chapter 5 regarding the standard deviation estimator (2.47), show that this estimator works well for state occupation probabilities for non-Markovian data. This result was not known to us in advance. Also the confidence intervals are ok for $s = 0$.

For $1 \rightarrow 2$ transition probabilities from time points later than the initial time ($s > 0$), the Aalen-Johansen estimator underestimates the probabilities, and performs worse the more the Markov condition is violated. For $s > 0$, also the standard deviation estimator does not perform well for non-Markovian data. Also for this estimator, the performance is worse the more the Markov condition is violated.

For real data we have seen how the Brier score can be used to find the method that fits the data best. For our two examples, we saw that the semi-Markov method seemed to overestimate the probabilities for small $t$ values, and underestimate for larger. This is in accordance with the over- and underestimation we found in Chapter 5 when the semi-Markov method was used in settings where the semi-Markov assumption was not fulfilled, and $1 \rightarrow 2$ transitions happened quite early. Based on this analysis we should say that the semi-Markov
method should only be used in situations where the Brier score is smallest for this method, for large parts of the interval we are considering. For both the data sets, the estimates calculated by the non-Markov method and by the Markov method, are close together for $s = 0$; the differences in Brier scores are of order $10^{-5}$. This is in agreement with the Datta and Satten (2001) result. For the CML data, the Markov method is the best method choice, at least for the $s$ values we are considering. For the platelet data, it seems that the Markov condition is violated. For these data, the Markov method could be used for state occupation probabilities, but calculation of transition probabilities for $s > 0$ should be handled with care. For $s$ close to 0, that means when there are still quite many individuals left in state 1, the Markov method performs well, but when $s$ increases, the non-Markov method seems to be a better choice. We will come back to this below.

The drawback of the non-Markov method is the larger uncertainty in the estimates, than for the two other methods. The positive element is that the method produces approximately unbiased estimates in all our settings in Chapter 5. As we saw in setting 4, in the case where $\delta_f = 0.5$, the large standard deviation of the non-Markov method, made the Markov method a better method. This points in the direction that the Markov method could be used for the platelet data, also for $s > 0$.

For our two data examples we saw that the confidence intervals estimated by the Markov method were more narrow, than the bootstrap confidence intervals for the non-Markov method. This is in accordance with what we have seen through the simulations in Chapter 5. The standard deviation estimator for the Markov method has a tendency to underestimate the variation, and there is also more variation related to the non-Markov estimator.

6.2 Further work and challenges

We will mention some moments that would be interesting to study further:

- In this thesis, we have been considering the progressive illness-death model, as given in Figure 2.3. This is a irreversible model, hence it could be interesting, also to consider a reversible three state model. The extension would then be that a $2 \rightarrow 1$ transition is possible. When the states are as in Figure 2.3; healthy, diseased and dead, we could then include the possibility to be cured of the disease. It is also of interest to include more states in the model. This is no problem for the Markov method, cf. Section 2.4.1, but for the non-Markov method, estimators for extended models are not available.

- We have not included any covariates in our study. For many cases in medicine, factors such as age, sex, lifestyle, use of drugs and size of tumor (for cancer patients), have a major impact on the transition probabilities. For the Markov method there is no problem to include fixed covariates.
6.2. FURTHER WORK AND CHALLENGES

\( x_i \) for each individual, \( i \), by using Cox regression to estimate the cumulative transition intensities. The non-Markov method does not have the possibility to include covariates.

- For the semi-Markov method, we could have made bootstrap confidence intervals for our two data examples. This is not difficult, but time consuming, due to the lack of software for this method.

- For the confidence intervals (2.48)-(2.50) in Chapter 5, it is of interest to figure out if one of these intervals is preferable. For the survival function \( S(t) \), it may be shown that the log-minus-log interval (2.22) is better than the standard interval (2.21) for small samples, (Borgan and Liestøl 1990). A better interval means that the sampling distribution of \( \log(-\log(\hat{S}(t))) \) is closer to the normal distribution, than is the case for the sampling distribution of \( \hat{S}(t) \). For the cumulative hazard function \( A(t) \), the log interval (2.16) is better than the standard interval (2.15). \( P_{12}(s,t) \) looks quite similar as \( A(t) \) in the beginning. However, when \( t \) increases, \( A(t) \) may take on values larger than 1, while that is not possible for the transition probability. For our simulations, we considered all the intervals (2.48) - (2.50) for estimation of coverage probabilities, but none of the intervals seemed to be better than the others, hence only the coverage probabilities based on the standard CI is included. It could be interesting to see how the confidence intervals works when \( N \), the number of individuals in each data set, is smaller.
References


Appendix A

Some calculations

A.1 The expression for $P_{12}(s, t)$ in Section 2.5

In (2.41) we give an expression for $P_{12}(s, t)$ in the illness-death model. This expression is the solution of the differential equation (2.40). Here we show how to solve the equation.

$$\frac{\partial}{\partial t} P_{12}(s, t) = \alpha_{12}(t) P_{11}(s, t) - \alpha_{23}(t) P_{12}(s, t)$$

$$\frac{\partial}{\partial t} P_{12}(s, t) + \alpha_{23}(t) P_{12}(s, t) = \alpha_{12}(t) P_{11}(s, t)$$

$$\frac{\partial}{\partial t} P_{12}(s, t) \exp \left( \int_s^t \alpha_{23}(u) du \right) + \alpha_{23}(t) P_{12}(s, t) \exp \left( \int_s^t \alpha_{23}(u) du \right) = \alpha_{12}(t) P_{11}(s, t) \exp \left( \int_s^t \alpha_{23}(u) du \right)$$

Recognizing the left-hand side as a partial derivative we have

$$\frac{\partial}{\partial t} \left[ P_{12}(s, t) \exp \left( \int_s^t \alpha_{23}(u) du \right) \right] = \alpha_{12}(t) P_{11}(s, t) \exp \left( \int_s^t \alpha_{23}(u) du \right),$$

and hence

$$P_{12}(s, t) \exp \left( \int_s^t \alpha_{23}(u) du \right) = \int_s^t \alpha_{12}(u) P_{11}(s, u) \exp \left( \int_s^u \alpha_{23}(v) dv \right) \ du,$$

$$P_{12}(s, t) = \exp \left( - \int_s^t \alpha_{23}(u) du \right) \int_s^t \alpha_{12}(u) P_{11}(s, u) \exp \left( \int_s^u \alpha_{23}(v) dv \right) \ du$$

$$= \int_s^t P_{11}(s, u) \alpha_{12}(u) \exp \left( - \int_u^t \alpha_{23}(v) dv \right) \ du$$

$$= \int_s^t P_{11}(s, u) \alpha_{12}(u) P_{22}(u, t) \ du.$$
APPENDIX A

A.2 Estimated variance of \( \hat{P}_{12}(s, t) \) in Section 2.5

The full expression of (2.46) is

\[
\text{Var}(\hat{P}_{12}(s, t)) = \sum_{s < T_j \leq t} \hat{P}_{12}(s, T_j)^2 [\hat{P}_{12}(T_j, t) - \hat{P}_{22}(T_j, t)]^2 \Delta \hat{\sigma}^2_{31}(T_j) \\
+ \sum_{s < T_j \leq t} \hat{P}_{13}(s, T_j)^2 [\hat{P}_{12}(T_j, t) - \hat{P}_{32}(T_j, t)]^2 \Delta \hat{\sigma}^2_{32}(T_j) \\
+ \sum_{s < T_j \leq t} \hat{P}_{11}(s, T_j)^2 [\hat{P}_{22}(T_j, t) - \hat{P}_{12}(T_j, t)]^2 \Delta \hat{\sigma}^2_{12}(T_j) \\
+ \sum_{s < T_j \leq t} \hat{P}_{11}(s, T_j)^2 [\hat{P}_{32}(T_j, t) - \hat{P}_{12}(T_j, t)]^2 \Delta \hat{\sigma}^2_{13}(T_j) \\
+ \sum_{s < T_j \leq t} \hat{P}_{12}(s, T_j)^2 [\hat{P}_{32}(T_j, t) - \hat{P}_{22}(T_j, t)]^2 \Delta \hat{\sigma}^2_{23}(T_j).
\]

Since \( N_{21}(t), N_{31}(t), N_{32}(t) \) are zero for all \( t \), those terms involving \( \hat{\sigma}^2_{21}(T_j), \hat{\sigma}^2_{31}(T_j), \hat{\sigma}^2_{32}(T_j) \) are zero. Also \( \hat{P}_{32}(T_j, t) = 0 \).

A.3 The expectation of the Brier score in Section 4.1

The expectation of the Brier score (4.7) is given by

\[
E\{BS_{1h}(s, t)\} = E\{(I(\hat{X}_i(t) = h) - \Pi_{1h}(s, t))^2w_i(t \mid s) \mid \hat{X}_i(s) = 1\} \\
= E \left\{ (I(\hat{X}_i(t) = h) - \Pi_{1h}(s, t))^2 \frac{I(s < \hat{T}_i \leq t, \hat{X}_i(t) \neq 0)}{G(\hat{T}_i \mid s)} \mid \hat{X}_i(s) = 1 \right\} \\
+ E \left\{ (I(\hat{X}_i(t) = h) - \Pi_{1h}(s, t))^2 \frac{I(\hat{T}_i > t)}{G(t \mid s)} \mid \hat{X}_i(s) = 1 \right\} \\
= \text{Part I} + \text{Part II},
\]
Part I:
\[
E \left\{ \left( I(\bar{X_i}(t) = h) - \Pi_{1h}(s,t) \right)^2 \frac{I(s < \bar{T}_i \leq t, \bar{X_i}(t) \neq 0)}{G(\bar{T}_i \mid s)} \, \bigg| \, \bar{X_i}(s) = 1 \right\}
\]
\[
= E \left\{ \left( I(X_i(t) = h) - \Pi_{1h}(s,t) \right)^2 \frac{I(s < T_i \leq t)I(C_i > T_i > s)}{G(T_i \mid s)} \, \bigg| \, X_i(s) = 1, C_i > s \right\}
\]
\[
= E \left\{ E \left\{ \left( I(X_i(t) = h) - \Pi_{1h}(s,t) \right)^2 \frac{I(s < T_i \leq t)I(C_i > T_i > s)}{G(T_i \mid s)} \, \bigg| \, T_i, X_i(s) = 1, C_i > s \right\} \right\}
\]
\[
= E \left\{ \left( I(X_i(t) = h) - \Pi_{1h}(s,t) \right)^2 \frac{I(s < T_i \leq t)I(C_i > T_i > s)}{G(T_i \mid s)} \, \bigg| \, X_i(s) = 1 \right\}
\]
\[
= E \left\{ \left( I(X_i(t) = h) - \Pi_{1h}(s,t) \right)^2 I(s < T_i \leq t) \, \bigg| \, X_i(s) = 1 \right\},
\]
Part II:
\[
E \left\{ \left( I(\bar{X_i}(t) = h) - \Pi_{1h}(s,t) \right)^2 \frac{I(\bar{T}_i > t)}{G(t \mid s)} \, \bigg| \, \bar{X_i}(s) = 1 \right\}
\]
\[
= E \left\{ \left( I(X_i(t) = h) - \Pi_{1h}(s,t) \right)^2 \frac{I(T_i > t)I(C_i > t)}{G(t \mid s)} \, \bigg| \, X_i(s) = 1, C_i > s \right\}
\]
\[
= E \left\{ \left( I(X_i(t) = h) - \Pi_{1h}(s,t) \right)^2 I(T_i > t) \, \bigg| \, X_i(s) = 1 \right\} \frac{E \{I(C_i > t \mid C_i > s)\}}{G(t \mid s)}
\]
\[
= E \left\{ \left( I(X_i(t) = h) - \Pi_{1h}(s,t) \right)^2 I(T_i > t) \, \bigg| \, X_i(s) = 1 \right\}.
\]
Hence
\[
E \{BS_{1h}^s(s,t)\} = E \{BS_{1h}(s,t)\}.
\]

A.4 A check of the simulation procedure in Section 5.1

We generate \(T_{12}, T_{13}\) and \(T_3 = T_{12} + T_{23}\) as described in section 5.1. Note that we generate \(T_3\) such that, given \(T_{12}\), the survival function of \(T_3\) is
\[
P(T_3 > t \mid T_{12}) = e^{-\int_{T_{12}}^{t} \alpha_{23}(u) \, du},
\]
for \(t > T_{12}\). This means that, given \(T_{12}\), the survival function of \(T_{23}\) is
\[
P(T_{23} > t \mid T_{12}) = e^{-\int_{0}^{t} \alpha_{23}(v + T_{12}) \, dv},
\]
for $t > 0$. As in section 3.2, we set $Z = \min(T_{12}, T_{13})$. Then we have
\[
P(Z > t) = P(T_{12} > t)P(T_{13} > t) \\
= \exp \left\{ -\int_0^t \alpha_{12}(u)du \right\} \exp \left\{ -\int_0^t \alpha_{13}(u)du \right\} \\
= \exp \left\{ -\int_0^t \alpha_{12}(u) + \alpha_{13}(u)du \right\}.
\]
By (3.3) we have
\[
P_{11}(s, t) = \frac{P(Z > t)}{P(Z > s)} = \exp \left\{ -\int_s^t \alpha_{12}(u) + \alpha_{13}(u)du \right\}. \quad (A.1)
\]
Further, by (3.4) we have
\[
P_{12}(s, t) = \frac{P(s < T_{12} \leq t, T_{12} \leq T_{13}, T_{12} + T_{23} > t)}{P(Z > s)}. \quad (A.2)
\]
The numerator of (A.2) is now
\[
P(s < T_{12} \leq t, T_{12} \leq T_{13}, T_{12} + T_{23} > t) \\
= \int_s^t f_{12}(u)P(T_{13} \geq u)P(T_{23} > t - u)du \\
= \int_s^t \left( \alpha_{12}(u)e^{-\int_u^\infty \alpha_{13}(v)dv} \right) e^{-\int_u^\infty \alpha_{23}(v+u)dv} du \\
= \int_s^t \alpha_{12}(u)e^{-\int_u^\infty \alpha_{13}(v)dv}e^{-\int_u^\infty \alpha_{23}(v)dv} du.
\]
Hence we get
\[
P_{12}(s, t) = \int_s^t \alpha_{12}(u)e^{-\int_u^\infty \alpha_{13}(v)dv}e^{-\int_u^\infty \alpha_{23}(v)dv} du \\
= \int_s^t e^{-\int_u^\infty \alpha_{12}(v)dv}e^{-\int_u^\infty \alpha_{23}(v)dv} \alpha_{12}(u) du. \quad (A.3)
\]
By (3.5) we have
\[
P_{22}(s, t) = \frac{P(T_{12} + T_{23} > t, T_{12} \leq s, T_{12} \leq T_{13})}{P(T_{12} + T_{23} > s, T_{12} \leq s, T_{12} \leq T_{13})}. \quad (A.4)
\]
The numerator of (A.4) is now
\[
P(T_{12} + T_{23} > t, T_{12} \leq s, T_{12} \leq T_{13}) \\
= \int_0^s P(T_{23} > t - u)f_{12}(u)P(T_{13} \geq u)du \\
= \int_0^s e^{-\int_u^\infty \alpha_{23}(v+u)dv} \left( \alpha_{12}(u)e^{-\int_u^\infty \alpha_{13}(v)dv} \right) e^{-\int_u^\infty \alpha_{23}(v)dv} du \\
= \int_0^s e^{-\int_u^\infty \alpha_{12}(v)dv}e^{-\int_u^\infty \alpha_{23}(v)dv} \alpha_{12}(u) du \\
= e^{-\int_u^\infty \alpha_{23}(v)dv} \int_0^s e^{-\int_u^\infty \alpha_{12}(v)dv}e^{-\int_u^\infty \alpha_{23}(v)dv} du,
We see that (A.1), (A.3), (A.5) gives us the right probabilities, cf. (2.38), (2.41), (2.39), respectively.

The probabilities involved in (5.10) are expressed by:

The product is then

Hence

We see that (A.1), (A.3), (A.5) gives us the right probabilities, cf. (2.38), (2.41), (2.39), respectively.

### A.5 Setting 4: Frailty non-Markov data in Section 5.2

The probabilities involved in (5.10) are expressed by:

The product is then

\[
\begin{align*}
P_{12}(s, t | \theta) &= \int_{t}^{s} P_{11}(s, t | \theta) \alpha_{12}(u | \theta) P_{22}(u, t | \theta) du \\
&= \int_{t}^{s} \left( e^{-(a_{12}+a_{13})(u-s)\theta} \right) a_{12} \theta \left( e^{-a_{23}(t-u)\theta} \right) du \\
&= a_{12} \theta e^{(a_{12}+a_{13})s\theta} e^{-a_{23}t\theta} \int_{t}^{s} e^{-(a_{12}+a_{13}-a_{23})u\theta} du \\
&= a_{12} e^{(a_{12}+a_{13})s\theta} e^{-a_{23}t\theta} \left[ \frac{1}{-(a_{12}+a_{13}-a_{23})} e^{-(a_{12}+a_{13}-a_{23})u\theta} \right]_{s}^{t} \\
&= \frac{a_{12}}{a_{12} + a_{13} - a_{23}} e^{(a_{12}+a_{13})s\theta} e^{-a_{23}t\theta} \left[ e^{-(a_{12}+a_{13}+a_{23})s\theta} e^{-(a_{12}+a_{13}+a_{23})t\theta} \right].
\end{align*}
\]
A.6 Setting 5: Non-Markov data in Section 5.2

The denominator of (5.13) is

\[ P(\min(T_{12}, T_{13}) > s) = P(T_{12} > s, T_{13} > s) = P(T_{12} > s)P(T_{13} > s) = \int_s^\infty \frac{1}{\lambda_{12}} e^{-\lambda_{12} t_{12}} dt_{12} \int_s^\infty \frac{1}{\lambda_{13}} e^{-\lambda_{13} t_{13}} dt_{13} = e^{-\lambda_{12} s} e^{-\lambda_{13} s} = e^{-(\lambda_{12} + \lambda_{13}) s}. \]

If \( \frac{t}{2.7} < s \), the numerator of (5.13) is:

\[ P(s < T_{12} \leq t, T_{12} \leq T_{13}, T_{12} > t/2.7) = P(s < T_{12} \leq t, T_{12} \leq T_{13}) = \int_s^t \int_{t_{12}}^\infty f(t_{12}, t_{13}) dt_{13} dt_{12} \]
\[ = \int_s^t \int_{t_{12}}^\infty \lambda_{12} \lambda_{13} e^{-\lambda_{12} t_{12}} e^{-\lambda_{13} t_{13}} dt_{13} dt_{12} = \lambda_{12} \int_s^t e^{-\lambda_{12} t_{12}} \left[ -e^{-\lambda_{13} t_{13}} \right]_{t_{12}}^\infty dt_{12} = \lambda_{12} \int_s^t e^{-(\lambda_{12} + \lambda_{13}) t_{12}} dt_{12} = \frac{\lambda_{12}}{\lambda_{12} + \lambda_{13}} \left[ -e^{-(\lambda_{12} + \lambda_{13}) t_{12}} \right]_s^t = \frac{\lambda_{12}}{\lambda_{12} + \lambda_{13}} \left( e^{-(\lambda_{12} + \lambda_{13}) s} - e^{-(\lambda_{12} + \lambda_{13}) t} \right). \]

Hence

\[ P_{12}(s, t) = \frac{1}{e^{-(\lambda_{12} + \lambda_{13}) s}} \frac{\lambda_{12}}{\lambda_{12} + \lambda_{13}} \left( e^{-(\lambda_{12} + \lambda_{13}) s} - e^{-(\lambda_{12} + \lambda_{13}) t} \right) = \frac{\lambda_{12}}{\lambda_{12} + \lambda_{13}} \left( 1 - e^{-(\lambda_{12} + \lambda_{13})(t-s)} \right). \]
If $\frac{t}{2.7} \geq s$, the numerator of (5.13) is:

$$P(s < T_{12} \leq t, T_{12} \leq T_{13}, T_{12} > t/2.7)$$

$$= P\left(\frac{t}{2.7} < T_{12} \leq t, T_{12} \leq T_{13}\right)$$

$$= \int_{t/2.7}^{t} \int_{t_{12}}^{\infty} f(t_{12}, t_{13}) dt_{13} dt_{12}$$

$$= \int_{t/2.7}^{t} \int_{t_{12}}^{\infty} \lambda_{12} \lambda_{13} e^{-\lambda_{12} t_{12}} e^{-\lambda_{13} t_{13}} dt_{13} dt_{12}$$

$$= \lambda_{12} \int_{t/2.7}^{t} e^{-\lambda_{12} t_{12}} \left[ -e^{-\lambda_{13} t_{13}} \right]_{t_{12}}^{\infty} dt_{12}$$

$$= \lambda_{12} \int_{t/2.7}^{t} e^{-\lambda_{12} t_{12}} \left[-e^{-\lambda_{13} t_{12}}\right] dt_{12}$$

$$= \frac{\lambda_{12}}{\lambda_{12} + \lambda_{13}} \left[e^{-\left(\lambda_{12} + \lambda_{13}\right) t_{12}}\right]_{t/2.7}^{t}$$

$$= \frac{\lambda_{12}}{\lambda_{12} + \lambda_{13}} \left(e^{-\left(\lambda_{12} + \lambda_{13}\right) t/2.7} - e^{-\left(\lambda_{12} + \lambda_{13}\right) t}\right).$$

Hence

$$P_{12}(s, t) = \frac{1}{e^{-\left(\lambda_{12} + \lambda_{13}\right)s}} \frac{\lambda_{12}}{\lambda_{12} + \lambda_{13}} \left(e^{-\left(\lambda_{12} + \lambda_{13}\right) t/2.7} - e^{-\left(\lambda_{12} + \lambda_{13}\right) t}\right)$$

$$= \frac{\lambda_{12}}{\lambda_{12} + \lambda_{13}} e^{\left(\lambda_{12} + \lambda_{13}\right)s} \left(e^{-\left(\lambda_{12} + \lambda_{13}\right) t/2.7} - e^{-\left(\lambda_{12} + \lambda_{13}\right) t}\right).$$
Appendix B

R-scripts

In this appendix, the most relevant parts of the R-scripts used in the thesis are included. In B.1, script for estimation of the transition probabilities and calculation of the Brier scores in Example 1 is given. In B.2, script for the simulation study when the data are Markovian (setting 1, Chapter 5), is included.

B.1 Estimation of transition probabilities and calculation of Brier scores

```r
library(mstate)
library(p3state.msm)
library("TPmsm")
data(ebmt1)

# Preparation of the data:

ebmt1data = matrix(0,1977,5)

ebmt1data[,1] = ebmt1$rel/365.25
ebmt1data[,2] = ebmt1$relstat
ebmt1data[,3] = (ebmt1$srv - ebmt1$rel)/365.25
ebmt1data[,4] = ebmt1data[,1]+ebmt1data[,3]

ebmt1dataframe = data.frame(ebmt1data)
colnames(ebmt1dataframe) = c("times1","delta","times2", "time","status")

# The code for the semi–Markov transition probabilities # and for the Brier score require the data to be a # p3state object:
```
ebmt1p3stateobj = p3state(ebmt1dataframe)
mydata = ebmt1p3stateobj$datafr

# Preparation for use of the TPmsm package:

ebmt1_tp_prep1 = as.vector(ebmt1data[,1])
ebmt1_tp_prep2 = as.vector(ebmt1data[,2])
ebmt1_tp_prep3 = as.vector(ebmt1data[,4])
ebmt1_tp_prep4 = as.vector(ebmt1data[,5])
cens = rep(1,dim(ebmt1)[1])
for (e in 1:(dim(ebmt1)[1]))
{
  if (((ebmt1_tp_prep3[e]-ebmt1_tp_prep1[e]) == 0 &
       ebmt1_tp_prep4[e] == 0)
  {cens[e] = 0}
}
fitTP = survTP(ebmt1_tp_prep1,cens,ebmt1_tp_prep3,
               ebmt1_tp_prep4)
tmat=transMat(list(c(2,3),c(3),c()),
               names=c("transplanted","relapse","dead"))

ebmt1long=msprep(time=c(NA,"times1","time"),
                 status=c(NA,"delta","status"),
                 data=ebmt1dataframe,trans=tmat)

# cumulative transition intensities
cox.ebmt1=coxph(Surv(Tstart,Tstop,status)~strata(trans),
                 data=ebmt1long, method="breslow")
haz.ebmt1=msfit(cox.ebmt1,trans=tmat)

# Set s value:
s = 0

# All the Markov transition probabilities and
# estimated std are contained in:
pt = probtrans(haz.ebmt1,predt=s)

# All the non-Markov transition probabilities
# and bootstrap CI are contained in:
pnon = transKMW(fitTP,s,3088/365.25,conf=TRUE,
           conf.level=0.95,n.boot=1000)
# Consider only those individuals who are still in state 1 at time s:
mydataS = mydata[which(mydata$times1 > s),]

# The transition probabilities:
# Markov P11:
times1 = pt[[1]]$time
markov11prob = pt[[1]]$pstate1
markov11se = pt[[1]]$se1

markov11 = cbind(times1, markov11prob)

# Semi-markov P11 (calculates the same as for Markov):
times1semi = mydataS$times1
d11 = as.numeric((mydataS$delta == 1) | (mydataS$delta == 0 & mydataS$status == 1))
fitsemi11 = survfit(Surv(times1semi, d11) ~ 1)

semi11prob = fitsemi11$surv

semi11 = cbind(c(s, unique(times1semi)), c(1, semi11prob))

# Non–Markov P11:
pnontime = pnon$time
pn11 = pnon$est[,1]
pnCIlower = pnon$inf[,1]
pnCIupper = pnon$sup[,1]

nonmarkov11 = cbind(pnontime, pn11)

# Markov P12:
times12 = pt[[1]]$time
markov12prob = pt[[1]]$pstate2
markov12se = pt[[1]]$se2

markov12 = cbind(times12, markov12prob)

# Semi–Markov P12:
d12 = as.numeric(mydataS$delta == 1)
fitsemi12 = survfit(Surv(times1semi, d12) ~ 1, type="f1")

Ahat = c(0, -log(fitsemi12$surv))
NAa = diff(Ahat)
times2 = mydata$times2[-1]
sorted_times_in_2 = sort(unique(times2))
ddd = as.numeric((mydata$delta[-1] == 1) &
                 (mydata$status[-1] == 1))
fit23 = survfit(Surv(times2, ddd)~1)
semi22 = fit23$surv

m1 = length(fitsemi11$time)
m2 = length(fit23$time)
newtime = rep(NA, m1 * m2)
i = 0
for (j in 1:m1)
    for (l in 1:m2)
    {
        i = i + 1
        newtime[i] = fitsemi11$time[j] + fit23$time[l]
    }
newtime1 = newtime[newtime <= max(fitsemi11$time)]
newtime2 = sort(unique(newtime1))

semi12prob = rep(NA, length(newtime2))
j = 1
for (t in newtime2)
{
    times = fitsemi11$time[fitsemi11$time <= t]
z = t - times
p = rep(0, length(z))
i = 1
for (e in z)
{
    p[i] = semi22[max(which(sorted_times_in_2 <= e))]
i = i + 1
}
semi12prob[j] = sum(c(1, semi11prob[fitsemi11$time <= t]
                     [-length(semi11prob[fitsemi11$time<=t])]) *
                     NAa[fitsemi11$time <= t] * p)
j = j + 1
}

semi12 = cbind(c(s, newtime2), c(0, semi12prob))

# Non-Markov P12:
pnontime = pnon$time
pnon12 = pnon$est[,2]
pnon12CIlower = pnon$inf[,2]
pnon12CIupper = pnon$sup[,2]

nonmarkov12 = cbind(pnon$time, pnon12)

# All times where there could be changes in the probabilities for one or more of the methods:
change_in_prob_times = sort(unique(c(markov12[,1], semi12[,1], nonmarkov12[,1])))
interesting_times = change_in_prob_times[change_in_prob_times <= 7]

# Brier score:
# Constructing G:
# I = mydata$status
# I[1] = 1
fit11 = survfit(Surv(mydata$time,1-I)~1)
G = cbind(fit11$time, fit11$surv)
# G at time s:
gs = G[max(which(fit11$time <= s)),2]
# Order the data after total time, to avoid problems in the weights in the Brier score:
mydataS_ordered = mydataS[order(mydataS$time),]

i=1
 g = rep(0,dim(mydataS_ordered)[1])
for (t in mydataS_ordered$time)
{
 g[i] = G[max(which(fit11$time <= t)),2]/gs
 i = i+1
}

# Number in state 1 at time s:
mydataminus1 = mydata[-1,]
mydataS_included_s = mydataminus1[which(mydataminus1$times1 >= s),]
number = dim(mydataS_included_s)[1]
# P12 for the three methods, to be put in the
# algorithm below, one at a time:

prob12 = markov12[max(which(markov12[,1] <= t)),2]
prob12 = semi12[max(which(semi12[,1] <= t)),2]
prob12 = nonmarkov12[max(which(nonmarkov12[,1] <= t)),2]

j = 1
Brier = rep(NA, length(interesting_times))
for (t in interesting_times)
{
  prob12 = markov12[max(which(markov12[,1] <= t)),2]
  X <- as.numeric(mydataS_ordered$times1 <= t)*
       as.numeric(mydataS_ordered$time > t)
  gt = G[max(which(fit11$time <= t)),2]/gs

  ind_with_max_time = which(mydataS_ordered$time ==
                           max(mydataS_ordered$time))
  iwmax = ind_with_max_time

  weight_firstpart = as.numeric(mydataS_ordered$time
                                <= t)[−iwmax]*mydataS_ordered$status[−iwmax]/g[−iwmax]

  weight = rep(0,(length(X)))
  weight[1:length(weight_firstpart)] = weight_firstpart

  weight = weight + (as.numeric(mydataS_ordered$time > t)/
                     (gt*rep(1,length(X))))

  Brier[j] = (1/number)*sum((X−(prob12*
                           rep(1,length(X))))^2*weight)
  j = j+1
}

# Integrated Brier score:
diff_tvalues = diff(interesting_times)
IBriermarkov = sum(Brier[length(Brier)]*
                   diff_tvalues)/(max(interesting_times))
B.2 Simulation study: Markovian data (setting 1)

```r
library(p3state msm)
library(mstate)
library("TPmsm")
require(matlab)

K = 1000; N = 400; s = 0
tau = 60; a = 160

a12 = 0.02; a13 = 0.01; a23 = 0.025
b12 = 0; b13 = 0; b23 = 0

set.seed(237985)
seed.vector = sample(1:999999,1000)

markov12 = matrix(0,6,(2*K)+7)
nonmarkov12 = matrix(0,6,K+6)
semimarkov12 = matrix(0,6,K+6)

for (k in 1:K)
{
  set.seed(seed.vector[k])
  # Start by generating data:
  U12 = runif(N,0,1)
  U13 = runif(N,0,1)
  V23 = runif(N,0,1)

  T12 = (-(b12+1)/a12)*log(U12)^(1/(b12+1))
  T13 = (-(b13+1)/a13)*log(U13)^(1/(b13+1))

  U = runif(N,0,a)
  C = rep(0,N)
  for (e in 1:N)
  {
    C[e] = min(tau,U[e])
  }

  T1 = rep(0,N)
  delta = rep(0,N)
  for (e in 1:N)
  {
    T1[e] = min(T12[e],T13[e],C[e])
    if (T1[e] == T12[e])
    {
      delta[e] = 1
    }
  }
```
MT23 = rep(0,N)
for (e in 1:N)
{
  MT23[e] = ((-b23+1)/a23*log(V23[e]) + T1[e]^(b23+1))^(1/(b23+1))
}
MT23status = (MT23–T1)*delta

T2 = rep(0,N)
for (e in 1:N)
{
  if (C[e] >= T1[e] & T1[e] != T13[e])
  { T2[e] = min(MT23status[e], C[e]–T1[e]) }
  else
  { T2[e] = 0 }
}

time = rep(0,N)
for (e in 1:N)
{
  time[e] = min(C[e], T1[e]+T2[e])
}

status = rep(0,N)
for (e in 1:N)
{
  if (min(T12[e],T13[e],C[e]) == T12[e])
  { if (T1[e] + MT23status[e] <= C[e])
    { status[e] = 1 }
  }
  if (min(T12[e],T13[e],C[e]) == T13[e])
  { status[e] = 1 }
}

# Generated data:
dataMarkov = matrix(0,N,5)
dataMarkov[,1] = T1
dataMarkov[,2] = delta
dataMarkov[,3] = T2
dataMarkov[,4] = time
dataMarkov[,5] = status
# Calculate the transition probabilities for the chosen \( t \) values:

```r
markovdataframe = data.frame(dataMarkov)
colnames(markovdataframe) = c("times1", "delta", "times2", "time", "status")
obj = p3state(markovdataframe)
mydata = obj$datafr

newd1 = as.vector(T1)
newd2 = as.vector(delta)
newd3 = as.vector(time)
newd4 = as.vector(status)

cens = rep(1,N)
for (e in 1:N)
{
  if ((newd3[e] - newd1[e]) == 0 & newd4[e] == 0)
  { cens[e] = 0 }
}

survtp = survTP(newd1, cens, newd3, newd4)

tmat2 = transMat(x = list(c(2, 3), c(3), c()),
  names = c("Tr", "Pr", "RelDeath"))
msbmt2 = msprep(time = c(NA, "times1", "time"),
  status = c(NA, "delta", "status"),
  data = markovdataframe, trans = tmat2)

cox.ebmt1 = coxph(Surv(Tstart, Tstop, status)~strata(trans),
  data = msbmt2, method="breslow")
haz.ebmt1 = msfit(cox.ebmt1, trans = tmat2)

pt = probtrans(haz.ebmt1, predt = s)
```

# Semi-Markov P12:

tid1 = mydata$times1[mydata$times1 > s]

dd = as.numeric((mydata$delta == 1) | (mydata$delta == 0 & mydata$status == 1))
[mydata$times1 > s]

fit = survfit(Surv(tid1, dd)~1)
p11 = fit$surv
d = as.numeric(mydata$delta == 1)[mydata$times1 > s]

fit12 = survfit(Surv(tid1, d)~1, type="fl")
Ahat = c(0, -log(fit12$surv))
```r
NAa = diff(Ahat)

tid2 = mydata$times2
ddd = as.numeric((mydata$delta == 1) &
                (mydata$status == 1))
fit23 = survfit(Surv(tid2, ddd)~1)
p22 = fit23$surv

m = sort(unique(tid2))

i = 1
for (t in seq(10, 60, 10)) {
  if (t <= s)
    semimarkov12[i, k] = 0
  else
    {
      ti = fit$time[fit$time <= t]
z = t - ti
p = rep(0, length(z))
  j = 1
  for (e in z)
    {
      p[j] = p22[max(which(m <= e))]
      j = j + 1
    }
  if (min(fit$time) > t)
    semimarkov12[i, k] = 0
  else
    semimarkov12[i, k] = sum(c(1, p11[fit$time <= t]
                              [-length(p11[fit$time<=t])])*NAa[fit$time <= t]*p))
  i = i + 1
} # Markov and non-Markov P12:
i = 1
for (t in seq(10, 60, 10)) {
  if (t <= s)
    {
      markov12[i, k] = 0
      markov12[i, (K+k)] = 0
      nonmarkov12[i, k] = 0
    }
  else
```


```r
{
num = max(which(pt[[1]]$time <= t))
markov12[i,k] = pt[[1]]$pstate2[num]
markov12[i,(K+k)] = pt[[1]]$se2[num]
nmarkov = transKMW(survtp,s,t)
nonmarkov12[i,k] = nmarkov$est[2][length(nmarkov$est[,2])]
}
i = i + 1
}

# Calculate the true transition probabilities:

g <- function(u) {
  exp(-((a12/(b12+1)*u^(b12+1))+(a13/(b13+1)*u^(b13+1))
+ ((a12/(b12+1)*s^(b12+1))+(a13/(b13+1)*s^(b13+1)))*
a12*u^b12*exp(-(a23/(b23+1)*t^(b23+1)) +
(a23/(b23+1)*u^(b23+1)))
}

P12 = rep(0,6)
e=1
for (t in seq(10,60,10)) {
  if (t <= s) {
P12[e] = 0
  } else {
P12[e] = integrate(g,s,t)$value
  }
  e = e+1
}
markov12[, (2*K)+1] = P12

# Similar for semi–Markov and non–Markov:

# mean of estimates:
i = 1
for (i in 1:6) {
  markov12[i, (2*K)+2] = mean(markov12[i,1:K])
i = i + 1
}
```
# mean of standard deviation (Only for Markov):
\[
i = 1 \\
\text{for } (i \text{ in } 1:6) \\
\{
markov12[i,(2*K)+7] = \text{mean}(\markov12[i,(K+1):(2*K)])
\}
\]

# bias:
\[
\markov12[,,(2*K)+3] = \markov12[,,(2*K)+2] - \markov12[,,(2*K)+1]
\]

# relative bias * 100:
\[
\markov12[,,(2*K)+4] = \markov12[,,(2*K)+3] / \markov12[,,(2*K)+1] * 100
\]

# MSE*1000:
\[
\text{for } (j \text{ in } 1:6) \\
\{
\markov12[j,(2*K)+5] = \text{mean}((\markov12[j,1:K] - \markov12[j,(2*K)+1])^2) * 1000
\}
\]

# Empirical standard deviation:
\[
\text{for } (j \text{ in } 1:6) \\
\{
\markov12[j,(2*K)+6] = \text{sqrt}(\text{sum}((\markov12[j,1:K] - \markov12[j,(2*K)+2])^2) / (K-1))
\}
\]
Appendix C

Plots

In Chapter 5 we are considering the bias of our three methods. There we are giving the bias only at every tenth \( t \) value. In the figures presented here, the bias is calculated at each integer, and we get a better glimpse of how the bias is evolving.

Figure C.1: Markov data with constant hazards. The bias is calculated at each integer. Green: Markov, Red: Semi-Markov, Blue: Non-Markov.
Figure C.2: Markov data with nonconstant hazards. The bias is calculated at each integer. Green: Markov, Red: Semi-Markov, Blue: Non-Markov.
Figure C.4: Frailty data with $\delta_f = 0.5$. Green: Markov, Red: Semi-Markov, Blue: Non-Markov.
Figure C.5: Frailty data. $\delta = 1.5$ Green: Markov, Red: Semi-Markov, Blue: Non-Markov.