Department of Mathematical Sciences

PROJECT

Non-markov modelling of transition probabilities

J.L. Bilyk*

University of Copengagen, Copenhagen, 2100, Denmark
*Corresponding author. Email: joakim.bilyk@outlook.com

Abstract

Insert abstract text here. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Keywords: keyword entry 1, keyword entry 2, keyword entry 3

1 Introduction

In this paper,...

We will occasionally suppress some notation when it is prudent to do so. Even so, we try to state definitions and statements as clear as possible. Henceforth we will use the following shorthand and *silence of notation.*

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

$$f(t-) = \lim_{h \to 0^+} f(t-h) \tag{1}$$

2. Whenever we introduce a random variable or stochastic process, we will assume that it lives on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

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 $X : (\Omega, \mathcal{F}, \mathbb{P}) \mapsto (\mathcal{X}, \mathcal{A}, \mu)$ 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

$$\forall A \in \mathcal{A} : \{X \in A\} \overset{\text{def}}{=} \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F} \tag{2}$$

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

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

$$\int_{\Omega} |g(X(\omega))| \, \mathrm{d}P(\omega) < \infty. \tag{3}$$

We define the object

$$\mathbb{E}\left[g(X)\right] \overset{\text{def}}{=} \int_{\Omega} g(X(\omega)) \, \mathrm{d}\mathbb{P}(\omega). \tag{4}$$

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

$$\mathbf{X} = (X_t)_{t\geq 0} \overset{\text{def}}{=} (X_t)_{t\in\mathbb{R}^+}. \tag{5}$$

Life insurance contracts are often modelled by a particular 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 $\mathcal{I} = \mathbb{R}^+$. Consider a finite space $\mathcal{Z}$ with bijective mapping $\psi : \mathcal{J} = \{1, ..., J\} \mapsto \mathcal{Z}$ where $J = \#\mathcal{Z}$. A stochastic process $(Z_t)_{t\geq 0}$ is called a pure jump process if

1. $Z_t \in \mathcal{Z}$ for all $t \geq 0$,
2. the sample paths $t \mapsto Z_t(\omega)$ are almost surely piece-wise constant.

We included the mapping ψ in the definition to enforce the idea that ψ is chosen simultaneously alongside $\mathcal{Z}$ and it works as a translator of the actual state and the position on $\mathcal{J}$. Of cause, one can simply define $\mathbf{Z}$ on $\mathcal{Z}$ and simply afterwards decide on ψ, as it is indeed not unique.

Example 1 (Pure jump process with no explosion)**.** One can also see that if we assume no explosion the process may be uniquely characterised by the process $(Z_n, \tau_n)_{n\in\mathbb{N}_0}$ by setting

$$\tau_n = \inf\{s \geq \tau_{n-1} : Z_s \neq Z_{s-}\}, \qquad n \geq 1, \tag{6}$$

and setting $\tau_0 = 0$. We use the convention that $\inf \emptyset = \infty$. If the set in (6) is empty for some smallest number $m \geq 1$ then we define

$$n_\infty = \inf\{n \in \mathbb{N} : \tau_n = \infty\}. \tag{7}$$

Again we have $n_\infty = \infty$ if after all jumps τ_i the process eventually jumps to another state. Often we will have at least one absorbing state in $\mathcal{Z}$ and in that case n_∞ will be finite almost surely. The value of Z_n is then defined by Z_{τ_n}

$$Z_n = \begin{cases} Z_{\tau_n} & n \leq n_\infty, \\ Z_{\tau_{n_\infty}} & \text{otherwise.} \end{cases} \tag{8}$$

In that case, the process $\mathbf{Z}$ has the unique representation

$$Z_t = \sum_{n=0}^{n_\infty} Z_{\tau_n} \mathbb{1}_{\{\tau_n \leq t < \tau_{n+1}\}}. \tag{9}$$

■

We will later see how we may restrict $\mathbf{Z}$ to a class such that with probability one it will not explode. However before this we need to introduce the related counting processes.

Definition 4 (Multivariate counting process)**.** Let $\mathbf{Z} = (Z_t)_{t\geq 0}$ be a 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 a matrix-stochastic process with entries

$$N_t^{jk} = \#\{0 \leq s \leq t : Z_s = k, Z_{s-} = j\}, \qquad j \neq k, \tag{10}$$

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

Remark. This process is only useful in the case of no explosion or at least at most one explosion. This is because if explosion occur we wont be able to deduce the behavior of $\mathbf{Z}$ from $\mathbf{N}$ beyond the point of explosion as at least two entrances in N_t will be infinite. We can however somewhat circumvent this by studying $N_t - N_\tau$ for $t \geq \tau$ where τ is the time of explosion. From this we can locate a *new* starting point for the process as $Z_{\tau+}$.

The link between $\mathbf{Z}$ and $\mathbf{N}$ is only one-to-one if $\mathbf{Z}$ does not explode. This may be achieved by assuming $\mathbb{E}[(N_t^{jk})^2] < \infty$ or assuming a certain structure on the transition probabilities.

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 $\mathbf{Z}$, for instance its value for some subset $A \subseteq \mathbb{B}(\mathbb{R}^+)$ ($\mathbb{B}(\cdot)$ is the operator giving the Borel σ-algebra on the argument), or some related information telling the distribution of $\mathbf{Z}$ given this knowledge.

In the following we let X be an arbitrary $\mathcal{X}$-valued random variable. By choosing $\mathcal{G}$, this gives rise to the measure

$$\mathbb{P}(A \mid \mathcal{G}) = \mathbb{P}(A) \qquad \text{on } A \in \mathcal{G}. \tag{11}$$

Furthermore, this gives rise to a new random variable defined as

$$\mathbb{E}[X \mid \mathcal{G}] = X \qquad \text{on } A \in \mathcal{G}. \tag{12}$$

Another way of defining this variable is under the integral condition below.

Definition 5 (Conditional expectation)**.** Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub σ-algebra of $\mathcal{F}$. The conditional expectation of X on $\mathcal{G}$ called $\mathbb{E}[X|\mathcal{G}]$, is an $\mathcal{G}$-$\mathcal{A}$ measurable random variable satisfying

$$\forall G \in \mathcal{G} : \int_G \mathbb{E}[X|\mathcal{G}](\omega) \, \mathrm{d}\mathbb{P}(\omega) = \int_G X(\omega) \, \mathrm{d}\mathbb{P}(\omega). \tag{13}$$

Remark. Such a variable does exist and it is furthermore almost surely unique. (Hansen 2021, p. 340)

This construction of conditioning leads to the obvious choice of conditioned probability measure. Notice that for a set $A \in \mathcal{A}$ we have

$$\mathbb{E}[\mathbb{1}\{X \in A\}] = \int_\Omega \mathbb{1}\{X \in A\} \, \mathrm{d}\mathbb{P}(\omega) = \int_{\{X \in A\}} \mathrm{d}\mathbb{P}(\omega) = \mathbb{P}(X \in A) \tag{14}$$

This gives rise to the definition $\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable the integral

$$\mathbb{P}(X \in A|\mathcal{G}) \overset{\text{def}}{=} \mathbb{E}[\mathbb{1}\{X \in A\}|\mathcal{G}]. \tag{15}$$

We will often use the shorthand $\mathbb{E}[\cdot|X]$ or $\mathbb{E}[\cdot|X = x]$ when referring to $\mathbb{E}[\cdot|\sigma(X)]$ and $\mathbb{E}[\cdot|\sigma(X = x)]$ where

$$\sigma(X) \overset{\text{def}}{=} \sigma\left(\bigcup_{A \in \mathcal{A}} \{\omega \in \Omega : X(\omega) \in A\}\right), \quad \sigma(X = x) \overset{\text{def}}{=} \sigma\left(\{\omega \in \Omega : X(\omega) = x\}\right). \tag{16}$$

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 6 (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 hope that the information available is enough to determine the state of the insured in order to pay the agreed upon payments. Such an assumption is not entirely trivial as one can imagine a *lack* between the jump and the reporting of the jump. Furthermore, a case may be reviewed or undergo approval before the jump is determined to be valid. This issue was recently discussed in Buchardt, Furrer **and** Sandqvist 2023. We will not be making such assumption and we will be studying the following setup.

Definition 7 (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 $\mathbf{Z}$ is adapted to the filtration $\mathbb{F}$. We say that $\mathbf{Z}$ is a non-explosive pure jump process adapted to $\mathbb{F}$ and we write $\mathbf{Z} \sim \text{PJP}(\mathbb{F})$.

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

Definition 8 (Markov property)**.** Let $\mathbb{F}$ be a filtration and let $\mathbf{Z} \sim \text{PJP}(\mathbb{F})$. $\mathbf{Z}$ is said to excipit the Markov property if

$$\forall 0 \leq t \leq s : \mathbb{P}(Z_s \in A|\mathcal{F}_t) = \mathbb{P}(Z_s \in A|Z_t). \tag{17}$$

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

Remark. We can think of this property as the distribution of $\mathbf{Z}$ is memory-less. The property states that if we are to say anything about the state of $\mathbf{Z}$ 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.

4 Counting processes

In this section, we briefly introduce some representations of the multivariate counting process $\mathbf{N}$ stemming from $\mathbf{Z} \sim \mathrm{PJP}(\mathbb{F})$. Recall that in (9) 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

$$N^{jk}(t) = \sum_{n=1}^{n_\infty} \mathbb{1}_{[\tau_n,\infty)} \mathbb{1}_{\{Z_{n-1}=j\}} \mathbb{1}_{\{Z_n=k\}}. \tag{18}$$

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

$$dN^{jk}(t) = \mathbb{1}_{\{Z_{t-}=j\}} \mathbb{1}_{\{Z_t=k\}} \mathrm{d}N(t), \quad N^{jk}(0) = 0. \tag{19}$$

In the above, $N(t)$ is simply $N(t) = \sum_{j,k\in\mathcal{Z}} N^{jk}(t)$. In other words, we can define $N^{jk}(t)$ as the Lebesgue-Stieltjes integral

$$N^{jk}(t) = \int_0^t \mathbb{1}_{\{Z_{t-}=j\}} \mathbb{1}_{\{Z_t=k\}} \mathrm{d}N(t). \tag{20}$$

Another equivalent way of formulating the dynamics is

$$dN^{jk}(t) = \mathbb{1}_{\{Z_{t-}=j\}} \mathrm{d}N^k(t), \quad N^{jk}(0) = 0, \tag{21}$$

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

5 Transition rates and probabilities

Let $\mathbb{F}$ be a given filtration and let $\mathbf{Z} \sim \mathrm{PJP}(\mathbb{F})$. We may study the 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, 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 $\mathcal{G}$ which most likely will include the information Z_t for some $t \geq 0$. As we will see choosing $\mathcal{G} = \sigma(Z_t)$ leads to the tractable as-if-Markov-model.

Definition 9. Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. Let $\mathbf{Z} \sim \mathrm{PJP}(\mathbb{F})$ for some filtration $\mathbb{F}$. The *transition probabilities given $\mathcal{G} \subseteq \mathcal{F}$* is defined as

$$p^{jk}(s,t|\mathcal{G}) \overset{\text{def}}{=} \mathbb{E}\left[\mathbb{1}_{\{Z_s=j\}}\mathbb{1}_{\{Z_t=k\}} \mid \mathcal{G}\right], \quad j,k \in \mathcal{Z}, s,t \geq 0. \tag{22}$$

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

$$p^{j}(t|\mathcal{G}) \overset{\text{def}}{=} \mathbb{E}\left[\mathbb{1}_{\{Z_t=j\}} \mid \mathcal{G}\right], \quad j \in \mathcal{Z}, t \geq 0. \tag{23}$$

Lastly, we define the *conditional counting process* as

$$\tilde{p}^{jk}(t|\mathcal{G}) \overset{\text{def}}{=} \mathbb{E}\left[N^{jk}(t) \mid \mathcal{G}\right], \quad j,k \in \mathcal{Z}, t \geq 0. \tag{24}$$

As mentioned above, we simply condition on any σ-algebra $\mathcal{G}$. We will often be interested in the transition probabilities $p^{jk}(s,t|Z_s)$ and $p^{j}(t|Z_s)$. We furthermore define the following called the *cumulative conditional transition rates*. Using this setup we introduced aboved.

Definition 10 (Cumulative conditional transition rates)**.** Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. Let $\mathbf{Z} \sim \mathrm{PJP}(\mathbb{F})$ for some filtration $\mathbb{F}$. Using the notation above the *cumulative conditional transition rates* is defined as

$$\Lambda^{jk}(t|\mathcal{G}) \overset{\text{def}}{=} \int_0^t \frac{1}{p^{j}(s- \mid \mathcal{G})} \, \mathrm{d}\tilde{p}^{jk}(s|\mathcal{G}), \quad j,k \in \mathcal{Z}, t \geq 0. \tag{25}$$

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

$$\mathbb{1}_{\{Z_t=j\}} = \mathbb{1}_{\{Z_0=j\}} + \sum_{k\in\mathcal{Z},k\neq j} \Big(N^{kj}(t) - N^{jk}(t)\Big). \tag{26}$$

Proof. Define the ordered jump times into j as

$$\Delta_j = \{s \geq 0 : Z_s = j, Z_{s-} \neq j\} := \{\tau_1^j, \tau_2^j, ...\}. \tag{27}$$

and out of j as

$$\Delta^{j\bullet} = \{s \geq 0 : Z_s \neq j, Z_{s-} = j\} := \{\tau_1^{j\bullet}, \tau_2^{j\bullet}, ...\}. \tag{28}$$

By assuming $Z_{0-} \neq j$ we see that 0 is included in Δ_j on the event $\{Z_0 = j\}$ giving that $\tau_1^j = 0$. Contrary to this we have that $\tau_1^{j\bullet} > 0$ always. Clearly, 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 $m \geq 1$ 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

$$\sum_{k\in\mathcal{Z},k\neq j} N^{kj}(t) = m - \mathbb{1}_{\{Z_0=j\}}, \quad \sum_{k\in\mathcal{Z},k\neq j} N^{jk}(t) = m - 1. \tag{29}$$

Hence

$$\begin{aligned}(*) :&= \mathbb{1}_{\{Z_0=j\}} + \sum_{k\in\mathcal{Z},k\neq j}\Big(N^{kj}(t) - N^{jk}(t)\Big) && (30)\\ &= \mathbb{1}_{\{Z_0=j\}} + m - \mathbb{1}_{\{Z_0=j\}} - (m-1) && (31)\\ &= 1 = \mathbb{1}_{\{Z_t=j\}}. && (32)\end{aligned}$$

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

$$\sum_{k\in\mathcal{Z},k\neq j} N^{kj}(t) = m - \mathbb{1}_{\{Z_0=j\}}, \qquad \sum_{k\in\mathcal{Z},k\neq j} N^{jk}(t) = m. \tag{33}$$

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.

Theorem 12. *Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. Let $\mathbf{Z} \sim PJP(\mathbb{F})$ for some filtration $\mathbb{F}$. Using the notation the processes $p^j(t|\mathcal{G})$ and $\Lambda^{jk}(t|\mathcal{G})$ relates through the dynamics*

$$dp^j(t|\mathcal{G}) = \sum_{k\in\mathcal{Z}} p^k(t-|\mathcal{G})\, d\Lambda^{kj}(t|\mathcal{G}). \tag{34}$$

In the above using the convention $d\Lambda^{jj}(t|\mathcal{G}) = -\sum_{k\in\mathcal{Z},k\neq j} d\Lambda^{jk}(t|\mathcal{G})$.

Proof. From proposition 11 we have that

$$\mathbb{1}_{\{Z_t=j\}} - \mathbb{1}_{\{Z_s=j\}} = \sum_{k\in\mathcal{Z},k\neq j}\Big(N^{kj}(t) - N^{jk}(t)\Big) - \sum_{k\in\mathcal{Z},k\neq j}\Big(N^{kj}(s) - N^{jk}(s)\Big) \tag{35}$$

Thus implying

$$\mathbb{E}\left[\mathbb{1}_{\{Z_t=j\}} - \mathbb{1}_{\{Z_s=j\}} \;\middle|\; \mathcal{G}\right] = \mathbb{E}\left[\sum_{k\in\mathcal{Z},k\neq j}\Big(N^{kj}(t) - N^{jk}(s)\Big) - \sum_{k\in\mathcal{Z},k\neq j}\Big(N^{kj}(t) - N^{jk}(s)\Big) \;\middle|\; \mathcal{G}\right] \tag{36}$$

Now, using the linearity of expectation and inserting definition 9 we have

$$p^j(t|\mathcal{G}) - p^j(s|\mathcal{G}) = \sum_{k\in\mathcal{Z},k\neq j}\Big(\tilde{p}^{kj}(t|\mathcal{G}) - \tilde{p}^{jk}(t|\mathcal{G})\Big) - \sum_{k\in\mathcal{Z},k\neq j}\Big(\tilde{p}^{kj}(s|\mathcal{G}) - \tilde{p}^{jk}(s|\mathcal{G})\Big). \tag{37}$$

Or on differential form

$$\mathrm{d}p^j(t|\mathcal{G}) = \sum_{k\in\mathcal{Z},k\neq j} \mathrm{d}\tilde{p}^{kj}(t|\mathcal{G}) - \mathrm{d}\tilde{p}^{jk}(t|\mathcal{G}) \tag{38}$$

Using that $p^j(t-|\mathcal{G})\mathrm{d}\Lambda^{jk}(t|\mathcal{G}) = \mathrm{d}\tilde{p}^{jk}(t|\mathcal{G})$ we have

$$\begin{aligned}
\mathrm{d}p^j(t|\mathcal{G}) &= \sum_{k\in\mathcal{Z},k\neq j} p^k(t-|\mathcal{G})\mathrm{d}\Lambda^{kj}(t|\mathcal{G}) - p^j(t-|\mathcal{G})\mathrm{d}\Lambda^{jk}(t|\mathcal{G}) && (39)\\
&= \left(\sum_{k\in\mathcal{Z},k\neq j} p^k(t-|\mathcal{G})\mathrm{d}\Lambda^{kj}(t|\mathcal{G})\right) - p^j(t-|\mathcal{G})\sum_{k\in\mathcal{Z},k\neq j}\mathrm{d}\Lambda^{jk}(t|\mathcal{G}) && (40)\\
&= \sum_{k\in\mathcal{Z}} p^k(t-|\mathcal{G})\mathrm{d}\Lambda^{kj}(t|\mathcal{G}). && (41)
\end{aligned}$$

Using the convention $\mathrm{d}\Lambda^{jj}(t|\mathcal{G}) = -\sum_{k\in\mathcal{Z},k\neq j}\mathrm{d}\Lambda^{jk}(t|\mathcal{G})$. □

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

$$p(t|\mathcal{G}) \overset{\text{def}}{=} \begin{pmatrix} p^{\psi(1)}(t|\mathcal{G}) \\ \vdots \\ p^{\psi(J)}(t|\mathcal{G}) \end{pmatrix}, \quad \Lambda(t|\mathcal{G}) \overset{\text{def}}{=} \begin{pmatrix} \Lambda^{\psi(1),\psi(1)}(t|\mathcal{G}) & \cdots & \Lambda^{\psi(1),\psi(J)}(t|\mathcal{G}) \\ \vdots & \ddots & \vdots \\ \Lambda^{\psi(J),\psi(1)}(t|\mathcal{G}) & \cdots & \Lambda^{\psi(J),\psi(J)}(t|\mathcal{G}) \end{pmatrix}. \tag{42}$$

Using again the convention $\Lambda^{jj}(t|\mathcal{G}) = -\sum_{k\in\mathcal{J},k\neq j}\Lambda^{jk}$. We will be using the product integral notation so the definition i given below.

Definition 13 (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

$$\mathbf{Y}(t) = \mathbf{C}\prod_{u\in(0,t]}\Big(\mathrm{Id}+\mathrm{d}\mathbf{A}(u)\Big) \overset{\text{def}}{=} \mathbf{C}\lim_{\max|t_i-t_{i-1}|\to 0}\prod_{i}\Big(\mathrm{Id}+\mathbf{A}(t_i)-\mathbf{A}(t_{i-1})\Big). \tag{43}$$

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.

We can think of this operator as a signed measure on $\mathbb{B}([0,\infty))$ and we can extend thte definition to an arbitrary Borel set by the following.

Definition 14. Using the assumptions in 13, let $B\in\mathbb{B}([0,\infty))$ an arbitrary Borel set. We define the product integral of $\mathbf{A}$ on B as

$$\mathbf{Y}(B) = \lim_{t\to\infty}\prod_{u\in(0,t]}\Big(\mathrm{Id}+\mathbb{1}_{\{t\in B\}}\,\mathrm{d}\mathbf{A}_B(u)\Big) \overset{\text{def}}{=} \prod_{u\in B}\Big(\mathrm{Id}+\mathrm{d}\mathbf{A}_B(u)\Big). \tag{44}$$

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

$$\mathbf{X}(t) = \begin{cases} \mathbf{C}\mathbf{Y}((s,t]), & t>s \\ \mathbf{C}\mathbf{Y}((t,s]), & t\leq s \end{cases}. \tag{45}$$

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

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

Lemma 15. *Let* $\mathbf{A}$ *be as given in definition 13 and* $\mathbf{Y}(B)$ *as given in definition 14. It holds that*

$$d_t\mathbf{Y}((s,t]) = \mathbf{Y}((s,t])\, d\mathbf{A}(t), \qquad \textit{(the forward equation)} \tag{46}$$

$$d_s\mathbf{Y}((s,t]) = -d\mathbf{A}(s)\mathbf{Y}((s,t]), \qquad \textit{(the backward equation)} \tag{47}$$

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

$$\mathcal{P}(s,t;\mathbf{A}) \overset{\text{def}}{=} \text{Id} + \sum_{n=1}^{\infty} \int_s^t \int_s^{u_n} \cdots \int_s^{u_3} \int_s^{u_2} \, \text{d}\mathbf{A}(u_1)\, \text{d}\mathbf{A}(u_2) \cdots \, \text{d}\mathbf{A}(u_{n-1})\, \text{d}\mathbf{A}(u_n) \tag{48}$$

$$\overset{\text{def}}{=} \text{Id} + \sum_{n=1}^{\infty} \mathbf{A}^{(n)}(s,t). \tag{49}$$

Now take n arbitrary and a fixed $u \in (s,t)$, then for a $0 \leq i \leq n$ it holds that

$$\mathbf{A}^{(n)}(s,t;u,i) := \underset{s<u_1<\cdots u_i \leq u < u_i < \cdots u_n \leq t}{\int_s^t \cdots \int_s^{u_2}} \, \text{d}\mathbf{A}(u_1) \cdots \, \text{d}\mathbf{A}(u_n) \tag{50}$$

$$= \left(\underset{s<u_1<\cdots u_i \leq u}{\int_s^u \cdots \int_s^{u_2}} \text{d}\mathbf{A}(u_1) \cdots \, \text{d}\mathbf{A}(u_i) \right) \left(\underset{u<u_{i+1}<u_{i+2}<\cdots u_n \leq t}{\int_u^t \cdots \int_s^{u_i}} \, \text{d}\mathbf{A}(u_i) \cdots \, \text{d}\mathbf{A}(u_n) \right). \tag{51}$$

and since $\mathbf{A}^{(n)}(s,t) = \sum_{i=0}^{n} \mathbf{A}^{(n)}(s,t;u,i)$ we have by

$$\sum_{n=1}^{\infty} \mathbf{A}^{(n)}(s,t) = \sum_{n=1}^{\infty} \sum_{i=0}^{n} \mathbf{A}^{(n)}(s,t;u,i) \tag{52}$$

$$= \sum_{n=1}^{\infty} \sum_{i=0}^{n} \mathbf{A}^{(i)}(s,u)\mathbf{A}^{(n-i)}(u,t) \tag{53}$$

$$= \left(\sum_{n=1}^{\infty} \mathbf{A}^{(n)}(s,u) \right) \left(\sum_{n=1}^{\infty} \mathbf{A}^{(n)}(u,t) \right). \tag{54}$$

Giving that the Péano series is multiplicative i.e.

$$\mathcal{P}(s,t;\mathbf{A}) = \mathcal{P}(s,u;\mathbf{A})\mathcal{P}(u,t;\mathbf{A}). \tag{55}$$

Now take any partitioning of the interval $s = t_1 < \cdots < t_n = t$. We have

$$\mathcal{P}(s,t;\mathbf{A}) - \prod_{i=1}^{n} \Big(\text{Id} + \mathbf{A}(t_i) - \mathbf{A}(t_{i-1}-) \Big) = \prod_{i=1}^{n} \mathcal{P}(t_{i-1},t_i;\mathbf{A}) - \prod_{i=1}^{n} \Big(\text{Id} + \mathbf{A}(t_i) - \mathbf{A}(t_{i-1}) \Big) \tag{56}$$

$$= \sum_{i=1}^{n} \prod_{j=1}^{i-1} \Big(\text{Id} + \mathbf{A}(t_j) - \mathbf{A}(t_{j-1}) \Big) \Big(\mathcal{P}(t_i,t_{i-1};\mathbf{A}) - \mathbf{A}(t_i) - \mathbf{A}(t_{i-1}) - \text{Id} \Big) \prod_{k=i+1}^{n} \mathcal{P}(t_k,t_{k-1};\mathbf{A}). \tag{57}$$

Using a domination argument on the Péano series and using a dominating function $\mathbf{A}_0$ we can obtain an inequality as follows

$$\left|\mathcal{P}(s,t;\mathbf{A}) - \prod_{i=1}^{n}\Big(\mathrm{Id} + \mathbf{A}(t_i) - \mathbf{A}(t_{i-1})\Big)\right| \leq \mathbf{K}_n\left(\mathbf{A}_0(t) - \mathbf{A}_0(s)\right)\exp\left(\mathbf{A}_0(t) - \mathbf{A}_0(s)\right)^2, \tag{58}$$

The constant $\mathbf{K}_n$ tends to zero giving that **the Pèano series is equal to the product integral.** We can now use Fubini to see that

$$\mathbf{A}^{(n+1)}(s,t) = \int_s^t \mathbf{A}^{(n-1)}(s,u-)\,\mathrm{d}\mathbf{A}(u) \tag{59}$$

giving that

$$\begin{aligned}\mathrm{d}_t\mathbf{Y}((s,t]) &= \sum_{n=1}^{\infty}\mathrm{d}_t\mathbf{A}^{(n)}(s,t) = \sum_{n=1}^{\infty}\mathbf{A}^{(n-1)}(s,t-)\,\mathrm{d}\mathbf{A}(t) &(60)\\ &= \left(\mathrm{Id} + \sum_{n=1}^{\infty}\mathbf{A}^{(n)}(s,t-)\right)\,\mathrm{d}\mathbf{A}(t) = \mathcal{P}(s,t-)\,\mathrm{d}\mathbf{A}(t) &(61)\\ &= \mathbf{Y}((s,t])\,\mathrm{d}\mathbf{A}(t). &(62)\end{aligned}$$

The backward equation is derived from the identity

$$\mathbf{A}^{(n+1)}(s,t) = \int_s^t \mathrm{d}\mathbf{A}(u)\mathbf{A}^{(n)}(u,t). \tag{63}$$

□

Giving this notation we have the following proposition.

Proposition 16. *Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. Let $\mathbf{Z} \sim$ PJP($\mathbb{F}$) for some filtration $\mathbb{F}$. Assuming for some $s \geq 0$ it holds that $p(s|\mathcal{G})$ is deterministic. Then*

$$p(t|\mathcal{G})^\top = p(s|\mathcal{G})^\top \prod_s^t \Big(Id + d\Lambda(u|\mathcal{G})\Big). \tag{64}$$

Proof. By theorem 12 it holds for all $j \in \mathcal{Z}$:

$$\mathrm{d}p^j(t|\mathcal{G}) = \sum_{k\in\mathcal{Z}} p^k(t-|\mathcal{G})\,\mathrm{d}\Lambda^{kj}(t|\mathcal{G}). \tag{65}$$

thus in particular we can translate states through ψ and write on matrix form

$$\mathrm{d}p(t|\mathcal{G})^\top = p(t-|\mathcal{G})^\top d\Lambda(t|\mathcal{G}). \tag{66}$$

Thus the jump points on $p(t|\mathcal{G})$ happens at the jump points of $\tilde{p}^{jk}(t|\mathcal{G})$ hence by the definition of the product integral using that $p(s|\mathcal{G})$ is deterministic we get the desired result. □

With these results we turn to estimation.

6 Estimation with censoring

We follow the setup of the one given i Bladt **and** Furrer 2023b. We start by defining what data we will be working with.

6.1 Observations and censoring

Life processes modelled as pure jump processes on a state-space $\mathcal{Z}$ is in many cases right-censored. This is obviously because many in the sample data is still alive and so the final transition into the absorbing death state is yet to be observed. Consider a pure jump process $\mathbf{Z} \sim \mathrm{PJP}(\mathbb{F})$ as described above. We order the related space $\mathcal{J}$ such that the last state $J \in \mathcal{J}$ is the absorbing state while all other states $j = 1, ..., J-1$ is transient. Thus the following $\mathbb{F}$-stopping time τ is well-defined. We define the absorption time τ as:

$$\tau = \inf\{s \geq 0 : \psi(Z_s) = J\}. \tag{67}$$

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

$$(X, (Z_t)_{0 \leq t \leq R}, \tau \wedge R), \tag{68}$$

where the $\wedge$-operator gives the minimum and X is a generic outcome from conditioning variable. We will be making the following assumption.

Assumption 1. The right-censoring mechanism R is entirely independent of $\mathbf{Z} \mid \mathcal{G}$, that is

$$R \perp\!\!\!\perp \mathbf{Z} \mid X. \tag{69}$$

We thus may introduce the following modified probabilities and transition rates.

Definition 17. The censored occupational probabilities and transition rates given $X = x$ is defined as

$$p_j^c(t|x) = \mathbb{E}[\mathbb{1}_{\{Z_t = j\}} \mathbb{1}_{\{t<R\}} | X = x], \tag{70}$$
$$\tilde{p}_{jk}^c(t|x) = \mathbb{E}[N^{jk}(t \wedge R) | X = x]. \tag{71}$$

Given this we arrive at the following.

Proposition 18. *Under assumption 1 it holds that*

$$\Lambda^{jk}(t|x) = \Lambda^{c,jk}(t|x) = \int_0^t \frac{1}{p_j^c(s-|x)} d_s \tilde{p}_{jk}^c(s|x). \tag{72}$$

Proof. Under assumption 1 it holds that

$$\begin{aligned} p_j^c(t|x) &= \mathbb{E}[\mathbb{1}_{\{Z_t = j\}} \mathbb{1}_{\{t<R\}} | X = x] = \mathbb{E}[\mathbb{1}_{\{Z_t = j\}} | X = x] \mathbb{E}[\mathbb{1}_{\{t<R\}} | X = x] \quad (73) \\ &= p^j(t|x) \mathbb{P}(t < R | X = x), \quad (74) \end{aligned}$$

and

$$\tilde{p}^c_{jk}(t|x) = \mathbb{E}[N^{jk}(t \wedge R)|X=x] = \mathbb{E}\left[\int_0^t \mathbb{1}_{\{s<R\}}\,\mathrm{d}N^{jk}(s) \,\middle|\, X=x\right] \tag{75}$$

$$= \int_0^t \mathbb{E}[\mathbb{1}_{\{s<R\}}|X=x]\,\mathrm{d}\mathbb{E}[N^{jk}(s)|X=x] = \int_0^t \mathbb{P}(s<R|X=x)\,\mathrm{d}_s\tilde{p}^{jk}(s|x). \tag{76}$$

Thus

$$\int_0^t \frac{1}{p^c_j(s-|x)}\mathrm{d}_s p^c_{jk}(s|x) = \int_0^t \frac{1}{p^j(t-|x)\mathbb{P}(t<R|X=x)}\mathbb{P}(s<R|X=x)\,\mathrm{d}_s\tilde{p}^{jk}(s|x) \tag{77}$$

$$= \int_0^t \frac{1}{p^j(t-|x)}\,\mathrm{d}_s\tilde{p}^{jk}(s|x) = \Lambda^{jk}(t|x) \tag{78}$$

as desired. □

This in particular means, that if we accept the assumption that censoring is independent of the conditioned process $(Z_t|X)_{t\geq 0}$, then we can one to one relate the cumulative transition rates in the censored case with the ones in the uncensored case.

6.2 Estimators

Assume that we have $\ell = 1, ..., L$ samples

$$(X^\ell, (Z^\ell_t)_{0\leq t\leq R}, \tau^\ell \wedge R^\ell)_{\ell=1,\dots,L}. \tag{79}$$

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

$$g^{(L)} = \frac{1}{n}\sum_{\ell=1}^{L} g(x^\ell), \tag{80}$$

$$\mathbb{N}^{(L)}_{jk}(t|x) = \frac{1}{L}\sum_{\ell=1}^{L} N^\ell_{jk}(t\wedge R^\ell)\frac{g(X^\ell)}{g^{(L)}}, \tag{81}$$

$$\mathbb{I}^{(L)}_j(t|x) = \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{1}_{\{t<R^\ell\}}\mathbb{1}_{\{Z^\ell_t=j\}}\frac{g(X^\ell)}{g^{(L)}}, \tag{82}$$

where g is a function yet to be defined. The function $g(\cdot)$ is chosen according to the conditioning variable and in such a way that

$$\mathbb{N}^{(L)}_{jk}(t|x) \overset{\text{a.s.}}{\to} \tilde{p}^c_{jk}(t|x), \quad \mathbb{I}^{(L)}_j(t|x) \overset{\text{a.s.}}{\to} p^c_j(t|x) \tag{83}$$

for $L \to \infty$.

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

$$\hat{\Lambda}^{(L)}_{jk}(t|x) = \int_0^t \frac{1}{\mathbb{I}^{(L)}_j(s-|X)}\,\mathrm{d}_s\mathbb{N}^{(L)}_{jk}(s|x). \tag{84}$$

Definition 20 (Aalen-Johansen estimator)**.** The conditional Aalen-Johansen estimator is defined as

$$\hat{p}^{(L)}(t|x)^\top = \hat{p}^{(L)}(0|x)^\top \prod_0^t \Big(\mathrm{Id} + \mathrm{d}_u \hat{\Lambda}^{(L)}(u|x)\Big) \tag{85}$$

with $\hat{p}_j^{(L)}(0|x) = \mathbb{I}_j^{(L)}(0|x)$.

The above definition above amounts to the differential form

$$\mathrm{d}_t \hat{p}_j^{(L)}(t|x) = \sum_{k \in \mathcal{Z}} \hat{p}_k^{(L)}(t|x) \frac{1}{\mathbb{I}_j^{(L)}(t-|x)} \mathrm{d}_t \mathbb{N}_{jk}^{(L)}(t|x) \tag{86}$$

with boundary condition $\hat{p}_j^{(L)}(0|x) = \mathbb{I}_j^{(L)}(0|x)$. Such a solution is actually attained rather easily. Consider the following lemma in this endavour.

Lemma 21. *Let* $\mathbf{Y}((s,t])$ *be the product integral of* $\mathbf{A}$ *on* $(s,t]$*. Then for any* $u \in \mathbb{R}$

$$\mathbf{Y}((s,t]) = \mathbf{Y}((s,u])\mathbf{Y}((u,t]). \tag{87}$$

Furthermore, if $\mathbf{A}$ *is CADLAG with at most countable infinite discontinuous points on* $(s,t]$ *say* $(t_i)_{i=1,\ldots,N}$ *(with* N *possible being* ∞*) being the ordered set of discontinuous points, then the product integral is*

$$\mathbf{Y}((s,t]) = \left(\prod_{i=1}^N \mathbf{Y}((s,t_{i-1}])\Big(Id + \Delta\mathbf{A}(t_i)\Big)\right)\mathbf{Y}((\sup_i t_i, t]), \tag{88}$$

using that $\Delta\mathbf{A}(t) = \mathbf{A}(t) - \mathbf{A}(t-)$*. In the case where* $d\mathbf{A}(t) = a(t)\mathbf{A}\ dt$ *(a being continuous) on a set* $(s,t]$ *the integral simplifies to* $\mathbf{Y}((s,t]) = \exp(\mathbf{A}\int_s^t a(u)\ du)$ *being the matrix exponential of* $\mathbf{A}\int_s^t a(u)\ du$*.*

Proof. We lay a rough sketch of the proof. Start by using that the product integral is equal to the Péano-series according to the definition in equation 48. Then using multiplicity of the series we can write the Péano-series as a product of all the ordered partitioning that $(t_i)_{i=1,\ldots,N}$ give rise to. Doing some calculations reveals that the jump points t_i may be extracted from the series and so the result follows.

Regarding the matrix exponential: Evaluate the Péano series to see that is is indeed the sum as in the definition of the exponential function as defined in the matrix case. □

We can then apply this lemma to the following special case.

Theorem 22 (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. Set* $X = \mathbb{1}_{\{Z_s = j\}}$*. 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* $(\mathbf{Z}, R)$*. Set furthermore* $g(x) = x$*. It holds that the estimators:*

$$p_i^c(t|Z_s = j) \overset{def}{=} p_i^c(t|\sigma(Z_s = j)), \quad \tilde{p}_{ik}^c(t|Z_s = j) \overset{def}{=} \tilde{p}_{ik}^c(t|\sigma(Z_s = j)), \tag{89}$$

Then the samples $\mathbb{N}_{ik}^{(L)}$ *and* $\mathbb{I}_i^{(L)}$ *simplifies to the below with the following almost sure convergence*

$$\mathbb{N}_{ik}^{(L)}(t|Z_s = j) = \frac{1}{Lg^{(L)}} \sum_{\ell=1}^{L} N_{ik}^{\ell}(t \wedge R^{\ell}) \mathbb{1}_{\{Z_s^{\ell}=j\}} \qquad \overset{a.s.}{\to} \quad \tilde{p}_{ik}^{c}(t|Z_s = j), \tag{90}$$

$$\mathbb{I}_i^{(L)}(t|Z_s = j) = \frac{1}{Lg^{(L)}} \sum_{\ell=1}^{L} \mathbb{1}_{\{t<R^{\ell}\}} \mathbb{1}_{\{Z_t^{\ell}=i\}} \mathbb{1}_{\{Z_s^{\ell}=j\}} \qquad \overset{a.s.}{\to} \quad p_i^{c}(t|Z_s = j). \tag{91}$$

Lastly, let $0 = \tau_0 \leq \tau_1 \leq ... \leq \tau_M$ *be the ordered set of all transitions across all L observations. Define* $\langle t \rangle = \sum_{i=1}^{M} \mathbb{1}_{[\tau_i,\infty)}(t)$ *as the number of transitions up until t. Then* $\hat{p}^{(L)}(t|Z_s = j)^{\top}$ *is given by*

$$\hat{p}^{(L)}(t|Z_s = j)^{\top} = L \cdot \mathbb{I}^{(L)}(0|Z_s = j)^{\top} \prod_{i=1}^{\langle t \rangle} \Big(Id + \Delta\hat{\Lambda}^{(L)}(\tau_i|Z_s = j) \Big) . \tag{92}$$

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

$$\hat{p}^{(L)}(t|Z_s = j) = \mathbf{e}_j^{\top} \prod_{i=\langle s \rangle}^{\langle t \rangle} \Big(Id + \Delta\hat{\Lambda}^{(L)}(t_u|Z_s = j) \Big) \tag{93}$$

for $t \geq s$.

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

$$\frac{1}{L} \sum_{\ell=1}^{L} \frac{N_{ik}^{\ell}(t \wedge R^{\ell}) \mathbb{1}_{\{Z_s^{\ell}=j\}}}{g^{(L)}} \overset{\text{a.s.}}{\to} \frac{\mathbb{E}\left[N_{ik}(t \wedge R) \mathbb{1}_{\{Z_s=j\}} \right]}{\mathbb{P}\left(Z_s = j \right)} = \mathbb{E}\left[N_{ik}(t \wedge R) | Z_s = j \right], \tag{94}$$

$$\frac{1}{L} \sum_{\ell=1}^{L} \frac{\mathbb{1}_{\{t<R^{\ell}\}} \mathbb{1}_{\{Z_t^{\ell}=i\}} \mathbb{1}_{\{Z_s^{\ell}=j\}}}{g^{(L)}} \overset{\text{a.s.}}{\to} \frac{\mathbb{E}\left[\mathbb{1}_{\{t<R\}} \mathbb{1}_{\{Z_t=i\}} \mathbb{1}_{\{Z_s=j\}} \right]}{\mathbb{P}\left(Z_s = j \right)} = \mathbb{P}\left(t < R, Z_t = i | Z_s = j \right). \tag{95}$$

Thus giving the desired convergence in 90 and 91. Then using that the product integral is multiplicative we have

$$\hat{p}^{(L)}(t|Z_s = j)^{\top} = \hat{p}^{(L)}(0|Z_s = j)^{\top} \prod_{0}^{t} \Big(\mathrm{Id} + \mathrm{d}_u \hat{\Lambda}^{(L)}(u|Z_s = j) \Big) \tag{96}$$

$$= \hat{p}^{(L)}(0|Z_s = j)^{\top} \prod_{0}^{s} \Big(\mathrm{Id} + \mathrm{d}_u \hat{\Lambda}^{(L)}(u|Z_s = j) \Big) \prod_{s}^{t} \Big(\mathrm{Id} + \mathrm{d}_u \hat{\Lambda}^{(L)}(u|Z_s = j) \Big) \tag{97}$$

$$= \hat{p}^{(L)}(s|Z_s = j)^{\top} \prod_{s}^{t} \Big(\mathrm{Id} + \mathrm{d}_u \hat{\Lambda}^{(L)}(u|Z_s = j) \Big). \tag{98}$$

Recall, that $\hat{\Lambda}^{(L)}$ is a pure jump process and thus due to lemma 21 the product integral becomes

$$\prod_{0}^{t} \Big(\mathrm{Id} + \mathrm{d}_u \hat{\Lambda}^{(L)}(u|Z_s = j) \Big) = \prod_{i=1}^{\langle t \rangle} \Big(\mathrm{Id} + \Delta\hat{\Lambda}^{(L)}(\tau_i|Z_s = j) \Big). \tag{99}$$

Hence if $\hat{p}^{(L)}(s|Z_s = j) = \mathbf{e}_j$ then 92 follows. We have

$$g^{(L)}\left(\hat{p}^{(L)}(s|Z_s = j)\right)_j = \left(\hat{p}^{(L)}(0|Z_s = j)\prod_{i=1}^{\langle s\rangle}\left(\mathrm{Id} + \Delta\hat{\Lambda}^{(L)}(\tau_i|Z_s = j)\right)\right)_j \tag{100}$$

$$= \sum_{i=1}^{J}\sum_{\ell=1}^{L}\mathbb{1}_{\{Z_s^\ell = j\}}\mathbb{1}_{\{Z_0^\ell = i\}}\left(\prod_{i=1}^{\langle s\rangle}\left(\mathrm{Id} + \Delta\hat{\Lambda}^{(L)}(\tau_i|Z_s = j)\right)\right)_{ij} \tag{101}$$

$$= \ldots \tag{102}$$

$$= g^{(L)}. \tag{103}$$

In particular this implies that $\hat{p}^{(L)}(s|Z_s = j) = \mathbf{e}_j$ as desired. Jeg er ikke sikker på det er produktivt det jeg har gang i her. Tænker intuitivt at det kan lade sige gøre at vise at $\hat{p}^{(L)}(s|Z_s = j) = {}_j$ ved at regne frem fra $\hat{p}^{(L)}(0|Z_s = j)$ og der er sikkert en smart måde at tælle differens ligningerne igennem og opnå de ønskede, men jeg ville ikke lede i blinde. □

We call the above modelling framework 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 the 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 up until the time of forecasting. Therefore, in the Markov case we have that the transition probabilities simplifies to

$$p(t,s)^\top = \pi_s^\top \prod_s^t \left(\mathrm{Id} + \mathrm{d}\Lambda(u)\right), \tag{104}$$

for some initial distribution π_s at time s and the cumulative transition rates given by the integral

$$\Lambda^{ij}(t) = \int_s^t \frac{1}{\mathbb{P}(Z_{u-} = i)}\,\mathrm{d}\mathbb{E}[N^{ij}(u)]. \tag{105}$$

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. Thus 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.

7 Establishing the estimators

In this section we study numerical examples of the Aalen-Johansen estimator using the results from theorem 22. We will in particular be using data-samples generated using the `AalenJohansen` package (Bladt **and** Furrer 2023a). We start the section with a few simple examples showing the method of estimation in the Markov case and the as-if-Markov case. Then,..

7.1 Implementing the estimator

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. We can thus identify the space $\mathcal{J} = \{1, 2, 3\}$ using the mapping ψ given by $1 \mapsto a$, $2 \mapsto b$ and $3 \mapsto c$ (see figure 1). We assume that we have L samples satisfying assumption 1:

$$(\mathbb{1}_{\{Z^\ell_s = j\}}, (Z^\ell_t)_{0\le t\le R}, \tau^\ell \wedge R^\ell)_{\ell=1,\dots,L}. \tag{106}$$

We also assume that $Z_0 = a$ almost surely. 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 from lemma 21, that the product integral of such a Matrix and using the boundary condtion $\pi_0 = (1, 0, 0)$ (3×1-matrix) we have that

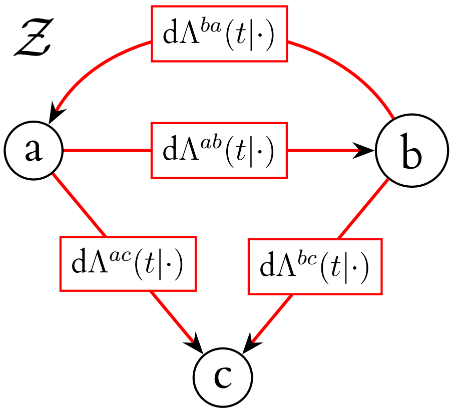
$\mathcal{Z}$
$\mathrm{d}\Lambda^{ba}(t|\cdot)$
a
$\mathrm{d}\Lambda^{ab}(t|\cdot)$
b
$\mathrm{d}\Lambda^{ac}(t|\cdot)$
$\mathrm{d}\Lambda^{bc}(t|\cdot)$
c

Figure 1: A representation of the chain discussed in sub-section *Implementing the estimator*

$$p(0,t)^\top = \pi_0^\top \exp\left(\mathbf{M}\int_0^t \lambda(u)\,\mathrm{d}u\right). \tag{107}$$

In the above $p(0, t) = (\mathbb{P}(Z_t = j))_{j\in\mathcal{Z}}$ and $\pi_0 = p(0, 0)$. In other words, we can easily compute the true values of the transition probabilities using regular matrix exponentiation. For this examples we choose

$$\lambda(t) = \frac{1}{1+\frac{1}{2}t} \quad \text{and} \quad \mathbf{M} = \begin{pmatrix} -3 & 2 & 1 \\ 3 & -4 & 1 \\ 0 & 0 & 0 \end{pmatrix}. \tag{108}$$

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 appendix A.1.

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 $\psi(Z_{3-}) = 2$.

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

$$\int_s^t \lambda(u)\,\mathrm{d}u = 2\int_{1+\frac{1}{2}s}^{1+\frac{1}{2}t} \frac{1}{y}\,\mathrm{d}y = 2\log\left(1+\frac{1}{2}t\right) - 2\log\left(1+\frac{1}{2}s\right). \tag{109}$$

Thus we can easily compare the Aalen-Johansen estimator with the true value of $p(0,t)$. In fact, we have that

$$p(0,t)^\top = (1,0,0)^\top \exp\left(2\log\left(1+\frac{1}{2}t\right)\mathbf{M}\right). \tag{110}$$

Regarding estimation, we use a matrix-based approach and transform a main dataframe. Recall from theorem 22 that the Aalen-Johansen estimator is uniquely given by the product integral of the cumulative transition rates. And since the rates only changes on jump times of the observed chains, we get the result

$$\mathrm{d}_t \underbrace{\left(\prod_0^t \left(\mathrm{Id} + \mathrm{d}_u\hat{\Lambda}^{(L)}(u|Z_s=j)\right)\right)}_{:=\hat{p}^{(L)}(s,j,(0,t])} = \hat{p}^{(L)}(s,j,(0,t])\Big(\mathrm{Id} + \Delta\hat{\Lambda}^{(L)}(u|Z_s=j)\Big). \tag{111}$$

Thus on discontinuity points τ_i we have

$$\hat{p}^{(L)}(s,j,(0,\tau_i]) = \hat{p}^{(L)}(s,j,(0,\tau_i-]) + \hat{p}^{(L)}(s,j,(0,\tau_i-])\Delta\hat{\Lambda}^{(L)}(\tau_i|Z_s=j). \tag{112}$$

Or entrywise we have the difference equation

$$\begin{aligned}\hat{p}^{(L)}_{lk}(s,j,(0,\tau_i]) &= \hat{p}^{(L)}_{lk}(s,j,(0,\tau_i-]) + \sum_{m=1}^{J}\hat{p}^{(L)}_{lm}(s,j,(0,\tau_i-])\Delta\hat{\Lambda}^{(L)}_{mk}(\tau_i|Z_s=j) &(113)\\ &= \hat{p}^{(L)}_{lk}(s,j,(0,\tau_i-]) + \sum_{m=1}^{J}\frac{\hat{p}^{(L)}_{lm}(s,j,(0,\tau_i-])}{\mathbb{I}^{(L)}_{m}(\tau_i-|Z_s=j)}\Delta\mathbb{N}^{(L)}_{mk}(\tau_i|Z_s=j). &(114)\end{aligned}$$

We of cause have that

$$\Delta\mathbb{N}^{(L)}_{mk}(\tau_i|Z_s=j) = \frac{1}{Lg^{(L)}}\sum_{\ell=1}^{L}\mathbb{1}_{\{Z^\ell_{\tau_i}=k\}}\mathbb{1}_{\{Z^\ell_{\tau_i-}=m\}}\mathbb{1}_{\{\tau_i<R^\ell\}}, \tag{115}$$

$$\mathbb{I}^{(L)}_{m}(\tau_i-|Z_s=j) = \frac{1}{Lg^{(L)}}\sum_{\ell=1}^{L}\mathbb{1}_{\{Z^\ell_{\tau_i-}=m\}}\mathbb{1}_{\{\tau_i<R^\ell\}}. \tag{116}$$

Although we can ignore the fraction as they cancel out in $\Delta\mathbb{N}^{(L)}_{mk}(\tau_i|Z_s=j)/\mathbb{I}^{(L)}_{m}(\tau_i-|Z_s=j)$. Then we have

$$\hat{p}^{(L)}_{lk}(s,j,(0,\tau_i]) = \hat{p}^{(L)}_{lk}(s,j,(0,\tau_i-]) + \sum_{m=1}^{J}\frac{\hat{p}^{(L)}_{lm}(s,j,(0,\tau_i-])}{\sum_{\ell=1}^{L}\mathbb{1}_{\{Z^\ell_{\tau_i-}=m\}}\mathbb{1}_{\{\tau_i<R^\ell\}}}\sum_{\ell=1}^{L}\mathbb{1}_{\{Z^\ell_{\tau_i}=k\}}\mathbb{1}_{\{Z^\ell_{\tau_i-}=m\}}\mathbb{1}_{\{\tau_i<R^\ell\}} \tag{117}$$

Thus by cleverly coding the function $\mathbb{I}_m^{(L)}(\tau_i - |Z_s = j)$ and $\Delta\mathbb{N}_{mk}^{(L)}(\tau_i|Z_s = j)$ we can easily implement the the function $\hat{p}_{lk}^{(L)}(s, j, (0, \tau_i])$ for all unique τ_i.

Table 7.1 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.

We start by calculating $\mathbb{I}_m^{(L)}(\tau_i - |Z_s = j)$ and $\mathbb{N}_{ml}^{(L)}(\tau_i - |Z_s = j)$ in the As-If-Markov case or $\mathbb{I}_m^{(L)}(\tau_i -)$ and $\mathbb{N}_{ml}^{(L)}(\tau_i -)$ in the Markov case. The code is given in appendix A.2.2 and A.2.3. Then we may calculate the increments for the Nelson-Aalen estimator (see appendix A.2.4) and finally compute the transition probabilities $\hat{p}^{(L)}(s, j, (0, t])$ and $p(0, t]$ for the as-if case and Markov case respectively (see appendix A.2.6). For the estimates $\hat{p}^{(L)}(t|Z_s = j)$ we simply follow 92.

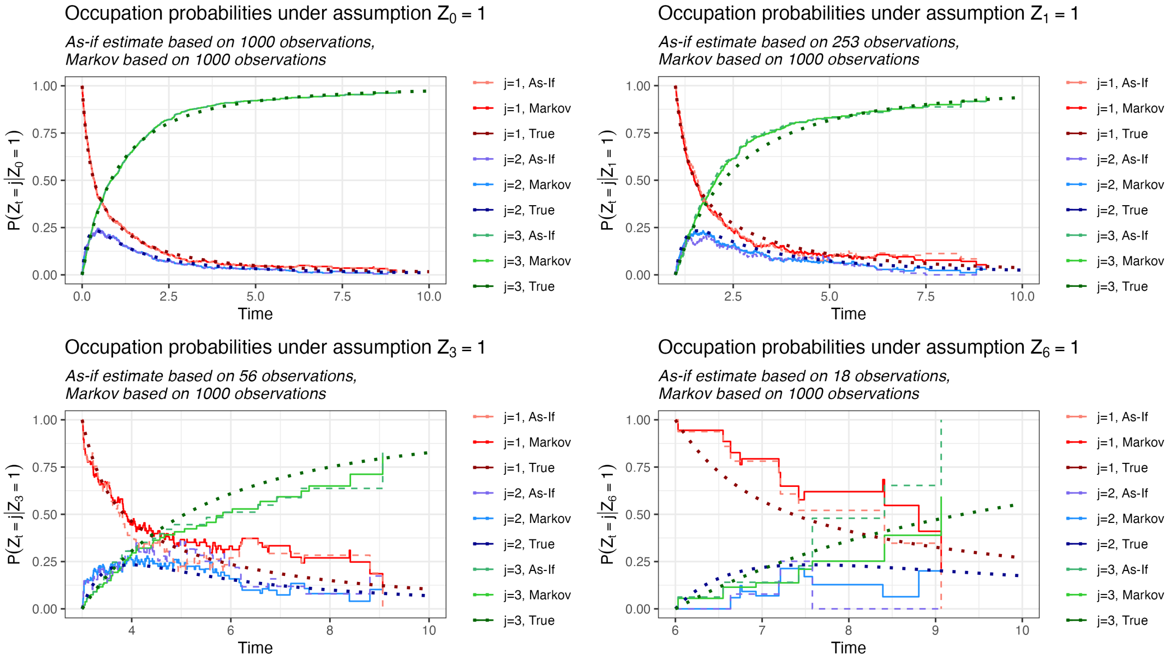

Figure 2: The figure shows the conditional probabilities under a Markov assumption and As-If-Markov assumption. The dashed line is estimate from the As-If model, the regular line is under the Markov assumption and the dots represent the true value. Based on $L = 1,000$ samples. (code given in appendix A.3)

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 $Z_s = 1$ for $s = 0, 1, 3, 6$).

References

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

– (2023b). *Conditional Aalen–Johansen estimation.* DOI: `10.48550/ARXIV.2303.02119`. URL: `https://arxiv.org/abs/2303.02119`.

Buchardt, Kristian, Christian Furrer **and** Oliver Lunding Sandqvist (**march** 2023). ?Transaction time models in multi-state life insurance? **in***Scandinavian Actuarial Journal*: **pages** 1–26. DOI: 10.1080/03461238.2023.2181708. URL: https://doi.org/10.1080/03461238.2023.2181708.

Christiansen, Marcus C. (**may** 2021). ?On the calculation of prospective and retrospective reserves in non-Markov models? **in***European Actuarial Journal*: 11.2, **pages** 441–462. DOI: 10.1007/s13385-021-00277-y. URL: https://doi.org/10.1007/s13385-021-00277-y.

Christiansen, Marcus C. **and** Christian Furrer (**november** 2022). ?Extension of as-if-Markov modeling to scaled payments? **in***Insurance: Mathematics and Economics*: 107, **pages** 288–306. DOI: 10.1016/j.insmatheco.2022.09.001. URL: https://doi.org/10.1016/j.insmatheco.2022.09.001.

Gill, Richard D. **and** Soren Johansen (**december** 1990). ?A Survey of Product-Integration with a View Toward Application in Survival Analysis? **in***The Annals of Statistics*: 18.4. DOI: 10.1214/aos/1176347865. URL: https://doi.org/10.1214/aos/1176347865.

Hansen, Ernst (2021). *Stochastic Processes*. en. 2 **edition**.

A Source code

A.1 Simulating time-in-homogeneous Markov Chain

```
M <- t(matrix(c(-3,2,1,
             3,-4,1,
             0,0,0),ncol=3,nrow=3))
lambda <- function(t){
  1/(1+0.5*t)
}
d_Lambda <- function(t,M,lambda){
  lambda(t)*M
}
jump_rate <- function(i,t,u){
  #the variable u is not used, it will be used later for time sojourned in the current state
  d_L_t <- d_Lambda(t,M,lambda)
  J <- dim(d_L_t)[1]
  vec <- (1:J==i)*1
  -vec%*%d_L_t%*%vec
}
mark_dist <- function(i,s,v){
  #the variable v is not used
  d_L_t <- d_Lambda(s,M,lambda)
  J <- dim(d_L_t)[1]
  vec <- (1:J==i)*1
  tmp <- (vec %*% d_L_t)* (1-vec)
  tmp / sum(tmp)
}

L <- 1000
set.seed(1)
R <- runif(L,0,10)
#Simulate paths
paths <- lapply(1:L,function(n) {
  sim_path(1,rates = jump_rate, dist = mark_dist,
        tn = R[n], bs= c(R[n]*3,R[n]*4,0))})
```

A.2 Calculation of estimators

A.2.1 Extraction of observations from `paths`, see A.1

```
#Convert path data to main_df
paths_to_df <- function(paths){
  times <- unlist(lapply(1:L, function(i){
    paths[[i]]$times
  }))
  states <- unlist(lapply(1:L, function(i){
    paths[[i]]$states
  }))
  obs <- unlist(lapply(1:L, function(i){
    rep(i,length(paths[[i]]$times))
  }))
  df <- data.frame(Id = obs, Start_Time = times, Start_State = states)
  #End time & end state
  df <- df %>%
    arrange(Start_Time) %>%
    group_by(Id) %>%
    mutate(End_Time = data.table::shift(Start_Time,-1),
           End_State = data.table::shift(Start_State,-1)) %>%
    ungroup() %>%
    replace(is.na(.), Inf) %>%
    arrange(Id, Start_Time) %>%
    mutate(End_State = ifelse(is.infinite(End_State),Start_State,End_State),
           Censored = ifelse(Start_State == End_State,
                             ifelse(is.finite(End_Time),TRUE,FALSE),FALSE)) %>%
    group_by(Id) %>%
    mutate(Censored = ifelse((!Censored) & is.infinite(End_Time) & (cumsum(Censored) >0),
        ↪ TRUE,Censored)) %>%
    ungroup()
  df <- df[,c("Id","Start_Time","End_Time","Start_State","End_State","Censored")]
  return(df)
}
main_df <- paths_to_df(paths)
```

A.2.2 Calculation of $\mathbb{I}$

```
#Convert main_df to I
df_to_I <- function(df,num_states) {
  Init <- df %>%
    group_by(State = Start_State) %>%
    summarise(Time = 0,
              Change = sum(Start_Time == 0))
  I <- suppressMessages(df %>%
    filter(End_Time < Inf) %>%
    mutate(Count_from = TRUE) %>%
    bind_rows(filter(.,Censored == FALSE) %>% mutate(Count_from = FALSE)) %>%
    arrange(Id,End_Time) %>%
    mutate(State = ifelse(Count_from, Start_State,End_State)) %>%
    group_by(Time = End_Time, State) %>%
    summarise(Change = sum((!Count_from)*1)-sum(Count_from*1)) %>%
    ungroup() %>%
    bind_rows(Init) %>%
    arrange(Time) %>%
    group_by(State) %>%
    mutate(I_j = cumsum(Change)) %>%
    reshape2::dcast(Time ~ State, value.var = "I_j") %>%
    zoo::na.locf(na.rm = FALSE) %>%
    replace(is.na(.), 0))
  if ( sum(!(1:num_states %in% colnames(I))) >0) {
    I[,1:num_states[!(1:num_states %in% colnames(I))]] <- 0
  }
  I <- I[,c("Time",1:num_states)]
  I_left_limit <- I
  I_left_limit[2:(dim(I)[1]-1),2:dim(I)[2]] <- I[3:dim(I)[1],2:dim(I)[2]]
  I_left_limit[dim(I)[1],2:dim(I)[2]] <- I[(dim(I)[1]-1),2:dim(I)[2]]
  return(list(I = I, I_left_limit = I_left_limit))
}
I_list <- df_to_I(main_df,3)
```

A.2.3 Calculation of $\mathbb{N}$

```
#Convert main_df to N
df_to_N <- function(df, num_states) {
  N <- df %>%
    filter((Censored == FALSE) & (End_Time < Inf)) %>%
    arrange(End_Time) %>%
    group_by(Start_State, End_State) %>%
    mutate(Transitions = cumsum(End_State >= 0),
           #j_k = paste0(Start_State,",", End_State)
    ) %>%
    select(Time = End_Time,Start_State, End_State,Transitions) %>%
    bind_rows(filter(., Transitions == 1) %>% mutate(Time = 0, Transitions = 0)) %>%
    reshape2::dcast(Time ~ Start_State * End_State, value.var = "Transitions") %>%
    zoo::na.locf(na.rm = FALSE) %>%
    replace(is.na(.), 0)
  c <- paste0(unlist(lapply( 1:num_states, function(i) rep(i,num_states))),"_",rep(1:num_states,num_states))
  if ( sum(!(c %in% colnames(N))) >0) {
    N[,c[!(c %in% colnames(N))]] <- 0
  }
  N <- N[,c("Time",c)]
  for (i in 1:num_states) {
    target <- paste0(i,"_",(1:num_states)[!(1:num_states %in% i)])
    N[,paste0(i,"_",i)] <- -rowSums(N[,target])
  }
  delta_N <- N
  delta_N[2:dim(N)[1],2:dim(N)[2]] <- delta_N[2:dim(N)[1],2:dim(N)[2]] - delta_N[1:(dim(N)[1]-1),2:dim(N)[2]]
  return(list(N = N, delta_N = delta_N))
}
N_list <- df_to_N(main_df, 3)
```

A.2.4 Calculation of Nelson-Aalen estimate

```
#Calculate Nelson-Aalen
N_I_to_NA <- function(I_list,N_list, num_states) {
  delta_N <- N_list$delta_N
  I_left_limit <- I_list$I_left_limit
  I_left_limit <- I_left_limit[I_left_limit$Time %in% delta_N$Time,]
  I_factor <- delta_N
  for (i in 1:num_states) {
    target <- paste0(i,"_",1:num_states)
    vec <- 1/I_left_limit[,colnames(I_left_limit) == i]
    vec <- ifelse(is.infinite(vec),0,vec)
    I_factor[,target] <- vec
  }
  delta_NelsonAalen <- delta_N
  delta_NelsonAalen[,2:dim(delta_N)[2]] <- delta_NelsonAalen[,2:dim(delta_N)[2]]*I_factor[,2:dim(delta_N)[2]]
  NelsonAalen <- delta_NelsonAalen
  NelsonAalen[,2:dim(delta_N)[2]] <- cumsum(delta_NelsonAalen[,2:dim(delta_N)[2]])
  return(list(NelsonAalen = NelsonAalen,delta_NelsonAalen=delta_NelsonAalen))
}
NelsonAalen_list <- N_I_to_NA(I_list,N_list,3)
```

A.2.5 Calculation of transition matrix $\prod_0^t \left(Id + d_u\hat{\Lambda}(u)\right)$

```
#Calculate p(0,t) i.e. Aalen-Johansen
#Calculate product integral
NA_to_p <- function(I_list,N_list,NelsonAalen_list, num_states) {
  #This may be slow
  identity <- diag(as.numeric(1), ncol = num_states,nrow=num_states)
  Nelson <- NelsonAalen_list$delta_NelsonAalen
  Nelson <- lapply(1:dim(Nelson)[1], function(i) matrix(Nelson[i,2:dim(Nelson)[2]],nrow=num_states,ncol = num_states,byrow=TRUE))
  Prod_int <- list()
  Prod_int[[1]] <- identity
  for (i in 2:length(Nelson)) {
    Prod_int[[i]] <- Prod_int[[i-1]] %*% (identity + as.numeric(Nelson[[i]]))
  }
  P <- NelsonAalen_list$delta_NelsonAalen
  P[,2:dim(P)[2]] <- matrix(unlist(Prod_int),ncol = num_states**2, byrow = TRUE)
  return(P)
}
transitionprobs <- NA_to_p(I_list,N_list,NelsonAalen_list, 3)
```

A.2.6 Calculation of transition probabilities $\mathbf{e}_j \prod_s^t \left(Id + d_u\hat{\Lambda}(u)\right)$

```
#Conditional probabilities
P_conditioned <- function(NelsonAalen_list,s,j,num_states) {
  Init <- (1:num_states==j)*1
  Nelson <- NelsonAalen_list$delta_NelsonAalen %>%
    filter(Time >= s)
  Nelson <- lapply(1:dim(Nelson)[1], function(i) matrix(Nelson[i,2:dim(Nelson)[2]],nrow=num_states,ncol = num_states,byrow=TRUE))
  Prod_int <- list()
  Prod_int[[1]] <- Init
  identity <- diag(as.numeric(1), ncol = num_states,nrow=num_states)
  for (i in 1:length(Nelson)) {
    Prod_int[[i+1]] <- Prod_int[[i]] %*% (identity + as.numeric(Nelson[[i]]))
  }
  p_con <- NelsonAalen_list$delta_NelsonAalen %>%
    filter(Time >= s) %>%
    bind_rows(filter(., row_number()==1) %>% mutate(Time = s)) %>%
    arrange(Time)
  p_con <- p_con[,1:(num_states + 1)]
  colnames(p_con) <- c("Time",1:num_states)
  p_con[,2:dim(p_con)[2]] <- matrix(unlist(Prod_int),ncol = num_states, byrow = TRUE)
  return(p_con)
}
```

A.2.7 Calculation of conditional estimate

```
#Putting it all together with as-if variant
Estimate <- function(paths,num_states,s= NA, j = NA, as_if = FALSE,debug = TRUE) {
  start_time <- Sys.time()
  # 1. Start by converting to data frame
  if (is.na(s)) {
    s <- 0
    j <- 1
  }
  main_df_tmp <- paths_to_df(paths)
  if(debug) {
    print(paste0("Convert␣paths␣to␣main_df:␣",round(Sys.time()-start_time,digits=2),"␣seconds."))
    start_time <- Sys.time()
  }
  if (as_if) {
    tmp <- main_df_tmp %>%
      filter(Start_Time <= s) %>%
      group_by(Id) %>%
      mutate(max_Time = max(Start_Time)) %>%
      filter(max_Time == Start_Time) %>%
      filter(((is.finite(End_Time) | (Censored == FALSE)))) %>%
      filter(Start_State == j)
    main_df_tmp <- main_df_tmp %>% filter(Id %in% tmp$Id)
  }
  # 2. Calculate I and N
  I_list_tmp <- df_to_I(main_df_tmp,num_states)
  if (debug) {
    print(paste0("Convert␣main_df␣to␣I:␣",round(Sys.time()-start_time,digits=2),"␣seconds."))
    start_time <- Sys.time()
  }
  N_list_tmp <- df_to_N(main_df_tmp,num_states)
  if (debug) {
    print(paste0("Convert␣main_df␣to␣N:␣",round(Sys.time()-start_time,digits=2),"␣seconds.")
       ↪ )
    start_time <- Sys.time()
  }
  # 3. Calculate Nelson-Aalen
  NelsonAalen_list_tmp <- N_I_to_NA(I_list_tmp,N_list_tmp,num_states)
  if (debug) {
    print(paste0("Convert␣I␣and␣N␣to␣Nelson-Aalen:␣",round(Sys.time()-start_time,digits=2),"
       ↪ ␣seconds."))
    start_time <- Sys.time()
  }
  # 4. Calculate Aalen-Johansen
  P_tmp <- NA_to_p(I_list_tmp,N_list_tmp,NelsonAalen_list_tmp, 3)
  if (debug) {
    print(paste0("Convert␣Nelson-Aalen␣to␣transition␣probs:␣",round(Sys.time()-start_time,
       ↪ digits=2),"␣seconds."))
    start_time <- Sys.time()
  }
  # 5. Calculate probabilties
  p_con_tmp <- P_conditioned(NelsonAalen_list_tmp,s,j,num_states)
  if (debug) {
    print(paste0("Convert␣transition␣probs␣to␣conditioned␣probs:␣",round(Sys.time()-start_
       ↪ time,digits=2),"␣seconds."))
    start_time <- Sys.time()
  }
  return(list(I = I_list_tmp$I, I_left_limit = I_list_tmp$I_left_limit,
              N = N_list_tmp$N,delta_N = N_list_tmp$delta_N,
              NelsonAalen = NelsonAalen_list_tmp$NelsonAalen, delta_NelsonAalen= NelsonAalen_
                 ↪ list_tmp$delta_NelsonAalen,
              P = P_tmp, p_con = p_con_tmp))
}
total <- Estimate(paths,3)
total <- Estimate(paths,3,s=1,j=1,as_if = TRUE)
```

A.2.8 Runtime

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

A.3 Plotting probabilities

```
plot_function1 <- function(paths,num_states,s= 0, j = 1, debug = TRUE) {
  #Markov estimate
  total <- Estimate(paths,num_states,s= s, j = j, as_if = FALSE, debug = debug)
  #As-If-Markov estimate
  total2 <- Estimate(paths,num_states,s= s, j = j, as_if = TRUE, debug = debug)
  plotdf <- total$p_con
  colnames(plotdf)[2:4] <- paste0("j=",1:3,",␣Markov")
  plotdf <- plotdf%>% reshape2::melt(., id = "Time")
  plotdf2 <- total2$p_con
  colnames(plotdf2)[2:4] <- paste0("j=",1:3,",␣As-If")
  plotdf2 <- plotdf2%>% reshape2::melt(., id = "Time")
  times <- s+0:1000*(10-s)/1000
  plotdf3 <- data.frame(Time = times,
                      matrix(unlist(lapply(times, function(t) ((1:num_states == j)*1)%*
                          ↪ %expm::expm(2*M*(log(1+0.5*t)-log(1+0.5*s))))),ncol=3,
                          ↪ byrow=TRUE))
  colnames(plotdf3)[2:4] <- paste0("j=",1:3,",␣True")
  plotdf3 <- plotdf3 %>% reshape2::melt(., id = "Time")
  asif_n <- sum(total2$I[1,])-total2$I[1,1]
  markov_n <- sum(total$I[1,])-total$I[1,1]
  ggplot() +
    geom_step(data = plotdf ,mapping = aes(x=Time, y = value,col = variable)) +
    geom_step(data = plotdf2,mapping = aes(x=Time, y = value,col = variable),linetype =
        ↪  "dashed") +
    geom_line(data = plotdf3,mapping = aes(x=Time, y = value,col = variable),linetype =
        ↪  "dotted",size=1) +
    theme_bw() +
    labs(title = TeX(paste0("Occupation␣probabilities␣under␣assumption␣$Z_",s,"=",j,"$"
        ↪ )),
         y = TeX(paste0("$P(Z_t=j|Z_",s,"=",j,")$")),
         subtitle = paste0("As-if␣estimate␣based␣on␣",asif_n,"␣observations,\nMarkov␣
             ↪ based␣on␣",markov_n,"␣observations")) +
    theme(legend.title = element_blank(),
          plot.title = element_text(face = "bold"),
          plot.subtitle = element_text(face = "italic")) +
    scale_color_manual(values=c("#FA8072", "#FF0000","#8B0000",
                                "#7B68EE", "#1E90FF","#00008B",
                                "#3CB371", "#32CD32","#006400"))
}
plot1 <- plot_function1(paths,3,s= 0, j = 1)
plot2 <- plot_function1(paths,3,s= 1, j = 1)
plot3 <- plot_function1(paths,3,s= 3, j = 1)
plot4 <- plot_function1(paths,3,s= 6, j = 1)
ggarrange(plotlist = list(plot1,plot2,plot3,plot4),ncol = 2,nrow=2)
```