Department of Mathematical Sciences

PROJECT

Non-Markov Modelling with Applications in Life Insurance

Author: Joakim Lavlund Bilyk[1]
Supervisor: Christian Furrer[1]

Submitted on: 18 July 2023

[1]Department of Mathematical Sciences, University of Copenhagen, Copenhagen Ø, Denmark

Abstract

Life insurance contracts are often defined contingent on the transitions of a state process based on the insured biometric state. The classical approach to estimating transition rates and occupation probabilities for such a process is the Nelson-Aalen estimator and Aalen-Johansen estimator by insisting that the process is Markovian. However, most real-life state processes does not satisfy the Markov property, thus making the classical approach not appropriate. In this project we offer a different approach to estimating probabilities and rates that is consistent with the underlying state process. This approach gives particular well-behaving structures similar to the classical estimators. The result is an estimator we call the As-If-Markov estimator, that significantly improve on the model risk inherent in the Markov estimator although with larger variance.

Keywords: Non-Markov modelling, conditional Aalen-Johansen estimator, conditional Nelson-Aalen estimator, As-If-Markov model, pure jump process, landmarking, model risk versus approximation risk, expected cash-flows

Contents

1 Introduction and Notation

In Life Insurance Mathematics we often model contracts between the insured and the insurance company contingent on an underlying biometric state process Z_t, where t represents age. The insurance company is interested in knowing the stochastic behaviour of the state process (or sometimes also called *chain*) or more importantly the expected flow of money. This knowledge ensures that the company sets aside appropriate amount of money to fulfill its future liabilities to its customers. It is moreover used when pricing new contracts such that on average the inflow of money is equal to or surpasses the outflow.

In order to give reasonable estimates of future payments from the company to the insured, called *benefits*, and from the insured to the company, called *premiums*, one has to make assumptions on the underlying chain and/or on the information present at the time of prognosis. Most actuarial students are familiar with the importance of reserving and master methods such as solving the Thiele's differential equation. These methods require the actuarial knowing the hazards/intensities of the chain. The actual calculation of these is sometimes taken for granted and so estimation is not always the strongest feature of an actuarial.

The classical methods in reserving require a rather strict assumption, the Markov assumption, to be made on Z_t which in most cases is not fulfilled. Therefore we often restrict ourselves to the small family of Makovian models, when in fact the underlying chain has a more exotic behaviour. In short terms, the Markov assumption tells us that any two similar aged individuals has equal probabilities of future transitions regardless of their respective pasts. Thus the Markov assumption ensures that on one hand the reserves is easily calculable (when knowing the hazards), while on the other hand estimating the hazards uses all data available.

In this project, we study how we may loosen this assumption to better fit the real-world processes the insurance company encounter. We will in detail discuss estimation methods to target the occupational probabilities, that the chain sojourn in some state k at some future time t given that today at time s the chain sojourns in state j. This require that we can estimate the cumulative intensities that Z, on a small interval, may transition from one state (i) to another (k). We call this the cumulative transition rate from i to k.

The estimator we will be discussing is the *conditional Aalen-Johnasen estimator* for the occupation probabilities and the *conditional Nelson-Aalen estimator* for the cumulative transition rates. Specifically we will be give examples on the version where we condition on Z sojourning in state j at time s. We call this technique *landmarking* or As-If-Markov modelling. In the examples, we show when the conditional estimates are appropriate and how they may drastically improve predictions compared to when one assume the Markov assumption. In particular, we will be discussing the choice between utilizing the Markov assumption and a landmarking approach in the light of *model risk* (associated with the Markov assumption) and *approximation risk* (associated with subsampling).

1.1 Structure

The project is structure in three parts; the first part deals with defining the processes we will be studying i.e. *the pure jump process*, the second part introduces the estimator for the occupation probabilities and transition rates and lastly the third part is a numerical study of estimators introduced in part two including errors decomposed in model risk and approximation risk.

The first part consist of Section 2 and 3. We start in Section 2 by in general defining stochastic processes as a family of random variable indexed by time. From this definition we define the main process of interest: *the pure jump process*. After introducing the pure jump process we outline the related counting processes, that counts visits in states and transitions from states. Lastly, in Section 3 we discuss the Markov property and what information we assume the insurance company has at its disposal.

In the second part, that includes Section 4 to 6. We start in Section 4 by defining the conditional transition rates and occupation probabilities and the relation between them and the counting processes introduced in Section 2. We show that the occupation probabilities is derived from the product integral of the transition rates and we give concrete solutions. Next in Section 5 we define the data we have for estimation. This gives rise to the definition of the Nelson-Aalen estimator and Aalen-Johansen estimator in section 5.2. We show how when we employ landmarking techniques, the estimator has a tractable closed solution. We call this estimator the As-If estimator. Next in Section 6 we explain how one in practice may implement the estimator and use them in approximating expected cash-flows.

Finally in the third part, which only consist of section 7, we conduct a numerical study of the impact of using landmarking estimates versus the classical Markov assumption. This is in practice, done by studying an actual Markov chain, thus the Markov estimator should be preferred and a Semi-Markov chain, where the As-If model should be preferred. The discussion is however a bit more nuanced in that the As-If estimates have much larger variance, than the Markov estimates, thus giving the Markov estimates an advantage on the approximation risk and the As-If estimates an advantage with respect to model risk.

1.2 Notation

We occasionally suppress some notation when it is prudent to do so. Henceforth we use the following shorthand.

1. For any function $f(t)$ we define the left-limit

$$f(t-) = \lim_{h \to 0^+} f(t-h)$$

2. Whenever we introduce a random variable or stochastic process, we assume that it is a measurable mapping from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to a measurable space $(\mathcal{X}, \mathcal{A}, \nu)$, that if not written explicitly will be obvious to the context.

2 Pure Jump Process

Definition 1 (Random variable)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random variable is a measurable function $X : (\Omega, \mathcal{F}, \mathbb{P}) \mapsto (\mathcal{X}, \mathcal{A}, \nu)$ and we say that X is a $\mathcal{X}$-valued random variable.

Remark. We never actually make any assumptions on the structure of the probability space it self. The only property we insist on is that X is measurable. That is

$$\forall A \in \mathcal{A} : \{X \in A\} \stackrel{\text{def}}{=} \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}$$

giving that $\mathbb{P}(\{X \in A\})$ exist for any choice of $A \in \mathcal{A}$. We will henceforth use the notation $\mathbb{P}(X \in A) \stackrel{\text{def}}{=} \mathbb{P}(\{X \in A\})$.

This definition gives us the ability to use the usual results from measure theory and define various operators such as the expectation, moments and expectation of transforms. Indeed given that

$$\int_\Omega |g(X(\omega))| \, \mathrm{d}P(\omega) < \infty.$$

We define the object

$$\mathbb{E}\left[g(X)\right] \stackrel{\text{def}}{=} \int_\Omega g(X(\omega)) \, \mathrm{d}\mathbb{P}(\omega).$$

In the general case of stochastic processes, we say that any collection of random variables is a stochastic processes.

Definition 2 (Stochastic process)**.** Consider an index-set $\mathcal{I}$. The collection $(X_i)_{i\in\mathcal{I}}$, with X_i being $\mathcal{X}$-valued random variables for all $i \in \mathcal{I}$, is called a $\mathcal{X}$-valued stochastic process on $\mathcal{I}$.

We will in general only consider time-indexed stochastic processes, that is $\mathcal{I} = \mathbb{R}^+ = [0, \infty)$ and we use the notation

$$\mathbf{X} = (X_t)_{t\geq 0} \stackrel{\text{def}}{=} (X_t)_{t\in\mathbb{R}^+}.$$

Life insurance contracts are often modelled by a particular underlying stochastic process, that gives rise to the multi-state contract. We model this process by defining a finite state space $\mathcal{Z}$ representing the different states the insured may sojourn in. We called this a pure jump process.

Definition 3 (Pure jump process)**.** Let $\mathcal{I} = \mathbb{R}^+$. A stochastic process $\mathbf{Z} = (Z_t)_{t\geq 0}$ with $Z_t : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{Z}, \mathcal{A}, \nu)$ is called a pure jump process if the sample paths $t \mapsto Z_t(\omega)$ are almost surely piece-wise constant. Furthermore, if the time-mark process $(\tau_n)_{n\in\mathbb{N}_0}$ defined by

$$\tau_n = \inf\{s \geq \tau_{n-1} : Z_s \neq Z_{s-}\}, \quad n \geq 1,$$

and $\tau_0 = 0$ satisfies

$$\mathbb{P}\left(\lim_{n\to\infty} \tau_n = \infty\right) = 1, \tag{1}$$

using the convention that $\inf \emptyset = \infty$, then we say that $\mathbf{Z}$ is non-explosive.

Remark. The non-explosive assumption in (1) is typically a technicality we need in order to define the integrals we will see later. The assumption ensures that for any bounded set $B \in \mathbb{B}(\mathbb{R}^+)$ (Borel σ-algebra on $\mathbb{R}^+$) the random variable

$$N(B) = \#\left\{s \in B : Z_s \neq Z_{-s}\right\} = \sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_n \in B\}}$$

is almost surely finite. Indeed we know that the event that $\tau_n \to \infty$ is equivalent to

$$\bigcap_{K=1}^{\infty} \bigcup_{M=K}^{\infty} (\tau_M \geq K),$$

i.e. for all $K \geq 1$ it holds that we can find τ_M that is larger than the index K i.e. τ_n diverges. This means that $N([0,t)) \leq K_t < \infty$ for some $K_t \geq 1$ since almost surely all $\tau_n \geq K_t$ for all $n \geq K_t$. This means that any increment $N([s,t)) = N([0,t)) - N([0,t))$ is bounded giving that integrating with respect to $N([0,t))$ is well-defined.

One can see that we can uniquely identify any pure jump process with no explosion with a point process $(Z_n, \tau_n)_{n \in \mathbb{N}_0}$ by setting $Z_n = Z_{\tau_n}$. For the n's for which $\tau_n = \infty$ we set $Z_n = Z_{\tau_\eta}$ with

$$\eta = \inf\{n \in \mathbb{N} : \tau_n = \infty\}. \tag{2}$$

Conversely, we can for any point process (Z_n, τ_n) construct a pure jump process by setting

$$Z_t = \sum_{n=0}^{\eta} Z_{\tau_n} \mathbb{1}_{\{\tau_n \leq t < \tau_{n+1}\}}. \tag{3}$$

with η defined as in (2). Note that we have $\eta = \infty$ if for all $i \geq 1$ the process $Z_t, t \geq \tau_i$ eventually jumps to another state. In this project, we will without loss of generality be studying the behaviour of a pure jump process on a finite state space $\mathcal{Z}$ with at least one absorbing state. In that case, η will be finite almost surely.

2.1 Counting Processes

In this section, we briefly introduce some representations of the multivariate counting process as given in Definition 4 below.

Definition 4 (Multivariate counting process)**.** Let $\mathbf{Z} = (Z_t)_{t \geq 0}$ be a non-explosive pure jump process on a finite state space $\mathcal{Z}$. The multivariate counting process $\mathbf{N} = (N_t)_{t \geq 0} = ((N_t^{jk})_{j,k \in \mathcal{Z}})_{t \geq 0}$ is the matrix-stochastic process with entries

$$N_t^{jk} = \#\big\{0 \leq s \leq t : Z_s = k, Z_{s-} = j\big\}, \qquad j \neq k,$$

and $N_t^{jk} = 0$ for all $t \geq 0$ if $j = k$.

Recall that in (3) we gave the representation of Z_t in terms of a marked point process (Z_n, τ_n). This in particular means that we can write

$$N^{jk}(t) = \sum_{n=1}^{\eta} \mathbb{1}_{[\tau_n, \infty)}(t) \mathbb{1}_{\{Z_{n-1} = j\}} \mathbb{1}_{\{Z_n = k\}}.$$

Then $N^{jk}(t)$ has dynamics on the form

$$\mathrm{d}N^{jk}(t) = \mathbb{1}_{\{Z_{t-} = j\}} \mathbb{1}_{\{Z_t = k\}} \mathrm{d}N(t), \quad N^{jk}(0) = 0.$$

In the above, $N(t)$ is simply $N(t) = \sum_{n=1}^{\eta} \mathbb{1}_{\{\tau_n,\infty)\}}(t)$ i.e. $\mathrm{d}N(t) = \sum_{n=1}^{\eta} \mathbb{1}_{\{t=\tau_n\}}$. In other words, we can define $N^{jk}(t)$ as the Lebesgue-Stieltjes integral

$$N^{jk}(t) = \int_0^t \mathbb{1}_{\{Z_{s-}=j\}} \mathbb{1}_{\{Z_s=k\}} \, \mathrm{d}N(s) = \int_0^t \mathbb{1}_{\{Z_{s-}=j\}} \, \mathrm{d}N^k(s)$$

with $N^k(t) = \sum_{j\in\mathcal{Z}} N^{jk}(t)$.

3 The Information Model

We can derive the notion of information by using a different probability measure by restricting the probability measure $\mathbb{P}$ to a smaller σ-algebra say $\mathcal{G} \subseteq \mathcal{F}$. We think of this σ-algebra as the information available about **Z**, for instance its value for some subset $A \subseteq \mathbb{B}(\mathbb{R}^+)$, or some related information telling the distribution of **Z** given this knowledge.

In insurance context, the insurance company is faced with a flow of information shedding light on the trajectory $t \mapsto Z_t(\omega)$. This information is increasing in time and we model this with a filtration.

Definition 5 (Filtration and adapted process)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A family $\mathcal{F}_t \subseteq \mathcal{F}$ of sub σ-algebras is called a filtration if for all $0 \leq s \leq t$ it holds that $\mathcal{F}_s \subseteq \mathcal{F}_t$. Furthermore, we say that a process $(Z_t)_{t\geq 0}$ is adapted to $(\mathcal{F}_t)_{t\geq 0}$ if for all $t \geq 0$ it holds that Z_t is $\mathcal{F}_t$-measurable.

We can thus define the adapted process in the case of the pure jump process.

Definition 6 (Adapted non-explosive pure jump process)**.** Let $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ be a filtration and $(Z_t)_{t\geq 0}$ be a non-explosive pure jump process. Assume that **Z** is adapted to the filtration $\mathbb{F}$. We say that **Z** is a non-explosive pure jump process adapted to $\mathbb{F}$ and we write $\mathbf{Z} \sim \mathrm{PJP}(\mathbb{F})$.

Of cause, any pure jump process is adapted with respect to the filtration generated by itself i.e. $\mathcal{F}_t^Z = \sigma(Z_u : 0 \leq u \leq t)$.

A sub-class of pure jump processes that are mathematically tractable and widely used is the ones that satisfy the Markov property.

Definition 7 (Markov property)**.** Let $\mathbb{F}$ be a filtration and let $\mathbf{Z} \sim \mathrm{PJP}(\mathbb{F})$. **Z** is said to exhibit the Markov property if

$$\forall 0 \leq t \leq s : \mathbb{P}(Z_s \in A|\mathcal{F}_t) = \mathbb{P}(Z_s \in A|Z_t). \tag{4}$$

In this case we call **Z** a *continuous time Markov chain on $\mathcal{Z}$*.

Remark. We can think of this property as the distribution of **Z** is memory-less. The property states that if we are to say anything about the state of **Z** at some future time $s \geq t$, the only thing that matters is the state at the current time, that is any path $\{Z_u : 0 \leq u \leq t\}$ will lead to the same expected trajectory in the future. This is indeed a strong property and most-likely not fulfilled by most real-life pure jump processes.

In this project, we will be considering the case, where we know the value of Z_t on $t \in [t_0, s]$ for some $t_0 \geq 0$ and $s \geq t_0$. We will be interested in how we may estimate the future behaviour of Z_t

for $t > s$ based on the sub-σ-algebra

$$\mathcal{G}_s = \sigma\left(Z_t : t_0 \leq t \leq s\right) \subseteq \mathcal{F}_t^Z.$$

We think of the information $\mathcal{G}_s$ as the information we gather purely based on the information from a contract commencing at time t_0. In particular, we want to study how we may reasonable calculate probabilities such as

$$\mathbb{P}(Z_s = k|\mathcal{G}_s)$$

and derive estimates of the reserve based on $\mathcal{G}_s$. We will in practice see what we can do when $t_0 \to s$ i.e. we base our future projections on the most recent known value of Z_t. Furthermore, our main objective is to study measurable effects on how the Markov assumption above may lead to enormous error in reserves and occupation probabilities and how we may mitigate these by loosening the assumption.

The main theme is that by loosening the Markov assumption we assume that Z_t is not memory less and therefore we need to either sub sample or smooth estimates such that the estimates of the occupation probabilities uses more information. In some cases, this sub sampling technique will lead to large approximation errors, that may outweigh the lower model error obtained compared to the model risk of a Markov estimate (that uses all available observations).

4 Transition Rates and Probabilities

Let $\mathbb{F}$ be a given filtration and let $\mathbf{Z} \sim \text{PJP}(\mathbb{F})$. We may study the expected behaviour of such a process by studying the transition probabilities. In the following we use an analogous notation and idea as presented in Bladt **and** Furrer (2023b)[1], Christiansen (2021) and Christiansen **and** Furrer (2022). Our primary goal is to study the *As-If-Markov*-model discussed in Christiansen **and** Furrer (2022).

In the insurance setting the company would have some information at it disposal regarding the process $\mathbf{Z}$. It is common to define transition probabilities based on the entire history of $\mathbf{Z}$ up until Z and then assume the Markov property to obtain an differential form for the transition rates giving rise to the product integral. However as discussed in *The Information Model* both the Markov property is probably not true and the entire history of Z is most likely not known. We will therefore develop a framework for estimating transition probabilities given some arbitrary sub-σ-algebra $\mathcal{G}_s$ which we think of containing at least the information Z_s.

Definition 8. Let $\mathcal{G}_s \subseteq \mathcal{F}$ be a sub-σ-algebra. Let $\mathbf{Z} \sim \text{PJP}(\mathbb{F})$ for some filtration $\mathbb{F}$. The *two mark probabilities given* $\mathcal{G}_s \subseteq \mathcal{F}$ is defined as

$$p^{ik}(t_0, t|\mathcal{G}_s) \overset{\text{def}}{=} \mathbb{E}\left[\mathbb{1}_{\{Z_{t_0}=i\}}\mathbb{1}_{\{Z_t=k\}} \;\middle|\; \mathcal{G}_s\right], \quad i, k \in \mathcal{Z}, t_0, t \geq 0.$$

Likewise, we define the *occupation probability given* $\mathcal{G}_s$ as

$$p^k(t|\mathcal{G}_s) \overset{\text{def}}{=} \mathbb{E}\left[\mathbb{1}_{\{Z_t=k\}} \;\middle|\; \mathcal{G}_s\right], \quad k \in \mathcal{Z}, t \geq 0.$$

1. Due to complications while compiling this document the word "and" is bold. Furthermore, the quotation marks around titles in the papers cited is rendered as question marks.

Lastly, we define the *conditional expected counting process* as

$$M^{ik}(t|\mathcal{G}_s) \stackrel{\text{def}}{=} \mathbb{E}\left[N^{ik}(t) \mid \mathcal{G}_s\right], \quad i,k \in \mathcal{Z}, t \geq 0.$$

As mentioned above, we can condition on any σ-algebra $\mathcal{G}_s$. However we are interested in the case where the only knowledge available is current state. In other words, we are interested in the occupation probabilities $p^i(t|Z_s = j)$. Therefore we will be without loss of generality be conditioning on the an event $Z_s = j$ from the state process.

We furthermore define the following called the *cumulative conditional transition rates*. Using this setup we introduced aboved.

Definition 9 (Cumulative conditional transition rates)**.** Let $\mathbf{Z} \sim \text{PJP}(\mathbb{F})$ for some filtration $\mathbb{F}$. Using the notation above the *cumulative conditional transition rates* is defined as

$$\Lambda^{ik}(t|Z_s = j) \stackrel{\text{def}}{=} \int_0^t \frac{1}{p^i(u- \mid Z_s = j)} \, \mathrm{d}_u M^{ik}(u|Z_s = j), \quad i,j,k \in \mathcal{Z}, t \geq 0. \tag{5}$$

Remark. The above cumulative transition rates are well-defined in the sense that for all $t \geq 0$ it holds that $\Lambda^{ik}(t|Z_s = j) < \infty$. We know that the process $N^{ik}(t)$ is dominated and piece-wise constant. Thus for any Borel set $B \in \mathbb{B}(\mathbb{R}^+)$ we have $\int_B \mathrm{d}N^{ik}(u) < \infty$.

We want to relate the above cumulative transition rates to the probability rates. Before this we are going to be needing the following proposition.

Proposition 10. *Let* $\mathbf{Z} \sim$ *PJP*$(\mathbb{F})$ *for some filtration* $\mathbb{F}$*. It holds that*

$$\mathbb{1}_{\{Z_t = j\}} = \mathbb{1}_{\{Z_0 = j\}} + \sum_{k \in \mathcal{Z}, k \neq j} \Big(N^{kj}(t) - N^{jk}(t)\Big).$$

Proof. Define the ordered jump times into j as

$$\Delta_j = \{s \geq 0 : Z_s = j, Z_{s-} \neq j\} := \{\tau_1^j, \tau_2^j, ...\}.$$

and out of j as

$$\Delta^{j\bullet} = \{s \geq 0 : Z_s \neq j, Z_{s-} = j\} := \{\tau_1^{j\bullet}, \tau_2^{j\bullet}, ...\}.$$

By construction we set $Z_{0-} = Z_0$ and so $0 \notin \Delta_j, \Delta^{j\bullet}$ thus giving that $\tau_1^j, \tau_1^{j\bullet} > 0$. However for the sake of argument, we define that $\tau_1^j = 0$ if $Z_0 = j$. Clearly for any $t \geq 0$ we have three scenarios: 1) $t \in [\tau_m^j, \tau_m^{j\bullet})$ for some $m \geq 1$ that is $Z_t = j$, 2) $t \in [\tau_m^{j\bullet}, \tau_{m+1}^j)$ for some $m \geq 1$, that is $Z_t \neq j$ and 3) $t \in [0, \tau_1^j)$ i.e. $Z_0 \neq j$ and the first jump is yet to arrive. Firstly, if (1) is satisfied, then

$$\sum_{k \in \mathcal{Z}, k \neq j} N^{kj}(t) = m - \mathbb{1}_{\{Z_0 = j\}}, \qquad \sum_{k \in \mathcal{Z}, k \neq j} N^{jk}(t) = m - 1.$$

Hence

$$\begin{aligned}(*) :&= \mathbb{1}_{\{Z_0=j\}} + \sum_{k\in\mathcal{Z},k\neq j}\Big(N^{kj}(t) - N^{jk}(t)\Big)\\ &= \mathbb{1}_{\{Z_0=j\}} + m - \mathbb{1}_{\{Z_0=j\}} - (m-1)\\ &= 1 = \mathbb{1}_{\{Z_t=j\}}.\end{aligned}$$

as desired. Secondly, if (2) is satisfied then the above amounts to

$$\sum_{k\in\mathcal{Z},k\neq j} N^{kj}(t) = m - \mathbb{1}_{\{Z_0=j\}}, \qquad \sum_{k\in\mathcal{Z},k\neq j} N^{jk}(t) = m.$$

Hence giving that $(*) = 0 = \mathbb{1}_{\{Z_t=j\}}$. Lastly, on the third case we see that $\sum_{k\in\mathcal{Z},k\neq j} N^{kj}(t) = 0$, $\sum_{k\in\mathcal{Z},k\neq j} N^{jk}(t) = 0$ and $\mathbb{1}_{\{Z_0=j\}} = 0$ giving that $(*) = 0$ as desired. □

Now we may state the important result, that relates the occupation probabilities with the cumulative transition rates.

Theorem 11. *Let* $\mathbf{Z} \sim$ *PJP*$(\mathbb{F})$ *for some filtration* $\mathbb{F}$. *Using the notation the processes* $p^k(t|Z_s = j)$ *and* $\Lambda^{ik}(t|Z_s = j)$ *relates through the dynamics*

$$dp^k(t|Z_s=j) = \sum_{i\in\mathcal{Z}} p^i(t-|Z_s=j)\, d\Lambda^{ik}(t|Z_s=j), \quad j,k\in\mathcal{Z}, t\geq 0.$$

In the above using the convention $d\Lambda^{kk}(t|Z_s=j) = -\sum_{i\in\mathcal{Z},i\neq k} d\Lambda^{ki}(t|Z_s=j)$.

Proof. From Proposition 10 we have that

$$\mathbb{1}_{\{Z_t=k\}} - \mathbb{1}_{\{Z_s=k\}} = \sum_{i\in\mathcal{Z},i\neq k}\Big(N^{ik}(t) - N^{ki}(t)\Big) - \sum_{k\in\mathcal{Z},k\neq j}\Big(N^{ik}(s) - N^{ki}(s)\Big),$$

thus implying

$$\begin{aligned}&\mathbb{E}\left[\mathbb{1}_{\{Z_t=k\}} - \mathbb{1}_{\{Z_s=k\}} \;\middle|\; Z_s=j\right]\\ &\quad= \mathbb{E}\left[\sum_{i\in\mathcal{Z},i\neq k}\Big(N^{ik}(t) - N^{ki}(s)\Big) - \sum_{k\in\mathcal{Z},k\neq j}\Big(N^{ik}(t) - N^{ki}(s)\Big) \;\middle|\; Z_s=j\right]\end{aligned}$$

Now, using the linearity of expectation and inserting Definition 8 we have

$$\begin{aligned}p^k(t|Z_s=j) - p^k(s|Z_s=j) = &\sum_{i\in\mathcal{Z},i\neq k}\Big(M^{ik}(t|Z_s=j) - M^{ki}(t|Z_s=j)\Big)\\ &- \sum_{i\in\mathcal{Z},i\neq k}\Big(M^{ik}(s|Z_s=j) - M^{ki}(s|Z_s=j)\Big).\end{aligned}$$

Or on differential form

$$\mathrm{d}p^k(t|Z_s=j) = \sum_{i\in\mathcal{Z},i\neq k} \mathrm{d}M^{ik}(t|Z_s=j) - \mathrm{d}M^{ki}(t|Z_s=j)$$

Using that $p^k(t-|Z_s=j)\mathrm{d}\Lambda^{ki}(t|Z_s=j) = \mathrm{d}M^{ki}(t|Z_s=j)$ we have

$$\begin{aligned}
\mathrm{d}p^k(t|Z_s=j) &= \sum_{i\in\mathcal{Z}, i\neq k} p^i(t-|Z_s=j)\mathrm{d}\Lambda^{ik}(t|Z_s=j) - p^k(t-|Z_s=j)\mathrm{d}\Lambda^{ki}(t|Z_s=j) \\
&= \left(\sum_{i\in\mathcal{Z}, i\neq k} p^i(t-|Z_s=j)\mathrm{d}\Lambda^{ik}(t|Z_s=j)\right) - p^k(t-|Z_s=j)\sum_{i\in\mathcal{Z}, i\neq k}\mathrm{d}\Lambda^{ki}(t|Z_s=j) \\
&= \sum_{i\in\mathcal{Z}} p^i(t-|Z_s=j)\mathrm{d}\Lambda^{ik}(t|Z_s=j).
\end{aligned}$$

Using the convention $\mathrm{d}\Lambda^{kk}(t|Z_s=j) = -\sum_{i\in\mathcal{Z}, i\neq k}\mathrm{d}\Lambda^{ki}(t|Z_s=j)$. □

In the following we define the probability occupation matrix and cumulative transition matrix as follows:

$$p(t|Z_s=j) \overset{\text{def}}{=} \begin{pmatrix} p^1(t|Z_s=j) \\ \vdots \\ p^J(t|Z_s=j) \end{pmatrix}, \quad \Lambda(t|Z_s=j) \overset{\text{def}}{=} \begin{pmatrix} \Lambda^{1,1}(t|Z_s=j) & \cdots & \Lambda^{1,J}(t|Z_s=j) \\ \vdots & \ddots & \vdots \\ \Lambda^{J,1}(t|Z_s=j) & \cdots & \Lambda^{J,J}(t|Z_s=j) \end{pmatrix}.$$

for some ordering of $\mathcal{Z}$ onto $\{1,...,J\} = \mathcal{J}$ with $J = \#\mathcal{Z}$. In the above, we use the convention $\Lambda^{kk}(t|Z_s=j) = -\sum_{i\in\mathcal{J}, i\neq k}\Lambda^{ki}(t|Z_s=j)$ and similar construction will later be used for the Nelson-Aalen estimator. We will be using the product integral notation as given below.

Definition 12 (Product integral/limit)**.** Let $\mathbf{A}(t) : \mathcal{M}_{n\times n}(\mathbb{R}) \to \mathcal{M}_{n\times n}(\mathbb{R})$ be a $m\times n$-matrix function taking values entry-wise in $\mathbb{R}$ that is of finite variation. Define the function $\mathbf{Y}(t) : \mathcal{M}_{m\times n}(\mathbb{R}) \to \mathcal{M}_{m\times n}(\mathbb{R})$ for a fixed real-valued $m\times n$ matrix $\mathbf{C}$ as

$$\mathbf{Y}(t) = \mathbf{C}\prod_{u\in(0,t]}\Big(\mathrm{Id} + \mathrm{d}\mathbf{A}(u)\Big) \overset{\text{def}}{=} \mathbf{C}\lim_{\max|t_i - t_{i-1}|\to 0}\prod_i\Big(\mathrm{Id} + \mathbf{A}(t_i) - \mathbf{A}(t_{i-1})\Big).$$

We say that $\mathbf{Y}(t)$ is the *product integral* or *product limit* of $\mathbf{A}$ on $(0,t]$ with boundary condition $\mathbf{C}$.

Remark. The limit above is on any partitioning of the interval $(0,t]$ with $0 = t_0 < t_1 < \cdots < t_n = t$ such that the largest increment goes to zero. The following product $\prod_i$ is over all $i = 1,...,n$ where n is the number of partitions. This definition is somewhat analogous to the one given in Definition 1 in Gill **and** Johansen (1990). The two definitions differ as we added the boundary condition as this will be used later as the boundary condition for the occupation probabilities. The boundary condition is not integral to the definition as with $\mathbf{C} = \mathrm{Id}$ we have $\mathbf{Y}(0) = \mathrm{Id}$.

We can think of the operator $\prod$ as a interval function $\prod : (0,t] \mapsto \mathcal{M}_{m\times n}(\mathbb{R})$ on $\mathbb{B}([0,\infty))$ and we can extend the definition to an arbitrary Borel set by the following.

Definition 13. Using the assumptions in Definition 12, let $B \in \mathbb{B}([0,\infty))$ an arbitrary Borel set. We define the product integral of $\mathbf{A}$ on B as

$$\mathbf{Y}(B) = \lim_{t\to\infty}\prod_{u\in(0,t]}\Big(\mathrm{Id} + \mathbb{1}_{\{t\in B\}}\,\mathrm{d}\mathbf{A}_B(u)\Big) \overset{\text{def}}{=} \prod_{u\in B}\Big(\mathrm{Id} + \mathrm{d}\mathbf{A}_B(u)\Big).$$

In particular, we say that $\mathbf{X}(t)$ is the product integral of $\mathbf{A}$ with boundary condition $\mathbf{C}$ at $s \geq 0$ if

$$\mathbf{X}(t) = \begin{cases} \mathbf{C}\mathbf{Y}((s,t]), & t > s \\ \mathbf{C}\mathbf{Y}((t,s]), & t \leq s \end{cases}.$$

In particular, $\mathbf{X}(s) = \mathbf{C}$.

In the framework above, we can derive the dynamics of the product integral.

Lemma 14. *Let $\mathbf{A}$ be as given in Definition 12 and $\mathbf{Y}(B)$ as given in Definition 13. It holds that*

$$\begin{aligned} d_t\mathbf{Y}((s,t]) &= \mathbf{Y}((s,t])\, d\mathbf{A}(t), && \textit{(the forward equation)} \\ d_s\mathbf{Y}((s,t]) &= -d\mathbf{A}(s)\mathbf{Y}((s,t]), && \textit{(the backward equation)} \end{aligned}$$

Proof. We roughly sketch the proof. The full proof may be read in Gill **and** Johansen (1990). The proof rely on the Péano series on $(s,t]$ given by:

$$\mathcal{P}(s,t;\mathbf{A}) \stackrel{\text{def}}{=} \text{Id} + \sum_{n=1}^{\infty} \int_s^t \int_s^{u_n} \cdots \int_s^{u_3} \int_s^{u_2} \text{d}\mathbf{A}(u_1)\, \text{d}\mathbf{A}(u_2) \cdots \text{d}\mathbf{A}(u_{n-1})\, \text{d}\mathbf{A}(u_n) \tag{6}$$

$$\stackrel{\text{def}}{=} \text{Id} + \sum_{n=1}^{\infty} \mathbf{A}^{(n)}(s,t). \tag{7}$$

Now one can show that the Péano series is multiplicative i.e. for some u it holds

$$\mathcal{P}(s,t;\mathbf{A}) = \mathcal{P}(s,u;\mathbf{A})\mathcal{P}(u,t;\mathbf{A}).$$

Now by partitioning of the interval $s = t_1 < \cdots < t_n = t$ one can cleverly show that

$$\left| \mathcal{P}(s,t;\mathbf{A}) - \prod_{i=1}^{n} \Big(\text{Id} + \mathbf{A}(t_i) - \mathbf{A}(t_{i-1}) \Big) \right| \leq \mathbf{K}_n,$$

for a matrix $\mathbf{K}_n$ that tends to $\mathbf{0}$ giving that **the Pèano series is equal to the product integral.** Finally, from studying the Péano series one can show the desired differential form hold for the Péano series and therefore the product integral. □

The above lemma gives directly the Kolmogorov's forward equations, when one identifies the transitions probabilities or occupation probabilities with $\mathbf{Y}$ and find the matrix function associated $\mathbf{A}$. It turns out, that the cumulative transition rates is the right choice.

Proposition 15. *Let $\mathbf{Z} \sim PJP(\mathbb{F})$ for some filtration $\mathbb{F}$, then*

$$p(t|Z_s = j)^\top = p(s-|Z_s = j)^\top \prod_s^t \Big(Id + d\Lambda(u|Z_s = j) \Big). \tag{8}$$

Proof. By Theorem 11 it holds for all $k \in \mathcal{Z}$:

$$\mathrm{d}p^k(t|Z_s = j) = \sum_{i \in \mathcal{Z}} p^i(t - |Z_s = j)\, \mathrm{d}\Lambda^{ik}(t|Z_s = j).$$

Thus in particular we can write on matrix form

$$\mathrm{d}p(t|Z_s = j)^\top = p(t - |Z_s = j)^\top d\Lambda(t|Z_s = j).$$

In particular, we see that $p(t|Z_s = j)$ is simply the product integral

$$p(t|Z_s = j)^\top = p(s - |Z_s = j)^\top \prod_s^t \Big(\mathrm{Id} + d\Lambda(u|Z_s = j)\Big).$$

as desired. □

As one can see this directly gives Kolmogorov's equations

$$\mathrm{d}p(t|Z_s = j)^\top = p(t - |Z_s = j)^\top \mathrm{d}\Lambda(t|Z_s = j).$$

5 Estimation with Censoring

We follow the setup in Bladt **and** Furrer (2023b). We start by defining what data we will be working with.

5.1 Observations and Censoring

We consider a pure jump process on a finite state-space $\mathcal{Z}$ with at least one absorbing state. In many cases we only have historic data for i.i.d. sample paths up until some time point R i.e. we have only observed Z_t on $[0, R]$. This is because the observations is based on people we in most cases only have observed at most up until their current age and so the final transition into the absorbing death-state is yet to be observed.

It is because of this fact, we need to be careful in estimating quantities such as the probability of transitioning from state j to k on a time interval $[0, t]$. We cannot simply calculate the number of people in state k at time t divided with the number of people in state j at time 0 as some of the population has not lived until time t.

Consider a pure jump process $\mathbf{Z} \sim \mathrm{PJP}(\mathbb{F})$ as described above. We set $\mathcal{Z}_0$ to be the absorbing states. Thus the following $\mathbb{F}$-stopping time τ is well-defined.

$$\tau = \inf\{s \geq 0 : Z_s \in \mathcal{Z}_0\}.$$

Let now $R > 0$ be a right-censored time such that we only observe $(Z_t)_{0 \leq t \leq R}$. One could also add a left-censored time, however we will assume that $\mathbf{Z}$ is known on the entire interval $[0, R]$. Thus we collect the following data-points

$$(\mathbb{1}_{\{Z_s = j\}}, (Z_t)_{0 \leq t \leq R}, \tau \wedge R),$$

where the $\wedge$-operator gives the minimum. We will be making the following assumption.

Assumption 1. Let $s \geq 0$ and $j \in \mathcal{Z}$. The right-censoring mechanism R is independent of $\mathbf{Z} \mid \mathbb{1}_{\{Z_s=j\}}$, that is

$$R \perp\!\!\!\perp \mathbf{Z} \mid \mathbb{1}_{\{Z_s=j\}}.$$

This assumption is imperative when studying Non-Markov jump processes. In essence, we need this condition to ensure, that whenever we sub-sample with $Z_s = j$ we do not obtain variables $\mathbf{Z}|Z_s = j$, where the censoring time is correlated with different behaviours of $\mathbf{Z}|Z_s = j$. If that was the case we would not be able to reasonable assume, that the non-censored observations past some right-censoring time $R = r$ would explain the behaviour of $Z_s|Z_s = j$ had the right-censoring time been larger.

We may introduce the following modified probabilities and transition rates.

Definition 16. The censored occupational probabilities and transition rates given $X = x$ is defined as

$$\begin{aligned} p_k^c(t|Z_s = j) &= \mathbb{E}[\mathbb{1}_{\{Z_t=k\}}\mathbb{1}_{\{t<R\}}|Z_s = j], \\ M_{ki}^c(t|Z_s = j) &= \mathbb{E}[N^{ik}(t \wedge R)|Z_s = j]. \end{aligned}$$

We now see why the assumption above is integral to the study.

Proposition 17. *Under Assumption 1 it holds that*

$$\Lambda^{ki}(t|Z_s = j) = \Lambda^{c,ki}(t|Z_s = j) = \int_0^t \frac{1}{p_k^c(s-|x)} d_s M_{ki}^c(s|Z_s = j).$$

Proof. Under Assumption 1 it holds that

$$\begin{aligned} p_k^c(t|Z_s = j) &= \mathbb{E}[\mathbb{1}_{\{Z_t=k\}}\mathbb{1}_{\{t<R\}}|Z_s = j] = \mathbb{E}[\mathbb{1}_{\{Z_t=k\}}|Z_s = j]\mathbb{E}[\mathbb{1}_{\{t<R\}}|Z_s = j] \\ &= p^k(t|Z_s = j)\mathbb{P}(t < R|Z_s = j), \end{aligned}$$

and

$$\begin{aligned} M_{ki}^c(t|Z_s = j) = \mathbb{E}[N^{ki}(t \wedge R)|Z_s = j] &= \mathbb{E}\left[\int_0^t \mathbb{1}_{\{s<R\}} \,\mathrm{d}N^{ki}(s) \,\middle|\, Z_s = j\right] \\ &= \int_0^t \mathbb{E}[\mathbb{1}_{\{s<R\}}|Z_s = j] \,\mathrm{d}\mathbb{E}[N^{ki}(s)|Z_s = j] \\ &= \int_0^t \mathbb{P}(s < R|Z_s = j) \,\mathrm{d}_s M^{ki}(s|Z_s = j). \end{aligned}$$

Thus

$$\begin{aligned} &\int_0^t \frac{1}{p_k^c(s-|Z_s = j)} \mathrm{d}_s M_{ki}^c(s|Z_s = j) \\ &\quad = \int_0^t \frac{1}{p^k(t-|x)\mathbb{P}(t < R|Z_s = j)} \mathbb{P}(s < R|Z_s = j) \,\mathrm{d}_s M^{ki}(s|Z_s = j) \end{aligned}$$

$$
\begin{aligned}
&= \int_0^t \frac{1}{p^k(t-|Z_s=j)} \, \mathrm{d}_s M^{ki}(s|Z_s=j) \\
&= \Lambda^{ki}(t|Z_s=j)
\end{aligned}
$$

as desired. □

This in particular means, that if we accept the independence assumption, then we can one to one relate the cumulative transition rates in the censored case with the ones in the uncensored case. In other words, if the data we collect are i.i.d. with Assumption 1 being satisfied, then we can define estimators, where we simply estimate $\hat{p}(t|Z_s=j)$ using observations with $R>t$.

5.2 The Nelson-Aalen and Aalen-Johansen Estimators

We studied above why the independence assumption in Assumption 1 is important when we want to define estimators for the occupation probabilities and cumulative transition rates. We therefore assume that we collect $\ell = 1, ..., L$ i.i.d. samples

$$
(\mathbb{1}_{\{Z_s^\ell=j\}}, (Z_t^\ell)_{0\le t\le R}, \tau^\ell \wedge R^\ell)_{\ell=1,\dots,L}
$$

satisfying Assumption 1. We the define the condtional Nelson-Aalen and Aalen-Johanson estimators below. We use the following shorthand notation:

$$
\begin{aligned}
L_0 &= \sum_{\ell=1}^L \mathbb{1}_{\{Z_s^\ell=j\}}, \\
\mathbb{N}_{ki}^{(L)}(t|Z_s=j) &= \frac{1}{L}\sum_{\ell=1}^L N_{ki}^\ell(t\wedge R^\ell)\frac{\mathbb{1}_{\{Z_s^\ell=j\}}}{L_0}, \\
\mathbb{I}_k^{(L)}(t|Z_s=j) &= \frac{1}{L}\sum_{\ell=1}^L \mathbb{1}_{\{t<R^\ell\}}\mathbb{1}_{\{Z_t^\ell=k\}}\frac{\mathbb{1}_{\{Z_s^\ell=j\}}}{L_0}.
\end{aligned}
$$

We furthermore define the diagonals

$$
\mathbb{N}_{kk}^{(L)}(t|Z_s=j) = -\sum_{i\in\mathcal{Z}, i\neq k} \mathbb{N}_{ki}^{(L)}(t|Z_s=j).
$$

The natural estimator for the cumulative transitions rates is then with inspiration from (5) the Nelson-Aalen estimator.

Definition 18 (Nelson-Aalen estimator)**.** The conditional Nelson-Aalen estimator is defined as

$$
\hat{\Lambda}_{ki}^{(L)}(t|Z_s=j) = \int_0^t \frac{1}{\mathbb{I}_k^{(L)}(s-|Z_s=j)} \, \mathrm{d}_s \mathbb{N}_{ki}^{(L)}(s|Z_s=j).
$$

We also have the Aalen-Johansen estimator for the occupation probabilities in (8) as follows.

Definition 19 (Aalen-Johansen estimator)**.** The conditional Aalen-Johansen estimator is defined as

$$\hat{p}^{(L)}(t|Z_s=j)^\top = \hat{p}^{(L)}(0|Z_s=j)^\top \prod_0^t \Big(\mathrm{Id} + \mathrm{d}_u\hat{\Lambda}^{(L)}(u|Z_s=j)\Big)$$

with $\hat{p}_k^{(L)}(0|Z_s=j) = \mathbb{I}_k^{(L)}(0|Z_s=j)$.

Using the differential form for the product integral in Lemma 14 we see that the Aalen-Johansen estimator takes the differential form

$$\mathrm{d}_t\hat{p}^{(L)}(t|Z_s=j)^\top = \hat{p}^{(L)}(t|Z_s=j)^\top \mathrm{d}_t\hat{\Lambda}^{(L)}(t|Z_s=j)$$

or entry-wise we have for all $k \in \mathcal{Z}$

$$\begin{aligned}\mathrm{d}_t\hat{p}_k^{(L)}(t|Z_s=j) &= \sum_{i\in\mathcal{Z}} \hat{p}_i^{(L)}(t|Z_s=j)\frac{1}{\mathbb{I}_i^{(L)}(t-|Z_s=j)}\mathrm{d}_t\mathbb{N}_{ik}^{(L)}(t|Z_s=j)\\
&= \sum_{i\in\mathcal{Z},i\neq k} \hat{p}_i^{(L)}(t|Z_s=j)\frac{1}{\mathbb{I}_i^{(L)}(t-|Z_s=j)}\mathrm{d}_t\mathbb{N}_{ik}^{(L)}(t|Z_s=j)\\
&- \hat{p}_k^{(L)}(t|Z_s=j)\frac{1}{\mathbb{I}_k^{(L)}(t-|Z_s=j)}\sum_{i\in\mathcal{Z},i\neq k}\mathrm{d}_t\mathbb{N}_{ki}^{(L)}(t|Z_s=j)\end{aligned}$$

with boundary condition $\hat{p}_j^{(L)}(0|Z_s=j) = L\cdot\mathbb{I}_j^{(L)}(0|Z_s=j)$.

Theorem 20 (As-If-Markov landmarking estimation)**.** *Assume that* $\mathbf{Z} \sim PJP(\mathbb{F})$ *for some filtration* $\mathbb{F}$. *Let* $s \geq 0$ *be a fixed time and* $j \in \mathcal{Z}$ *be a fixed state. Assume that* $(\mathbb{1}_{\{Z_s^\ell=j\}}, (Z_t^\ell)_{0\leq t\leq R^\ell}, \tau^\ell \wedge R^\ell)$ *for* $\ell = 1, .., L$ *are i.i.d samples from* $(\mathbf{Z}, R)$.

Then the samples $\mathbb{N}_{ik}^{(L)} \overset{a.s.}{\to} M_{ik}^c(t|Z_s=j)$ *and* $\mathbb{I}_i^{(L)} \overset{a.s.}{\to} p_i^c(t|Z_s=j)$. *Lastly, define* $\tau_i = \inf\bigcup_{\ell=1}^L\{t > \tau_{i-1} : Z_t^\ell \neq Z_{\tau_i}^\ell\}$ *for* $i \geq 1$ *and* $\tau_0 = 0$. *Set* $\langle t\rangle = \sum_{i=1}^M \mathbb{1}_{[\tau_i,\infty)}(t)$ *as the number of jump-times up until* t. *Then* $\hat{p}^{(L)}(t|Z_s=j)^\top$ *is given by*

$$\hat{p}^{(L)}(t|Z_s=j)^\top = L\cdot\mathbb{I}^{(L)}(0|Z_s=j)^\top\prod_{i=0}^{\langle t\rangle}\Big(Id + \Delta\hat{\Lambda}^{(L)}(\tau_i|Z_s=j)\Big). \tag{9}$$

with $\hat{p}^{(L)}(s|Z_s=j) = \mathbf{e}_j = (\mathbb{1}_{\{i=j\}})_{i\in\mathcal{Z}}$.

Proof. We have that by the strong law of large numbers that

$$\frac{1}{L}\sum_{\ell=1}^L \frac{N_{ik}^\ell(t\wedge R^\ell)\mathbb{1}_{\{Z_s^\ell=j\}}}{L_0} \overset{\text{a.s.}}{\to} \frac{\mathbb{E}\left[N_{ik}(t\wedge R)\mathbb{1}_{\{Z_s=j\}}\right]}{\mathbb{P}(Z_s=j)} = \mathbb{E}\left[N_{ik}(t\wedge R)|Z_s=j\right],$$

$$\frac{1}{L}\sum_{\ell=1}^L \frac{\mathbb{1}_{\{t<R^\ell\}}\mathbb{1}_{\{Z_t^\ell=i\}}\mathbb{1}_{\{Z_s^\ell=j\}}}{L_0} \overset{\text{a.s.}}{\to} \frac{\mathbb{E}\left[\mathbb{1}_{\{t<R\}}\mathbb{1}_{\{Z_t=i\}}\mathbb{1}_{\{Z_s=j\}}\right]}{\mathbb{P}(Z_s=j)} = \mathbb{P}(t<R, Z_t=i|Z_s=j).$$

Thus giving the desired convergence.

In order to show (9) we use that the product integral is multiplicative i.e. we have

$$\begin{aligned}\hat{p}^{(L)}(t|Z_s=j)^\top &= \hat{p}^{(L)}(0|Z_s=j)^\top \prod_0^t \Big(\mathrm{Id} + \mathrm{d}_u\hat{\Lambda}^{(L)}(u|Z_s=j)\Big) \\ &= \hat{p}^{(L)}(0|Z_s=j)^\top \prod_0^s \Big(\mathrm{Id} + \mathrm{d}_u\hat{\Lambda}^{(L)}(u|Z_s=j)\Big) \prod_s^t \Big(\mathrm{Id} + \mathrm{d}_u\hat{\Lambda}^{(L)}(u|Z_s=j)\Big) \\ &= \hat{p}^{(L)}(s|Z_s=j)^\top \prod_s^t \Big(\mathrm{Id} + \mathrm{d}_u\hat{\Lambda}^{(L)}(u|Z_s=j)\Big).\end{aligned}$$

Hence if $\hat{p}^{(L)}(s|Z_s=j)=\mathbf{e}_j$ then (9) follows.

Now, consider all unique jump-times occurring on $[0,s]$: $0=\tau_0\le\tau_1\le ...\le\tau_{\langle s\rangle}\le s$. Recall, that $\hat{\Lambda}^{(L)}$ is a pure jump function due to $\mathbb{I}$ and $\mathbb{N}$ being jump functions with discontinuity points τ_i for $i\ge 1$. We thus have the product integral

$$\prod_0^t \Big(\mathrm{Id} + \mathrm{d}_u\hat{\Lambda}^{(L)}(u|Z_s=j)\Big) = \prod_{i=0}^{\langle t\rangle} \Big(\mathrm{Id} + \Delta\hat{\Lambda}^{(L)}(\tau_i|Z_s=j)\Big).$$

where for $i=0$ we have $\Delta\hat{\Lambda}^{(L)}(\tau_i|Z_s=j)=\mathbf{0}$. We want to show the statement

$$\forall t\in[0,s]:\hat{p}_k^{(L)}(t|Z_s=j) = \frac{1}{L_0}\sum_{\ell=1}^{L}\mathbb{1}_{\{Z_s^\ell=j\}}\mathbb{1}_{\{Z_t^\ell=k\}} = \frac{1}{L_0}\sum_{\ell\in\mathbb{L}_0}\mathbb{1}_{\{Z_t^\ell=k\}} \tag{10}$$

with $\mathbb{L}_0=\{\ell:Z_s^\ell=j\}$. Obviously, if (10) holds then

$$\hat{p}_j^{(L)}(s|Z_s=j) = \frac{1}{L_0}\sum_{\ell\in\mathbb{L}_0}\mathbb{1}_{\{Z_s^\ell=j\}} = \frac{L_0}{L_0} = 1$$

being the desired conclusion. Equation (10) is shown using induction on $i=0,...,\langle s\rangle$. First notice that for all $t\in[0,s]$ it holds that

$$L\cdot\mathbb{I}_k^{(L)}(t) = \frac{1}{L_0}\sum_{\ell=1}^{L}\mathbb{1}_{\{Z_s^\ell=j\}}\mathbb{1}_{\{Z_t^\ell=k\}}\mathbb{1}_{\{t<R^\ell\}} = \frac{1}{L_0}\sum_{\ell\in\mathbb{L}_0}\mathbb{1}_{\{Z_t^\ell=k\}}.$$

as $R^\ell>s$ for all $\ell\in L_0$. Then (10) amounts to showing

$$\forall t\in[0,s]:\hat{p}_k^{(L)}(t|Z_s=j) = L\cdot\mathbb{I}_k^{(L)}(t).$$

The base case $i=0$: According to the definition of the conditional Aalen-Johansen estimator we have

$$\hat{p}_k^{(L)}(0|Z_s=j) = L\left(\mathbb{I}^{(L)}(0)\prod_0^0\Big(\mathrm{Id}+\mathrm{d}_u\hat{\Lambda}(u|Z_s=j)\Big)\right)_k = L\cdot\mathbb{I}_k^{(L)}(0).$$

Showing that (10) holds for the base case $t = 0$.

Induction step: We now turn to the induction step on $i = 1, ..., \langle s \rangle$. Assume that (10) holds for some $1 \leq i < \langle s \rangle$. Then we have that

$$\begin{aligned}
\hat{p}^{(L)}(\tau_{i+1}|Z_s = j) &= L \cdot \mathbb{I}^{(L)}(0) \prod_{m=0}^{i+1} \Big(\mathrm{Id} + \Delta\hat{\Lambda}(\tau_m|Z_s = j)\Big) \\
&= L \cdot \mathbb{I}^{(L)}(0) \left(\prod_{m=0}^{i} \Big(\mathrm{Id} + \Delta\hat{\Lambda}(\tau_m|Z_s = j)\Big)\right) \Big(\mathrm{Id} + \Delta\hat{\Lambda}(\tau_{i+1}|Z_s = j)\Big) \\
&= \hat{p}^{(L)}(\tau_i|Z_s = j)\Big(\mathrm{Id} + \Delta\hat{\Lambda}(\tau_{i+1}|Z_s = j)\Big).
\end{aligned}$$

We thus have for $k \in \mathcal{Z}$:

$$\hat{p}_k^{(L)}(\tau_{i+1}|Z_s = j) = \hat{p}_k^{(L)}(\tau_i|Z_s = j) + \sum_{i\in\mathcal{Z}} \hat{p}_i^{(L)}(\tau_i|Z_s = j)\Delta\hat{\Lambda}_{ik}^{(L)}(\tau_{i+1}|Z_s = j).$$

Using the induction assumption we have

$$\begin{aligned}
&\hat{p}_k^{(L)}(\tau_{i+1}|Z_s = j) = \frac{1}{L_0}\sum_{\ell\in\mathbb{L}_0} \mathbb{1}_{\{Z^\ell_{\tau_i}=k\}} + \sum_{i\in\mathcal{Z}} \left(\frac{1}{L_0}\sum_{\ell\in\mathbb{L}_0} \mathbb{1}_{\{Z^\ell_{\tau_i}=i\}}\right) \frac{\mathbb{N}_{ik}^{(L)}(\tau_{i+1}) - \mathbb{N}_{ik}^{(L)}(\tau_i)}{\mathbb{I}_i^{(L)}(\tau_i)} \\
&= \frac{1}{L_0}\left(\sum_{\ell\in\mathbb{L}_0} \mathbb{1}_{\{Z^\ell_{\tau_i}=k\}} + \sum_{i\in\mathcal{Z}}\left(\sum_{\ell\in\mathbb{L}_0} \mathbb{1}_{\{Z^\ell_{\tau_i}=i\}}\right) \frac{\frac{1}{L\cdot L_0}\sum_{\ell\in\mathbb{L}_0} N^\ell_{ik}(\tau_{i+1}) - N^\ell_{ik}(\tau_i)}{\frac{1}{L\cdot L_0}\sum_{\ell\in\mathbb{L}_0} \mathbb{1}_{\{Z^\ell_{\tau_i}=i\}}}\right) \\
&= \frac{1}{L_0}\left(\sum_{\ell\in\mathbb{L}_0} \mathbb{1}_{\{Z^\ell_{\tau_i}=k\}} + \sum_{i\in\mathcal{Z}}\sum_{\ell\in\mathbb{L}_0} N^\ell_{ik}(\tau_{i+1}) - N^\ell_{ik}(\tau_i)\right) \\
&= \frac{1}{L_0}\left(\sum_{\ell\in\mathbb{L}_0} \mathbb{1}_{\{Z^\ell_{\tau_i}=k\}} + \sum_{i\in\mathcal{Z}, i\neq k}\sum_{\ell\in\mathbb{L}_0} \mathbb{1}_{\{Z^\ell_{\tau_{i+1}}=k\}}\mathbb{1}_{\{Z^\ell_{\tau_i}=i\}} - \sum_{m\in\mathcal{Z}, m\neq k}\sum_{\ell\in\mathbb{L}_0} \mathbb{1}_{\{Z^\ell_{\tau_{i+1}}=m\}}\mathbb{1}_{\{Z^\ell_{\tau_i}=k\}}\right) \\
&= \frac{1}{L_0}\left(\sum_{\ell\in\mathbb{L}_0} \mathbb{1}_{\{Z^\ell_{\tau_i}=k\}} + \sum_{i\in\mathcal{Z}, i\neq k}\sum_{\ell\in\mathbb{L}_0} \Big(\mathbb{1}_{\{Z^\ell_{\tau_{i+1}}=k\}}\mathbb{1}_{\{Z^\ell_{\tau_i}=i\}} - \mathbb{1}_{\{Z^\ell_{\tau_{i+1}}=i\}}\mathbb{1}_{\{Z^\ell_{\tau_i}=k\}}\Big)\right)
\end{aligned}$$

using that $\mathbb{I}_i^{(L)}(\tau_{i+1}-) = \mathbb{I}_i^{(L)}(\tau_i)$. From the above we see that during the step $(\tau_i, \tau_{i+1}]$ we add one for each of the transitions into k from any state $i \neq k$ and subtract on for each transitions from k to any other state $i \neq k$, thus collecting the total in and outflow for state k on $(\tau_i, \tau_{i+1}]$. In total we have

$$\hat{p}_k^{(L)}(\tau_{i+1}|Z_s = j) = \frac{1}{L_0}\sum_{\ell\in\mathbb{L}_0} \mathbb{1}_{\{Z^\ell_{\tau_{i+1}}=k\}}.$$

This shows the induction step. Thus giving that (10) holds for all $t \in [0, s]$. This concludes the proof. $\square$

We call the modelling framework in Theorem 20 the *As-If-Markov*-model. However, although this seem to imply some Markov-like assumption on the underlying chain $\mathbf{Z}$, this term actually refers to the structure of how we define the conditional occupation probabilities.

Recall, that the Markov property states that the future trajectory only depends on the state at the time of forecasting not the entire path of Z_t up until the time of forecasting. Therefore, in the Markov case we have that the transition probabilities simplifies to

$$\Big(\mathbb{P}(Z_t = k|Z_s = j)\Big)_{j,k\in\mathcal{Z}} = p(s,t) = \prod_s^t \Big(\mathrm{Id} + \mathrm{d}\Lambda(u)\Big),$$

with the cumulative transition rates given by the integral

$$\Lambda^{jk}(t) = \int_s^t \frac{1}{\mathbb{P}(Z_{u-} = j)} \,\mathrm{d}\mathbb{E}[N^{jk}(u)].$$

Thus in the As-If-Markov model we can treat the transition probabilities as in the Markov case however with the important difference: the transition probabilities depends on some landmark of the distribution of $\mathbf{Z}$ and evolves with the associated cumulative transition rates given this landmark. Due to this the As-If-Markov model gives rise to a $\mathcal{Z}\times\mathbb{R}^+$ indexed family of transition probabilities each with its own unique transition rates $\mathrm{d}\Lambda(\cdot|Z_s = j)$. In the Markov model the transition probabilities all evolve with the same transition rates $\mathrm{d}\Lambda(\cdot)$. In particular, this means that the data required for a converging model in the As-If-Markov model is many times larger using purely landmarking techniques. That is if the underlying chain indeed is Markov!

In the following we will heavily rely on benchmarks of the regular Markov estimates when assessing estimation risk and/or model risk. Therefore we need to standardise some notation for what we mean by the Markov estimate and the As-If-Markov estimate. The As-If-Markov estimate, the estimators given in Theorem 20, will be denoted with superscript $^\bullet$ while the Markov estimate will be denoted with a $^\circ$ superscript. We introduce the Markov estimator as follows:

$$\mathbb{N}_{ik}^{\circ(L)}(t) = \frac{1}{L}\sum_{\ell=1}^{L} N_{ik}^{\ell}(t\wedge R^\ell), \quad \mathbb{I}_i^{\circ(L)}(t) = \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{1}_{\{t<R^\ell\}}\mathbb{1}_{\{Z_t^\ell = i\}}$$

and

$$\hat{p}^{\circ(L)}(s,t) = \prod_{i=\langle s\rangle+1}^{\langle t\rangle}\Big(\mathrm{Id} + \mathrm{d}\hat{\Lambda}^{\circ(L)}(u)\Big), \quad \hat{\Lambda}_{jk}^{\circ(L)}(t) = \sum_{i=1}^{\langle t\rangle}\frac{\Delta\mathbb{N}_{jk}^{\circ(L)}(\tau_i)}{\mathbb{I}_j^{\circ(L)}(\tau_{i-1})}$$

with $\hat{p}^{\circ(L)}(t|Z_s = j)^\top = \mathbf{e}_j^\top \hat{p}^{\circ(L)}(s,t)$ for $t\geq s$. Furthermore we state the As-If-Markov estimator explicitly as

$$\mathbb{N}_{ik}^{\bullet(L)}(t|Z_s = j) = \frac{1}{L_0}\sum_{\ell\in\mathbb{L}_0} N_{ik}^{\ell}(t\wedge R^\ell), \quad \mathbb{I}_i^{\bullet(L)}(t|Z_s = j) = \frac{1}{L_0}\sum_{\ell\in\mathbb{L}_0} \mathbb{1}_{\{t<R^\ell\}}\mathbb{1}_{\{Z_t^\ell = i\}}$$

In the above we use the notation $\mathbb{L}_0 = \{\ell : Z_s^\ell = j\}$. We also have the Aalen-Johansen estimator and Nelson-Aalen estimator respectively

$$\hat{p}^{\bullet(L)}(s,t|Z_s = j) = \prod_{i=\langle s\rangle+1}^{\langle t\rangle}\Big(\mathrm{Id} + \Delta\hat{\Lambda}^{\bullet(L)}(\tau_i|Z_s = j)\Big),$$

$$\hat{\Lambda}^{\bullet(L)}_{jk}(t|Z_s = j) = \sum_{i=1}^{\langle t\rangle} \frac{\Delta\mathbb{N}^{\bullet(L)}_{jk}(\tau_i|Z_s = j)}{\mathbb{I}^{\bullet(L)}_{j}(\tau_{i-1}|Z_s = j)}.$$

As in the Markov estimator we see that $\hat{p}^{\bullet(L)}(t|Z_s = j)^\top = \mathbf{e}_j^\top \hat{p}^{\bullet(L)}(s,t|Z_s = j)$.

Of cause we set $\hat{p}^{\bullet(L)}(s,t|Z_s = j) = \hat{p}^{\circ(L)}(s,t) = \mathrm{Id}$ when $\langle t\rangle < \langle s\rangle + 1$. Furthermore, our analysis will only consider $t \geq s$ as this reflects the task of predicting based on the knowledge that $Z_s = j$.

Whenever we suppress the superscript $^\bullet$ or $^\circ$ we state statements that hold for both the Markov and As-If-Markov estimator. For instance when we write $\hat{p}^{(L)}(s,t|Z_s = j)$ we state statements that holds for both $\hat{p}^{\bullet(L)}(s,t|Z_s = j)$ and $\hat{p}^{\circ(L)}(s,t)$.

One clear fact one sees is that the Markov estimator jumps more frequently than the As-If estimator. This is because essentially the two estimators only differ on the sample on which they are calculated on. In particular we have that if we let $\mathcal{D}$ be the set of observations then if the subset $\mathcal{D}_{\{Z_s=j\}}$ being the observations satisfying $Z_s^\ell = j$ and $R^\ell > s$ is in fact equal to $\mathcal{D}$, then the two estimates are equal.

5.3 Norms

In the discussions of risk in estimation we will be evaluating the following norms. Let $\mu : [a,b] \to \mathbb{R}$ be an integrable function. We define the *supremum norm* as follows

$$|\mu|_\infty = \sup_{t\in[a,b]} \mu(t).$$

We can thus evaluate distance between two function μ and ν as $|\mu - \nu|_\infty$. We will also provide plots of the function $|\mu - \nu|(t)$ i.e. the function

$$|\mu - \nu|(t) = |\mu(t) - \nu(t)|.$$

The supremum norm is important when showing uniform convergence, however it may fail in communicating the point where the convergence is slower than other points. To this we can plot the function $|\mu - \nu|$, but we can also evaluate the mean p-difference by evaluating the following

$$|\mu|_p = \frac{1}{b-a}\left(\int_a^b |\mu(t)|^p \, dt\right)^{1/p}.$$

6 Estimation of Probabilities and Cash-Flows

In this section we implement the Aalen-Johansen estimator using the results from Theorem 20. This will be done with a familiar active-disability model. We start the section with a practical method for estimating the Markov and the As-If-Markov estimators. Afterwards we estimate cash-flow for insurance contract.

6.1 *Implementing the Nelson-Aalen and Aalen-Johansen Estimators*

We start by implementing the estimators for a simple model with three states say $\mathcal{Z} = \{a, b, c\}$ with c being an absorbing state (see Figure 1). We assume that we have L samples satisfying Assumption 1:

$$(\mathbb{1}_{\{Z_s^\ell = j\}}, (Z_t^\ell)_{0 \le t \le R}, \tau^\ell \wedge R^\ell)_{\ell=1,\dots,L}.$$

We also assume that $Z_0 = a$ almost surely. Notice that we use a numbering system ($a = 1$, $b = 2$, $c = 3$) when implementing the calculations in statistical software (see appendicies).

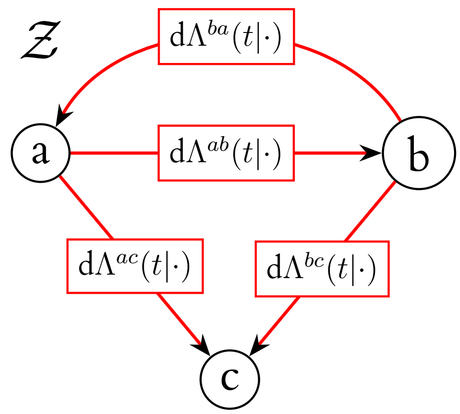

Figure 1: A representation of the chain discussed in sub-section *Implementing the Nelson-Aalen and Aalen-Johansen Estimators*

Table 1: The observations stored in `main_df`

Id	Start_Time	End_Time	Start_State	End_State	Censored
1	0	0.6	1	2	FALSE
1	0.6	2	2	3	FALSE
1	2	∞[a]	3	3	FALSE
2	0	1	1	2	FALSE
2	1	3	2	2	TRUE[b]
2	3	∞	2	2	TRUE[b]

a The infinity sign indicates that the chain was absorbed at time 2.
b This indicates that the chain was censored at time 3 while $Z_{3-} = b$.

We use a matrix-based approach in calculating the increments of the estimators. The observations will be stored in a table called `maindf`. Table 1 show an example of an observation dataset of two observations. The first jumps from state 1 to state 2 at time 0.6 and then at time 2 to state 3, where i sojourns indefinetely. The other observation jumps to state 2 at time 1 and then is censored at time 3.

Recall, from Theorem 20 that the Aalen-Johansen estimator is given by the product integral of the cumulative transition rates. However, as we expained in the rates only changes on distinct jump-times of the observed sample paths Z_t^ℓ. Thus by cleverly coding the function $\mathbb{I}_m^{(L)}(\tau - |Z_s = j)$ and

$\Delta\mathbb{N}^{(L)}_{mk}(\tau|Z_s = j)$ we can easily implement the the function $\hat{p}^{(L)}(\tau|Z_s = j)$ for all unique τ. In practice, we use matrix multiplication of the occupation probabilities: that is we use the equation

$$\hat{p}^{(L)}(\tau_{i+1}|Z_s = j) = \hat{p}^{(L)}(\tau_i|Z_s = j)\Big(\mathrm{Id} + \Delta\hat{\Lambda}(\tau_{i+1}|Z_s = j)\Big).$$

6.2 Approximating Cash-Flows

In life insurance the contractual payments are defined conditional on the behaviour of a pure jump process $\mathbf{Z}$. The classic example is the pure endowment life insurance, where the policy holder is payed a sum at time T if the holder is alive at time T. This could be modelled with at two state process commencing in state a (alive state) and the payment is made if $Z_T = a$.

In general we can define the payment process as follows:

$$\mathrm{d}B(t) \stackrel{\text{def}}{=} \sum_{k\in\mathcal{Z}} \left(\mathbb{1}_{\{Z_t=k\}}\mathrm{d}B^k(t) + \sum_{i\in\mathcal{Z}, i\neq k} b^{ki}(t)\,\mathrm{d}N^{ki}(t) \right),$$

where b^{ki} are deterministic functions. Typically we would construct the occupational payments we continuously payed rates that is $\mathrm{d}B^k(t) = b^k(t)\,\mathrm{d}t$ with b^k being deterministic.

The above is of cause dependent on ω and thus we are in practice interested in the conditional expectation of such process wrt. realisation of the state process $Z_s = j$. Thus we may define the cumulative (expected) cash-flow conditioned on $Z_s = j$ as

$$\begin{aligned}
\mathrm{d}A(t|x) &\stackrel{\text{def}}{=} \mathbb{E}[\mathrm{d}B(t)|Z_s = j] \\
&= \sum_{k\in\mathcal{Z}} \left(\mathbb{E}\left[\mathbb{1}_{\{Z_t=k\}}|Z_s = j\right]\mathrm{d}B^k(t) + \sum_{i\in\mathcal{Z}, i\neq k} b^{ki}(t)\mathbb{E}\left[\mathrm{d}N^{ki}(t)|Z_s = j\right] \right).
\end{aligned}$$

We recognise the above in terms of the conditional occupation probabilities and cumulative transition rates as defined in *Transition Rates and Probabilities*. We thus have the process

$$\mathrm{d}A(t|Z_s = j) = \sum_{k\in\mathcal{Z}} \left(p^k(t|Z_s = j)\mathrm{d}B^k(t) + \sum_{i\in\mathcal{Z}, i\neq k} b^{ki}(t)\,\mathrm{d}M^{ki}(t|Z_s = j) \right).$$

Inserting $\mathrm{d}M^{ki}(t|Z_s = j) = p^k(t-|Z_s = j)\,\mathrm{d}\Lambda^{ki}(t|Z_s = j)$ we obtain the familiar equation

$$\mathrm{d}A(t|Z_s = j) = \sum_{k\in\mathcal{Z}} \left(p^k(t|Z_s = j)\mathrm{d}B^k(t) + \sum_{i\in\mathcal{Z}, i\neq k} b^{ki}(t)p^k(t-|Z_s = j)\,\mathrm{d}\Lambda^{ki}(t|Z_s = j) \right).$$

For reserving and pricing we need to calculate the conditional expected reserve as follows

$$V(t|Z_s = j) \stackrel{\text{def}}{=} \mathbb{E}\left[\int_t^\infty e^{-\int_t^u \mathrm{d}\kappa(v)}\,\mathrm{d}B(u) \,\middle|\, Z_s = j \right],$$

with κ being the cumulative rate function. We know from Fubini that, the above is not equal to the expected integrand integrated wrt. expected integrator. However by assuming $\mathbf{Z} \perp\!\!\!\perp (\kappa(t))_{t\geq 0}$ we have

$$V(t|Z_s = j) = \int_t^\infty \mathbb{E}\left[e^{-\int_t^u \mathrm{d}\kappa(v)} \middle| Z_s = j \right] \mathbb{E}\left[\mathrm{d}B(u) | Z_s = j\right] = \int_t^\infty e^{-\int_t^u \mathrm{d}\rho(v|Z_s=j)} \,\mathrm{d}A(u|Z_s = j) \tag{11}$$

with $\rho(u|Z_s = j)$ being the forward cumulative rate function defined by the above equation. If indeed the assumption that rates are independent of the state process, then we can decompose the problem of reserving into market risk and mortality risk. In other words, we can study the cash-flow independently and only worry ourselves with the convergence of $\mathrm{d}\hat{A}(t|Z_s = j)$ to $\mathrm{d}A(t|Z_s = j)$.

To the end of studying cash-flows we consider landmark estimates as in the As-If Markov with $Z_s = j$, thus studying $\mathrm{d}A(t|Z_s = j)$. If the Markov property is fulfilled we furthermore know that $V(t|Z_s) = V(t|\mathcal{F}_s^Z)$ being the regular reserve. In this project, we study the cash-flow and will occasionally refer to the reserve as a measure of the economic risk associated with each estimator. Therefore, we will assume deterministic and constant rates that is $\mathrm{d}\rho(u|Z_s = j) = r\,\mathrm{d}u$ and so inserting in (11) we have

$$V(t|Z_s = j) = \int_t^\infty e^{-(u-t)r}\,\mathrm{d}A(u|Z_s = j)$$

and we will be referring to the reserve $V(s|Z_s = j)$ whenever we study the impact of the estimator on reserve. In practice, we assume that $\mathrm{d}A(u|Z_s = j) = 0$ for all $u \geq \max\{\tau_i : i = 0, ..., M\}$ where M is the number of total unique transitions in the dataset used in estimation.

7 Numerical Examples

In the previous two section we introduced the estimators $\hat{p}^{\bullet(L)}(s,t|Z_s = j)$, $\hat{p}^{\circ(L)}(s,t|Z_s = j)$, $\hat{\Lambda}^{\bullet(L)}(t|Z_s = j)$ and $\hat{\Lambda}^{\circ(L)}(t|Z_s = j)$ being the conditional Aalen-Johansen and Nelson-Aalen estimators respectively. In this section, we give comprehensive examples of Non-Markov chains, that give rise to a nuanced discussion of model risk versus estimation risk.

We will throughout the section assume that we have L sample paths that are independently right-censored by some mechanism R. We thus study an i.i.d. data sample of the following structure.

$$(\mathbb{1}_{\{Z_s^\ell = j\}}, (Z_t^\ell)_{0\leq t\leq R}, \tau^\ell \wedge R^\ell)_{\ell=1,...,L}.$$

We will be using the package `AalenJohansen` from Bladt **and** Furrer (2023a) to simulate the samples.

7.1 *Time In-Homogeneous Markov Chain*

We start by generating samples from a Markov chain that is time-inhomogeneous but on the form $\mathrm{d}\Lambda(t) = \lambda(t)\mathbf{M}\,\mathrm{d}t$. We know that the product integral of such a Matrix and we have that

$$p(s,t) = \exp\left(\mathbf{M} \int_0^t \lambda(u)\,\mathrm{d}u \right). \tag{12}$$

In the above $p(s,t) = (\mathbb{P}(Z_t = j | Z_s = i))_{i,j \in \mathcal{Z}}$ and $p(s,s) = \mathrm{Id}$. In other words, we can easily compute the true values of the transition probabilities using regular matrix exponentiation. For this examples we choose

$$\lambda(t) = \frac{1}{1+\frac{1}{2}t} \quad \text{and} \quad \mathbf{M} = \begin{pmatrix} -3 & 2 & 1 \\ 3 & -4 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

We start by sampling from this distribution using a censoring mechanism with $R \sim \mathrm{Unif}(0,10)$ and an initial distribution $\pi_0 = (1,0,0)$. We have layed out the code used in simulating the sample paths in appendices.

Recall, that the cumulative transition rates is the matrix exponent of the product of the integral of λ over the interval and the matrix $\mathbf{M}$. We see that using a substitution argument with $y = 1 + \frac{1}{2}u$ hence $\frac{dy}{du} = \frac{1}{2}$ and thus

$$\int_s^t \lambda(u)\,\mathrm{d}u = 2\int_{1+\frac{1}{2}s}^{1+\frac{1}{2}t} \frac{1}{y}\,\mathrm{d}y = 2\log\left(1+\frac{1}{2}t\right) - 2\log\left(1+\frac{1}{2}s\right).$$

Thus we can easily compare the Aalen-Johansen estimator with the true value of $p(s,t)$. In fact, we have that

$$p(s,t) = \exp\left\{2\left(\log\left(1+\frac{1}{2}t\right) - \log\left(1+\frac{1}{2}s\right)\right)\mathbf{M}\right\}.$$

7.1.1 Comparing Occupation Probabilities and Transition Rates

We start by calculating $\mathbb{I}_m^{(L)}(\tau - | Z_s = j)$ and $\mathbb{N}_{mk}^{(L)}(\tau - | Z_s = j)$ in the As-If-Markov case or $\mathbb{I}_m^{(L)}(\tau-)$ and $\mathbb{N}_{mk}^{(L)}(\tau-)$ in the Markov case. The code is given in the appendices. Then we may calculate the increments for the Nelson-Aalen estimator (see appendices) and finally compute the transition probabilities $\hat{p}^{(L)}(s,t|Z_s = j)$ for the As-If case and Markov case respectively (see appendices). For the estimates $\hat{p}^{(L)}(t|Z_s = j)$ we simply follow $\hat{p}^{(L)}(t|Z_s = j)^\top = \mathbf{e}_j^\top \hat{p}^{(L)}(s,t|Z_s = j)$.

As we can see in Figure 2 both the Markov model and As-If Markov model does well in estimating the true transition probabilities $\mathbb{P}(Z_t = j|Z_s = i)$ (here tested for $Z_s = a$ for $s = 0,1,3,6$). It is obvious to the eye that the both the Markov model and As-If Markov model approximate the transition probabilities.

However by construction the As-If model uses only data from observations that has was sojourning in state j at time s to asses the probability that the chain sojourns in some other state i at time $t \geq s$, while the Markov model uses all observations. Thus when the Markov property indeed is fulfilled both estimates would tend to the same limit, but the As-If model with approximately a $\mathbb{P}(Z_s = j, R > s)/\mathbb{P}(R > s)$'th the rate of the Markov estimate.

We can investigate this proposed convergence property by simulating a large amount of sample paths, say $L_{\max} = 100,000$, and then for an increasing sequence $L = L_1, ..., L_M = L_{\max}$ calculate $|\Lambda_{jk}(\cdot|Z_s = j) - \hat{\Lambda}^{(L_i)}(\cdot|Z_s = j)|_\infty$. If the above mentioned convergence hold we would assume that

$$\left|\Lambda_{jk}(\cdot|Z_s = j) - \hat{\Lambda}^{\bullet(L)}(\cdot|Z_s = j)\right|_\infty = \left|\Lambda_{jk}(\cdot|Z_s = j) - \hat{\Lambda}^{\circ(\lfloor L\cdot\mathbb{P}(Z_s=j)\rfloor)}(\cdot|Z_s = j)\right|_\infty$$

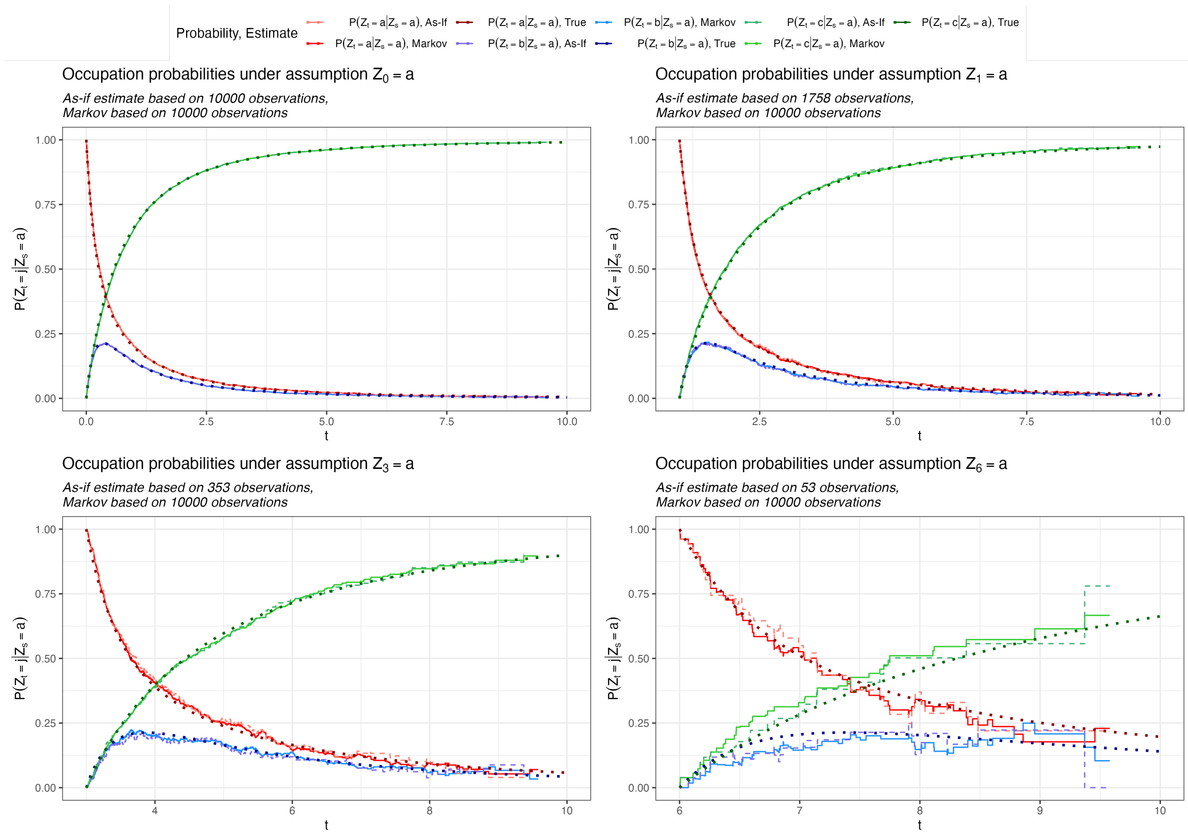

Figure 2: The figure shows the conditional probabilities under a Markov assumption and As-If-Markov assumption. The dashed line is estimate from the As-If model, the regular line is under the Markov assumption and the dots represent the true value (based on fourth order Runge-Kutta approximation). Based on $L = 10,000$ samples. (code given in appendices)

as $L \to \infty$. In the above, we use that by construction $\mathbf{Z}$ and R is independent. Furthermore, Z_0 has distribution $\pi_0 = (1, 0, 0)$ and with the structure of Λ we easily have optain the probabilities from equation (12) by $\mathbb{P}(Z_s = j) = \Big((1, 0, 0)^\top p(0, t)\Big)_j$.

Let us take $s = 6$ and $j = a, b$. We then have

$$\mathbb{P}(Z_6 = a) \approx 0.0155, \quad \mathbb{P}(Z_6 = b) \approx 0.0114.$$

In other words, the As-If estimate should asymptotically need around $1/0.0155 \approx 65$ or $1/0.0114 \approx 88$ as many observations as the Markov estimate.

In order to test this hypothesis, we could run Monte Carlo simulations of the supremum norms of the difference of the estimate and the true value and take the average. Thus we consider the following. Let N be the number of Monte Carlo simulations. We draw the following.

$$\Big(1_{\{Z_s^{\ell,n}=j\}}, (Z_t^{\ell,n})_{0\le t\le R^{\ell,n}}, \tau^{\ell,n} \wedge R^{\ell,n}\Big)_{\ell=1,\dots,L,n=1,..,N}.$$

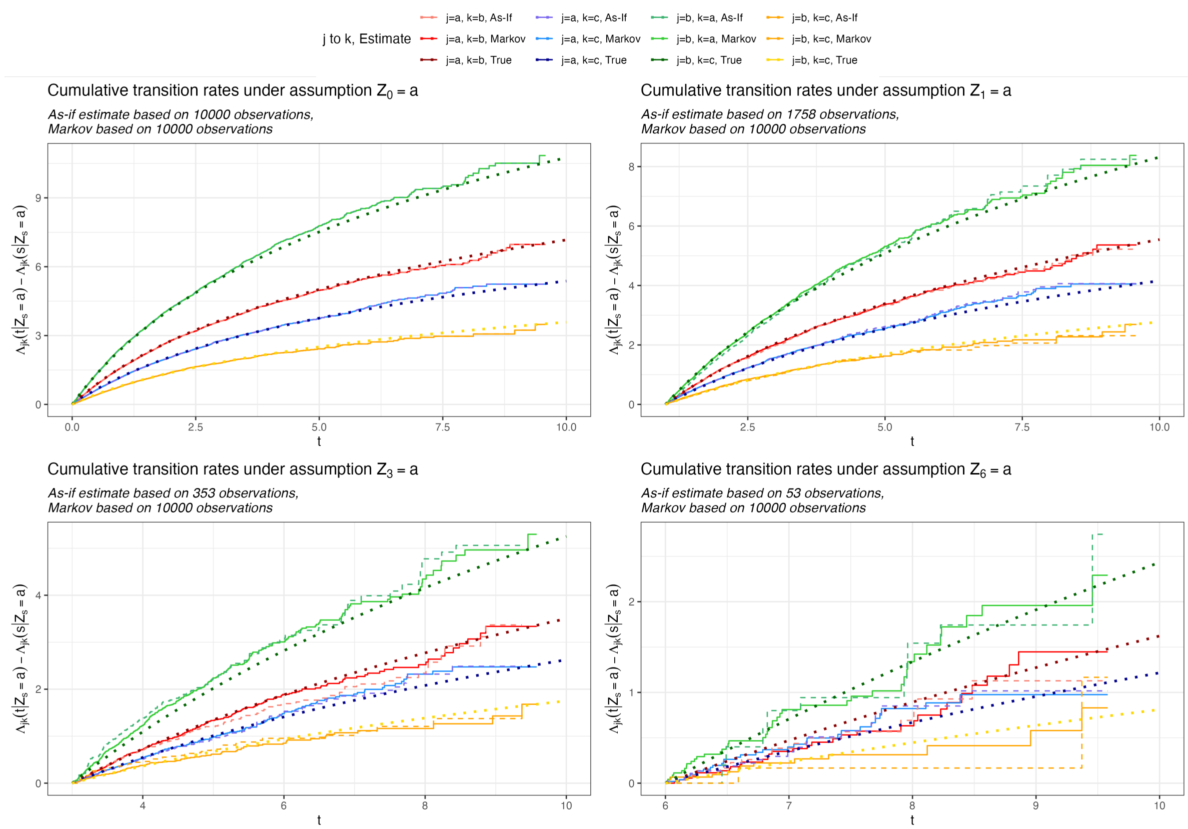

Figure 3: The figure shows the cumulative transition rates $\Lambda_{jk}(t|Z_s = j) - \Lambda_{jk}(s|Z_s = j)$ under a Markov assumption and As-If-Markov assumption. The dashed line is estimate from the As-If model, the regular line is under the Markov assumption and the dots represent the true value (based on fourth order Runge-Kutta approximation). Based on $L = 10,000$ samples. (code given in appendices)

We can then base our estimate of the supremum norm on a grid of different sample sizes $0 < L_1 < ... < L_M = L$ and consider the following

$$S(L_m, v) = \left|\Lambda^{ik}(\cdot|Z_s = j) - \hat{\Lambda}_{ik}^{(L_m,v)}(\cdot|Z_s = j)\right|_\infty, \quad \text{for} \quad i = a, b, \ k = a, b, c, \ k \neq i.$$

where the estimate (L_m, v) is based on the observations with $\ell = 1, ..., L_m$ and $n = v$. We see that for all $m = 1, ..., M$ it holds that

$$S^{(N)}(L_m) = \frac{1}{N}\sum_{v=1}^{N} S(L_m, v) \to \mathbb{E}\left|\Lambda^{ik}(\cdot|Z_s = j) - \hat{\Lambda}_{ik}^{(L_m)}(\cdot|Z_s = j)\right|_\infty$$

as $N \to \infty$. We know how to calculate the difference since Λ and $\hat{\Lambda}$ is non-decreasing. We also know that $\hat{\Lambda}$ is piecewise constant and so the supremum becomes

$$S(L_m, v) = \max\left\{\max_\tau \left|\Lambda^{ik}(\tau|Z_s = j) - \hat{\Lambda}_{ik}^{(L_m,v)}(\tau|Z_s = j)\right|,\right.$$

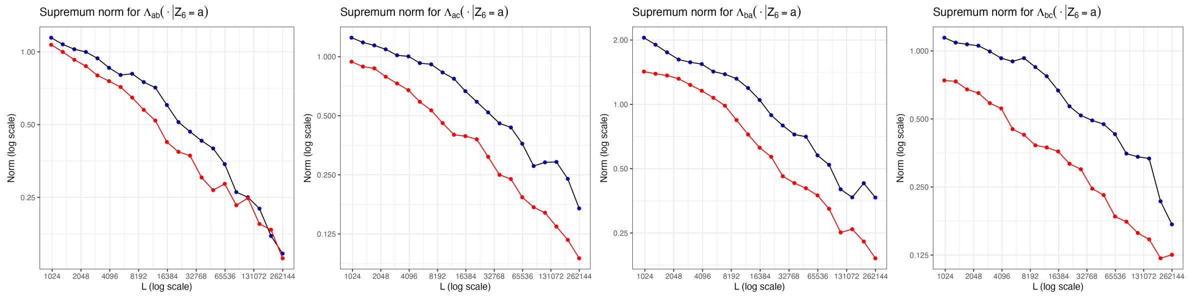

Figure 4: The figure shows the value $\mathbb{E}|\Lambda^{jk}(\cdot|Z_6 = 1) - \hat{\Lambda}^{(L)}_{jk}(\cdot|Z_6 = 1)|_\infty$ for both the As-If estimate (blue) and the Markov estimate (red). Based on 20 Monte Carlo estimates. Each increment in the grid is a multiple of two.

$$\max_{\tau} \left| \Lambda^{ik}(\tau - |Z_s = j) - \hat{\Lambda}^{(L_m,v)}_{ik}(\tau - |Z_s = j) \right| \Big\}.$$

This is because between jump point the difference between the true value of Λ^{ik} and the estimate is non-decreasing and at jump points the estimate will jump and either have a greater distance to the true value just prior og at the jump point. In total giving, that we can narrow our search to all discontinuity points of $\hat{\Lambda}^{ik}$.

In Figure 4 we see simulated values of $S^{(N)}(L_m)$ for $N = 20$ and L_m being chosen as $2 \cdot 65^{i/25+1}$ for $i = 0, .., 20$ on $Z_6 = a$ (according to the discussion above). We can see, that the average norm is decreasing for larger L_m and the relationship is linear on a $\log - \log$ transform with the two norm being roughly a $\log(L)$ translated version of each other.

7.1.2 *Cash-Flows and Reserving*

We extend the example in *Implementing the Nelson-Aalen and Aalen-Johansen Estimators* by constructing a disability model defined by the following

$$\mathrm{d}B^i(t) = \begin{cases} \left[\mathbb{1}_{\{t>T\}} - \mathbb{1}_{\{t\leq T\}}\pi\right] \ \mathrm{d}t & i = a, \\ \mathrm{d}t & i = b, \\ 0 & i = c. \end{cases}, \quad b^{ij}(t) = \mathbb{1}_{\{j=c\}}.$$

The above is a contract where the insured pay an equivalence premium π until pension $T \geq 0$ and pay a rate of one during pension so long the insured is alive. Then pays a rate of one if the insured is disabled and a lump sum of one if the insured dies. We can of cause decompose the payment stream into the sum of four payment streams. The cash-flow in this model is then

$$\mathrm{d}A(t|Z_s = j) = \sum_{m=1}^{4} \mathrm{d}A_m(t|Z_s = j)$$

with

$$\mathrm{d}A_1(t|Z_s = j) = \mathbb{1}_{\{t>T\}} p^a(t|Z_s = j) \ \mathrm{d}t,$$

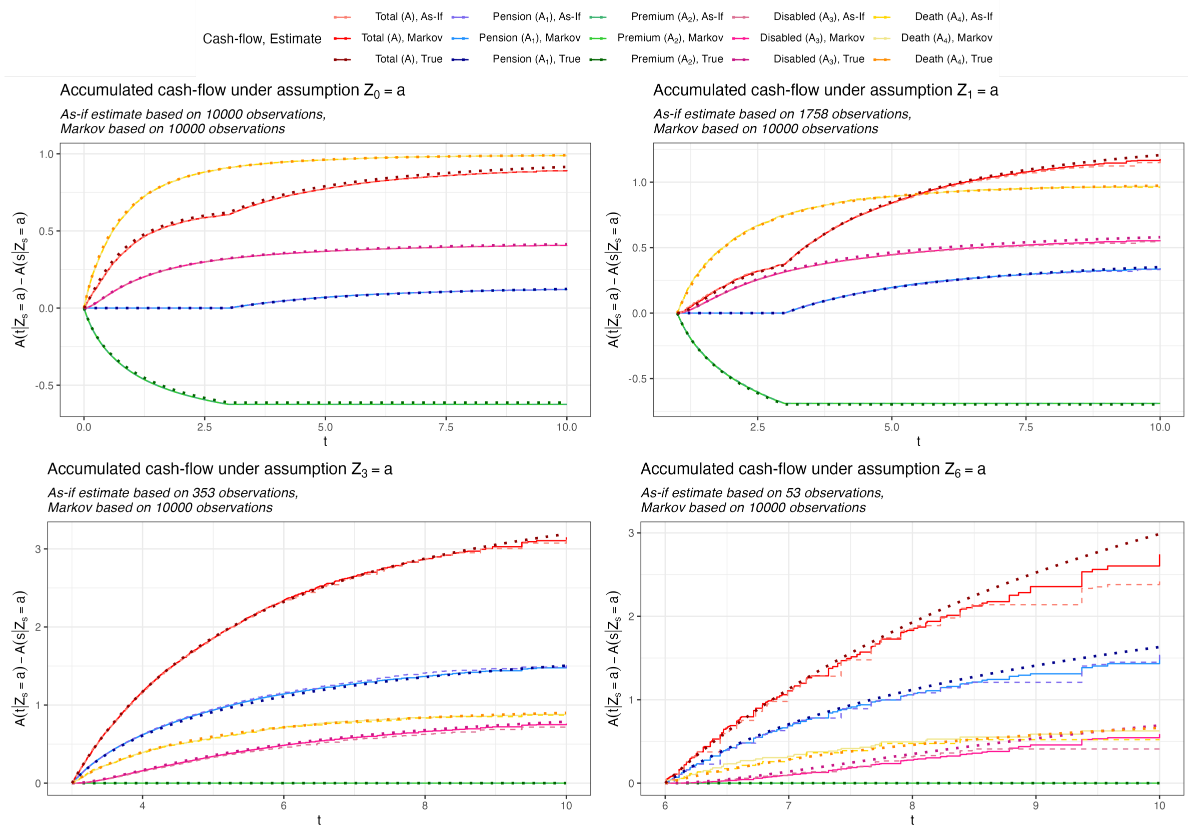

Figure 5: The figure shows the cumulative cash-flow $A_m(t|Z_s = j) - A_m(s|Z_s = j)$ under a Markov assumption and As-If-Markov assumption. The dashed line is estimate from the As-If model, the regular line is under the Markov assumption and the dots represent the true value (based on fourth order Runge-Kutta approximation). Based on $L = 10,000$ samples. (code given in appendices)

$$
\begin{aligned}
\mathrm{d}A_2(t|Z_s = j) &= -\mathbb{1}_{\{t \le T\}} \pi p^a(t|Z_s = j)\, \mathrm{d}t, \\
\mathrm{d}A_3(t|Z_s = j) &= p^b(t|Z_s = j)\, \mathrm{d}t, \\
\mathrm{d}A_4(t|Z_s = j) &= p^a(t-|Z_s = j)\, \mathrm{d}\Lambda^{ac}(t|Z_s = j) + p^b(t-|Z_s = j)\, \mathrm{d}\Lambda^{bc}(t|Z_s = j).
\end{aligned}
$$

We can estimate these cash-flows rather easily in Markov vase from the previous section as the solutions to $p(t, s)$ is easily calculated. Using the Aalen-Johansen estimator and the Nelson-Aalen estimator, we know that the Nelson-Aalen estimator only jumps when a jump occurs in the data and that the Aalen-Johansen estimator is constant between jumps. With this in mind, we have that A_1, A_2 and A_3 has piece-wise constant derivatives i.e. linear between jumps. The cash-flow A_4 can be computed as it is piece-wise constant and the jumps amounts to

$$
\mathbb{I}_a^{(L)}(\tau - |Z_s = j) \Delta \hat{\Lambda}_{ac}^{(L)}(\tau|Z_s = j) + \mathbb{I}_b^{(L)}(\tau - |Z_s = j) \Delta \hat{\Lambda}_{bc}^{(L)}(\tau|Z_s = j),
$$

on jump points τ. We compare the accumulated cash-flows in Figure 5, where we estimate using both estimators and compare to the true value. The value of the reserves together with the calculated equivalence premium is shown in Table 2.

Table 2: Estimated reserves for $\pi = 1$, $T = 3$, $s = 2$ and $r = 0.04$

Estimate	Z_s	$\hat{\pi}$[a]	$\hat{V}(s\|Z_s)$	$\hat{V}_1(s\|Z_s)$	$\hat{V}_2(s\|Z_s)$	$\hat{V}_3(s\|Z_s)$	$\hat{V}_4(s\|Z_s)$
Markov	a	3.850	1.572	0.648	-0.552	0.598	0.878
Markov	b	10.319	2.350	0.705	-0.252	1.033	0.864
As-If	a	3.763	1.505	0.611	-0.545	0.571	0.868
As-If	b	9.954	2.338	0.728	-0.261	1.015	0.856
True	a	3.966	1.633	0.660	-0.550	0.638	0.885
True	b	10.262	2.371	0.701	-0.256	1.051	0.874

a $\hat{\pi}$ is the equivalence premium such that $\hat{V}(s|Z_s) = 0$.

7.2 Semi-Markov Model

Let us first consider a Semi-Markov model, where the assumption layes on the pair of the current state and the time spend in that state. Recall, that a non-explosive pure jump process is uniquely determined by the collection $((Z_n, \tau_n))_{n=1,\dots,n_\infty}$. We can thus define the state duration process as follows:

$$U_t = \sum_{n=0}^{\eta} (t - \tau_n) \mathbb{1}_{\tau_n \leq t < \tau_{n+1}},$$

or simply

$$U_t = t - \tau_{\langle t \rangle}$$

where $\langle t \rangle$ is the number of jumps at time t. We will often in real life see that life processes have some duration dependence i.e. Z_t is not independent of U_t. In fact, one could argue that almost all life processes probably are state dependent, at least to some degree. Take for instance a disability chain such as the one in the previous section. It is likely true that the probabilities

$$\mathbb{P}(Z_t = a | Z_s = b, U_s = u_1) \quad \text{and} \quad \mathbb{P}(Z_t = a | Z_s = b, U_s = u_2)$$

for $u_1 < u_2$ is very different. One would indeed assume that for a larger duration spend as disabled the smaller the probability of being reactivated. This is where the Semi-Markov model assumption comes into consideration.

In the Semi-Markov model we do not make a Markov assumption on Z_t but rather on the pair (Z_t, U_t). In other words, we assume that the future trajectory of Z_t does depend on the past, but only on the time spend in the current state. Thus we only keep track on Z_t and U_t at any time-point. To be specific we have that if Z_t is Semi-Markov, then

$$\mathbb{P}(Z_t = i, U_t \leq u | \mathcal{F}_s) = \mathbb{P}(Z_t = i, U_t \leq u | Z_s, U_s)$$

for all $s \leq t$. In particular, this means that if the transition probabilities are absolutely continuous then

$$p(s, t; u, v) = \big(\mathbb{P}(Z_t = j, U_t \leq v | Z_s = i, U_s = u)\big)_{ij \in \mathcal{Z}}$$

satisfies the forward integro differential equation the differential form

$$\frac{\mathrm{d}}{\mathrm{d}t} p^{ij}(s,t;u,u+t-s) = \sum_{k\in\mathcal{Z},k\neq j} \int_0^{u+t-s} p^{ik}(s,t;u,\nu) - p^{ij}(s,t;u,\nu)\,\mathrm{d}_\nu \mu^{kj}(s,\nu)$$

with boundary condition $p^{ij}(s,s;u,u) = \mathbb{1}_{\{i=j\}}$. Notice that the duration model fits in the pure jump process framework by considering that the process $(Z_t, \tau_{\langle t\rangle})$ consitutes a non-explosive piece-wise constant process.

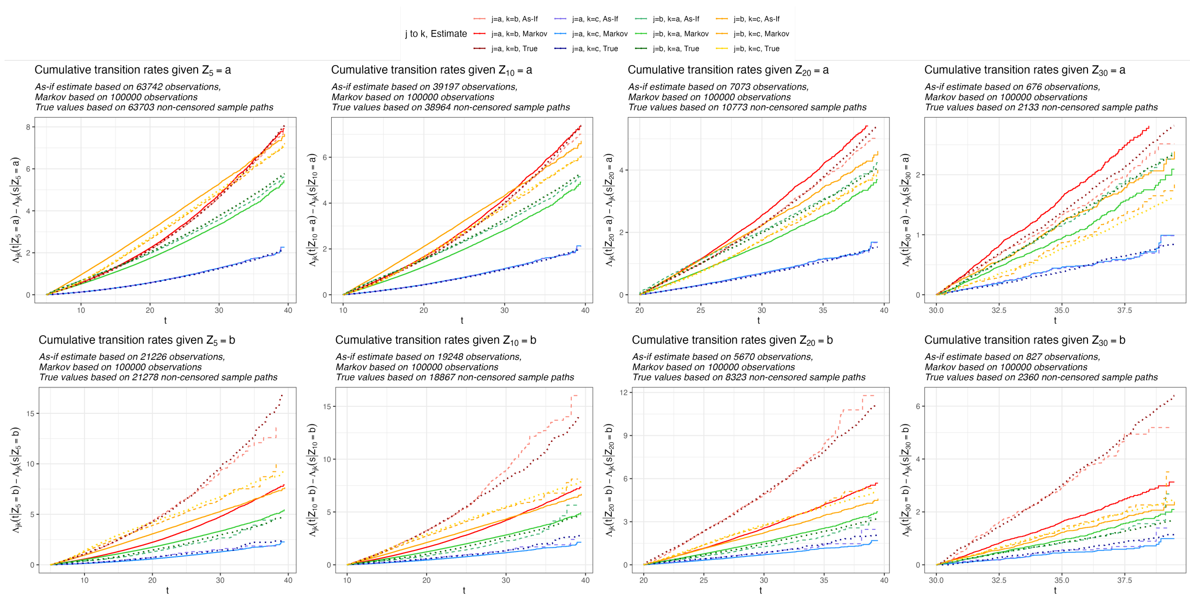

Figure 6: The figure shows the conditional Nelson-Aalen estimator under a Markov assumption and As-If-Markov assumption. The dashed line is estimate from the As-If model, the regular line is under the Markov assumption. The estimate of the true values is based on Monte Carlo simulation of 100,000 non-censored samples. (code given in appendices)

We can thus by choosing an appropriate cumulative transition rate matrix construct a realistic model, where we would prefer an estimator, that takes some of the past information into account. We choose the below intensities in order to obtain a model where: 1) the likelihood that Z_t transition to the active state a from b decay the longer Z_t stays in b, 2) the likelihood of death from state b is larger than from a, 3) the difference in intensity from b to c and a to c (given no reactivation) diverges for larger u and 4) if one becomes reactivated one has permanent increased intensity from a to b. We consider the example:

$$\begin{aligned}
\lambda^{ab}(t,u) &= 0.09 + 0.001t + \mathbb{1}_{\{(t-u)^+>0\}} \cdot 0.015t,\\
\lambda^{ac}(t,u) &= 0.01 + 0.002t + \mathbb{1}_{\{(t-u)^+>0\}} \cdot 0.001t,\\
\lambda^{ba}(t,u) &= 0.04 + 0.005t + 0.1 \cdot 0.5^u,\\
\lambda^{bc}(t,u) &= 0.09 + 0.001t + 0.01 \cdot 2^u.
\end{aligned}$$

We see that all the criteria above is satisfied. Obvious we set $\lambda^{ca} = \lambda^{cb} = 0$.

7.2.1 Comparing Occupation Probabilities and Transition Rates

We can calculate the conditional Aalen-Johansen estimator and Nelson-Aalen estimator as we did with the time in-homogeneous Markov chain i the previous section. In that model, the only predictor of the probability of a transition occurring in the near future was the current time and the state in which the chain sojourns in. In other words, we can say that the model had only one covariate, namely t.

In the Semi-Markov model, we say that short-term predictions takes two covariates. We here assume that λ is a function of two variable; the current time t, the state Z_t and the time spend in the current state U_t. Thus for precise predictions for the immediate future we need to keep track on Z_t and U_t. This means that, we would expect an estimator to be superior to another estimator in predicting the future trajectory if the estimator somehow accounts for U_t.

As we know, the As-If estimator gives us information on how the expected trajectory is for another i.i.d. path with the same landmark, say $Z_s = j$. Thus in the immediate future after time s we base the intensities on a population that almost surely all sojourn in j. Meanwhile, the Markov estimator will base the intensities in the immediate future on the at time s non-censored population.

If we only consider transitions from from j to another state k on $(s, s+h]$ for some small h, then we know that the information of these transitions will come from the same observations in both estimator. The difference is that the Markov estimator has in addition to those observations all other at time s non-censored sample paths sojourning in any other state $k \neq j$.

If we let

$$\overline{\mathbb{I}}^{\circ(L)}(s) = \sum_{i \in \mathcal{Z}} \mathbb{I}_i^{\circ(L)}(s) \quad \text{and} \quad \overline{\mathbb{I}}^{\bullet(L)}(s) = \sum_{i \in \mathcal{Z}} \mathbb{I}_i^{\bullet(L)}(s).$$

Then the above is the total non-censored individuals at time s used for the Markov estimator and As-If estimator respectively. We can then heuristically state some important observations. As mentioned the transitions j to k on s to $s+h$ will be based on the same observations. We do however have that in the Markov estimator the number of expected individuals in percent of the non-censored observations that will jump from j to k is

$$q_j^{\circ(L)}(s) = \frac{1}{\overline{\mathbb{I}}^{\circ(L)}(s)} \sum_{\ell=1}^{L} \mathbb{1}_{\{R^\ell > s\}} \mathbb{1}_{\{Z_s = j\}} \Big(\mu^{jk}(s, U^\ell(s))h + o(h) \Big).$$

We then have that for the As-If estimator the equivalent amount becomes

$$q_j^{\bullet(L)}(s) = \frac{\overline{\mathbb{I}}^{\circ(L)}(s)}{\overline{\mathbb{I}}^{\bullet(L)}(s)} q_j^{\circ(L)}(s)$$

as the data samples on aggregate only differ by composition. Thus we see that $q_j^{\bullet(L)}(s) \geq q_j^{\circ(L)}(s)$ giving that the As-If estimate should give a greater probability of staying in state j in the immediate future. We also see from the above that both estimate will include a mix of different duration's spend in j. However, strictly after s the observations that may jump from j to k will differ. In general, we would expect the duration for the observations jumping from j to k at some future time $s_0 > s$ to be composed of larger duration in the As-If estimator than the Markov estimator. This is because the Markov population will be saturated with observations jumping from i to j on $(s, s_0]$ and then from j to k, therefore having a duration of at most $s_0 - s$. Therefore in that sense, the Markov estimate will be biased to the downside on duration for transitions $j \to k$.

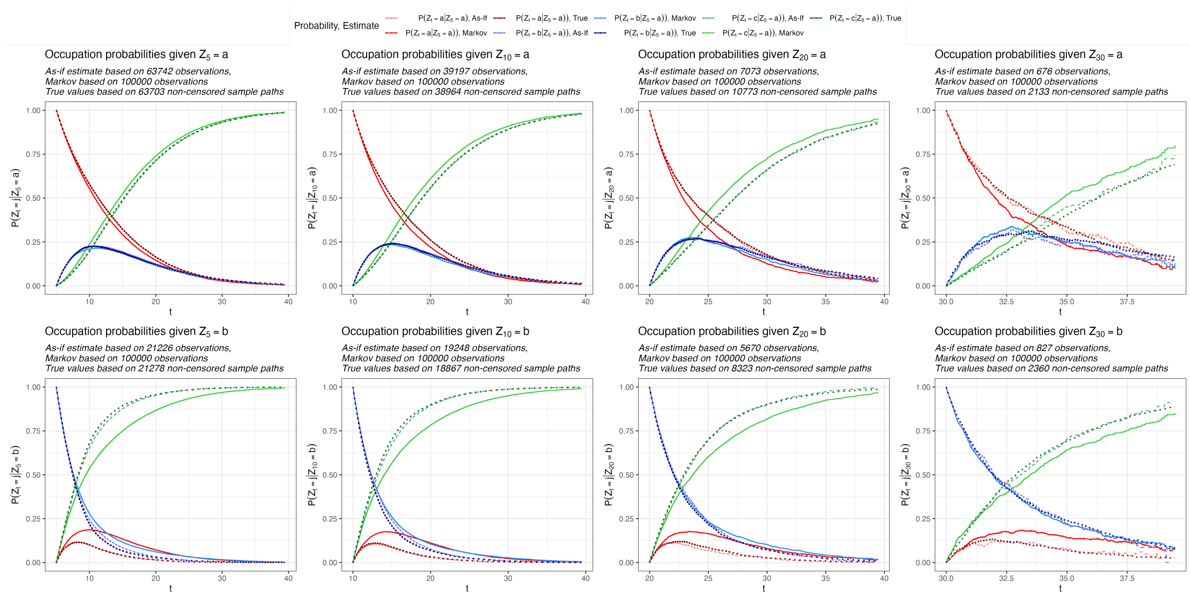

Figure 7: The figure shows the occupation probabilities under a Markov assumption and As-If-Markov assumption. The dashed line is estimate from the As-If model, the regular line is under the Markov assumption. The estimate of the true values is based on Monte Carlo simulation of 100,000 non-censored samples. (code given in appendices)

We can outline similar heuristic argument for the other transitions that may occur i.e. from $i \neq j$ to k. In the immediate future, the Markov estimator counts all non-censored transitions $i \to k$ on $(s, s+h]$. On the contrary for the As-If estimator any transition from $i \neq j$ to k to be used we will have to have the occurrence $j \to i$ on $(s, s+h_1]$ and then from $i \to k$ on $(s+h_1, s+h]$ for some $0 < h_1 < h$. Furthermore, heuristically such a transition will happen with probability

$$o(h) + h \int_0^h p^{ji}(s, s+h_1; U^\ell(s), h_1)\mu^{ik}(s+h_1, 0)\, \mathrm{d}h_1$$

if $Z_s^\ell = j$. We thus have that the above will constitutes the expected transitions from i to k on $(s, h]$ of the non-censored observations with $Z_s^\ell = j$. This means that the Markov estimator gives a larger probability of a transition $i \to k$ in the short term.

Additional to the above, the large majority of transition $i \to k$ in the Markov estimator will have duration in i strictly larger than zero, while the As-If Markov model will by construction have duration in i of at most h. In total, this means that the Markov estimate will systematic estimate the transition rates wrong since the transitions from i will have the same or larger duration than possible under the condition $Z_s = j$. This effect will however in the long run and as $L \to \infty$ and t tends to infinity go to zero as the two populations would likely converge in distribution.

We see in Figure 7 that the two estimates $\hat{p}^{\bullet}$ (As-If) and $\hat{p}^{\circ}$ (Markov) in this example differs dramatically. Obviously, we see that the As-If estimate in this case is not just a version of the Markov estimate with fewer observations (and therefore an estimate with higher variance). In particular, we see that the As-If estimate appear to systematically approximate the true values of $p^k(t|Z_s = j)$ significantly better than under the Markov assumption.

If we study the eight plots in more details we see a trend. We simulated the sample paths under the initial distribution $\pi_0 = (1, 0, 0)$ i.e. $Z_0 = a$ with probability one. Then for small t the quantity $\sum \mathbb{1}_{\{Z_t = b\}}$ is small. This explains why the estimator $\hat{p}^{\bullet}(\cdot|Z_5 = b)$ did not converge with the available observations. In particular, we see a rather large deviation on $t \in [7.5, 20]$ for the quantity

$$\left| p^k(t|Z_5 = b) - \hat{p}^{\bullet}_k(t|Z_5 = b) \right|.$$

As we increase the landmarking time s past $s = 5$ we have a large amount of observations in the state b and so $\hat{p}^{\bullet}_k(\cdot|Z_s = b)$ does better systematically. However, due to censoring we get more and more erratic estimates for larger s (say $s = 30$ in the lower right plot). Thus the mentioned trend is loosely formulated *the As-If model systematically outperforms the Markov model despite data restrictions and on higher probability landmark populations the As-If estimate converges rapidly*. This is clearly a vastly different conclusion than if the underlying chain is Markov as in the previous example.

7.2.2 Cash-Flows and Reserving

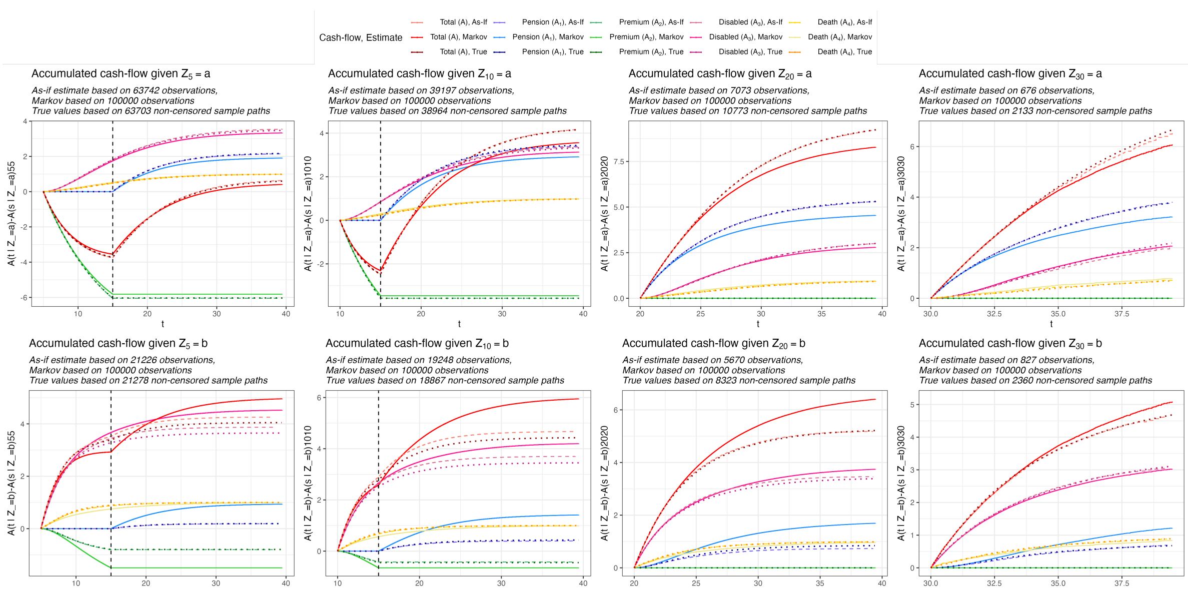

Figure 8: The figure shows the accumulated cash-flows under a Markov assumption and As-If-Markov assumption. The dashed line is estimate from the As-If model, the regular line is under the Markov assumption. The estimate of the true values is based on Monte Carlo simulation of 100,000 non-censored samples. (code given in appendices)

We discussed in the previous section, hoe the occupation probabilities and cumulative transition rates differed from the conditional estimators. In particular, we noted that the As-If estimate approximates the true value of the cumulative rates and occupation probabilities especially on high probability sets $\{Z_s = j\}$. We also noted that the Markov estimator systematically does not converge to the desired rates and probabilities.

We saw that given that $Z_s = a$ for some $s > 0$ the Markov estimator approximate $\hat{\Lambda}^{ak}(t|Z_s = a)$ larger than the true rates and so we would expect a higher reserve for the disability payments A_3 conditioned on $Z_s = a$ if one base the reserve estimate on the Markov estimate. Furthermore, this also means that the reserve for the premium payments A_2 and pension payments A_1 will be smaller than the true value. This in total will mean that the insurance company will greatly misspecify the equivalence premium with a higher premium than needed. The As-If estimator appear to approximate the true rates and occupational probabilities well and thus the equivalence premium and reserves will likely have a smaller difference than for the Markov estimator.

Table 3: Estimated reserves for $\pi = 1$, $T = 15$, $s = 5$ and $r = 0.04$

Estimate	Z_s	$\hat{\pi}$[a]	$\hat{V}(s\|Z_s)$	$\hat{V}_1(s\|Z_s)$	$\hat{V}_2(s\|Z_s)$	$\hat{V}_3(s\|Z_s)$	$\hat{V}_4(s\|Z_s)$
Markov	a	0.788	-1.055	1.034	-4.985	2.229	0.667
Markov	b	4.118	3.771	0.505	-1.209	3.692	0.782
As-If	a	0.807	-0.995	1.165	-5.165	2.356	0.650
As-If	b	6.358	3.581	0.107	-0.668	3.307	0.836
True	a	0.790	-1.088	1.196	-5.171	2.232	0.656
True	b	5.471	3.162	0.117	-0.707	2.902	0.851

a $\hat{\pi}$ is the equivalence premium such that $\hat{V}(s|Z_s) = 0$.

In the case where we condition on $Z_s = b$ for some $s > 0$ we saw the most striking difference between all three values. Although they all yield different paths for the probabilities and rates, we see that the As-If model does much better in targeting the true probabilities and rates. Concretely, we see that the Markov estimate has lower transition rates out of state b. This is due to the model not catching the high transition rates from b to a for the low duration observations and simply apply a *mean*-like estimate of the duration. This in particular means that the estimate of reserve for the disability payments should be much larger for the Markov estimator compared to the true value and the As-If estimate. Incidentally, this also means the the pension and premium reserves will be smaller for the Markov estimate compared to the As-If and true values.

Moreover, like with the short-term high transition rates from b to a for small duration, the Markov estimate does not catch the dynamic with the transitions to the death state. Duration does for impact rates from b to c exponentially in u the Markov estimator estimate lower transition rates to the death state. Implying that the reserve estimate for the death insurance will be far too large.

In Figure 8 and Table 3, we recognise the above consideration taking shape in the estimated cash-flows (Figure 8) and the impact on reserving (Table 3). Indeed, we see that for $Z_s = b$ for

$s = 5$ we have a noticeable difference between the estimated reserve for a premium π

$$\hat{V}_2(s|Z_s = b) = -\pi \int_s^{40} e^{-r(t-s)} \hat{p}_a^{(L)}(t|Z_s = b)\, \mathrm{d}t.$$

The calculated passive (i.e. for $\pi = 1$) is for the Markov estimate -1.209 and -0.668 for the As-If estimate. Meanwhile the true reserve is approximately -0.707. This is with a short term deterministic rate $r = 0.04$. This is a huge difference as it means that using the Markov estimate, the insurance company would price the equivalence premium as a factor of $-1.209/-0.707 \approx 1.71$ of the true equivalence premium. This is obviously not optimal for the insurance company nor for the insured. First, the pricing means that the insurance company would be less attractable to customers and the customers have less money in the short term thus lowering their utility. As time progresses however the company have to transfer the money back to the insured or offer the insured a higher coverage. The large difference between the V_2 reserve is also easily seen by comparing the green lines in the lower left plot in Figure 8 representing the function $A_2(t|Z_s = b) - A_s(s|Z_s = b)$.

We also notice that the reserve for the pension benefits i.e. V_1 differs greatly. We know from Figure 7 and Figure 6 and the above discussion, that the Markov estimate overestimate the alive probability $p_a(t|Z_s = b)$. Therefore the reserve for the pension benefits

$$\hat{V}_1(s|Z_s = b) = \int_s^{40} \mathbb{1}_{\{t \geq 15\}} e^{-r(t-s)} \hat{p}_a^{(L)}(t|Z_s = b)\, \mathrm{d}t$$

is exaggerated. We see that the true value is around 0.117 and the Markov estimate gives an estimate of 0.505 i.e. 332% higher. Simultaneously, we see that the As-If model gives a more realistic estimate of 0.107, which is an error of -8.5%. This have a great impact on the price of pension benefits and the Markov estimate would lead to enormous bonus and/or too high coverage with respect to the customers needs.

Although the As-If estimator is painted as the clear winner in targeting the true reserves, we still see some error worth noting. The actual reserve for the disability benefits while the insured is disabled is too high. This due to the probabilities $\hat{p}_b^{(L)}(t|Z_s = b)$ is higher than the true occupation probability. This obviously means that the reserve

$$\hat{V}_3(s|Z_s = b) = \int_s^{40} e^{-r(t-s)} \hat{p}_b^{(L)}(t|Z_s = b)\, \mathrm{d}t$$

is priced higher than the true value. As we discussed in the previous section, the error in the As-If estimator could be due to not enough data points and thus $\hat{p}_b^{\bullet(L)}(t|Z_s = b)$ did not converge. In total we see that the As-If estimator does well in targeting the true conditional cash-flow due to its efficiency in approximating the cumulative rates and occupations probabilities.

7.2.3 Model Risk versus Approximation Risk

We can calculate the risk of the Markov and As-If estimator by studying the model risk and approximation risk we define as follows. Let $|\cdot|$ be a norm and μ be the target object of the

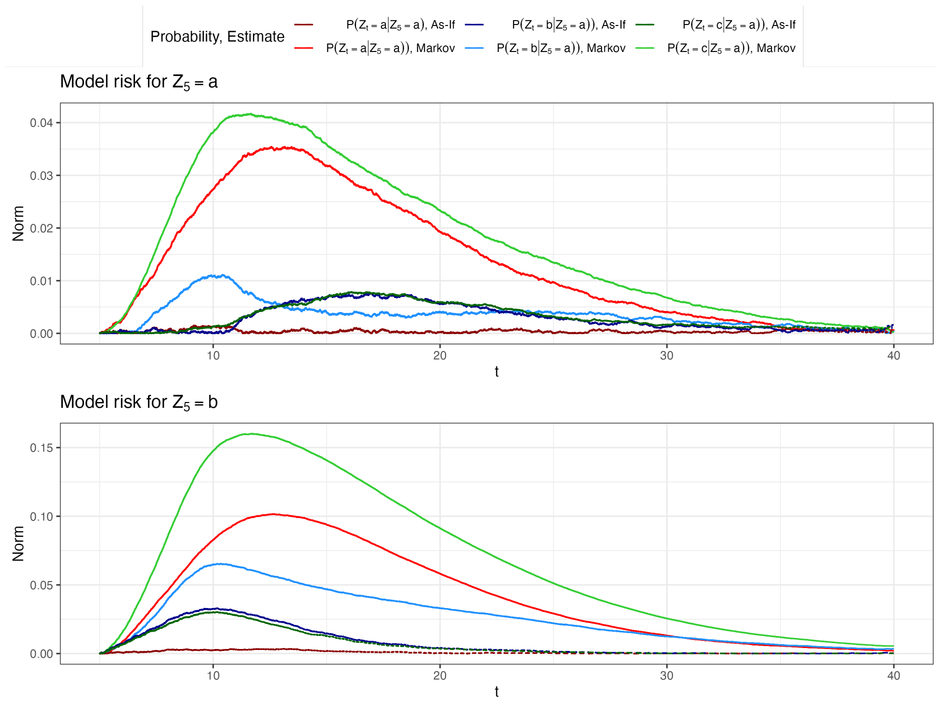

Figure 9: The figure shows the point wise distance between $\hat{p}_k^{(L_{\max})}(t|Z_5 = b)$ and the true value $p_k(t|Z_5 = b)$ for $k = a, b$. (code given in appendices)

estimator $\hat{\mu}^{(L)}$. We define the *risk* of $\hat{\mu}^{(L)}$ as

$$\mathrm{R}(\hat{\mu}^{(L)}) = \left|\hat{\mu}^{(L)} - \mu\right|.$$

We furthermore define the *model risk* of the estimator family $\hat{\mu}$ as

$$\mathrm{MR}(\hat{\mu}) = \lim_{L\to\infty} \mathbb{E}\left[\mathrm{R}(\hat{\mu}^{(L)})\right].$$

Lastly, we define the *approximation risk* of $\hat{\mu}^{(L)}$ as

$$\mathrm{AR}(\hat{\mu}^{(L)}) = \left|\hat{\mu}^{(L)} - \lim_{M\to\infty} \hat{\mu}^{(M)}\right|.$$

We can thus by drawing sufficiently many sample paths Z_s, say $L_{\max}$, and consider decompose the risk of a smaller model with $L \leq L_{\max}$ into the model risk, approximated through $\hat{\mu}^{(L_{\max})}$, and the approximation risk i.e.

$$\mathrm{MR}(\hat{\mu}) \approx \widehat{\mathrm{MR}}(\hat{\mu}) = \mathrm{R}(\hat{\mu}^{(L_{\max})}), \quad \widehat{\mathrm{AR}}(\hat{\mu}^{(L)}) \approx (\hat{\mu}^{(L)}) = \left|\hat{\mu}^{(L)} - \hat{\mu}^{(L_{\max})}\right|.$$

We will henceforth use $\widehat{\mathrm{MR}}(\hat{\mu})$ when referring to the approximated version $\mathrm{R}(\hat{\mu}^{(L_{\max})})$.

We start by simulating $L_{\max} = 1,000,000$ observations to base the model risk for the Markov estimator and the As-If estimator. We will purely be calculating for the occupation probabilities. We consider the supremum norm $|\cdot|_\infty$ and the p–norm $|\cdot|_p$ for $p = 2$ and as defined in *Norms*.

Table 4: Model risk for Markov and As-If-Markov estimates for $s = 5$

Norm	Estimate	j	$\widehat{\mathrm{MR}}(\hat{p}_a(\cdot\|Z_5 = j))$[a]	$\widehat{\mathrm{MR}}(\hat{p}_b(\cdot\|Z_5 = j))$	$\widehat{\mathrm{MR}}(\hat{p}_c(\cdot\|Z_5 = j))$
p-norm	As-If	a	30.01	6.27	6.54
p-norm	As-If	b	82.38	21.91	19.62
p-norm	Markov	a	1065.76	7.30	37.09
p-norm	Markov	b	3151.85	55.29	144.72
sup-norm	As-If	a	15.34	76.65	78.86
sup-norm	As-If	b	35.42	330.98	303.93
sup-norm	Markov	a	354.15	111.20	417.23
sup-norm	Markov	b	1015.65	653.82	1600.80

a All risk estimates is multiplied with 1,000.

We see in Figure 9 and Table 4 the approximated model risk of the optimal estimator for both the Markov and As-If estimators. In the mentioned figure, we clearly see that the As-If estimator has inherent lower risk as it catches some of the duration dynamics embedded in the Semi-Markov chain, while the Markov estimator has larger distance to the true occupation probabilities. We especially see that the As-If estimator decreases at a larger rate after both estimators distance tops around $t = 10$.

One also strikingly sees, that the active probability $Z_t = a$ given $Z_5 = b$ has close to zero risk in the As-If case. The p-norm for the active occupation probability is 0.0005 with $Z_5 = b$. In other words, the As-If model catches correctly the probability that $Z_t = a$ given $Z_5 = b$, but somehow misplaces the remaining probabilities wrong.

We can likewise calculate the approximation risk by comparing estimators $\hat{p}_k^{(L)}(t|Z_s = j)$ to the approximate limit estimator $\hat{p}_k^{(L_{\max})}(t|Z_s = j)$ for $L = L_1, ..., L_N$ for $L_1 < L_2 < ... < L_N$. We can compare the accumulating function

$$\widehat{\mathrm{AR}}(\hat{p}_k^{(L)}(\cdot|Z_s = j); t) = \frac{1}{40 - s}\left(\int_s^t |\hat{p}^{(L)}(u|Z_s = j) - \hat{p}^{(L_{\max})}(u|Z_s = j)|^p \,\mathrm{d}u\right)^{1/p},$$

for both the Markov and As-If Markov estimators for $L = 5000, 10000, 50000, 100000, 500000$. From these we may compare the approximation risk associated with one data sample of the sizes above. We use $p = 1$ as to not encounter diminishing increments from the exponentiation to the $1/p$th power.

We can from Figure 10 see the values of $t \mapsto \widehat{\mathrm{AR}}(\hat{p}_k^{(L)}(\cdot|Z_5 = b); t)$. We see that, in general we have

$$\widehat{\mathrm{AR}}(\hat{p}_k^{(L)}(\cdot|Z_5 = b); t) \geq \widehat{\mathrm{AR}}(\hat{p}_k^{\circ(L)}(\cdot|Z_5 = b); t)$$

i.e. the Markov estimator has less approximation risk than the As-If estimator. We can think of the value of $\widehat{\mathrm{AR}}(\hat{p}_k^{(L)}(\cdot|Z_5 = b); t)$ as the accumulated norm of the argument here the occupation

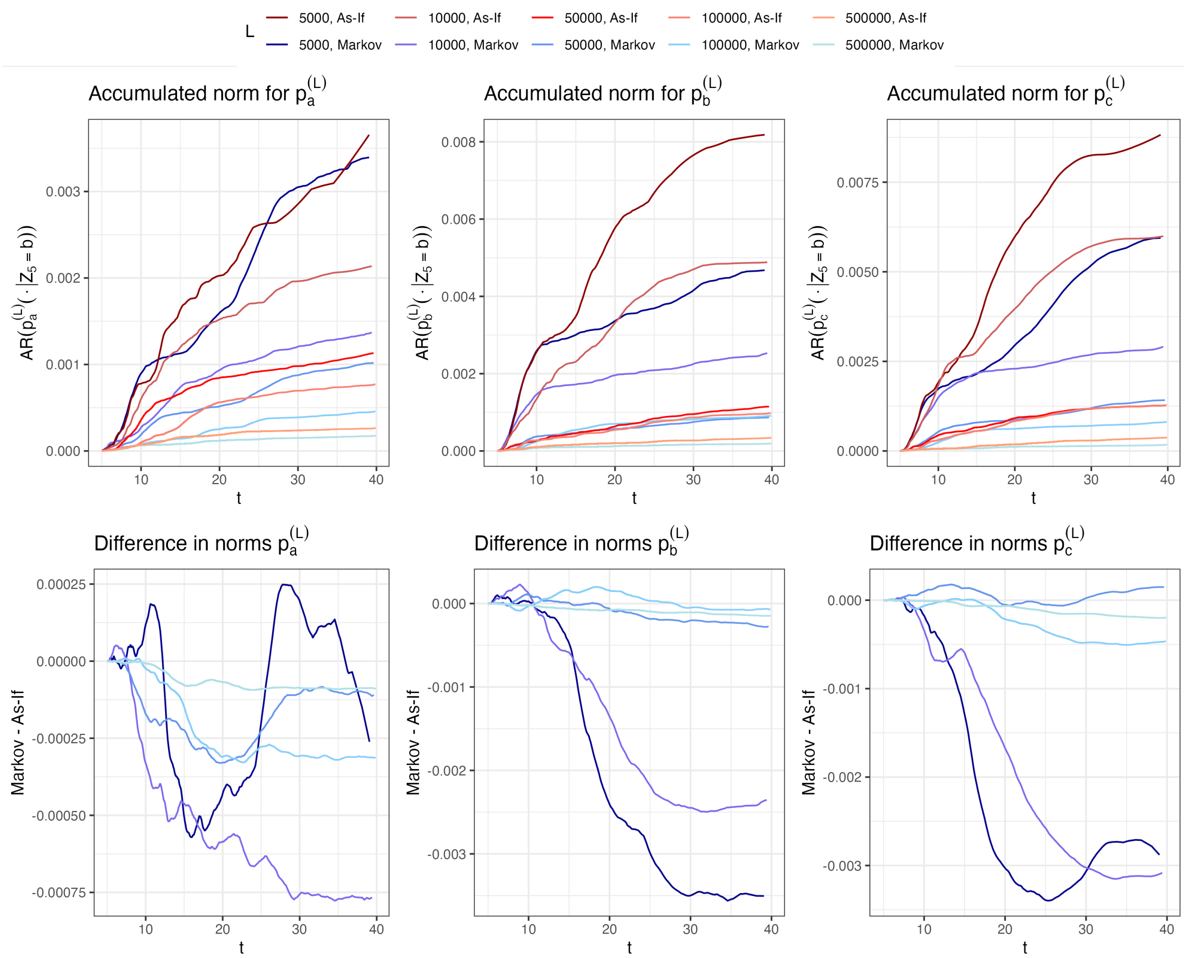

Figure 10: The figure shows the value of $\widehat{\mathrm{AR}}(\hat{p}_k^{(L)}(\cdot|Z_5 = b); t)$ (top) and $\widehat{\mathrm{AR}}(\hat{p}_k^{\circ(L)}(\cdot|Z_s = j); t) - \widehat{\mathrm{AR}}(\hat{p}_k^{\bullet(L)}(\cdot|Z_s = j); t)$ (bottom) for $k = a, b, c$. (code given in appendices)

probability that $Z_t = k$ given $Z_5 = b$. We thus see graphically in which intervals each estimator gains the most approximation risk.

In total, we observe that the As-If estimator gives a well description of the occupation probabilities and transition rates of the Semi-Markov model we studied. We do however see that although the As-If model is the obvious choice in minimising model risk it comes with large amount of approximation risk if L is not sufficiently large.

References

Bladt, Martin, **and** Christian Furrer. 2023a. *AalenJohansen: Conditional Aalen-Johansen Estimation.* R package version 1.0. https://CRAN.R-project.org/package=AalenJohansen.

Bladt, Martin, **and** Christian Furrer. 2023b. *Conditional Aalen–Johansen estimation.*

Christiansen, Marcus C. 2021. ?On the calculation of prospective and retrospective reserves in non-Markov model? *European Actuarial Journal* 11:441–462.

Christiansen, Marcus C., **and** Christian Furrer. 2022. ?Extension of as-if-Markov modeling to scaled payments? *Insurance: Mathematics and Economics* 107:288–306.

Gill, Richard D., **and** Soren Johansen. 1990. ?A Survey of Product-Integration with a View Toward Application in Survival Analysis? *The Annals of Statistics* 18.

A Additional Tables

Table 5: Approximation risk for Markov and As-If-Markov estimates for $s = 5$ ($p = 2$ norm)

j	Estimate	L	$\widehat{\text{AR}}(\hat{p}_a(\cdot\|Z_5 = j))$[a]	$\widehat{\text{AR}}(\hat{p}_b(\cdot\|Z_5 = j))$	$\widehat{\text{AR}}(\hat{p}_c(\cdot\|Z_5 = j))$
a	As-If	5000	14.78	10.73	19.14
a	As-If	10000	6.63	7.30	6.76
a	As-If	50000	5.48	3.13	6.34
a	As-If	100000	3.88	2.44	3.16
a	As-If	500000	0.77	0.71	0.75
a	Markov	5000	13.42	10.44	18.58
a	Markov	10000	6.02	7.09	7.53
a	Markov	50000	4.78	2.58	5.41
a	Markov	100000	3.04	2.06	2.33
a	Markov	500000	0.65	0.64	0.58
b	As-If	5000	8.26	18.64	18.90
b	As-If	10000	5.12	10.84	13.30
b	As-If	50000	2.80	2.47	3.03
b	As-If	100000	1.79	2.32	2.77
b	As-If	500000	0.62	0.77	0.75
b	Markov	5000	7.55	13.52	12.52
b	Markov	10000	3.11	7.48	8.03
b	Markov	50000	2.22	2.20	3.25
b	Markov	100000	1.03	2.13	2.04
b	Markov	500000	0.43	0.50	0.41

a All risk estimates is multiplied with 1,000.

B Source Code

The code used in this project may be downloaded from Git-Hub repository joakim-bilyk/act-math under projects -> projliv -> `Appendicies.html` with associated markdown file `Appendicies.Rmd`.

In the below table the reader sees the calculation time for some of the functions developed for this project. The calculation times are based on the in-homogeneous Markov chain.

Task	$n = 1,000$	$n = 10,000$	$n = 100,000$	$n = 1,000,000$
Simulation	0.248	2.469	29.448	267.96
`paths_to_df`	0.066	0.465	2.249	33.675
`df_to_I`	0.059	0.176	2.437	42.078
`df_to_N`	0.029	0.049	0.702	5.884
`N_I_to_NA`	0.002	0.011	0.489	2.082
`NA_to_p`	0.195	1.156	12.088	296.94
`P_conditioned`	0.186	1.136	12.947	605.22
Total	0.785	5.462	60.36	1550.779

Table 6: In seconds. Run on R v. 4.3.0 on an Apple ARM64 M1 system with 8 GB ram.