Carl von Ossietzky
Universität
Oldenburg

Non-Markov modeling in life insurance

Von der Fakultät V für Mathematik und Naturwissenschaften
der Carl von Ossietzky Universität Oldenburg
zur Erlangung des Grades und Titels eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

angenommene Dissertation

von

Herrn Theis Bathke
geboren am 31.07.1998 in Oldenburg.

Betreuender Gutachter: Prof. Dr. Marcus C. Christiansen
Zweiter Gutachter: Prof. Dr. Mogens Bladt
Dritter Gutachter: Prof. Dr. Peter Ruckdeschel
Tag der Disputation: 28.01.2025

Acknowledgements

I would like to begin this acknowledgment by expressing my deepest gratitude to my main supervisor, Marcus Christiansen. From the very start of this journey, you have been a source of support, encouragement, and guidance. Your belief in me and this project has been truly invaluable. I am particularly grateful for the funding you provided, which not only enabled me to carry out my research at the Carl von Ossietzky Universität Oldenburg but also allowed me to attend numerous great conferences. These experiences have been impactful in my academic and personal development. Beyond the academic and financial support, I also appreciate your supervision throughout the entire project, as well as your valuable advice regarding my future development. Your constructive feedback, patience, and insights have played a crucial role in shaping this work, and I am very happy to have had you as my supervisor.

I would also like to extend my sincere thanks to Peter Ruckdeschel for the many instructive discussions, particularly in relation to statistical theory and the practical implementation of my research. Your willingness to share your expertise is a great inspiration to me. Additionally, I am grateful to Mogens Bladt for kindly agreeing to be the second examiner of this thesis. Your time and effort in reviewing my work and visiting us in Oldenburg are appreciated.

In addition to my supervisors, I want to express my deepest gratitude to my co-author, Christian Furrer. Our collaboration is one of my greatest memories during this time, and I sincerely appreciate your invitation to Copenhagen, where I had an incredibly productive time working with you. Beyond our professional relationship, your personal counsel during these times has meant a great deal to me. You have been an immense support, and upon reflection, I can say that you have grown to be somewhat of a mentor to me.

Furthermore, I would like to thank Hauke, Jannes and Miguel as well as my other fellow PhD students and friends for making this time much more enjoyable. The time we spent together, discussing mathematics and all the other great topics life has to offer were truly important to me.

Last but certainly not least, to Heike, Lars, Thore, and the rest of my wonderful family: Thank you for your unconditional support and for all the times you taught me how to keep my head up during challenging times. Finally, I want to thank my partner Marit for her unwavering support and for always helping me see the light in the dark.

Zusammenfassung

(German)

Diese Dissertation behandelt Fortschritte in non-Markov Modellen mit einem Schwerpunkt auf zweidimensionalen Erweiterungen von bedingten Übergangsraten und deren Anwendungen in der Lebensversicherung. Das Konzept der bedingten Übergangsraten wurde ursprünglich entwickelt, um die Schätzungen der Reserven von Versicherungspolicen zu verbessern, indem auf die Markov-Annahme verzichtet wird, und ermöglicht dadurch die Berechnung bedingter Erwartungswerte. Die neu entwickelten zweidimensionalen bedingten Übergangsraten werden zur Berechnung von bedingten Momenten zweiter Ordnung verwendet, die für die Risikoanalyse in der Versicherungsmathematik von entscheidender Bedeutung sind. Um die statistische Grundlage zu schaffen, wird ein Schätzer für zweidimensionale bedingte Übergangsraten eingeführt, der die Theorie bivariater Survivalfunktionen mit der Theorie der Landmark-Nelson-Aalen- und Landmark-Aalen-Johansen-Schätzung kombiniert. Diese zweidimensionalen bedingten Übergangsraten sind nicht nur nützlich für die Berechnung von Momenten zweiter Ordnung, sondern auch zur Erfassung der intertemporalen Abhängigkeit von pfadabhängigen Zahlungsströmen, wie sie beispielsweise durch optionale Vertragsänderungen entstehen.

Des Weiteren wird ein alternativer Ansatz zur Schätzung der Reserve für Policen mit pfadabhängigen Zahlungsströmen untersucht. Dieses Verfahren verwendet skalierte Wahrscheinlichkeiten, die die stochastische Natur der Zahlungsfunktionen erfassen, indem diese als Skalierungsfaktor in die Aalen-Johansen-Schätzung integriert werden. Diese Schätzmethode eignet sich auch für die Analyse allgemeinerer stochastischer Zahlungen, wie sie bei stochastischen Zinssätzen auftreten.

Abstract

(English)

This thesis investigates advances in non-Markov models, focusing on two-dimensional extensions of conditional forward transition rates and their applications in life insurance. The concept of conditional forward transition rates, originally developed to improve reserve estimation by eliminating the need for the Markov assumption, enables the calculation of conditional expected values. The newly developed two-dimensional conditional transition rates are used to calculate second-order conditional moments, which are crucial for risk analysis in actuarial applications. To further support this framework, an estimator for two-dimensional conditional transition rates is introduced, combining the theory of bivariate survival functions with the theory of landmark Nelson-Aalen and landmark Aalen-Johansen estimation. These two-dimensional conditional transition rates prove useful not only for calculating second-order moments but also for capturing the intertemporal dependency of path-dependent cash flows, such as those resulting from incidental policyholder behavior upon contract modifications.

Furthermore, an alternative approach to reserve estimation for policies with path-dependent cash flows is examined. This method uses scaled transition rates and scaled probabilities, which capture the stochastic nature of the payment functions by incorporating it as a scaling factor into the Aalen-Johansen estimation. This estimation procedure is also beneficial for analyzing more general stochastic payments, such as those involving stochastic interest rates.

Contents

List of Papers

Besides Chapter 1, which has been prepared as the introduction to this thesis, the three remaining chapters each present a self-contained and published or submitted manuscript. This is a list of the corresponding manuscripts and a disclosure illustrating the contributions of the author of this thesis:

Chapter 2: Bathke, T., Christiansen, M.C. Two-dimensional forward and backward transition rates. Eur. Actuar. J. 14, 411–436 (2024). https://doi.org/10.1007/s13385-023-00363-3
Marcus Christiansen, as Theis Bathke's doctoral supervisor, wrote the introduction and conclusion of the manuscript and supervised Theis Bathke's work on the remainder of the project.

Chapter 3: Bathke, T., On the estimation of bivariate conditional transition rates. Scandinavian Actuarial Journal, (2024). https://doi.org/10.1080/03461238.2024.2420067

Chapter 4: Bathke, T., and Furrer, C., Non-parametric estimators of scaled cash flows. arXiv preprint arXiv:2408.13176 (2024). https://doi.org/10.48550/arXiv.2408.13176
Both authors contributed equally to the conception of the project, the derivation and interpretation of the main results, and the writing of the manuscript. Theis Bathke did the relevant programming, but under Christian Furrer's supervision.

Minor modifications in spelling and formatting have been made in the chapters compared to the corresponding manuscripts.

List of Figures

List of Implementations

The thesis contains numerical studies in Chapter 3 and Chapter 4. The implementations and simulated data for these studies can be found under the following links:

Chapter 3 (Estimation of bivariate transition rates):
`https://github.com/Theis-Bathke/Estimation-of-bivariate-transition-rates`

Chapter 4 (Estimation of scaled cash flows):
`https://github.com/Theis-Bathke/Non-parametrix-estimators-of-scaled-cash-flows`

Chapter 1

Introduction

This thesis concerns multi-state modeling in actuarial mathematics, with a particular focus on applications in life insurance. Multi-state models operate on a finite state space representing the various states an insured individual can occupy. In practice, these models are frequently used alongside the Markov assumption, which broadly implies that future developments depend only on the current state and not the past states. However, recent actuarial and statistical literature has raised questions about the validity of the Markov assumption in certain contexts. This thesis primarily investigates challenges associated with non-Markov modeling in actuarial mathematics. Non-Markov models include all multi-state models, regardless of whether the Markov assumption holds.

The first chapter serves to motivate the research questions for the subsequent three chapters by illustrating the current state of research on different actuarial topics. See the following table for a short overview of the theoretical content in this introduction:

	Markov modeling	Non-Markov modeling
Reserve calculation and estimation	Subsections 1.1.1–1.1.2	Section 1.2
Calculation of higher moments	Subsection 1.1.3	Subsections 1.3.1–1.3.2
Modeling of stochastic payments	Subsection 1.1.4	Subsections 1.3.1–1.3.3

Section 1.1 introduces and discusses various known aspects of Markov modeling for actuarial applications, beginning with the calculation of conditional expected values of payment streams with deterministic payment functions and estimation via Nelson-Aalen and Aalen-Johansen estimators. Following this, the calculation of higher moments of these cash flows is illustrated. Lastly, generalizations to stochastic payment functions in Markov models are discussed. The section is not intended to provide a comprehensive overview of the extensive theory of Markov modeling, but rather to offer a concise introduction to the key concepts that are essential for the subsequent work.

Section 1.2 illustrates the challenges that arise when the Markov model is used without verifying the validity of the assumptions, along with the resulting systematic model risk. Following this, the section introduces the recently proposed non-Markov models, described in the literature as all multi-state models on finite state spaces. An overview of current non-Markov research is provided, covering topics such as the calculation of prospective reserves and estimation techniques such as landmark Nelson-Aalen and Aalen-Johansen estimations. The section concludes by identifying gaps in current non-Markov literature and establishing the research questions addressed in this thesis, which focus on two main areas: The calculation of higher moments of discounted accumulated future cash flows in non-Markov models and the modeling of cash flows with stochastic payment functions in non-Markov models.

Building on this, Section 1.3 continues with an overview of the main contributions of the subsequent chapters, emphasizing the construction and applications of two-dimensional conditional transition rates and scaled transition rates. It highlights the relevance of these developments to the research questions, while keeping the technical aspects to a minimum.

Lastly, Section 1.4 provides a conclusion of the thesis and an outlook on possible future research questions.

1.1 Markov modeling

Historically, future cash flows of insurance contracts have been modeled using Markov chain models, see the innovative work of Hoem [Hoe69]. The statistical basis of non-parametric estimators for these models, including the Nelson-Aalen and Aalen-Johansen estimators, was first introduced by Aalen [Aal78] and Aalen and Johansen [AJ78]. This approach was further explored in Norberg [Nor91] and Milbrodt and Stracke [MS97], while its continuing significance can still be seen in Buchardt and Møller [BM15]. Today, different models build on the idea of the multi-state Markov models, including the semi-Markov model; see the work of Hoem [Hoe72], Janssen and de Dominicis [JD84], Christiansen [Chr12], and Buchardt, Møller, and Schmidt [BMS15], as well as other extensions, see the work of Buchardt [Buc17] and Ahmad, Bladt and Furrer [ABF23].

This section is mainly an adaptation of classical Markov literature, as well as the statistical literature on Nelson-Aalen and Aalen-Johansen estimators. It gives an overview of the theory and introduces the basis for current research in multi-state life insurance modeling.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space with sigma algebras $(\mathcal{F}_t)_{t\geq 0}$ representing the information at the corresponding times. These do not need to form a filtration, since the information can be non-monotone. In actuarial mathematics and especially life insurance mathematics, the state of the insured is often governed by a non-explosive and càdlàg jump process $Z = (Z(t))_{t\geq 0}$ on a finite state space $\mathcal{Z}$. The elements, which are called states, represent

different, typically biometric and behavioral states of the insured. In the remainder of this chapter, the indices z, i, j, k, l will all be states in the state space $\mathcal{Z}$ as long as they are not defined otherwise. In addition to the state process, a multivariate counting process $(N_{ij}(t))_{t\geq 0}$ is introduced to count transitions between the different states in $\mathcal{Z}$, defined as

$$N_{ij}(t) := \#\{s \in (0,t] : Z(s-) = i, Z(s) = j\}, \quad t \geq 0, i \neq j.$$

For the remainder of this section, the state process Z is assumed to be a Markov chain. The goal of this modeling is the calculation of conditional expected values of cash flows of insurance policies. In Markov models, these calculations are usually based on transition probabilities $(P(s,t))_{t\geq s\geq 0}$ and transition rates $(\lambda(t))_{t\geq 0}$, whose components are defined as

$$\begin{aligned} P_{i,k}(s,t) &= \mathbb{P}(Z(t) = k | Z(s) = i), \\ \lambda_{ij}(t) &= \lim_{h \searrow 0} \frac{\mathbb{P}(Z(t+h) = j | Z(t) = i)}{h}, \quad 0 \leq s \leq t, i \neq j, \end{aligned}$$

whenever the limit exists. Define cumulative transition rates $\tilde{\Lambda}$ as the right-continuous process with

$$\tilde{\Lambda}_{jk}(t) := \int_{(0,t]} \frac{1}{\mathbb{P}(Z(u-) = j)} \mathrm{d}\mathbb{E}[N_{jk}(u)], \quad t \geq 0, k \neq j,$$

where $\frac{0}{0} := 0$, which is used for the rest of this thesis. For the rest of this section, $\tilde{\Lambda}$ is assumed to be absolutely continuous. One can show that

$$\tilde{\Lambda}_{jk}(t) = \int_{(0,t]} \lambda_{jk}(u) \mathrm{d}u, \quad t \geq 0, k \neq j,$$

which shows the reason for the name cumulative transition rates. If the transition rates exist, Kolmogorov's forward differential equations can be used to calculate the transition probabilities from the transition rates by solving the differential equations

$$\frac{\partial}{\partial t} P_{i,j}(s,t) = \sum_{\substack{k \in \mathcal{Z} \\ k \neq j}} P_{i,k}(s,t) \lambda_{kj}(t) - \sum_{\substack{k \in \mathcal{Z} \\ k \neq j}} P_{i,j}(s,t) \lambda_{jk}(t), \quad t \geq s,$$

where $P_{i,j}(s,s) = \mathbb{1}_{\{i=j\}}$. The solution to this differential equation can be computed with the help of the product integral and the cumulative transition rates

$$(P_{i,k}(s,t))_{k\in\mathcal{Z}} = (\mathbb{1}_{\{i=k\}})_{k\in\mathcal{Z}} \Big[\prod_s^t \Big(\mathrm{Id} + \tilde{\Lambda}(\mathrm{d}u) \Big) \Big], \quad t \geq s, \tag{1.1.1}$$

where the families of functions are interpreted as column vectors and matrices, and the diagonal of $\tilde{\Lambda}$ reads $\tilde{\Lambda}_{jj} = -\sum_{k:k\neq j} \tilde{\Lambda}_{jk}$.

1.1.1 Reserve calculation with canonical cash flows

In practice, an insurance policy can be characterized by a cash flow, including premium and benefit payments. These cash flows are usually stochastic because they depend on the development of the insured, meaning the state process Z. For classic insurance contracts, the cash flow $(B(t))_{t\geq 0}$, which refers to all payments that are part of the insurance policy up until the time point t, can be expressed by

$$B(t) := b_0 + \sum_{j\in\mathcal{Z}} \int_{(0,t]} \mathbb{1}_{\{Z(u-)=j\}} \mathrm{d}B_j(u) + \sum_{\substack{j,k\in\mathcal{Z}\\ j\neq k}} \int_{(0,t]} b_{jk}(u)\mathrm{d}N_{jk}(u). \tag{1.1.2}$$

Here, suitably regular, bounded and deterministic cumulative payment functions $(B_j)_{j\in\mathcal{Z}}$ and transition payment functions $(b_{jk})_{j,k:j\neq k}$ are considered, along with a deterministic lump sum payment b_0. When the cash flow of an insurance contract follows the form given in Equation (1.1.2), it is said to have a canonical cash flow representation and deterministic payment functions. The time value of payments is modeled with a deterministic short rate r, such that the value of a deterministic savings account is

$$\kappa(t) := e^{\int_0^t r(u)\mathrm{d}u},$$

which is assumed to be bounded away from zero on compacts. Non-deterministic short rates are not considered in this subsection, but will be discussed in Subsection 1.1.4.

Fixing a time point $s \geq 0$ as the current time point, the discounted accumulated future cash flow is given by

$$Y(s) := \int_{(s,\infty)} \frac{\kappa(s)}{\kappa(t)} \mathrm{d}B(t).$$

The discounted accumulated future payments $Y(s)$ are generally not $\mathcal{F}_s$ measurable. Therefore, for reserving and premium calculation, the calculation of the prospective reserve

$$V(s) := \mathbb{E}[Y(s)|\mathcal{F}_s]$$

is appropriate. Now assume that

$$\sigma(Z(s)) \subset \mathcal{F}_s \subset \sigma((Z(u))_{u\leq s}),$$

then the prospective reserve simplifies, due to the Markov assumption, to $V(s) = \mathbb{E}[Y(s)|\sigma(Z(s))]$. Thus, it is sufficient to calculate the state-wise prospective reserve $V_i(s) := \mathbb{E}[Y(s)|Z(s) = i]$ for all $i \in \mathcal{Z}$ with $\mathbb{P}(Z(s) = i) > 0$. Using transition probabilities and transition rates, the state-wise prospective reserve of a cash flow with a canonical cash flow representation follows the well-known formula

$$V_i(s) = \sum_{j\in\mathcal{Z}} \int_{(s,\infty)} \frac{\kappa(s)}{\kappa(t)} P_{i,j}(s,t)\mathrm{d}B_j(t) + \sum_{\substack{j,k\in\mathcal{Z}\\ j\neq k}} \int_{(s,\infty)} \frac{\kappa(s)}{\kappa(t)} b_{jk}(t)P_{i,j}(s,t)\lambda_{jk}(t)\mathrm{d}t, \tag{1.1.3}$$

where it is assumed that the transition rates exist everywhere. The use of the transition rates $(\lambda_{jk}(t))_{t\geq 0}$, which are independent of the condition at time s, is made possible by the Markov assumption.

1.1.2 Nelson-Aalen and Aalen-Johansen estimation

The central mathematical objects in Section 1.1 are the transition probabilities and transition rates, which are normally estimated with Nelson-Aalen and Aalen-Johansen estimators, see the work of Aalen [Aal78] and Aalen and Johansen [AJ78]. This subsection motivates and illustrates the use of Nelson-Aalen and Aalen-Johansen estimation in Markov models.

Both transition probabilities and the expected number of jumps could be targeted through simple averages. This would be sufficient for the estimation of conditional expected values of discounted accumulated future cash flows. To formalize this, assume that n independent observations $(Z^{(j)})_{j=1,\dots,n}$ of the state process Z are observed. A natural estimator for transition probabilities is given by

$$P^{\star}_{z,i}(s,t) = \frac{\sum_{j=1}^{n} \mathbb{1}_{\{Z^{(j)}(s)=z, Z^{(j)}(t)=i\}}}{\sum_{j=1}^{n} \mathbb{1}_{\{Z^{(j)}(s)=z\}}}, \quad t \geq s.$$

This estimator, however, does not fully utilize all available data because it sub-samples at time s and it is biased, when censored data is involved. This poses a challenge, especially when working with survival data, which is commonly encountered in actuarial mathematics.

For a basic inclusion of censoring into the model, consider entirely random right-censoring times $R^{(j)}$. Suppose that n independent individuals $((Z^{(j)}(t))_{0\leq t\leq R^{(j)}}), \eta^{(j)} \wedge R^{(j)})_{j=1,\dots,n}$ are observed, where $\eta^{(j)}$ is the absorption time for individual j. A classic estimation procedure in this case is the use of the Nelson-Aalen estimator in combination with the Aalen-Johansen estimator. For the definition of these estimators, start with

$$I^{\star}_{j}(t) := \sum_{i=1}^{n} \mathbb{1}_{\{Z^{(j)}(t)=j\}} \mathbb{1}_{\{t\leq R^{(j)}\}},$$
$$N^{\star}_{jk}(t) := \sum_{i=1}^{n} N^{(j)}_{jk}(t \wedge R^{(j)}),$$

for $t \geq 0$ and $j \neq k$. This leads to the Nelson-Aalen estimator

$$\Lambda^{\star}_{jk}(t) := \int_{(0,t]} \frac{1}{I^{\star}_{j}(u-)} \mathrm{d}N^{\star}_{jk}(u), \quad t \geq 0, k \neq j,$$

which is a uniformly consistent estimate for the cumulative transition rates $\tilde{\Lambda}_{jk}(t)$. This estimator can then be plugged into Kolmogorov's forward equation, which is then solved

via the product integral, resulting in the Aalen-Johansen estimator:

$$(P^{\star}_{i,k}(s,t))_{k\in\mathcal{Z}} = (\mathbb{1}_{\{i=k\}})_{k\in\mathcal{Z}}\Big[\prod_{s}^{t}\Big(\mathrm{Id}+\Lambda^{\star}(\mathrm{d}u)\Big)\Big], \quad t\geq s,$$

where $\Lambda^{\star}_{jj} = -\sum_{k:k\neq j}\Lambda^{\star}_{jk}$, using the usual matrix interpretation. This combination of estimation techniques is one reason why transition rates are so prevalent in the actuarial literature.

Using Duhamel's equation, see Gill and Johansen [GJ90], it can be shown that the product integral is a continuous operator on the class of additive interval functions with respect to the variation norm on bounded intervals. As a result, the consistency of the Aalen-Johansen estimator follows immediately. Both the Nelson-Aalen and the Aalen-Johansen estimators can be plugged into Equation (1.1.3) to obtain a uniformly consistent estimator for the prospective reserve.

Moreover, both the Nelson-Aalen and the Aalen-Johansen estimators are asymptotically normal, which allows for a general assessment of the statistical risk associated with estimating prospective reserves within the Markov model framework.

1.1.3 Higher moments of canonical cash flows

This subsection illustrates, why higher moments of cash flows are useful for the actuarial application and introduces the theoretical basis for these calculations.

After calculating the prospective reserve of a cash flow with a canonical representation, a risk analysis of the cash flow should be considered. The realized losses or gains of the portfolio are stochastic and thus, do not necessarily match the calculated conditional expected value, i.e. the prospective reserve. This risk can best be understood by calculating variances or even higher moments of the discounted accumulated future payments. Hoem and Aalen [HO78] show how to calculate higher moments of cash flows in the Markov model. Their approach is based on the ability to calculate the conditional expected value of the product of two counting processes at different times by applying Markov techniques. They derive the formula

$$\begin{aligned}
&\mathbb{E}[(N_{ij}(t_1)-N_{ij}(s))(N_{kl}(t_2)-N_{kl}(s))|Z(s)=z] \\
&\qquad = \int_s^{t_1}\int_s^{t_2} \mathbb{1}_{\{x\leq y\}} P_{z,i}(s,x)\lambda_{ij}(x)P_{j,k}(x,y)\lambda_{kl}(y) \\
&\qquad\qquad + \mathbb{1}_{\{x>y\}} P_{z,k}(s,y)\lambda_{kl}(y)P_{l,i}(y,x)\lambda_{ij}(x)\mathrm{d}y\mathrm{d}x \\
&\qquad + \mathbb{1}_{\{(i,j)=(k,l)\}} \int_s^{t_1\wedge t_2} P_{z,i}(s,x)\lambda_{ij}(x)\mathrm{d}x
\end{aligned}$$

for $t_1, t_2 \geq s$ and $i \neq j, k \neq l$. Combining this with the equation

$$\mathbb{1}_{\{Z(t)=i\}} = \mathbb{1}_{\{Z(0)=i\}} + \sum_{j \neq i} \big(N_{ji}(t) - N_{ij}(t)\big), \quad t \geq 0, \tag{1.1.4}$$

the second moment of a discounted accumulated future cash flow with a canonical cash flow representation can be calculated using only the transition probabilities and transition rates. This implies that the estimation of transition probabilities and rates with the Nelson-Aalen and Aalen-Johansen estimators is sufficient to calculate these variances. This is the basis for the rather simple analysis of the risk of fluctuations in Markov models.

This result can be generalized: In a Markov model, any moment of a discounted accumulated future cash flow with a canonical cash flow representation, can be calculated solely with transition probabilities and transition rates. This is a significant property of the Markov model but also highlights the loss of intertemporal dependency structure inherent in the use of the Markov assumption.

1.1.4 Modeling of stochastic payment functions

This subsection introduces extensions to canonical cash flows by incorporating stochastic payment functions. The deterministic payment functions and discounting discussed previously are limited in their ability to model complex actuarial payment schemes. To address this, payment functions that allow dependency on the evolving state of the policyholder are included. This subsection aims to show challenges and corresponding solutions in current literature related to the modeling of stochastic payment functions in Markov models.

In this thesis, cash flows with stochastic payment functions are modeled by introducing a stochastic scaling factor $H(t)_{t\geq 0}$. The corresponding cash flow $B(t)$ reads

$$B(t) := b_0 + \sum_{j \in \mathcal{Z}} \int_{(0,t]} \mathbb{1}_{\{Z(u-)=j\}} H(u) \mathrm{d}B_j(u) + \sum_{\substack{j,k \in \mathcal{Z} \\ j \neq k}} \int_{(0,t]} H(u) b_{jk}(u) \mathrm{d}N_{jk}(u),$$

for $t \geq 0$. Throughout, H is assumed to be a right-continuous stochastic process of finite variation. Different forms of stochastic processes H can represent different types of applications, as illustrated below:

(i) **Stochastic interest rates:**
The process H can model stochastic interest rates as follows:

$$H(t) = \exp\left\{ \int_0^t \delta(u) \mathrm{d}u \right\}, \quad t \geq 0$$

for a stochastic interest rate process δ. This would mean that $B(t)$ already models the discounted accumulated cash flow and therefore, $Y(s) = \int_{(s,\infty)} B(\mathrm{d}t)$ needs to be used for the calculation of the prospective reserves.

(ii) **Duration-dependent payments:**
Payments dependent on the duration since the last jump are used to model different practices, for example deferred payments. For that, define the right-continuous stochastic processes representing the time duration in the current state:

$$U(t) = t - \sup\{u \in (0,t] : Z(u) \neq Z(t)\}, \quad t \geq 0,$$

where the supremum over the empty set is defined to be zero. A deferred payment could then be defined by

$$H(t) = \mathbb{1}_{\{U(t) \geq t_1\}}, \quad t \geq 0,$$

where $t_1 \geq 0$ is defined according to the waiting period. Duration-dependent payments, defined using this stochastic scaling factor, include nearly all standard duration-dependent payments, with lump sum payments at stochastic times being one exception.

(iii) **Policyholder behavior and free policy option:**
In the case of the free policy option the policyholder is able to forgo all future premium payments, while reducing future benefits by a scaling factor ρ. This scaling factor is defined by partitioning the state space $\mathcal{Z}$ into the subsets: $\mathcal{Z}_0 \subset \mathcal{Z}$ (pre-option states) and $\mathcal{Z}_1 \subset \mathcal{Z}$ (post-option states). The scaling factor depends on the hitting time τ of $\mathcal{Z}_1$ as well as the states before and after the transition. The process H is then given by:

$$H(t) = \rho(\tau, Z(\tau)-, Z(\tau))^{\mathbb{1}_{\{\tau \leq t\}}}, \quad t \geq 0.$$

There is a rich theory on the modeling of stochastic payment functions in Markov models. Below are some modeling approaches for the previously discussed stochastic payment functions:

(i) **Stochastic interest rates:**
Stochastic interest rates are often modeled through forward interest rates, represented by a real-valued and $\mathcal{F}_t-$measurable stochastic process $(\rho(t,u))_{t,u\geq 0}$, which satisfies

$$\mathbb{E}_Q\left[\exp\left\{\int_t^T \delta(u)\mathrm{d}u\right\}\middle|\mathcal{F}_t\right] = \exp\left\{\int_t^T \rho(t,u)\mathrm{d}u\right\},$$

for an equivalent market measure Q and the stochastic interest rate δ. This forward interest rate concept known from the financial literature generalizes well to life insurance mathematics, when independence of interest rate and transition rates is assumed. When stochastic mortality is included, forward rates are employed, see Christiansen and Niemeyer [CN15] and Buchardt, Furrer and Steffensen [BFS19]. Building on the idea of forward interest and forward transition rates, even dependency via affine processes has been examined, see Buchardt [Buc14]. Various definitions of forward rates have been suggested, yet a unified definition answering all the questions posed by Norberg [Nor10] has not been found, see Subsection 5.4 in Buchardt, Furrer, and Steffensen [BFS19].

(ii) **Duration-dependent payments:**
Duration-dependent payments are traditionally modeled via semi-Markov modeling, see for example Møller [Møl93], Helwich [Hel08], and Christiansen [Chr12]. This is due to the fact, that the semi-Markov models assume that the tuple $(Z(t), U(t))_{t\geq 0}$ is Markov. This then allows the inclusion of dependence on duration into the calculation of transition probabilities through the simultaneous modeling of duration and time. As a result, calculating prospective reserves in semi-Markov models requires solving two-dimensional differential equations and evaluating two-dimensional integrals, which is numerically challenging. When the state process is Markovian, a semi-Markov approach can be used to allow payment functions to depend on duration.

(iii) **Policyholder behavior and free policy option:**
Modeling policyholder behavior, such as the free policy option, has been explored using multiple approaches:

- Henriksen et al. [Hen+14] use two state processes, which are assumed to be combined Markov. They show that the modeling of policyholder options is possible, though challenges remain. For instance, in disability models, they show that reserve calculations remain challenging, since required transition probabilities cannot be calculated explicitly when dependency between both state processes is assumed and reactivation from disability is permitted.
- Buchardt, Møller, and Schmidt [BMS15] use semi-Markov modeling and define an additional waiting time W, which is used to define the scaling factor. This dependency on W is managed by adjusting the differential equations for the prospective reserve.
- Furrer [Fur22] addresses policyholder behavior in multi-state models by using a change of measure technique to calculate prospective reserves for cash flows with scaled payment functions. This method maintains the simplicity of reserve calculations but introduces an artificial state process. Finally, the general approach is applied to Markov models, giving a computational scheme for reserve calculations.

1.2 Non-Markov modeling

Recently, the idea of non-Markov models has gained considerable attention in biostatistics and actuarial sciences. Notable works include those of Christiansen [Chr21] and Christiansen and Furrer [CF22] on the actuarial side, as well as Putter and Spitoni [PS18] and Bladt and Furrer [BF24b] from the statistical perspective. This section gives an introduction into the current research on non-Markov models but mainly focuses on the actuarial side. It begins by exploring the use of the Markov assumption in multi-state models, particularly in cases where this assumption may not be justified. This is followed by an overview of the current non-Markov literature and an illustration of the research questions of this thesis.

The previously introduced Markov model is widely used in actuarial practice. One reason for this is the possibility of simple and effective risk analysis provided in the model. Both the statistical risk and the portfolio risk can be estimated, and they decrease as the amount of data or the number of insured individuals increases. However, this is not the case for the so-called systematic model risk, which arises due to the use of the Markov assumption. This risk is characterized by a potential bias in the estimation of the prospective reserve when the data does not conform to the Markov assumption.

Examples where the Markov assumption may not be justified include disability or long term care insurance contracts, where degenerative diseases could affect the insured. Other reasons for the potential failure of the Markov assumption include unobserved heterogeneity, which can appear, when the portfolio is impacted by exogenous effects, as discussed by Guibert and Planchet [GP18], or when artificial aggregation of states is used, see Ahmad, Bladt and Furrer [ABF23].

The problem with naive use of the Markov assumption is that Equation (1.1.3) only holds, when the Markov assumption is actually justified. Bear in mind that $s \geq 0$ is a fixed time point. Now take $i \in \mathcal{Z}$ such that $\mathbb{P}(Z(s) = i) > 0$ and define the cumulative conditional transition rates Λ_i as the right-continuous processes with

$$\Lambda_{i,jk}(s,t) := \int_{(s,t]} \frac{1}{P_{i,j}(s,u-)} \mathrm{d}\mathbb{E}[N_{jk}(u)|Z(s) = i], \quad t \geq s,\ j \neq k.$$

Then Equation (1.1.3) correctly reads

$$V_i(s) = \sum_{j \in \mathcal{Z}} \int_{(s,\infty)} \frac{\kappa(s)}{\kappa(t)} P_{i,j}(s,t) \mathrm{d}B_j(t) + \sum_{\substack{j,k \in \mathcal{Z} \\ j \neq k}} \int_{(s,\infty)} \frac{\kappa(s)}{\kappa(t)} b_{jk}(t) P_{i,j}(s,t) \Lambda_{i,jk}(s, \mathrm{d}t), \tag{1.2.1}$$

without the Markov assumption, when transition probabilities are defined just as before. Assuming the Markov assumption holds, the equation $\Lambda_{i,jk}(s, \mathrm{d}t) = \mathrm{d}\tilde{\Lambda}_{jk}(t)$ is valid for $t \geq s$ and $j \neq k$, see Milbrodt and Stracke [MS97]. This draws the connection between Equation (1.1.3) and Equation (1.2.1). However, if a model uses the Markov property when it does not hold in reality, the use of this equality introduces a systematic modeling error.

This issue becomes apparent when analyzing the calculation of the transition probabilities. If one uses $\tilde{\Lambda}$ instead of Λ_i and Equation (1.1.1) for $(P_{i,j})_{j \in \mathcal{Z}}$, the resulting probabilities are incorrect, when the Markov assumption is not satisfied. The outcome is a function $(\tilde{P}_{i,j})_{j \in \mathcal{Z}}$, which solves Kolmogorov's forward equation for $(\mathbb{P}(Z(t) = j))_{j \in \mathcal{Z}, t \geq s}$ but with wrong boundary condition $(\tilde{P}_{i,j}(s,s))_{j \in \mathcal{Z}} = (\mathbb{1}_{\{i=j\}})_{j \in \mathcal{Z}}$, instead of the correct boundary condition $(\tilde{P}_{i,j}(s,s))_{j \in \mathcal{Z}} = (\mathbb{P}(Z(s) = j))_{j \in \mathcal{Z}}$. This comes from the fact that:

$$\mathbb{P}(Z(t) = j) = \mathbb{P}(Z(s) = j) + \sum_{k \neq j} \int_s^t \mathbb{P}(Z(u) = k) \mathrm{d}\tilde{\Lambda}_{kj}(u)$$

$$- \sum_{k \neq j} \int_s^t \mathbb{P}(Z(u) = k) \mathrm{d}\tilde{\Lambda}_{jk}(u),$$

which immediately follows from Equation (1.1.4) for $t \geq s$. In practice, this $(\tilde{P}_{i,j})_{j \in \mathcal{Z}}$ would be used in Equation (1.1.3), leading to inaccurate reserving or premium calculations. This is especially concerning because currently, there is no theoretical understanding of the differences between the correct transition probabilities and the resulting $(\tilde{P}_{i,j})_{j \in \mathcal{Z}}$.

To better understand these systematic risks, a modeling framework that weakens the restrictive Markov assumption is of essence. There are several extensions of the Markov model, such as the previously mentioned semi-Markov models or the doubly stochastic Markov models from Buchardt [Buc17]. However, these still maintain some form of Markov assumption, which may continue to pose a basis for systematic model risk. For this reason, the so-called non-Markov model is introduced.

The non-Markov models use the same multi-state framework as the Markov models, but also include all the models that do not satisfy the Markov assumption. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space with sigma algebras $(\mathcal{F}_t)_{t \geq 0}$ representing the information at the corresponding times. In a non-Markov model, different information structures containing external information are considered. The state process $Z = (Z(t))_{t \geq 0}$ takes values in a finite state space $\mathcal{Z}$, while N is the corresponding counting process, which is assumed to have finite second order moments, and $s \geq 0$ is the current time point. Without the Markov assumption, the whole information needs to be considered. For the rest of this section, it is assumed that $\sigma(Z(s)) \subset \mathcal{F}_s$. Define the right-continuous processes

$$\begin{aligned} P_i(s,t) &:= \mathbb{E}[\mathbb{1}_{\{Z(t)=i\}} | \mathcal{F}_s], \\ Q_{ij}(s,t) &:= \mathbb{E}[N_{ij}(t) | \mathcal{F}_s], \end{aligned}$$

for $t \geq s$ and $i \neq j$. The processes $(P_i)_{i \in \mathcal{Z}}$ are called conditional state occupation probabilities. On this basis, cumulative conditional transition rates are defined for $i \neq j$ and $t \geq s$ as follows:

$$\Lambda_{ij}(s,t) := \int_{(s,t]} \frac{1}{P_i(s, u-)} Q_{ij}(s, \mathrm{d}u).$$

$\Lambda_{ij}(s, \mathrm{d}t)$ is called conditional transition rate for the rest of this chapter. The link between the cumulative conditional transition rates in the non-Markov model and the cumulative transition rates in the Markov model is that they coincide if the Markov assumption holds, transition rates exist, and the information models are identical. Note that it is assumed that $\Lambda_{ij}(s,t) < \infty$ for $i \neq j,\ t \geq s$.

Similar to the Markov model, a relationship between the conditional state occupation probabilities and the conditional transition rates can be derived. The differential equations read:

$$P_j(s, \mathrm{d}t) = \sum_{k : k \neq j} P_k(s, t-) \Lambda_{kj}(s, \mathrm{d}t) - \sum_{k : k \neq j} P_j(s, t-) \Lambda_{jk}(s, \mathrm{d}t), \quad t \geq s, \tag{1.2.2}$$

with the boundary condition $P_j(s,s) = \mathbb{1}_{\{Z(s)=j\}}$. This equation serves as a substitution to Kolmogorov's forward equation from the Markov model.

Because of the more general information model, the definition of the prospective reserve in the non-Markov model needs to be adjusted. The modified prospective reserve is $V(s) := \mathbb{E}[Y(s)|\mathcal{F}_s]$, where again $s \geq 0$ is fixed as the current time point. Similar to the formulas for the calculation of prospective reserves in the Markov model, the conditional state occupation probabilities and transition rates can be used to calculate the prospective reserve of insurance contracts with a canonical cash flow:

$$V(s) = \sum_{j \in \mathcal{Z}} \int_{(s,\infty)} \frac{\kappa(s)}{\kappa(t)} P_j(s,t) \mathrm{d}B_j(t) + \sum_{\substack{j,k \in \mathcal{Z} \\ j \neq k}} \int_{(s,\infty)} \frac{\kappa(s)}{\kappa(t)} b_{jk}(t) P_j(s,t) \Lambda_{jk}(s, \mathrm{d}t). \tag{1.2.3}$$

The equation is similar to Equation (1.1.3), but with the key difference that the conditional transition rates now dependent on the full information $\mathcal{F}_s$.

1.2.1 Landmark Nelson-Aalen and Aalen-Johansen estimation

This subsection introduces the estimation techniques used for the estimation of conditional transition rates and conditional state occupation probabilities in non-Markov models.

The non-parametric estimation of conditional state occupation probabilities and conditional transition rates is generally more challenging than the Nelson-Aalen and Aalen-Johansen estimation in the Markov model. This difficulty arises from the general information model introduced in the non-Markov model. The model with the simplest information structure and the one that is most akin to standard practice in actuarial mathematics is the as-if Markov model. The model is based on the information model, $\mathcal{F}_s = \sigma(Z(s))$, constituting the term as-if Markov, because the condition used for the prospective reserves in the non-Markov model is identical to that in the Markov model. From a theoretical perspective, this model already finds practical application, though often with biased estimators when Markov estimators are applied when the Markov assumption is not justified.

Typically, the prospective reserve of a life insurance contract is defined as the expected value of the discounted accumulated future cash flow, conditioned on the whole past of the insured. In practice, however, this is difficult to achieve for legal, computational, and theoretical reasons. The as-if Markov model, with its simple information model, approximates the true prospective reserve with a partially averaged state-wise prospective reserve, and is practically observed and therefore theoretically significant, see Christiansen and Djehiche [CD24].

The rest of this subsection introduces the estimation of the prospective reserve in the as-if Markov model. This approach is taken solely to simplify the notation. Again, the calculation of the prospective reserve can be simplified to the calculation of the state-wise

prospective reserve V_i.

The estimation of the state-wise prospective reserves in the as-if Markov model is based on estimating the conditional state occupation probabilities and transition rates, see equation (1.2.3). Let $z \in \mathcal{Z}$, so that $\mathbb{P}(Z(s) = z) > 0$ and define

$$\begin{aligned} P_{z,i}(s,t) &:= \mathbb{E}[\mathbb{1}_{\{Z(t)=i\}} | Z(s) = z], \\ Q_{z,ij}(s,t) &:= \mathbb{E}[N_{ij}(t) | Z(s) = z], \\ \Lambda_{z,ij}(s,t) &:= \int\limits_{(s,t]} \frac{1}{P_{z,i}(s,u-)} Q_{z,ij}(s, \mathrm{d}u), \end{aligned}$$

for $i \neq j$, $t \geq s$. These conditional expected values can be estimated with sub-sampling. Assume that n independent individuals $((Z^{(k)}(t))_{0 \leq t \leq R^{(k)}}, \eta^{(k)} \wedge R^{(k)})_{k=1,\dots,n}$ are observed, where the censoring $R^{(k)}$ remains entirely random. The landmark Nelson-Aalen estimator is defined by:

$$\begin{aligned} \hat{I}_{z,j}(s,t) &:= \sum_{i=1}^{n} \mathbb{1}_{\{Z^{(i)}(s)=z\}} \mathbb{1}_{\{Z^{(i)}(t)=j\}} \mathbb{1}_{\{t < R^{(i)}\}}, \\ \hat{N}_{z,jk}(s,t) &:= \sum_{i=1}^{n} \mathbb{1}_{\{Z^{(i)}(s)=z\}} N_{jk}^{(i)}(t \wedge R^{(i)}), \\ \hat{\Lambda}_{z,jk}(s,t) &:= \int\limits_{(s,t]} \frac{1}{\hat{I}_{z,j}(s,u-)} \hat{N}_{z,jk}(s, \mathrm{d}u), \end{aligned}$$

for $j \neq k$ and $t \geq s$. The only difference to the Nelson-Aalen estimation in Markov models is the sub-sampling on information at time s. This sub-sampling is the reason, the estimators are called landmark estimators. In the same vein as the Aalen-Johansen estimator, the landmark Aalen-Johansen estimator combines the landmark Nelson-Aalen estimator and the differential Equation (1.2.2). The estimator reads

$$\hat{P}_z(s,t) = P_z(s,s) \Big[\prod_{s}^{t} \Big(\mathrm{Id} + \hat{\Lambda}_z(s, \mathrm{d}u) \Big) \Big],$$

for $t \geq s$ with the usual matrix interpretation and diagonal elements $\hat{\Lambda}_{z,jj} = -\sum_{k:k \neq j} \hat{\Lambda}_{z,jk}$.

These estimators are uniformly consistent and asymptotically normal. This was first postulated by Data and Satten [DS01]. However, recent works have questioned their approach and proposed different proofs, see Overgaard [Ove19] and Nießl et al. [Nie+23]. The main consistency results differ in their realization, but the key finding is: For the classical convergence results known in the Markov case, there are corresponding convergence results in the landmarking literature as well.

Landmark Nelson-Aalen and Aalen-Johansen estimators can be applied not only to the as-if Markov model, but to any non-Markov model, where the sigma algebra $\mathcal{F}_s$ is generated

by a discrete random variable. In practice, these discrete conditions may be sufficient, however, for a more comprehensive estimation in non-Markov models, including conditions generated by a continuous random variable, Bladt and Furrer [BF24b] propose kernel estimation methods. In theory, this allows the classical prospective reserve to be estimated conditioned on the entire past.

1.2.2 Research questions

As of today, the literature on Markov models reflects the deep understanding of the general theory and practical application of Markov chains in actuarial sciences and biostatistics. The previous section 1.1 has outlined important developments in Markov modeling specific to actuarial science, with a focus on practical challenges encountered when applying this theory to reserve and premium calculation for insurance products. The initial illustration of calculation and estimation of prospective reserves was followed by variance analysis through the calculation of higher moments of future discounted accumulated cash flows. Lastly, the modeling of stochastic payments, which occur in practice via state dependent discounting, policyholder behavior, or duration-dependent payments, was addressed.

This review does not exhaust the extensive theory regarding Markov modeling. Some key developments not mentioned before are advancements in biostatistics, which offer parametric or semi-parametric estimation, and additional actuarial contributions, such as the Thiele differential equation, improving the calculation of reserves, or the Cantelli theorem, enabling the comparison of reserves of different Markov models.

In contrast, the understanding of the more general non-Markov models remains less developed. Although, as discussed in this section, recent advances have established methods for calculating and estimating conditional expected values in non-Markov models, current statistical and actuarial theory has made limited progress in further analyzing non-Markov modeling. This thesis aims to shed light on non-Markov theory by mainly considering two research questions regarding the impact of the Markov assumption on actuarial applications. In the following a short introduction into these research questions is given:

(i) **Higher moments of canonical cash flows:**
When working with canonical cash flows, the problem of the calculation of higher moments of discounted accumulated cash flows can be reduced to the calculation of $\mathbb{E}[N_{ij}(t_1)N_{kl}(t_2)|\mathcal{F}_s]$. Different from the Markov case, this cannot be calculated with one-dimensional conditional transition rates and state occupation probabilities. This is due to the fact that the Chapman-Kolmogorov equations do not generally hold in the non-Markov case and that the intertemporal dependency present in the non-Markov model does not allow the decomposition into multiple one-dimensional transition rates and probabilities. This thesis proposes an approach based on the use of two-dimensional transition rates, with Kolmogorov like differential equations and landmark estimation of the Nelson-Aalen and Aalen-Johansen type.

(ii) **Stochastic payment functions:**
Modeling of stochastic payments presents significant challenges, evident from the variety of approaches even within the Markov model. For non-Markov models, the recent paper of Christiansen and Furrer [CF22] models policyholder behavior with scaled payments, using the change of measure techniques developed by Furrer [Fur22]. They introduce an artificial state process with an additional state to get the correctly scaled transition rates and probabilities. Beyond this, stochastic payments in the non-Markov theory remain largely unexplored. This thesis seeks to develop new approaches for the stochastic payments introduced in Subsection 1.1.4, focusing on modeling policyholder behavior through two-dimensional transition rates and generalized scaled transition rates and probabilities. Additionally, it explores connections to stochastic interest rates and searches for an answer to the question posed by Norberg [Nor10] regarding a unified definition of forward rates.

1.3 Overview and contributions

This thesis consists of four chapters, this introductory chapter and three further chapters, each composed as a scientific paper. Due to the nature of these chapters, there are significant differences between model assumptions and notation across chapters. The different chapters are described in detail, while similarities and differences of the model assumptions are discussed.

1.3.1 Two-dimensional forward and backward transition rates

Chapter 2, which contains the paper by Bathke and Christiansen [BC24], establishes two-dimensional conditional forward and backward transition rates to capture the intertemporal dependency structure of the state process. These dependencies are modeled to calculate conditional expected values of cash flows with a two-dimensional canonical representation. To ensure correct calculations, even with a censored state process, two-dimensional integral equations that connect conditional state occupation probabilities with conditional forward and backward transition rates are introduced. Finally, a procedure is derived, that allows for the calculation of conditional variances of future liabilities and conditional expected values of specific path-dependent cash flows.

In the following, a detailed account of the setup and results of Chapter 2 is given. The focus lies on the main contributions, while omitting some general but cumbersome equations by presenting only forward calculations and summarizing the reserve calculation without detailed formulas. The notation and naming has been adjusted to align with that of Chapter 3 to ensure consistency throughout the whole introduction.

The state of the insured is governed by a jump process Z, which is assumed to be càdlàg, with values in a finite state space $\mathcal{Z}$. The corresponding transition counting processes are denoted by N and assumed to have finite fourth order moments. The multi-state model is

not assumed to be Markovian or constrained by any specific intertemporal dependency structure. Information available at time $s \geq 0$ is represented by the sigma algebra $\mathcal{F}_s$. In this chapter, it is assumed that $\sigma(Z(s)) \subset \mathcal{F}_s$. Additionally, two-dimensional time points use bold variables $\boldsymbol{t} = (t_1, t_2)$ and a vector of indices also uses bold symbols $\boldsymbol{i} = (i_1, i_2)$. The basis for the ordering of time points in $\mathbb{R}^2$ is the component wise partial ordering.

The state of the insured is used to define cash flows $\boldsymbol{A}$ that have a two-dimensional canonical representation. A cash flow $\boldsymbol{A}$ is considered to have such a representation, when it can be expressed as

$$\begin{aligned}\boldsymbol{A}(\boldsymbol{t}) = &\sum_{i,j\in\mathcal{Z}} \int_{(\boldsymbol{0},\boldsymbol{t}]} I_i(u_1-)I_j(u_2-)A_{ij}(\mathrm{d}u_1,\mathrm{d}u_2)\\ &+ \sum_{\substack{i,k,l\in\mathcal{Z}\\ k\neq l}} \int_{(\boldsymbol{0},\boldsymbol{t}]} I_i(u_1-)a_{ikl}(u_1,u_2)A_i(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2)\\ &+ \sum_{\substack{i,j,k,l\in\mathcal{Z}\\ i\neq j,k\neq l}} \int_{(\boldsymbol{0},\boldsymbol{t}]} a_{ijkl}(u_1,u_2)N_{ij}(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2),\end{aligned}$$

for sufficiently regular payment functions a_{ijk}, a_{ijkl}, A_i, and A_{ij}.

Cumulative two-dimensional conditional transition rates $(\boldsymbol{\Lambda_{ij}})_{\boldsymbol{i},\boldsymbol{j}}$ are the random surfaces, that are càdlàg in both variables, and satisfy:

$$\boldsymbol{\Lambda_{ij}}(t_1,t_2) = \int_{(s,t_1]\times(s,t_2]} \frac{1}{\boldsymbol{P_i}(u_1-,u_2-)}\boldsymbol{Q_{ij}}(\mathrm{d}u_1,\mathrm{d}u_2),$$

for $\boldsymbol{t} \geq (s,s)$, where

$$\begin{aligned}\boldsymbol{P_i}(u_1,u_2) &:= \mathbb{E}[\mathbb{1}_{\{Z(u_1)=i_1,Z(u_2)=i_2\}}|\mathcal{F}_s],\\ \boldsymbol{Q_{ij}}(u_1,u_2) &:= \mathbb{E}[N_{i_1j_1}(u_1)N_{i_2j_2}(u_2)|\mathcal{F}_s],\end{aligned}$$

for $\boldsymbol{u} \geq (s,s)$. In the cases $i_1 = j_1$ and $i_2 = j_2$, the cumulative two-dimensional conditional transition rates $\boldsymbol{\Lambda_{ij}}$ are defined appropriately. $\boldsymbol{\Lambda_{ij}}(\mathrm{d}\boldsymbol{u})$ is called the two-dimensional conditional transition rate and $\boldsymbol{P}$ is called the two-dimensional conditional state occupation probability. It is assumed that these processes are almost surely finite. They are then linked by the integral equations

$$\begin{aligned}\boldsymbol{P_i}(t_1,t_2) = &\mathbb{1}_{\{Z(s)=i_1=i_2\}} + \mathbb{1}_{\{Z(s)=i_2\}} \int_{(s,t_1]} \sum_{j\in\mathcal{Z}} P_j(u-)\Lambda_{ji_1}(\mathrm{d}u)\\ &+ \mathbb{1}_{\{Z(s)=i_1\}} \int_{(s,t_2]} \sum_{l\in\mathcal{Z}} P_l(u-)\Lambda_{li_2}(\mathrm{d}u)\\ &+ \int_{(s,t_1]\times(s,t_2]} \sum_{\boldsymbol{j}\in\mathcal{Z}^2} \boldsymbol{P_j}(u_1-,u_2-)\boldsymbol{\Lambda_{ji}}(\mathrm{d}u_1,\mathrm{d}u_2),\end{aligned}$$

where $t_1, t_2 \geq s$ and P_k and Λ_{ji} are conditional state occupation probabilities and cumulative conditional transition rates from Section 1.2, where the dependency on the time point s is suppressed in the notation. Using these integral equations and Campbell's theorem, a formula for the conditional expected value of a cash flow with a two-dimensional canonical representation can be derived.

Cash flows with a two-dimensional canonical representation are useful for modeling payment processes of specific life insurance policies. One prevalent example is the cash flow of a life insurance policy with the canonical payment process

$$C(t) = \sum_{j \in \mathcal{Z}} \int_{(0,t]} \Big[\mathbb{1}_{\{Z(u-)=i\}} \mathrm{d}C_i(u) + \sum_{i,j:j \neq i} c_{ij}(u) \mathrm{d}N_{ij}(u) \Big],$$

for suitably regular payment functions C_i and c_{jk}, with an included option for the policyholder, that introduces scaling factors for future payments. A prominent example of such a setup is the inclusion of the so-called free policy option. The resulting payment process B for the insurance policy with the policyholder option takes the form

$$B(t) = \int_{(0,t]} \rho(\tau, Z(\tau-), Z(\tau))^{\mathbb{1}_{\{\tau \leq u\}}} \mathrm{d}C(u),$$

where τ is the time the option has been exercised and ρ defines the scaling factors. Dividing the state space into subsets $\mathcal{Z}_0$ (pre-option states) and $\mathcal{Z}_1$ (post-option states), the payment process in $\mathcal{Z}_1$ has the two-dimensional canonical representation

$$B(t) = \sum_{i \in \mathcal{Z}_0} \sum_{j \in \mathcal{Z}_1} \int_{(0,t]^2} \rho(u_2, i, j) \mathrm{d}\tilde{C}(u_1) \mathrm{d}N_{ij}(u_2),$$

if $\tilde{C}$ is the payment process C restricted on the states in $\mathcal{Z}_1$. This could then be used to calculate prospective reserves for such an insurance policy.

This capability to model policyholder options is one reason why the chapter is not constrained to forward calculation. If τ can occur before the current time s, then the outer counting process must search in the interval $(0, t]$, which requires dependencies on past developments as well.

Another important example for the use of cash flows with a two-dimensional canonical representation is the modeling of squared payment processes, which are fundamental for variance calculation. The best example is the previous payment process C. Squaring this payment process returns a process with exactly a two-dimensional canonical representation.

Thus, the theoretical contributions of Chapter 2 provide formulas for the calculation of second order moments of future discounted accumulated cash flows in non-Markov models, giving an answer to the first research question on higher moment calculations in non-Markov

models. Furthermore, a novel approach to the calculation of conditional expected values of cash flows, which emerge from policyholder options in non-Markov models, is given. This provides a partial answer to the second research question on the modeling of stochastic payment functions.

1.3.2 On the estimation of bivariate conditional transition rates

Chapter 3, which contains the paper by Bathke [Bat24], is concerned with statistical estimation of two-dimensional conditional transition rates and state occupation probabilities. The approach to the estimation follows the lines of classical landmark Nelson-Aalen and Aalen-Johansen estimation but departs from the statistical methods shown in Christiansen [Chr21] and Christiansen and Furrer [CF22] by demonstrating uniform convergence with ideas similar to Bladt and Furrer [BF24b] and Gill, Van Der Laan, and Wellner [GLW95]. In the following, a more detailed illustration of the setup and results in Chapter 3 is given.

The basic setup, including the state space, the state process, and the multivariate counting process is the same as in the previous subsection 1.3.1. The difference is that no backward or mixed modeling is considered. For the estimation of two-dimensional conditional transition rates additional assumptions are needed. First, the censoring R is modeled as entirely random and almost surely greater than the current time point s. Assume that the following triple is observed:

$$((Z(t))_{s\leq t\leq R}, \eta \wedge R, \xi),$$

where η is the absorption time of Z and ξ is a discrete random variable, which generates the information $\mathcal{F}_s$ at time s. This means that state-wise estimates are sufficient, which is the basis of the landmarking approach. The classic example is the as-if Markov model, where $\mathcal{F}_s = \sigma(Z(s))$, which is practical, since the only information needed is usually available in practice.

Following the approach from Bladt and Furrer [BF24b], the estimation targets perturbed cumulative two-dimensional conditional transition rates

$$\boldsymbol{\Lambda}_{\boldsymbol{ij}}^{\varepsilon}(t_1, t_2) = \int\limits_{(s,t_1]\times(s,t_2]} \frac{1}{\boldsymbol{P_i}(u_1-, u_2-) \vee \varepsilon} \boldsymbol{Q_{ij}}(\mathrm{d}u_1, \mathrm{d}u_2),$$

for $\varepsilon > 0$ and $t_1, t_2 \geq s$. Similar to the Nelson-Aalen estimator, these can be targeted by estimators that plug in simple averages of sub-sampled data for $\boldsymbol{Q_{ij}}$ and $\boldsymbol{P_i}$. This works even with right-censored data, as long as t_1 and t_2 are left to the right endpoint of the support of R. The emerging estimator $\hat{\boldsymbol{\Lambda}}_{\boldsymbol{ij}}^{\varepsilon}$ is uniformly convergent and converges to $\boldsymbol{\Lambda}_{\boldsymbol{ij}}^{\varepsilon}$. Thus,

$$\lim_{n\to\infty} \sup_{\boldsymbol{t}\in((s,s),\boldsymbol{\theta}]} \left|\hat{\boldsymbol{\Lambda}}_{\boldsymbol{ij}}^{\varepsilon}(\boldsymbol{t}) - \boldsymbol{\Lambda}_{\boldsymbol{ij}}^{\varepsilon}(\boldsymbol{t})\right| \overset{a.s}{=} 0,$$

where θ_1 and θ_2 are left to the right endpoint of the support of R. These estimators are then used to estimate the conditional state occupation probabilities by plugging them into the two-dimensional integral equation

$$\begin{aligned}\hat{\boldsymbol{P}}_{\boldsymbol{i}}^{\varepsilon}(t_1,t_2) = \boldsymbol{P}_{\boldsymbol{i}}(s,s) &+ P_{i_2}(s) \int\limits_{(s,t_1]} \sum_{j\in\mathcal{Z}} \hat{P}_j^{\varepsilon}(u-)\hat{\Lambda}_{ji_1}^{\varepsilon}(\mathrm{d}u) \\ &+ P_{i_1}(s) \int\limits_{(s,t_2]} \sum_{l\in\mathcal{Z}} \hat{P}_l^{\varepsilon}(u-)\hat{\Lambda}_{li_2}^{\varepsilon}(\mathrm{d}u) \\ &+ \int\limits_{(s,t_1]\times(s,t_2]} \sum_{\boldsymbol{j}\in\mathcal{Z}^2} \hat{\boldsymbol{P}}_{\boldsymbol{j}}^{\varepsilon}(u_1-,u_2-)\hat{\boldsymbol{\Lambda}}_{\boldsymbol{ji}}^{\varepsilon}(\mathrm{d}u_1,\mathrm{d}u_2),\end{aligned}$$

where the corresponding one-dimensional perturbed landmark Nelson-Aalen and Aalen-Johansen estimators from Bladt and Furrer [BF24b] are used as well. The solution to this integral equation can be written as

$$\hat{\boldsymbol{P}}^{\varepsilon}(\boldsymbol{t}) = \phi^{\varepsilon}(\boldsymbol{t}) + \int\limits_{((s,s),\boldsymbol{t}]} \phi^{\varepsilon}(\boldsymbol{u}-)\hat{\boldsymbol{\Lambda}}^{\varepsilon}(\mathrm{d}\boldsymbol{u})\mathcal{P}((\boldsymbol{u},\boldsymbol{t}],\hat{\boldsymbol{\Lambda}}^{\varepsilon}),$$

where ϕ^{ε} is defined by the boundary conditions and $\mathcal{P}$ is called the two-dimensional Peano-series and is defined by

$$\mathcal{P}((\boldsymbol{u},\boldsymbol{t}],\hat{\boldsymbol{\Lambda}}^{\varepsilon}) := Id + \sum_{n=1}^{\infty} \int\limits_{\boldsymbol{u}<\boldsymbol{x}^{(1)}<\ldots<\boldsymbol{x}^{(n)}\leq\boldsymbol{t}} \cdots \int \hat{\boldsymbol{\Lambda}}^{\varepsilon}(\mathrm{d}\boldsymbol{x}^{(1)})\cdots\hat{\boldsymbol{\Lambda}}^{\varepsilon}(\mathrm{d}\boldsymbol{x}^{(n)}).$$

The convergence results for $\hat{\boldsymbol{\Lambda}}^{\varepsilon}$ extend to $\hat{\boldsymbol{P}}^{\varepsilon}$, because the two-dimensional Peano-series can be proven to be Hadamard differentiable.

The payment functions in the definition of the two-dimensional canonical representation are suitably regular and bounded, so the operator that calculates the conditional expected value from the two-dimensional conditional transition rates and state occupation probabilities is also Hadamard differentiable. Thus, the convergence results for both estimators can be carried over when using the conditional expectation formulas from Chapter 2.

This approach offers a solution for the estimation of conditional second order moments of future discounted accumulated cash flows and conditional expected values of payment streams arising from policyholder options in non-Markov models. With that it addresses the first research question on higher moment calculations in non-Markov models and partially answers the second research question on modeling stochastic payments.

The chapter also introduces a numerical study, which gives a proof of concept for the estimation with these two-dimensional landmark Nelson-Aalen and Aalen-Johansen estimators. The implementation and the simulated data for this numerical study can be found on GitHub[1] or in the List of Implementations.

[1] `https://github.com/Theis-Bathke/Estimation-of-bivariate-transition-rates`

1.3.3 Non-parametric estimators of scaled cash flows

Chapter 4, which contains the manuscript by Bathke and Furrer [BF24a], explores the estimation of conditional expected values of scaled cash flows, particularly focusing on the inclusion of policyholder behavior, including the free policy option. The novel concept of scaled transition rates adds to the change of measure technique used by Furrer [Fur22] and further by Christiansen and Furrer [CF22]. The chapter also presents a numerical simulation comparing the performance of the change of measure technique and the new approach involving scaled transition rates for the estimation of scaled probabilities. Furthermore, the estimation of expected values of scaled cash flows using two-dimensional landmark estimators and scaled estimators are compared.

In the following, a review of the framework and the primary results of Chapter 4 is provided. The focus lies on the difference in setup to the previous chapters and the primary contributions regarding the estimation of expected values of scaled cash flows. Again, the notation is adjusted to ensure consistency throughout the whole introduction.

The basic setup, including the state space, the state process, the censoring, and the available data for estimation mirrors Subsection 1.3.2. The main difference is the deterministic starting point $Z(0) = z_0$ with an initial time point $s = 0$, making the initial state of the insured at time s deterministic. This eliminates the need for sub-sampling when estimating the conditional expected values of future discounted accumulated cash flows, simplifying the notation in the process. The extension to discrete landmarks works by introducing sub-sampling of the available data.

The chapter considers cash flows of insurance policies with the free policy option, as already introduced in Subsection 1.1.4:

$$B(t) = \int_{(0,t]} \rho(\tau, Z(\tau-), Z(\tau))^{\mathbb{1}_{\{\tau \leq u\}}} \Big(\sum_{j \in \mathcal{Z}} \mathbb{1}_{\{Z(u-)=j\}} B_j(\mathrm{d}u) + \sum_{\substack{j,k \in \mathcal{Z} \\ j \neq k}} b_{jk}(u) N_{jk}(\mathrm{d}u) \Big).$$

The expected values of accumulated scaled cash flows can then be calculated with the formula

$$A(t) = \sum_{j \in \mathcal{Z}} \int_{(0,t]} P_j^{\rho}(u-) \Big(B_j(\mathrm{d}u) + \sum_{\substack{k \in \mathcal{Z} \\ k \neq j}} b_{jk}(u) \Lambda_{jk}^{\rho}(\mathrm{d}u) \Big), \tag{1.3.1}$$

where the functions $(P_j^{\rho})_{j \in \mathcal{Z}}$ and $(Q_{jk}^{\rho})_{j,k:j \neq k}$ are defined as

$$\begin{aligned} P_j^{\rho}(t) &= \mathbb{E}[\rho(\tau, Z(\tau-), Z(\tau))^{\mathbb{1}_{\{\tau \leq t\}}} \mathbb{1}_{\{Z(t)=j\}}], \\ Q_{jk}^{\rho}(t) &= \mathbb{E}\Big[\int_0^t \rho(\tau, Z(\tau-), Z(\tau))^{\mathbb{1}_{\{\tau \leq u\}}} N_{jk}(\mathrm{d}u) \Big], \end{aligned}$$

with the cumulative scaled transition rates $(\Lambda^{\rho}_{jk})_{j,k:j\neq k}$ defined as

$$\Lambda^{\rho}_{jk}(t) = \int\limits_{(0,t]} \frac{1}{P^{\rho}_{j}(s-)} Q^{\rho}_{jk}(\mathrm{d}s).$$

Similar to the two-dimensional estimators, only perturbed cumulative scaled transition rates are targeted in the estimation:

$$\Lambda^{\rho,\varepsilon}_{jk}(t) = \int\limits_{(0,t]} \frac{1}{P^{\rho}_{j}(s-) \vee \varepsilon} Q^{\rho}_{jk}(\mathrm{d}s).$$

Similar to the ordinary landmark Nelson-Aalen estimator, the scaled Nelson-Aalen estimator $\hat{\Lambda}^{\rho,\varepsilon}_{jk}$ is constructed by plugging in averages. Provided that ρ is uniformly bounded on the support of R, the estimator is shown to converge uniformly and almost surely, i.e.

$$\lim_{n\to\infty} \sup_{0\leq t\leq\theta} |\hat{\Lambda}^{\rho,\varepsilon}_{jk}(t) - \Lambda^{\rho,\varepsilon}_{jk}(t)| \overset{a.s}{=} 0,$$

even when the used data is censored. Now, the concept of scaled transition rates is only useful when the connection between scaled transition rates and scaled state occupation probabilities is preserved. This connection comes in the form of the equation

$$P^{\rho}_{j}(t) = \left[\prod_{0}^{t} \left(Id + \Lambda^{\rho}(\mathrm{d}s) \right) \right]_{z_0 j},$$

where the product integral is the solution to a Kolmogorov type differential equation and the diagonals are defined specifically as

$$\begin{aligned} -\Lambda^{\rho}_{jj} &:= -\Lambda_{jj}, && j \in \mathcal{Z}_0, \\ -\Lambda^{\rho}_{jj} &:= \sum_{\substack{k\in\mathcal{Z} \\ k\neq j}} \Lambda^{\rho}_{jk}, && j \in \mathcal{Z}_1, \end{aligned}$$

where Λ_{jj} are the cumulative conditional transition rates. The perturbed scaled state occupation probabilities are given by

$$P^{\rho,\varepsilon}_{j}(t) = \left[\prod_{0}^{t} \left(Id + \Lambda^{\rho,\varepsilon}(\mathrm{d}s) \right) \right]_{z_0 j},$$

with estimator $\hat{P}^{\rho,\varepsilon}$ obtained via plug-in of the scaled Nelson-Aalen estimator and diagonals that are defined according to the previous definition. This estimator $\hat{P}^{\rho,\varepsilon}$ then converges uniformly and almost surely to $P^{\rho,\varepsilon}$. Additionally, asymptotic normality, here meaning weak convergence to a tight Gaussian process of both estimators $\hat{\Lambda}^{\rho,\varepsilon}$ and $\hat{P}^{\rho,\varepsilon}$ is proven. For comparison with the two-dimensional method, asymptotic normality of the two-dimensional landmark Nelson-Aalen estimator $\hat{\boldsymbol{\Lambda}}^{\varepsilon}$ and landmark Aalen-Johansen estimator $\hat{\boldsymbol{P}}^{\varepsilon}$ is also established.

The chapter concludes this comparison with numerical simulation studies comparing the different estimators for scaled state occupation probabilities and expected values of scaled cash flows. First, the estimation with the change of measure technique by Christiansen and Furrer [CF22] is compared to the scaled Nelson-Aalen and Aalen-Johansen estimators, confirming that the change of measure technique introduces artificial variance into the estimation of scaled occupation probabilities. Overall, the scaled Nelson-Aalen and Aalen-Johansen estimators exhibit lower variance. Secondly, the numerical and statistical effectiveness of the scaled estimator is compared with the two-dimensional landmark estimator, both targeting expected values of scaled cash flows. The study indicates that scaled estimators are numerically and statistically superior, particularly due to the higher variance of the two-dimensional estimators. The main selling point of the two-dimensional estimators is the separation of contractual modeling, including scaling factors, and statistics of biometric and behavioral risk. The implementation and the simulated data for this numerical study can be found on GitHub[2] or in the List of Implementations.

When comparing the numerical studies in Chapter 3 and Chapter 4, a noticeable difference in the variance of the two-dimensional estimators may prompt interest. This contrast arises because the estimation target differs between the two chapters. In Chapter 3, the estimation target is the prospective reserve, which includes interest rates of 3%, effectively diminishing the impact of later payments. However, this discounting factor is not included in the simulation study in the latter Chapter 4, since there, the estimation target is the expected accumulated future cash flow, which does not include discounting.

Lastly, the chapter extends the theory of scaled transition rates and state occupation probabilities to include various scaling factors. The setup works for stochastic processes H, which are right-continuous, of finite variation, uniformly bounded on compacts, and adapted to the information generated by Z. While the estimation procedure for this generalized setting is not fully developed in Chapter 4, the proof works similarly to the special case involving free policy options. This includes all the stochastic payments introduced for the Markov model in Subsection 1.1.4, providing the flexibility known from the Markov model in the non-Markov model, while mitigating systematic model risk and without substantially increasing numerical complexity.

In conclusion, the main contribution of Chapter 4 is the definition of scaled transition rates and scaled estimators. This approach provides a flexible solution to open questions on the modeling of actuarial cash flows with non-Markov models. It partially addresses the search for a unifying approach to forward rates and allows the modeling of the different stochastic payment functions introduced in Subsection 1.1.4 in non-Markov models. The application to the modeling of policyholder options is shown to work well in a numerical example. This answers the second research question on the specific stochastic payment functions introduced in the current Chapter. Additionally, this approach holds potential for broader applications.

[2] `https://github.com/Theis-Bathke/Non-parametrix-estimators-of-scaled-cash-flows`

1.4 Conclusion and further research opportunities

This thesis is concerned with complications arising from the departure of the Markov assumption, meaning, the use of non-Markov models. These models are valuable because they replace systematic model risk with unsystematic estimation risk, which has more favorable asymptotic properties. Overall, there is a bias-variance trade-off, warranting further investigation for practical applications.

One of the main challenges in these non-Markov models is understanding the intertemporal dependency structure of the state process. Such dependencies play a major role in actuarial mathematics, when dealing with life insurance cash flows that are path-dependent or when higher moments of life insurance cash flows are of interest. The understanding of these dependencies is improved by the investigation of two-dimensional conditional transition rates. These transition rates can be estimated by using two-dimensional landmark Nelson-Aalen and Aalen-Johansen estimators, combining techniques from survival analysis and actuarial mathematics. While these estimations give valuable insights into dependencies in the data, they are prone to high variances, especially in the tail when censoring is introduced.

For specific cases, the need for two-dimensional conditional transition rates can be circumvented by applying the newly introduced method of scaled conditional transition rates. This avoids the curse of dimensionality while still correctly capturing dependencies in non-Markov models. Numerical examples indicate that estimations of expected accumulated future payments converge better with scaled Nelson-Aalen and Aalen-Johansen estimators than with the two-dimensional counterparts and that the one-dimensional scaled estimators are significantly more feasible numerically. In these cases, the scaled estimators depend on the contractual setup of the insurance product, whereas two-dimensional estimations achieve true separation between contractual choices and statistical estimation.

There are numerous open research questions within the field of non-Markov modeling and multi-dimensional conditional transition rates. Below follows a brief and non-exhaustive list:

(i) For a full understanding of the probability distribution of the state space, it is necessary to calculate all multi-dimensional conditional transition rates. This can be achieved by using the tools introduced in this manuscript. The remaining task involves identifying the integral equations that connect multi-dimensional conditional transition rates to the corresponding transition probabilities.

(ii) The high variances and non-monotonicity in the tail of plug-in estimators using two-dimensional conditional transition rates are generally not a concern in actuarial applications due to discounting. However, an estimator with improved tail performance is still desirable. This could be achieved using parametric or semi-parametric estimation methods or by extending the work of Gill, Van Der Laan and Wellner [GLW95], where they improve on the Volterra estimator, which is an estimator that uses a similar method to Chapter 3 but for survival functions.

(iii) Non-Markov modeling and the associated landmark Nelson-Aalen and Aalen-Johansen estimators have been shown to reduce bias and provide a positive bias-variance trade-off in some simulations, as demonstrated in survival contexts by Putter and Spitoni [PS18]. To encourage practitioners to adopt non-Markov modeling, especially for risk analysis, a statistical evaluation of these bias-variance trade-offs should be conducted for an actuarial example, where real reserves are estimated.

Chapter 2

Two-dimensional forward and backward transition rates

Abstract

Forward transition rates were originally introduced with the aim to evaluate life insurance liabilities market-consistently. While this idea turned out to have its limitations, recent literature repurposes forward transition rates as a tool for avoiding Markov assumptions in the calculation of life insurance reserves. While life insurance reserves are some form of conditional first-order moments, the calculation of conditional second-order moments needs an extension of the forward transition rate concept from one dimension to two dimensions. Two-dimensional forward transition rates are also needed for the calculation of path-dependent life insurance cash-flows as they occur upon contract modifications. Forward transition rates are designed for doing prospective calculations, and by a time-symmetric definition of so-called backward transition rates one can do retrospective calculations.

Keywords: life & health insurance; non-Markov modelling; prospective & retrospective reserves; second-order moments; free policy option

2.1 Introduction

Forward transition rates describe the expected number of future transitions conditional on the currently available information. If the current information is incomplete, then backward transition rates serve as a proxy for the number of past transitions. However, forward and backward transition rates do not describe the inter-temporal dependency structure of two successive jump events, so they are unsuitable for the calculation of second-order moments or for dealing with path-dependent insurance cash-flows as they occur upon contract modifications. For this reason, this paper introduces two-dimensional forward and backward transition rates and explains their use in life insurance calculations.

The research on forward transition rates originally emerged from the desire to calculate market values for life insurance liabilities. Miltersen and Persson [MP05] suggested to define mortality rates implicitly in such a way that the classical actuarial formulas reproduce market values instead of real-world expectations. They denoted these implicit mortality

rates as forward mortality rates, inspired by the concept of forward interest rates from financial mathematics. Norberg [Nor10] developed generalizations of the forward concept to multi-state life insurance but observed that the implicit definitions are lacking uniqueness. Christiansen and Niemeyer [CN15] observed that the implicit definitions often do not have any solution at all. So it turned out that the implicit concept has serious limitations. Buchardt, Furrer and Steffensen [BFS19] suggested to overcome these limitations by introducing additional artificial states, but their concept is restricted to insurance contracts that have sojourn payments only and no transition payments.

Buchardt [Buc17] discarded the implicit concept and presented an explicit definition that works for a doubly stochastic Markov framework. Buchardt actually drifts away from market-consistent valuation but without clearly mentioning.

Christiansen [Chr21] discovered that Buchardt's approach can be extended to a general recipe for calculating real-world expectations in fully non-Markovian life insurance frameworks. Christiansen and Furrer [CF22] enriched this non-Markov concept with a change of measure technique that makes it possible to deal also with path-dependent insurance payments. To put it into a nutshell, we find two divergent concepts of forward transition rates in the actuarial literature. First, there is the implicit concept which aims to reproduce market values, first suggested by Miltersen and Persson [MP05] and taken on by numerous further authors. Second, there are the explicit definitions of Buchardt [Buc17], Christiansen [Chr21], and Christiansen and Furrer [CF22] that put the focus back on real-world expectations and use forward transition rates as a tool to cope with complex inter-temporal dependency structures in life insurance models. This paper follows the second line of thought.

The change of measure technique in Christiansen and Furrer [CF22] transfers the complexity of path-dependent insurance cash-flows to auxiliary probabilistic models. This way the forward transition rates may stay one-dimensional, but each insurance cash-flow needs another probabilistic model. Our aim is to have one probabilistic model only, and we achieve that by expanding the concept of forward transition rates to two dimensions. The price for having just one probabilistic model is an increased numerical effort that comes with the extra dimension. So the approach of Christiansen and Furrer [CF22] is beneficial when many scenarios are calculated for one and the same cash-flow, whereas the results of this paper are favorable when many different cash-flows ought to be calculated.

One-dimensional forward transition rates are designed as a tool for calculating expectations. Our two-dimensional concept can be used as a tool for calculating second-order moments. The calculation of second-order moments in classical Markov models was first outlined in Hoem [Hoe69]. Helwich [Hel08] presents general formulas for the calculation of variances in semi-Markov models.

Calculation formulas for second-order moments in fully non-Markovian models do not exist yet in the literature. By introducing two-dimensional forward and backward transition rates we help to close that gap.

Markov modeling has a long tradition in life insurance. The classical concept to model the random pattern of states of the insured as a Markov process has been extended to semi-Markov modeling and further sophisticated Markov structures by numerous authors.

The problem with any kind of Markov assumptions is that they come with model risk, while at the same time there is a lack of tools for quantifying the model error. This motivates the search for non-Markovian calculation techniques, and this paper is a major step into that direction.

The paper is structured as follows: Section 2.2 introduces the fundamental life insurance modeling framework. Section 2.3 introduces the definition of two-dimensional forward and backward transitions rates and develops a corresponding integral equation that generalizes Kolmogorov's forward equation. For the latter integral equation Section 2.4 verifies the uniqueness of solutions. Section 2.5 then turns to the main purpose of two-dimensional forward and backward transitions rates, namely the calculation of certain conditional moments. In Section 2.6 we explain the calculation of conditional variances. Section 2.7 illustrates the calculation of path-dependent cash-flows. Section 2.8 concludes and gives an outlook on open research questions.

2.2 Life insurance modeling framework

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$. We consider an individual life insurance contract and describe the status of the individual insured by an adapted càdlàg jump process

$$Z = (Z(t))_{t\geq 0}$$

on a finite state space $\mathcal{Z}$. Additionally, we set $Z_{0-} := Z_0$. Throughout this paper we assume that we are currently at time $s \geq 0$. So the time interval $[0, s]$ represents to the past and the present, and the time interval (s, ∞) represents the future. The real number $s \geq 0$ is arbitrary but fixed. Based on Z we define state indicator processes $(I_i)_{i\in\mathcal{Z}}$ by

$$I_i(t) := \mathbb{1}_{\{Z(t)=i\}}, \quad t \geq 0,$$

and transition counting processes $(N_{ij})_{i,j\in\mathcal{Z}:i\neq j}$ by

$$N_{ij}(t) := \#\{u \in (0, t] : Z(u-) = i, Z(u) = j\}, \quad t \geq 0.$$

We generally assume that

$$\mathbb{E}[N_{ij}(t)] < \infty, \quad t \geq 0\,, i, j \in \mathcal{Z},\ i \neq j, \tag{2.2.1}$$

which in particular implies that Z has almost surely no explosions. Let

$$\begin{aligned} N_{ii}(t) &:= -\sum_{\substack{j\in\mathcal{Z}\\ j\neq i}} (N_{ij}(t) - N_{ij}(s)), \quad t > s, \\ N_{ii}(t) &:= -\sum_{\substack{j\in\mathcal{Z}\\ j\neq i}} (N_{ji}(t) - N_{ji}(s)), \quad t \leq s. \end{aligned} \tag{2.2.2}$$

This construction is made so that $N_{ii}(t)$ satisfies

$$\mathrm{d}N_{ii}(t) = -\sum_{\substack{j\in\mathcal{Z}\\ j\neq i}} \mathrm{d}N_{ij}(t), \quad t > s,$$
$$\mathrm{d}N_{ii}(t) = -\sum_{\substack{j\in\mathcal{Z}\\ j\neq i}} \mathrm{d}N_{ji}(t), \quad t \leq s,$$

and is càdlàg everywhere. Definition (2.2.2) and many further definitions following below involve a case differentiation between $t > s$ and $t \leq s$. As time s is fixed, we omit it in the notation, but one should keep in mind the dependence of many of our definitions on the parameter s. The following equations show a useful direct link between the processes $(N_{ij})_{i,j\in\mathcal{Z}}$ and $(I_i)_{i\in\mathcal{Z}}$:

$$I_i(t) = I_i(s) + \sum_{j\in\mathcal{Z}} \int_{(s,t]} N_{ji}(\mathrm{d}u), \quad t \geq s, i \in \mathcal{Z}, \tag{2.2.3}$$

and

$$I_i(t) = I_i(s) + \sum_{j\in\mathcal{Z}} \int_{(t,s]} N_{ij}(\mathrm{d}u), \quad s \geq t, i \in \mathcal{Z}. \tag{2.2.4}$$

The latter integrals and all following integrals in this paper are meant as path-wise Lebesgue-Stieltjes integrals.

The sigma-algebra $\mathcal{F}_s$ represents the available information at time s. In insurance practice, the insurer often uses a reduced information set $\mathcal{G}_s$ for actuarial evaluations. In this paper we generally assume that

$$\sigma(Z(s)) \subseteq \mathcal{G}_s \subseteq \mathcal{F}_s. \tag{2.2.5}$$

The special case $\mathcal{G}_s = \sigma(Z(s))$ is known as the as-if Markov model, since it uses information of Markov-type only. The choice of $\mathcal{G}_s$ can be influenced by many factors. Some of these are listed here:

- By cutting down the information used, one can reduce the complexity of numerical calculations.
- A lack of data for statistical inference may make it necessary to simplify the information model.
- For some actuarial tasks it is sufficient to study mean portfolio values only, and then it is convenient to minimize the individual information.
- Anti-discrimination regulation may be a limiting factor for the use of information, such as unisex calculation.
- Data privacy regulation can restrict the information that the insurer actually gathers and stores. For example, the General Data Protection Regulation of the European Union gives individuals considerable control and rights over their personal data.

Let B be the insurance cash-flow of the individual life insurance contract, here assumed to be an adapted càdlàg process with paths of finite variation. We assume that the insurance contract has a maximum contract horizon of T, which means that

$$B(\mathrm{d}t) = 0, \quad t > T.$$

Throughout this paper we assume that

$$0 \leq s \leq T < \infty.$$

Let κ be a càdlàg function that describes the value development of a savings account. We assume that κ is strictly positive so that the corresponding discounting function $1/\kappa$ exists. The random variable

$$Y^+ := \int\limits_{(s,T]} \frac{\kappa(s)}{\kappa(u)} B(\mathrm{d}u) \tag{2.2.6}$$

describes the discounted accumulated future cash-flow seen from time s, and the random variable

$$Y^- := \int\limits_{[0,s]} \frac{\kappa(s)}{\kappa(u)} B(\mathrm{d}u) \tag{2.2.7}$$

is the compounded accumulated past cash-flow. In order to ensure integrability, we generally assume that $1/\kappa(\cdot)$ is bounded on $[0, T]$.

Classical Markov modeling focuses on benefit payment functions that may depend on current states or current jumps of the insured but not on past random events. This means that B is supposed to have a so-called one-dimensional canonical representation.

Definition 2.2.1.
A stochastic process A is said to have a *one-dimensional canonical representation* if there exist real-valued functions $(A_i)_i$ on $[0, \infty)$ that generate finite signed measures $A_i(\mathrm{d}u)$ and there exist measurable and bounded real functions $(a_{ij})_{ij:i\neq j}$ such that

$$A(t) = \sum_{i \in \mathcal{Z}} \int\limits_{[0,t]} I_i(u^-) A_i(\mathrm{d}u) + \sum_{\substack{i,j \mathcal{Z} \\ i \neq j}} \int\limits_{[0,t]} a_{ij}(u) N_{ij}(\mathrm{d}u), \quad t \geq 0. \tag{2.2.8}$$

Insurance cash-flows that involve contract modifications, such as a free policy option, can not be brought into the form (2.2.8). For this reason we need to allow for more complex structures.

Definition 2.2.2.
A stochastic process A is said to have a *two-dimensional canonical representation* if there exist real-valued functions $(A_i)_i$ on $[0, \infty)$ that generate finite signed measures $A_i(\mathrm{d}u_1)$, real-valued functions $(A_{ij})_{ij}$ on $[0, \infty)^2$ that generate finite signed measures $A_{ij}(\mathrm{d}u_1, \mathrm{d}u_2)$,

and measurable and bounded real-valued functions $(a_{ikl})_{ikl:k\neq l}$, $(a_{ijkl})_{ijkl:i\neq j,k\neq l}$ on $[0,\infty)^2$ such that

$$\begin{aligned} A(t) = &\sum_{i,j\in\mathcal{Z}} \int_{[0,t]^2} I_i(u_1^-)I_j(u_2^-)A_{ij}(\mathrm{d}u_1,\mathrm{d}u_2) \\ &+ \sum_{\substack{i,k,l\in\mathcal{Z}\\ k\neq l}} \int_{[0,t]^2} I_i(u_1^-)a_{ikl}(u_1,u_2)A_i(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2) \\ &+ \sum_{\substack{i,j,k,l\in\mathcal{Z}\\ i\neq j,k\neq l}} \int_{[0,t]^2} a_{ijkl}(u_1,u_2)N_{ij}(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2), \quad t_1,t_2\geq 0. \end{aligned} \tag{2.2.9}$$

Example 2.2.3.
Suppose that B is an insurance cash flow that has a one-dimensional canonical representation with respect to suitable functions $(B_i)_i$ and $(b_{ij})_{ij:i\neq j}$. Then the squared process B^2 has a two-dimensional canonical representation:

$$\begin{aligned} B^2(t) = &\sum_{i,j\in\mathcal{Z}} \int_{[0,t]^2} I_i(u_1^-)I_j(u_2^-)B_i(\mathrm{d}u_1)B_j(\mathrm{d}u_2) \\ &+ \sum_{\substack{i,k,l\in\mathcal{Z}\\ k\neq l}} \int_{[0,t]^2} 2I_i(u_1^-)b_{kl}(u_2)B_i(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2) \\ &+ \sum_{\substack{i,j,k,l\in\mathcal{Z}\\ i\neq j,k\neq l}} \int_{[0,t]^2} b_{ij}(u_1)b_{kl}(u_2)N_{ij}(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2), \quad t\geq 0. \end{aligned} \tag{2.2.10}$$

In order to see this, use integration by parts according to Proposition 2.A.2 and Fubini's theorem.

Example 2.2.4.
Suppose that $\mathcal{Z}=\mathcal{S}\times\{0,1\}$, where the elements of $\mathcal{S}$ describe the health status of the individual insured and $\{0,1\}$ indicates whether the policyholder has exercised a free policy option. We assume here that the free policy option can be exercised only once and that the policy cannot move back to a premium-paying status. Let τ be the random time where the free policy option is actually exercised, i.e. τ gives the time where Z moves from the set $\mathcal{S}_0:=\mathcal{S}\times\{0\}$ to the set $\mathcal{S}_1:=\mathcal{S}\times\{1\}$. At time τ the insurance payment scheme is re-scaled by a factor $\rho(\tau,Z_{\tau^-},Z_\tau)$ in order to maintain actuarial equivalence, cf. [CF22]. So, by writing C for the insurance payment scheme, the insurance cash-flow B equals

$$B(t)=\int_{[0,t]}\rho(\tau,Z_{\tau^-},Z_\tau)^{\mathbb{1}_{\{\tau\leq u\}}}C(\mathrm{d}u).$$

We assume that C has a finite horizon of T. For the sake of simplicity, let

$$\mathbb{1}_{\{\tau=u\}}C(\mathrm{d}u)=0, \quad \forall i\in\mathcal{Z}, \tag{2.2.11}$$

almost surely, so that we almost surely have no lump sum payments at time τ. Suppose that the payment scheme C has a one-dimensional canonical representation with respect to suitable functions $(C_i)_i$ and $(c_{ij})_{ij:i\neq j}$. The cash-flow B can be decomposed to

$$B(t) = \int_{[0,t]} \mathbb{1}_{\{u<\tau\}} C(\mathrm{d}u) + \int_{[0,t]} \mathbb{1}_{\{u\geq\tau\}} \rho(\tau, Z_{\tau^-}, Z_\tau) C(\mathrm{d}u). \tag{2.2.12}$$

We will now analyze both integrals separately. Beginning with the first one, we have

$$\int_{[0,t]} \mathbb{1}_{\{u<\tau\}} C(\mathrm{d}u) = \sum_{i\in\mathcal{S}_0} \int_{[0,t]} \mathbb{1}_{\{u<\tau\}} I_i(u^-) C_i(\mathrm{d}u) + \sum_{\substack{i,j\in\mathcal{S}_0 \\ i\neq j}} \int_{[0,t]} \mathbb{1}_{\{u<\tau\}} c_{ij}(u) N_{ij}(\mathrm{d}u)$$

since $\mathbb{1}_{\{u<\tau\}} I_i(u^-) = 0$ and $\mathbb{1}_{\{u<\tau\}} N_{ij}(\mathrm{d}u) = 0$ for all $(i,j) \notin \mathcal{S}_0^2$. Because of (2.2.11) and the fact that $\mathbb{1}_{\{u\leq\tau\}} I_i(u^-) = I_i(u^-)$ and $\mathbb{1}_{\{u<\tau\}} N_{ij}(\mathrm{d}u) = N_{ij}(\mathrm{d}u)$ for $i,j \in \mathcal{S}_0$, we furthermore get

$$\int_{[0,t]} \mathbb{1}_{\{u<\tau\}} C(\mathrm{d}u) = \sum_{i\in\mathcal{S}_0} \int_{[0,t]} I_i(u^-) C_i(\mathrm{d}u) + \sum_{\substack{i,j\in\mathcal{S}_0 \\ i\neq j}} \int_{[0,t]} c_{ij}(u) N_{ij}(\mathrm{d}u)$$

almost surely. So the insurance cash-flow prior to τ almost surely has a one-dimensional canonical representation. Since τ is the unique jump time of the counting process $\sum_{k\in\mathcal{S}_0}\sum_{l\in\mathcal{S}_1} N_{kl}$, by using assumption (2.2.11) the second integral in (2.2.12) can be almost surely transformed to

$$\begin{aligned}
&\int_{[0,t]} \mathbb{1}_{\{u\geq\tau\}} \rho(\tau, Z_{\tau^-}, Z_\tau) C(\mathrm{d}u) \\
&= \sum_{k\in\mathcal{S}_0} \sum_{l\in\mathcal{S}_1} \int_{[0,t]^2} \mathbb{1}_{\{u_1\geq u_2\}} \rho(u_2,k,l) C(\mathrm{d}u_1) N_{kl}(\mathrm{d}u_2) \\
&= \sum_{k\in\mathcal{S}_0} \sum_{i,l\in\mathcal{S}_1} \int_{[0,t]^2} I_i(u_1^-) \rho(u_2,k,l) C_i(\mathrm{d}u_1) N_{kl}(\mathrm{d}u_2) \\
&\quad + \sum_{k\in\mathcal{S}_0} \sum_{\substack{i,j,l\in\mathcal{S}_1 \\ i\neq j}} \int_{[0,t]^2} \rho(u_2,k,l) c_{ij}(u_1) N_{ij}(\mathrm{d}u_1) N_{kl}(\mathrm{d}u_2).
\end{aligned}$$

So, at and after time τ the insurance cash-flow almost surely has a two-dimensional canonical representation. All in all, we obtain for the insurance cash-flow B the almost sure representation

$$\begin{aligned}
B(t) = &\sum_{i\in\mathcal{S}_0} \int_{[0,t]} I_i(u^-) C_i(\mathrm{d}u) + \sum_{\substack{k,l\in\mathcal{S}_0 \\ k\neq l}} \int_{[0,t]} c_{kl}(u) N_{kl}(\mathrm{d}u) \\
&+ \sum_{k\in\mathcal{S}_0} \sum_{i,l\in\mathcal{S}_1} \int_{[0,t]^2} I_i(u_1^-) \rho(u_2,k,l) C_i(\mathrm{d}u_1) N_{kl}(\mathrm{d}u_2) \\
&+ \sum_{k\in\mathcal{S}_0} \sum_{\substack{i,j,l\in\mathcal{S}_1 \\ i\neq j}} \int_{[0,t]^2} \rho(u_2,k,l) c_{ij}(u_1) N_{ij}(\mathrm{d}u_1) N_{kl}(\mathrm{d}u_2), \quad t \geq 0.
\end{aligned} \tag{2.2.13}$$

2.3 Two-dimensional transition rates

This section generalizes the forward and backward transition rates of [Chr21] from one dimension to two dimensions. We still suppose that we are currently at time s and have the information $\mathcal{G}_s$ available. As the parameter s is fixed we omit in the notation.

Let $P_i = (P_i(t))_{t\geq 0}$ and $Q_{ij} = (Q_{ij}(t))_{t\geq 0}$ for $i, j \in \mathcal{Z}$ be the almost surely unique càdlàg processes that satisfy

$$\begin{aligned} P_i(t) &= \mathbb{E}[I_i(t)|\mathcal{G}_s], \\ Q_{ij}(t) &= \mathbb{E}[N_{ij}(t)|\mathcal{G}_s]. \end{aligned}$$

As we already mentioned after definition (2.2.2), the process N_{ii} is càdlàg. Let $P_{ik} = (P_{ik}(t_1, t_2))_{t_1,t_2\geq 0}$ and $Q_{ijkl} = (Q_{ijkl}(t_1, t_2))_{t_1,t_2\geq 0}$ for $i, j, k, l \in \mathcal{Z}$ be the almost surely unique random surfaces that are càdlàg in each variable and satisfy

$$\begin{aligned} P_{ik}(t_1, t_2) &= \mathbb{E}[I_i(t_1)I_k(t_2)|\mathcal{G}_s], \\ Q_{ijkl}(t_1, t_2) &= \mathbb{E}[N_{ij}(t_1)N_{kl}(t_2)|\mathcal{G}_s]. \end{aligned}$$

In order to find suitable modifications of the conditional expectations so that the processes have the postulated càdlàg path properties, one can calculate the conditional expectations on the basis of a fixed regular conditional distribution $\mathbb{P}(\,\cdot\,|\mathcal{G}_s)$. Then the càdlàg path properties follow directly from the càdlàg path properties of $(I_i)_i$ and $(N_{ij})_{ij}$ and the dominated convergence theorem. The càdlàg path properties imply that the processes P_i, Q_{ij} and the surfaces P_{ik}, Q_{ijkl} are uniquely given by their values at rational time points only, which are countably many, so the whole processes and surfaces are almost surely unique.

In the following the superscript '$\pm$' is a short notation for

$$f(u^{\pm}) := \begin{cases} f(u^-) & : u > s, \\ f(u) & : u \leq s. \end{cases}$$

Moreover, for the sake of brief formulas, we introduce the processes

$$\tilde{P}_{ij}(u) := \begin{cases} P_i(u) & : u > s, \\ P_j(u) & : u \leq s, \end{cases}$$

and

$$\tilde{P}_{ijkl}(u_1, u_2) := \begin{cases} P_{ik}(u_1, u_2) & : u_1, u_2 > s, \\ P_{jk}(u_1, u_2) & : u_1 \leq s, u_2 > s, \\ P_{il}(u_1, u_2) & : u_1 > s, u_2 \leq s, \\ P_{jl}(u_1, u_2) & : u_1, u_2 \leq s. \end{cases}$$

Definition 2.3.1.
For $i, j \in \mathcal{Z}$ let the stochastic process $(\Lambda_{ij}(t))_{t\geq 0}$ be defined by

$$\Lambda_{ij}(t) = \int\limits_{(s\wedge t, s\vee t]} \frac{\mathbb{1}_{\{\tilde{P}_{ij}(u^{\pm})>0\}}}{\tilde{P}_{ij}(u^{\pm})} Q_{ij}(\mathrm{d}u), \tag{2.3.1}$$

and for $i, j, k, l \in \mathcal{Z}$ let the random surface $(\Lambda_{ijkl}(t_1, t_2))_{t_1,t_2\geq 0}$ be defined by

$$\Lambda_{ijkl}(t_1, t_2) = \int\limits_{(s\wedge t_1, s\vee t_1]\times(s\wedge t_2, s\vee t_2]} \frac{\mathbb{1}_{\{\tilde{P}_{ijkl}(u_1^{\pm}, u_2^{\pm})>0\}}}{\tilde{P}_{ijkl}(u_1^{\pm}, u_2^{\pm})} Q_{ijkl}(\mathrm{d}u_1, \mathrm{d}u_2). \tag{2.3.2}$$

In case of $t_1, t_2 > s$ we denote $\Lambda_{ij}(\mathrm{d}t_1)$ as (one-dimensional) forward transition rate and $\Lambda_{ijkl}(\mathrm{d}t_1, \mathrm{d}t_2)$ as two-dimensional forward transition rate. In case of $t_1, t_2 \leq s$ we call them backward transition rates. In order to ensure existence of (2.3.1) and (2.3.2) we generally assume that

$$\begin{aligned} &\int\limits_{[0,t]} \frac{\mathbb{1}_{\{\tilde{P}_{ij}(t^{\pm})>0\}}}{\tilde{P}_{ij}(t^{\pm})} Q_{ij}(\mathrm{d}t) < \infty, \quad t \geq 0, \\ &\int\limits_{[0,t]^2} \frac{\mathbb{1}_{\{\tilde{P}_{ijkl}(t_1^{\pm}, t_2^{\pm})>0\}}}{\tilde{P}_{ijkl}(t_1^{\pm}, t_2^{\pm})} Q_{ijkl}(\mathrm{d}t_1, \mathrm{d}t_2) < \infty, \quad t \geq 0, \end{aligned} \tag{2.3.3}$$

almost surely for all $i, j, k, l \in \mathcal{Z}$. According to [Chr21] it holds that

$$P_i(t) = I_i(s) + \sum_{j\in\mathcal{Z}} \int\limits_{(s\wedge t, s\vee t]} P_j(u^{\pm}) \tilde{\Lambda}_{ji}(\mathrm{d}u), \quad t \geq 0, \tag{2.3.4}$$

for all $i \in \mathcal{Z}$, where

$$\tilde{\Lambda}_{ij}(u) := \begin{cases} \Lambda_{ij}(u) & : u > s, \\ \Lambda_{ji}(u) & : u \leq s. \end{cases}$$

This is a generalization of Kolmogorov's forward equation to non-Markov models.

Remark 2.3.2.
We introduce $u^{\pm}$ as it allows us to use forward calculation for $t > s$ and backward calculation for $t \leq s$. Consider the case $t > s$, where $u^{\pm} = u^{-}$. Then Definition 2.3.1 implies

$$P_i(t) = \sum_{\substack{j\in\mathcal{Z} \\ j\neq i}} P_j(t^-)\Delta\Lambda_{ji}(t) - P_i(t^-)\Big(\sum_{\substack{j\in\mathcal{Z} \\ j\neq i}} \Delta\Lambda_{ij}(t) - 1\Big),$$

when $\Delta\Lambda = \Lambda(t) - \Lambda(t^-)$ contains the jump-discontinuities of Λ. Which allows us to obtain $P_i(t)$, $i \in \mathcal{Z}$, from $P_i(t^-)$, $i \in \mathcal{Z}$. If we replaced Definition 2.3.1 by

$$\Lambda_{ij}(t) = \int\limits_{(s,t]} \frac{\mathbb{1}_{\{\tilde{P}_j(u)>0\}}}{\tilde{P}_j(u)} Q_{ij}(\mathrm{d}u), \quad t > s, \tag{2.3.5}$$

then we would get

$$P_i(t) = \frac{P_i(t^-) - \sum_{\substack{j\in\mathcal{Z}\\ j\neq i}} P_j(t)\Delta\Lambda_{ij}(t)}{1 - \sum_{\substack{j\in\mathcal{Z}\\ j\neq i}} \Delta\Lambda_{ji}(t)}, \quad t > s.$$

which fails as a forward formula in case that

$$\sum_{\substack{j\in\mathcal{Z}\\ j\neq i}} \Delta\Lambda_{ji}(t) = 1.$$

This case can occur for example when semi-Markov effects are modeled by state space expansions as in [Ams68]. If we replaced Definition 2.3.1 by

$$\Lambda_{ij}(t) = \int\limits_{(s,t]} \frac{\mathbb{1}_{\{\tilde{P}_i(u)>0\}}}{\tilde{P}_i(u)} Q_{ij}(\mathrm{d}u), \quad t < s,$$

then Theorem 6.3 in [Chr21] fails. In the case $t < s$ analogous problems occur. Our definition allows us to do backward calculation on $[0, s]$.

We now extend (2.3.4) from one to two dimensions.

Theorem 2.3.3.
The processes $(P_{ik})_{i,k\in\mathcal{Z}}$, $(\Lambda_{ijkl})_{i,j,k,l\in\mathcal{Z}}$,$(P_i)_{i\in\mathcal{Z}}$ and $(\Lambda_{ij})_{i,j\in\mathcal{Z}}$ almost surely satisfy the equation

$$\begin{aligned} P_{ik}(t_1,t_2) = I_i(s)I_k(s) &+ I_k(s) \int\limits_{(s\wedge t_1, s\vee t_1]} \sum_{j\in\mathcal{Z}} P_j(u^{\pm})\tilde{\Lambda}_{ji}(\mathrm{d}u) \\ &+ I_i(s) \int\limits_{(s\wedge t_2, s\vee t_2]} \sum_{l\in\mathcal{Z}} P_l(u^{\pm})\tilde{\Lambda}_{lk}(\mathrm{d}u) \\ &+ \int\limits_{(s\wedge t_1, s\vee t_1]\times(s\wedge t_2, s\vee t_2]} \sum_{j,l\in\mathcal{Z}} P_{jl}(u_1^{\pm}, u_2^{\pm})\tilde{\Lambda}_{jilk}(\mathrm{d}u_1, \mathrm{d}u_2) \end{aligned} \tag{2.3.6}$$

for $t_1, t_2 \geq 0$ and $i, k \in \mathcal{Z}$, where

$$\tilde{\Lambda}_{ijkl}(u_1,u_2) := \begin{cases} \Lambda_{ijkl}(u_1,u_2) & u_1,u_2 > s \\ \Lambda_{jikl}(u_1,u_2) & u_1 \leq s, u_2 > s \\ \Lambda_{ijlk}(u_1,u_2) & u_1 > s, u_2 \leq s \\ \Lambda_{jilk}(u_1,u_2) & u_1,u_2 \leq s. \end{cases}$$

Proof. We show the proof for the case $t_1, t_2 > s$ only. For the other cases the proof is similar. Let $\mathbb{P}_\omega(\cdot)$ be a regular version of the conditional distribution $\mathbb{P}(\cdot|\mathcal{G}_s)$ and let $\mathbb{E}_\omega[\cdot]$ be the Lebesgue-Stieltjes integral of the argument with respect to the integrator $\mathbb{P}_\omega$. By applying Campbell's theorem, see Section 2.A.1, and Fubini's theorem, we get for $(a,b],(c,d] \subset [s,\infty)$ and almost all $\omega \in \Omega$

$$\begin{aligned}
&\int_{(a,b]\times(c,d]} \mathbb{1}_{\{P_{ik}(u_1^-,u_2^-)(\omega)>0\}} Q_{ijkl}(\mathrm{d}(u_1,u_2))(\omega) \\
&\quad = \mathbb{E}_\omega\left[\int_{(a,b]\times(c,d]} \mathbb{1}_{\{P_{ik}(u_1^-,u_2^-)(\omega)>0\}} N_{ij}(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2)\right] \\
&\quad = \mathbb{E}_\omega\left[\int_{(a,b]\times(c,d]} I_i(u_1^-)I_k(u_2^-)N_{ij}(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2)\right] \\
&\quad = \int_{(a,b]\times(c,d]} Q_{ijkl}(\mathrm{d}(u_1,u_2))(\omega) \qquad (2.3.7)
\end{aligned}$$

since $P_{ik}(u_1^-,u_2^-)(\omega) = 0$ implies $I_i(u_1^-)I_k(u_2^-) = 0$ $\mathbb{P}_\omega(\cdot)$-almost surely and $I_i(u_1^-)N_{ij}(\mathrm{d}u_1) = N_{ij}(\mathrm{d}u_1)$. By applying (2.2.3), we get for $t_1, t_2 \geq s$ and $i,k \in \mathcal{Z}$

$$\begin{aligned}
P_{ik}(t_1,t_2) &= \mathbb{E}[I_i(t_1)I_k(t_2)|\mathcal{G}_s] \\
&= \mathbb{E}\Big[I_i(s)I_k(s)\Big|\mathcal{G}_s\Big] + \mathbb{E}\Big[I_k(s)\int_{(s,t_1]}\sum_{j\in\mathcal{Z}} N_{ji}(\mathrm{d}t)\Big|\mathcal{G}_s\Big] \\
&\quad + \mathbb{E}\Big[I_i(s)\int_{(s,t_2]}\sum_{l\in\mathcal{Z}} N_{lk}(\mathrm{d}t)\Big|\mathcal{G}_s\Big] \\
&\quad + \mathbb{E}\Big[\sum_{j,l\in\mathcal{Z}}\int_{(s,t_1]\times(s,t_2]} N_{ji}(\mathrm{d}t_1)N_{lk}(\mathrm{d}t_2)\Big|\mathcal{G}_s\Big].
\end{aligned}$$

By using the assumption (2.2.5), the definition of $(Q_{ijkl})_{ijkl}$, the definition of $(Q_{ij})_{i,j}$, Fubini's theorem, and Campbell's theorem, we can conclude that

$$\begin{aligned}
P_{ik}(t_1,t_2) = I_i(s)I_k(s) &+ I_k(s)\int_{(s,t_1]}\sum_{j\in\mathcal{Z}} Q_{ji}(\mathrm{d}u) + I_i(s)\int_{(s,t_2]}\sum_{l\in\mathcal{Z}} Q_{lk}(\mathrm{d}u) \\
&+ \sum_{j,l\in\mathcal{Z}}\int_{(s,t_1]\times(s,t_2]} Q_{jilk}(\mathrm{d}u_1,\mathrm{d}u_2)
\end{aligned}$$

almost surely. The assertion follows now from (2.3.2) and (2.3.7). □

Remark 2.3.4.
The integral equation (2.3.6) can be solved in different ways. For example one could use an

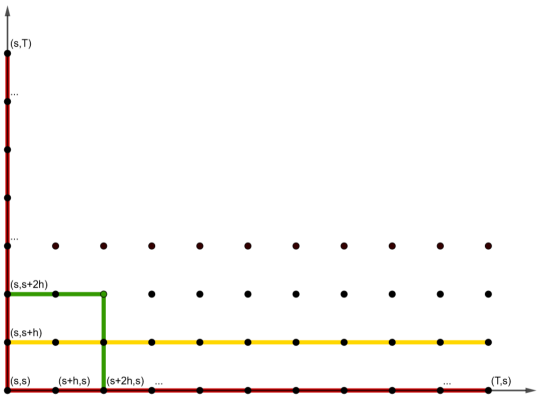


Figure 2.1: Sketch of an algorithm to calculate $P_{ik}(t_1, t_2)$.

algorithm similar to Finite Element methods. The idea is to use a simple box integration method to approximate the integral on small sub-rectangles of $(s,T] \times (s,T]$. Let $t_1, t_2, ...$ be a partition of $(s,T]$ we use the following approximation procedure:

$$\int\limits_{(t_i,t_{i+1}]\times(t_j,t_{j+1}]} f(x,y)G(\mathrm{d}x,\mathrm{d}y) \approx f(t_i,t_j) \int\limits_{(t_i,t_{i+1}]\times(t_j,t_{j+1}]} G(\mathrm{d}x,\mathrm{d}y).$$

This leaves us with the following formula:

$$\begin{aligned}
&P_{ik}(t_q,t_p)\\
&\approx I_i(s)I_k(s) + I_k(s)\sum_{j\in\mathcal{Z}}\sum_{n=0}^{q-1} P_j(t_n)\Lambda_{ji}((t_n,t_{n+1}]) + I_i(s)\sum_{l\in\mathcal{Z}}\sum_{\tilde{m}=0}^{p-1} P_l(t_m)\Lambda_{lk}((t_m,t_{m+1}])\\
&\quad + \sum_{j,l\in\mathcal{Z}}\sum_{m=0}^{p-1}\sum_{n=0}^{q-1} P_{jl}(t_n,t_m)\Lambda_{jilk}((t_n,t_{n+1}]\times(t_m,t_{m+1}]),
\end{aligned}$$

under the conventions $\Lambda_{ij}((a,b]) := \int_{(a,b]}\Lambda_{ij}(\mathrm{d}x)$, and $\Lambda_{ijkl}((a,b]\times(c,d]) := \int_{(a,b]\times(c,d]}\Lambda_{ijkl}(\mathrm{d}x,\mathrm{d}y)$. Combining this with an analogous procedure for the one-dimensional case, we obtain a recursion scheme.
We first calculate the points on the red lines in Figure 2.1 from the center to the edge. Then we calculate every point on the yellow line from left to right and proceed with the same notion for the points above. The reason this works is that the calculation of every point only depends on Points in the bottom left rectangle of the point, see the green rectangle for reference. This allows us to approximate P_{ik} on all the grid points, and then we approximate P_{ik} on the whole rectangle.
For an empirical Λ as a pure jump process, the approximation of the integral is exact, and our algorithm calculates P_{ij}, $i,j \in \mathcal{Z}$ accurately on the whole square.

2.4 Uniqueness of solutions of the integral equations

Equation (2.3.4) is commonly used for the calculation of the state occupation probabilities $(P_i)_i$ from given one-dimensional transition rates. Likewise, equation (2.3.6) may be used in order to calculate the two-dimensional state occupation probabilities $(P_{ik})_{ik}$ from the one-dimensional and two-dimensional transition rates, but it is crucial then that $(P_{ik})_{ik}$ is the only solution.

Theorem 2.4.1.
There exists an almost surely unique solution $(P_i)_i$, $(P_{ik})_{ik}$ to the stochastic integral equation system formed by equations (2.3.4) and (2.3.6).

Proof. The existence of a solution follows from Theorem 2.3.3, so it remains to show that the solution is almost surely unique. As the equations (2.3.4) and (2.3.6) are almost surely path-wise integral equations, in the remaining proof we identify without loss of generality all stochastic processes and random surfaces with just one of their paths. These paths have to be chosen such that (2.3.3) holds.

We show the proof for the case $t_1, t_2 > s$ only. For the other cases the proof is similar. We are going to use a fixed-point argument. For any choice of $T \in (s, \infty)$, the set

$$BV_2^{|\mathcal{Z}|} := \Big\{ f : [s,T] \times [s,T] \to \mathbb{R}^{|\mathcal{Z}|^2} \Big| \text{ there exist finite signed measures } \mu_{ik} \text{ with } f_{ik}(x,y) = \mu_{ik}([s,x] \times [s,y]), x, y \in [s,T] \times [s,T] \Big\}$$

is a linear space. The Hahn-Jordan decomposition offers for any finite signed measure a unique decomposition into a difference of two finite measures, and this decomposition can be furthermore used to decompose also any $f \in BV_2^{|\mathcal{Z}|}$ into a difference $f = f^+ - f^-$ of non-negative mappings $f^+, f^- \in BV_2^{|\mathcal{Z}|}$. Based on this unique construction of f^+ and f^-, we then define $|f| := f^+ + f^-$. By equipping $BV_2^{|\mathcal{Z}|}$ with the norm

$$\|(f_{ik})_{i,k\in\mathcal{Z}}\| := \sum_{i,k\in\mathcal{Z}} \int_{[s,T]\times[s,T]} |f_{ik}|(\mathrm{d}t_1, \mathrm{d}t_2) + \sum_{i,k\in\mathcal{Z}} |f_{ik}(s,s)|$$

we obtain a metric space. On this metric space, we define an operator $O : BV_2^{|\mathcal{Z}|} \to BV_2^{|\mathcal{Z}|}$ as follows:

$$(O((f_{jl})_{j,l\in\mathcal{Z}}))_{ik}(t_1,t_2) := \sum_{j,l\in\mathcal{Z}} \int_{(s,t_1]\times(s,t_2]} f_{jl}(u_1^-, u_2^-)\Lambda_{jilk}(\mathrm{d}u_1, \mathrm{d}u_2)$$

for $t_1, t_2 \in [s,T]$. We want this operator to be a contraction, but unfortunately this is not true, so we need to replace our norm for $BV_2^{|\mathcal{Z}|}$ by another equivalent norm. Let

$$\nu(t_1,t_2) := 4 \sum_{\substack{i,j,k,l\in\mathcal{Z} \\ i\neq j \\ k\neq l}} \Lambda_{ijkl}(t_1,t_2), \quad t_1, t_2 \in [s,T].$$

Because of assumption (2.3.3), each Λ_{ijkl}, $i \neq j$, $k \neq l$, has an associated measure $\mu_{\Lambda_{ijkl}}$ that is almost surely finite on $[s,T]^2$. Hence, there exists a finite measure μ that satisfies

$$\nu(t_1,t_2) = \mu([s,t_1] \times [s,t_2]), \quad t_1,t_2 \in [s,T]. \tag{2.4.1}$$

We moreover define

$$\begin{aligned} \nu_1(u) &:= \mu([s,u] \times [s,T]), \\ \nu_2(u) &:= \mu([s,T] \times [s,u]), \quad u \in [s,T], \end{aligned}$$

which are càdlàg by construction. For each $K \in (0,\infty)$ the mapping $\|\cdot\|_K$ defined by

$$\|(f_{ij})_{i,j\in\mathcal{Z}}\|_K := \sum_{i,j\in\mathcal{Z}} \int_{[s,T]\times[s,T]} e^{-K(\nu_1(u_1)+\nu_2(u_2))} |f_{ij}|(\mathrm{d}u_1,\mathrm{d}u_2) + \sum_{i,k\in\mathcal{Z}} |f_{ik}(s,s)|$$

is a norm on $BV_2^{|\mathcal{Z}|}$ that is equivalent to the norm $\|\cdot\|$. This construction of $\|\cdot\|_K$ is inspired by [Chr10] but is here extended to the two-dimensional case. For any $f \in BV_2^{|\mathcal{Z}|}$, $(a,b] \times (c,d] \subset [s,T] \times [s,T]$ and $k,i \in \mathcal{Z}$, the definitions of operator O and the two-dimensional transition rates $(\Lambda_{ijkl})_{i,j,k,l\in\mathcal{Z}}$ in conjunction with the triangle inequality yield that

$$\begin{aligned} \int_{(a,b]\times(c,d]} |O(f)_{ik}|(\mathrm{d}u_1,\mathrm{d}u_2) \leq \sum_{\substack{l:l\neq k \\ j:j\neq i}} \Bigg(& \int_{(a,b]\times(c,d]} |f_{jl}(u_1^-,u_2^-)| \Lambda_{jilk}(\mathrm{d}u_1,\mathrm{d}u_2) \\ & + \int_{(a,b]\times(c,d]} |f_{jk}(u_1^-,u_2^-)| \Lambda_{jikl}(\mathrm{d}u_1,\mathrm{d}u_2) \\ & + \int_{(a,b]\times(c,d]} |f_{il}(u_1^-,u_2^-)| \Lambda_{ijlk}(\mathrm{d}u_1,\mathrm{d}u_2) \\ & + \int_{(a,b]\times(c,d]} |f_{ik}(u_1^-,u_2^-)| \Lambda_{ijkl}(\mathrm{d}u_1,\mathrm{d}u_2) \Bigg). \end{aligned}$$

Summation over i,k and a reordering of some of the resulting sums lead to the inequality

$$\begin{aligned} \sum_{i,k} \int_{(a,b]\times(c,d]} |O(f)_{ik}|(\mathrm{d}u_1,\mathrm{d}u_2) &\leq 4 \sum_{\substack{i,j,k,l \\ i\neq j, k\neq l}} \int_{(a,b]\times(c,d]} |f_{ik}(u_1^-,u_2^-)| \Lambda_{ijkl}(\mathrm{d}u_1,\mathrm{d}u_2) \\ &\leq 4 \sum_{i,k} \int_{(a,b]\times(c,d]} |f_{ik}(u_1^-,u_2^-)| \, \nu(\mathrm{d}u_1,\mathrm{d}u_2). \end{aligned}$$

Suppose now that $f(s,s)$ is zero. Then it holds that

$$\sum_{i,k} \int_{(a,b]\times(c,d]} |O(f)_{ik}|(\mathrm{d}u_1,\mathrm{d}u_2) \leq 4 \sum_{k,i} \int_{(a,b]\times(c,d]} \int_{[s,u_1)\times[s,u_2)} |f_{ik}|(\mathrm{d}r_1,\mathrm{d}r_2) \, \nu(\mathrm{d}t_1,\mathrm{d}t_2).$$

As a consequence, the norm $\|\cdot\|_K$ of $O(f)$ has an upper bound of

$$\begin{aligned}\|O(f)\|_K &= \sum_{i,k} \int\limits_{[s,T]\times[s,T]} e^{-K(\nu_1(u_1)+\nu_2(u_2))}|O(f)_{ik}|(\mathrm{d}u_1,\mathrm{d}u_2) \\ &\leq \sum_{i,k} \int\limits_{[s,T]\times[s,T]} e^{-K(\nu_1(u_1)+\nu_2(u_2))} \int\limits_{[s,u_1)\times[s,u_2)} |f_{ik}|(\mathrm{d}r_1,\mathrm{d}r_2)\,\nu(\mathrm{d}u_1,\mathrm{d}u_2) \\ &= \sum_{i,k} \int\limits_{[s,T]\times[s,T]} \int\limits_{(r_1,T]\times(r_2,T]} e^{-K(\nu_1(u_1)+\nu_2(u_2))}\nu(\mathrm{d}u_1,\mathrm{d}u_2))|f_{ik}|(\mathrm{d}r_1,\mathrm{d}r_2),\end{aligned} \tag{2.4.2}$$

where the last equality uses Fubini's theorem. Moreover, for arbitrary but fixed $u_1, u_2 \in [s,T]$ let

$$\tilde{\nu}(r_1,r_2) := \frac{\mu((u_1,r_1]\times(u_2,r_2])}{\mu((u_1,T]\times(u_2,T])}, \quad r_1 \in (u_1,T],\, r_2 \in (u_2,T],$$

for the same μ is in equation (2.4.1). Without loss of generality we assume that $\mu((u_1,T]\times(u_2,T]) > 0$. Otherwise the conclusion that we want to draw is trivial. Then

$$\begin{aligned}\tilde{\nu}_1(r_1) &:= \tilde{\nu}(r_1,T), \\ \tilde{\nu}_2(r_2) &:= \tilde{\nu}(T,r_2), \quad (r_1,r_2) \in (u_1,T]\times(u_2,T],\end{aligned}$$

correspond to cumulative distribution functions. Let $C := \tilde{\nu}(u_1,u_2) > 0$ and let (A,B) be a random vector that has $\tilde{\nu}$ as its two-dimensional cumulative distribution function. Then, by applying Sklar's theorem, see e.g. Theorem 2.3.3 in [Nel06], we can show that

$$\begin{aligned}&\int\limits_{(u_1,T]\times(u_2,T]} e^{-K(\nu_1(r_1)+\nu_2(r_2))}\nu(\mathrm{d}r_1,\mathrm{d}r_2) \\ &\quad\leq Ce^{-K(\nu_1(u_1)+\nu_2(u_2)} \int\limits_{(u_1,T]\times(u_2,T]} e^{-CK(\tilde{\nu}_1(r_1)+\tilde{\nu}_2(r_2))}\tilde{\nu}(\mathrm{d}r_1,\mathrm{d}r_2) \\ &\quad= Ce^{-K(\nu_1(u_1)+\nu_2(u_2)}\,\mathbb{E}\left[e^{-CK(\tilde{\nu}_1(\tilde{\nu}_1^{-1}(U))+\tilde{\nu}_2(\tilde{\nu}_2^{-1}(V)))}\right]\end{aligned} \tag{2.4.3}$$

for a suitable random vector (U,V) whose components are uniformly distributed on $(0,1)$ and such that (A,B) and $(\tilde{\nu}_1^{-1}(U),\tilde{\nu}_2^{-1}(V))$ have the same distribution. Note here that the copula of (U,V) may be non-trivial. The inverse functions $\tilde{\nu}_1^{-1}$ and $\tilde{\nu}_2^{-1}$ are here defined as $\tilde{\nu}_n^{-1}(t) := \inf\{x : \tilde{\nu}_n(x) \geq t\}$, $n = 1,2$ for $t \in (0,1)$. Since $\tilde{\nu}_1(\tilde{\nu}_1^{-1}(t)) \geq t$ for $t \in (0,1)$, see e.g. Theorem 3.1 in [Sho00], the inequality (2.4.3) has an upper bound of

$$\int\limits_{(u_1,T]\times(u_2,T]} e^{-K(\nu_1(r_1)+\nu_2(r_2))}\nu(\mathrm{d}r_1,\mathrm{d}r_2) \leq Ce^{-K(\nu_1(u_1)+\nu_2(u_2)}\,\mathbb{E}\left[e^{-CK(U+V)}\right].$$

By applying Theorem 10.6.4 in [KGJD02], we can show that the latter expectation has an upper bound of

$$\mathbb{E}\left[e^{-CK(U+V)}\right] \leq \mathbb{E}\left[e^{-CK(2U)}\right]$$

since $x \mapsto \exp\{-CKx\}$ is a convex function. As U is uniformly distributed on $(0,1)$, we moreover have

$$\begin{aligned}\mathbb{E}\left[e^{-CK(2U)}\right] &= \int\limits_{(0,1)} e^{-CK(2u)}\mathrm{d}u \\ &= \frac{1}{2KC}\left(1-e^{-2KC}\right) \\ &\leq \frac{1}{2KC}.\end{aligned}$$

We set $K=1$. Then, all in all we can conclude that the inequality (2.4.2) has an upper bound of

$$\begin{aligned}\|O(f)\|_K &\leq \frac{1}{2K}\sum_{i,k}\int\limits_{[s,T]\times[s,T]} e^{-K(\nu_1(u_1)+\nu_2(u_2))}|f_{ik}|(\mathrm{d}u_1,\mathrm{d}u_2) \\ &= \frac{1}{2}\|f\|_K\end{aligned}$$

whenever $f(s,s)$ equals zero. Suppose now that we have two solutions $R=(P_{ik})_{ik}$ and $R'=(P'_{ik})_{ik}$ of (2.3.6) for given $(P_i)_i$. Then $R-R' \in BV_2^{|\mathcal{Z}|}$ is a fixed point of the operator O and $(R-R')(s,s)$ is zero. So we have

$$\|R-R'\|_K \leq \frac{1}{2}\|R-R'\|_K,$$

which necessarily implies that $R-R'$ is zero on $[s,T]\times[s,T]$. In an analogous way it is possible to proof $R=R'$ also on the three rectangles $[0,s]^2$, $[0,s]\times(s,T]$ and $(s,T]\times[0,s]$. Since $T\in(s,\infty)$ was arbitrary, we can expand the uniqueness property to infinity.

With the same ideas that we used in this proof, one can also show the uniqueness of a solution $(P_i)_{i\in\mathcal{Z}}$ of (2.3.4) with respect to $(\Lambda_{ij})_{i,j\in\mathcal{Z}}$. The one-dimensional case is actually even simpler. The equation system formed by equations (2.3.4) and (2.3.6) can then have only one solution $(P_i)_i$, $(P_{ik})_{ik}$. This completes the proof. □

2.5 Conditional expectations of canonical representations

According to [Chr21], for any stochastic process A that has a one-dimensional canonical representation of the form (2.2.8), it holds that

$$\mathbb{E}\Big[\int\limits_{(s,T]} A(\mathrm{d}t)\Big|\mathcal{G}_s\Big] = \sum_{i\in\mathcal{Z}}\int\limits_{(s,T]} P_i(t^-)A_i(\mathrm{d}t) + \sum_{\substack{i,j\in\mathcal{Z}\\ i\neq j}}\int\limits_{(s,T]} a_{ij}(t)P_i(t^-)\Lambda_{ij}(\mathrm{d}t) \tag{2.5.1}$$

and

$$\mathbb{E}\Big[\int\limits_{[0,s]} A(\mathrm{d}t)\Big|\mathcal{G}_s\Big] = \sum_{i\in\mathcal{Z}}\int\limits_{[0,s]} P_i(t^-)A_i(\mathrm{d}t) + \sum_{\substack{i,j\in\mathcal{Z}\\ i\neq j}}\int\limits_{[0,s]} a_{ij}(t)P_j(t)\Lambda_{ij}(\mathrm{d}t) \tag{2.5.2}$$

almost surely. These formulas are classical in the life insurance literature in case that Z is a Markov process, and the recent contribution of [Chr21] was to show that Markov assumptions are actually not needed here. The formulas (2.5.1) and (2.5.2) are used in life insurance for the calculation of prospective and retrospective reserves at time s. They are typically applied as follows: For given one-dimensional transition rates, calculate the corresponding one-dimensional state occupation probabilities by solving (2.3.4), and then solve the integrals in (2.5.1) and (2.5.2) in order to obtain the desired conditional expectations. The solving usually happens numerically.

The following three propositions will allow us to generalize (2.5.1) and (2.5.2) to stochastic processes A that have a two-dimensional canonical representation according to (2.2.9).

Proposition 2.5.1.
Let A_{ij}, $i, j \in \mathcal{Z}$, be a real-valued function on $[0,\infty)^2$ that generates a finite signed measure. Then we almost surely have

$$\mathbb{E}\Big[\int\limits_{[0,T]^2} I_i(u_1^-)I_j(u_2^-)A_{ij}(\mathrm{d}u_1,\mathrm{d}u_2)\Big|\mathcal{G}_s\Big] = \int\limits_{[0,T]^2} P_{ij}(u_1^-,u_2^-)A_{ij}(\mathrm{d}u_1,\mathrm{d}u_2).$$

Proof. The assertion follows from Fubini's theorem. □

Proposition 2.5.2.
Let A_i, $i \in \mathcal{Z}$, be a real-valued function on $[0,\infty)$ that generates a finite signed measure, and let a_{ikl}, $i,k,l \in \mathcal{Z}$, $k \neq l$, be a measurable and bounded real-valued function on $[0,\infty)^2$. Then we almost surely have

$$\begin{aligned}
&\mathbb{E}\Big[\int\limits_{[0,s]\times[0,T]} I_i(u_1^-)a_{ikl}(u_1,u_2)A_i(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2)\Big|\mathcal{G}_s\Big]\\
&= I_i(s)\int\limits_{[0,s]\times[0,T]} a_{ikl}(u_1,u_2)A_i(\mathrm{d}u_1)\tilde{P}_{kl}(u_2^{\pm})\Lambda_{kl}(\mathrm{d}u_2)\\
&\quad+\sum_{j\in\mathcal{Z}}\int\limits_{[0,s]}\int\limits_{[0,T]\times[u_1,s]} a_{ikl}(u_1,u_2)\tilde{P}_{klij}(u_2^{\pm},u_3)\Lambda_{klij}(\mathrm{d}u_2,\mathrm{d}u_3)A_i(\mathrm{d}u_1)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}\Big[\int\limits_{(s,T]\times[0,T]} I_i(u_1^-)a_{ikl}(u_1,u_2)A_i(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2)\Big|\mathcal{G}_s\Big]\\
&= I_i(s)\int\limits_{(s,T]\times[0,T]} a_{ikl}(u_1,u_2)A_i(\mathrm{d}u_1)\tilde{P}_{kl}(u_2^{\pm})\Lambda_{kl}(\mathrm{d}u_2)\\
&\quad+\sum_{j\in\mathcal{Z}}\int\limits_{(s,T]}\int\limits_{[0,T]\times(s,u_1)} a_{ikl}(u_1,u_2)\tilde{P}_{klji}(u_2^{\pm},u_3^-)\Lambda_{klji}(\mathrm{d}u_2,\mathrm{d}u_3)A_i(\mathrm{d}u_1).
\end{aligned}$$

Proof. From (2.2.3), (2.2.4) we get

$$I_i(t^-) = I_i(s) + \int\limits_{W(t)} \sum_{j\in\mathcal{Z}} N_{ji}^{W(t)}(du)$$

for $W(t) := (s,t)$ in case of $t > s$ and $W(t) := [t,s]$ in case of $t \le s$ and $N_{ij}^{[t,s]}(t) := N_{ji}(t)$ and $N_{ij}^{(s,t)}(t) := N_{ij}(t)$. This equation and assumption (2.2.5) imply for $D \in \{[0,s],(s,T]\}$ the equation

$$\begin{aligned}
&\mathbb{E}\Big[\int\limits_{D\times[0,T]} I_i(u_1^-)a_{ikl}(u_1,u_2)A_i(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2)\Big|\mathcal{G}_s\Big]\\
&= \mathbb{E}\Big[\int\limits_{D\times[0,T]} I_i(s)a_{ikl}(u_1,u_2)A_i(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2)\Big|\mathcal{G}_s\Big]\\
&\quad + \sum_{j\in\mathcal{Z}} \mathbb{E}\Big[\int\limits_D \int\limits_{[0,T]\times W(u_1)} a_{ikl}(u_1,u_2)N_{kl}(\mathrm{d}u_2)N_{ji}^{W(u_1)}(\mathrm{d}u_3)A_i(\mathrm{d}u_1)\Big|\mathcal{G}_s\Big].
\end{aligned}$$

Now apply Fubini's theorem and Campbell's theorem in order to obtain

$$\begin{aligned}
&\mathbb{E}\Big[\int\limits_{D\times[0,T]} I_i(u_1^-)a_{ikl}(u_1,u_2)A_i(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2)\Big|\mathcal{G}_s\Big]\\
&= \int\limits_{D\times[0,T]} I_i(s)a_{ikl}(u_1,u_2)A_i(\mathrm{d}u_1)Q_{kl}(\mathrm{d}u_2)\\
&\quad + \int\limits_D \int\limits_{[0,T]\times W(u_1)} a_{ikl}(u_1,u_2)Q_{klji}^{W(u_1)}(\mathrm{d}u_2,\mathrm{d}u_3)A_i(\mathrm{d}u_1),
\end{aligned}$$

for $Q_{klji}^{[0,s]}(t_1,t_2) = Q_{klij}(t_1,t_2)$ and $Q_{klji}^{(s,T]}(t_1,t_2) = Q_{klji}(t_1,t_2)$. Finally, use the equations

$$\begin{aligned}
Q_{kl}(\mathrm{d}u_2) &= \tilde{P}_{kl}(u_2^{\pm})\Lambda_{kl}(\mathrm{d}u_2),\\
Q_{klij}(\mathrm{d}u_2,\mathrm{d}u_3) &= \tilde{P}_{klij}(u_2^{\pm},u_3^{\pm})\Lambda_{klij}(\mathrm{d}u_2,\mathrm{d}u_3),
\end{aligned}$$

which follow from the definitions (2.3.1) and (2.3.2). □

Proposition 2.5.3.
Let a_{ijkl}, $i,j,k,l \in \mathcal{Z}$, $i \neq j$, $k \neq l$, be a measurable and bounded real-valued function on $[0,\infty)^2$. Then we almost surely have

$$\begin{aligned}
&\mathbb{E}\Big[\int\limits_{[0,T]^2} a_{ijkl}(u_1,u_2)N_{ij}(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2)\Big|\mathcal{G}_s\Big]\\
&= \int\limits_{[0,T]^2} a_{ijkl}(u_1,u_2)\tilde{P}_{ijkl}(u_1^{\pm},u_2^{\pm})\Lambda_{ijkl}(\mathrm{d}u_1,\mathrm{d}u_2).
\end{aligned}$$

Proof. The result follows directly from Campbell's theorem and (2.3.2). □

Corollary 2.5.4.
Suppose that A is a stochastic process that has a two-dimensional canonical representation according to (2.2.9). Then we almost surely have

$$\begin{aligned}
&\mathbb{E}\Big[\int_{[0,T]^2} A(\mathrm{d}u_1,\mathrm{d}u_2)\Big|\mathcal{G}_s\Big] \\
&= \sum_{i,j\in\mathcal{Z}} \int_{[0,T]^2} P_{ij}(u_1^-,u_2^-)A_{ij}(\mathrm{d}u_1,\mathrm{d}u_2) \\
&\quad + \sum_{\substack{i,k,l\\k\neq l}} I_i(s) \int_{[0,T]^2} a_{ikl}(u_1,u_2)A_i(\mathrm{d}u_1)\tilde{P}_{kl}(u_2^{\pm})\Lambda_{kl}(\mathrm{d}u_2) \\
&\quad + \sum_{\substack{i,j,k,l\\k\neq l}} \int_{[0,s]} \int_{[0,T]\times[u_1,s]} a_{ikl}(u_1,u_2)\tilde{P}_{klij}(u_2^{\pm},u_3)\Lambda_{klij}(\mathrm{d}u_2,\mathrm{d}u_3)A_i(\mathrm{d}u_1) \\
&\quad + \sum_{\substack{i,j,k,l\\k\neq l}} \int_{(s,T]} \int_{[0,T]\times(s,u_1)} a_{ikl}(u_1,u_2)\tilde{P}_{klji}(u_2^{\pm},u_3^-)\Lambda_{klji}(\mathrm{d}u_2,\mathrm{d}u_3)A_i(\mathrm{d}u_1) \\
&\quad + \sum_{\substack{i,j,k,l\\i\neq j,k\neq l}} \int_{[0,T]^2} a_{ijkl}(u_1,u_2)\tilde{P}_{ijkl}(u_1^{\pm},u_2^{\pm})\Lambda_{ijkl}(\mathrm{d}u_1,\mathrm{d}u_2).
\end{aligned} \tag{2.5.3}$$

From formula 2.5.3 we can immediately obtain general expressions for the expected values of discounted future liabilities as well as the pure expected values. The discounting needs to be included in the functions A_{ij}, A_i, a_{ikl}, and a_{ijkl}.

In insurance practice, formula 2.5.3 may be used as follows: For given one-dimensional and two-dimensional transition rates, calculate the corresponding one-dimensional and two-dimensional state occupation probabilities by solving (2.3.4) and (2.3.6). Then solve the integrals in (2.5.3) in order to obtain the desired conditional expectation.

2.6 Conditional variance of the future liabilities

We still think of s as an arbitrary but fixed parameter and omit this parameter in the notation. The discounted accumulated future payments Y^+ according to definition (2.2.6) are usually not measurable with respect to $\mathcal{G}_s$, so actuaries use the so-called *prospective reserve* at time s

$$V^+ := \mathbb{E}[Y^+|\mathcal{G}_s] \tag{2.6.1}$$

as a proxy for Y^+. In order to quantify the dispersion of the approximation error $Y^+ - V^+$, the actuary is also interested in the conditional variance

$$\mathrm{Var}[Y^+|\mathcal{G}_s] := \mathbb{E}[(Y^+)^2|\mathcal{G}_s] - (\mathbb{E}[Y^+|\mathcal{G}_s])^2. \tag{2.6.2}$$

This section discusses the calculation of V^+ and

$$S^+ := \mathbb{E}[(Y^+)^2|\mathcal{G}_s], \tag{2.6.3}$$

from which we can then get the conditional variance as

$$\mathrm{Var}[Y^+|\mathcal{G}_s] = S^+ - (V^+)^2.$$

Our results here are limited to insurance cash-flows B that have a one-dimensional canonical representation,

$$B(t) = \sum_i \int_{[0,t]} I_i(u^-)B_i(\mathrm{d}u) + \sum_{i,j:j\neq i} \int_{[0,t]} b_{ij}(u)N_{ij}(\mathrm{d}u), \quad t \geq 0. \tag{2.6.4}$$

From (2.2.6) and (2.5.1) we almost surely obtain

$$V^+ = \sum_{i\in\mathcal{Z}} \int_{(s,T]} \frac{\kappa(s)}{\kappa(t)} P_i(t^-)B_i(\mathrm{d}t) + \sum_{\substack{i,j\in\mathcal{Z}\\ i\neq j}} \int_{(s,T]} \frac{\kappa(s)}{\kappa(t)} b_{ij}(t)P_i(t^-)\Lambda_{ij}(\mathrm{d}t).$$

This formula and equation (2.3.4) allow us to calculate V^+ from given one-dimensional transition rates $(\Lambda_{ij}(t))_{t>s}$, $i, j \in \mathcal{Z}$.

We now turn to the calculation of S^+. Analogously to (2.2.10), one can show that

$$\begin{aligned}(Y^+)^2 = &\sum_{i,j\in\mathcal{Z}} \int_{(s,T]^2} \frac{2\kappa(s)^2}{\kappa(u_1)\kappa(u_2)} I_i(u_1^-)I_j(u_2^-)B_i(\mathrm{d}u_1)B_j(\mathrm{d}u_2)\\ &+ \sum_{\substack{i,j,k\in\mathcal{Z}\\ j\neq k}} \int_{(s,T]^2} \frac{2\kappa(s)^2}{\kappa(u_1)\kappa(u_2)} I_i(u_1^-)b_{jk}(u_2)B_i(\mathrm{d}u_1)N_{jk}(\mathrm{d}u_2)\\ &+ \sum_{\substack{i,j,k,l\in\mathcal{Z}\\ i\neq j,k\neq l}} \int_{(s,T]^2} \frac{2\kappa(s)^2}{\kappa(u_1)\kappa(u_2)} b_{ij}(u_1)b_{kl}(u_2)N_{ij}(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2).\end{aligned}$$

By using definition (2.6.3), interchanging $\int_{(s,T]^2}$ and $\int_{[0,T]^2} \mathbb{1}_{(s,T]^2}$, and applying Corollary 2.5.4, we obtain that

$$\begin{aligned}S^+ = &\sum_{i,j\in\mathcal{Z}} \int_{(s,T]^2} \frac{2\kappa(s)^2}{\kappa(u_1)\kappa(u_2)} P_{ij}(u_1^-, u_2^-)B_i(\mathrm{d}u_1)B_j(\mathrm{d}u_2)\\ &+ \sum_{\substack{i,k,l\in\mathcal{Z}\\ k\neq l}} I_i(s) \int_{(s,T]^2} \frac{2\kappa(s)^2}{\kappa(u_1)\kappa(u_2)} b_{kl}(u_2)P_k(u_2^-)B_i(\mathrm{d}u_1)\Lambda_{kl}(\mathrm{d}u_2)\end{aligned}$$

$$+\sum_{\substack{i,j,k,l\in\mathcal{Z}\\k\neq l}}\int_{(s,T]}\int_{(s,T]\times(s,u_1)}\frac{2\kappa(s)^2}{\kappa(u_1)\kappa(u_2)}b_{kl}(u_2)P_{kj}(u_2^-,u_3^-)\Lambda_{klji}(\mathrm{d}u_2,\mathrm{d}u_3)B_i(\mathrm{d}u_1)$$

$$+\sum_{\substack{i,j,k,l\in\mathcal{Z}\\i\neq j,k\neq l}}\int_{(s,T]^2}\frac{2\kappa(s)^2}{\kappa(u_1)\kappa(u_2)}b_{ij}(u_1)b_{kl}(u_2)P_{ik}(u_1^-,u_2^-)\Lambda_{ijkl}(\mathrm{d}u_1,\mathrm{d}u_2)$$

almost surely. This formula and the equations (2.3.4) and (2.3.6) allow us to calculate S^+ from given transition rates $(\Lambda_{ij}(t))_{t\in(s,T]}$ and $(\Lambda_{ijkl}(t))_{t\in(s,T]^2}$, $i,j,k,l\in\mathcal{Z}$.

For the conditional expectation and the conditional variance of Y^- we can obtain a similar result. We leave these analogous calculations to the reader.

2.7 Prospective and retrospective reserve for a path-dependent cash-flow

This section continues with Example 2.2.4. Recall that (2.6.1) is the so-called prospective reserve at time s. The *retrospective reserve* at time s is defined as

$$V^-:=\mathbb{E}[Y^-|\mathcal{G}_s] \tag{2.7.1}$$

for Y^- defined by (2.2.7). Analogously to (2.2.13) one can show that

$$\begin{aligned}Y^-=&\sum_{i\in\mathcal{S}_0}\int_{[0,s]}\frac{\kappa(s)}{\kappa(u)}I_i(u^-)C_i(\mathrm{d}u)+\sum_{\substack{k,l\in\mathcal{S}_0\\k\neq l}}\int_{[0,s]}\frac{\kappa(s)}{\kappa(u)}c_{kl}(u)N_{kl}(\mathrm{d}u)\\&+\sum_{k\in\mathcal{S}_0}\sum_{i,l\in\mathcal{S}_1}\int_{[0,s]^2}\frac{\kappa(s)}{\kappa(u_1)}I_i(u_1^-)\rho(u_2,k,l)C_i(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2)\\&+\sum_{k\in\mathcal{S}_0}\sum_{\substack{i,j,l\in\mathcal{S}_1\\i\neq j}}\int_{[0,s]^2}\frac{\kappa(s)}{\kappa(u_1)}\rho(u_2,k,l)c_{ij}(u_1)N_{ij}(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2).\end{aligned}$$

By using (2.5.2), interchanging $\int_{[0,s]}$ and $\int_{[0,T]}\mathbb{1}_{[0,s]}$, interchanging $\int_{[0,s]^2}$ and $\int_{[0,T]^2}\mathbb{1}_{[0,s]^2}$, and applying Corollary 2.5.4, we can conclude that

$$\begin{aligned}V^-=&\sum_{i\in\mathcal{S}_0}\int_{[0,s]}\frac{\kappa(s)}{\kappa(u)}P_i(u^-)C_i(\mathrm{d}u)+\sum_{\substack{k,l\in\mathcal{S}_0\\k\neq l}}\int_{[0,s]}\frac{\kappa(s)}{\kappa(u)}c_{kl}(u)P_l(u)\Lambda_{kl}(\mathrm{d}u)\\&+\sum_{k\in\mathcal{S}_0}\sum_{l,i\in\mathcal{S}_1}I_i(s)\int_{[0,s]^2}\frac{\kappa(s)}{\kappa(u_1)}\rho(u_2,k,l)P_l(u_2)C_i(\mathrm{d}u_1)\Lambda_{kl}(\mathrm{d}u_2)\\&+\sum_{k\in\mathcal{S}_0}\sum_{l,i\in\mathcal{S}_1}\sum_{j\in\mathcal{Z}}\int_{[0,s]}\int_{[0,s]\times[u_1,s]}\frac{\kappa(s)}{\kappa(u_1)}\rho(u_2,k,l)P_{lj}(u_2,u_3)\Lambda_{klij}(\mathrm{d}u_2,\mathrm{d}u_3)C_i(\mathrm{d}u_1)\end{aligned}$$

$$+\sum_{k\in\mathcal{S}_0}\sum_{\substack{l,i,j\in\mathcal{S}_1\\ i\neq j}}\int_{[0,s]^2}\frac{\kappa(s)}{\kappa(u_1)}\rho(u_2,k,l)c_{ij}(u_1)P_{jl}(u_1,u_2)\Lambda_{ijkl}(\mathrm{d}u_1,\mathrm{d}u_2)$$

almost surely. This formula and the equations (2.3.4) and (2.3.6) allow us to calculate V^- from given transition rates $(\Lambda_{ij}(t))_{t\leq s}$ and $(\Lambda_{ijkl}(t))_{t\leq s}$, $i,j,k,l\in\mathcal{Z}$.

By arguing analogously to (2.2.13), one can moreover show that the discounted future liabilities Y^+ have the representation

$$\begin{aligned}Y^+ &= \sum_{i\in\mathcal{S}_0}\int_{(s,T]}\frac{\kappa(s)}{\kappa(u)}I_i(u^-)C_i(\mathrm{d}u)+\sum_{\substack{k,l\in\mathcal{S}_0\\ k\neq l}}\int_{(s,T]}\frac{\kappa(s)}{\kappa(u)}c_{kl}(u)N_{kl}(\mathrm{d}u)\\ &\quad+\sum_{k\in\mathcal{S}_0}\sum_{i,l\in\mathcal{S}_1}\int_{(s,T]\times[0,T]}\frac{\kappa(s)}{\kappa(u_1)}I_i(u_1^-)\rho(u_2,k,l)C_i(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2)\\ &\quad+\sum_{k\in\mathcal{S}_0}\sum_{\substack{i,j,l\in\mathcal{S}_1\\ i\neq j}}\int_{(s,T]\times[0,T]}\frac{\kappa(s)}{\kappa(u_1)}\rho(u_2,k,l)c_{ij}(u_1)N_{ij}(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2).\end{aligned}$$

By applying (2.5.1), interchanging $\int_{(s,T]}$ and $\int_{[0,T]}\mathbb{1}_{(s,T]}$, interchanging $\int_{(s,T]\times[0,T]}$ and $\int_{[0,T]^2}\mathbb{1}_{(s,T]\times[0,T]}$, and applying Corollary 2.5.4, we almost surely get

$$\begin{aligned}V^+ &= \sum_{i\in\mathcal{S}_0}\int_{(s,T]}\frac{\kappa(s)}{\kappa(u)}P_i(u^-)C_i(\mathrm{d}u)+\sum_{\substack{i,j\in\mathcal{S}_0\\ i\neq j}}\int_{(s,T]}\frac{\kappa(s)}{\kappa(u)}c_{ij}(u)P_i(u^-)\Lambda_{ij}(\mathrm{d}u)\\ &\quad+\sum_{k\in\mathcal{S}_0}\sum_{l,i\in\mathcal{S}_1}I_i(s)\int_{(s,T]\times[0,T]}\frac{\kappa(s)}{\kappa(u_1)}\rho(u_2,k,l)\tilde{P}_{kl}(u_2^{\pm})C_i(\mathrm{d}u_1)\Lambda_{kl}(\mathrm{d}u_2)\\ &\quad+\sum_{k\in\mathcal{S}_0}\sum_{l,i\in\mathcal{S}_1}\sum_{j\in\mathcal{Z}}\int_{(s,T]}\int_{[0,T]\times(s,u_1)}\frac{\kappa(s)}{\kappa(u_1)}\rho(u_2,l,k)\tilde{P}_{klji}(u_2^{\pm},u_3^-)\Lambda_{klji}(\mathrm{d}u_2,\mathrm{d}u_3)C_i(\mathrm{d}u_1)\\ &\quad+\sum_{k\in\mathcal{S}_0}\sum_{\substack{l,i,j\in\mathcal{S}_1\\ i\neq j}}\int_{(s,T]\times[0,T]}\frac{\kappa(s)}{\kappa(u_1)}\rho(u_2,k,l)c_{ij}(u_1)\tilde{P}_{ijkl}(u_1^-,u_2^{\pm})\Lambda_{ijkl}(\mathrm{d}u_1,\mathrm{d}u_2).\end{aligned}$$

This formula and the equations (2.3.4) and (2.3.6) allow us to calculate V^- from given transition rates Λ_{ij} and Λ_{ijkl}, $i,j,k,l\in\mathcal{Z}$.

2.8 Conclusion and outlook

So far, the non-Markov calculation technique of Christiansen [Chr21] has been limited to the calculation of first-order moments. By introducing two-dimensional forward and backward transition rates, we make second-order moments accessible as well. The two-dimensional rates capture intertemporal dependency structures that the one-dimensional rates miss. In order to calculate also third-order or even higher-order moments, one may envision further

extensions of the forward and backward transition rate concept to three dimensions or even higher dimensions. Such extensions are beyond the scope of this paper and left for future research.

Intertemporal dependency structures play a major role when life insurance cash-flows are path-dependent. Such path dependencies occur for example upon contract modifications. We illustrated how two-dimensional forward and backward transition rates are of help here for actuarial calculations by studying insurance policies with free policy option.

The non-Markov approach exchanges systematic model risk for unsystematic estimation risk. The latter risk has much nicer asymptotic properties. For the one-dimensional case, the landmark Nelson-Aalen estimator of Putter and Spitoni [PS18] can be suitably adapted, see [Chr21]. The idea is to use sub-sampling, so in order to achieve a small estimation error the information model $\mathcal{G}_s$ should be rather small. A positive example is the as-if Markov model, which records only the current state of the insured. For the two-dimensional case, efficient generalizations of the landmark Nelson-Aalen estimator have yet to be explored.

Acknowledgments

We would like to thank the reviewers for insightful remarks that helped to improve the paper.

2.A Appendix

Theorem 2.A.1. *Campbell theorem*
Let η be a point process on $(\mathbb{R}^d, \mathbb{B}(\mathbb{R}^d))$ with intensity measure λ and let $u : \mathbb{R}^d \to \mathbb{R}$ be a measurable function. Then

$$\int u(x)\eta(\mathrm{d}x)$$

is a random variable and

$$\mathbb{E}\left[\int u(x)\eta(\mathrm{d}x)\right] = \int u(x)\lambda(\mathrm{d}x),$$

whenever $u \geq 0$ or $\int |u(x)|\lambda(\mathrm{d}x) < \infty$.

For the proof see section 2.2 in [LP17].

Proposition 2.A.2. *integration by parts*
Suppose that $(F(t))_{t\geq 0}$ and $(G(t))_{t\geq 0}$ are real-valued càdlàg processes with paths of finite variation. Then

$$\int\limits_{(a,b]} F(x)\mathrm{d}G(x) = F(b)G(b) - F(a)G(a) - \int\limits_{(a,b]} G(x^-)\mathrm{d}F(x)$$

for $(a, b] \subset (0, \infty)$.

Proof. For the proof see Corollary 2 following Theorem 22 in [Pro05]. □

Chapter 3

On the estimation of bivariate conditional transition rates

Abstract

Recent literature has found conditional transition rates to be a useful tool for avoiding Markov assumptions in multi-state models. While the estimation of univariate conditional transition rates has been extensively studied, the intertemporal dependencies captured in the bivariate conditional transition rates still require a consistent estimator. We provide an estimator that is suitable for censored data and emphasize the connection to the rich theory of the estimation of bivariate survival functions. Bivariate conditional transition rates are necessary for various applications in the survival context but especially in the calculation of moments in life insurance mathematics.

Keywords: bivariate survivor function; censoring; non-Markov models; non-parametric estimation; life & health insurance; free policy option

3.1 Introduction

In survival statistics, the Kaplan-Meier estimator, as seen in Kaplan and Meier [KM58], is widely used to obtain non-parametric estimation of survival functions for various modeling purposes when censoring is involved. This applies to various scenarios, such as life insurance modeling, time to machine failure, or the time to illness after the application of medicine. However, this estimator can only take into account a single event and therefore cannot capture dependencies between different events included in the data. Consequently, its suitability for analyzing data sets, particularly those arising from genetic epidemiological studies, for disease events of family members, or other similar dependencies, is limited. Many different estimators have been proposed to fill this gap. Dabrowska [Dab88] uses two-dimensional product integrals to estimate the bivariate survival function, whilst Prentice and Cai [PC92] employ Volterra integral equation techniques to connect the covariance function and the marginal survivor functions with the bivariate survivor function. Van Der Laan [Laa96] uses non-parametric maximum likelihood estimation, and recently, Prentice and Zhao [PZ18] extended the work of Dabrowska with a recursive formula for arbitrary

numbers of failure times. Gill [Gil93] gives a reason why there are so many different estimators: the fact that in two and higher dimensions there are different ways to get from one point in the two-dimensional plane to another point in the two-dimensional plane.
Another important concept in survival analysis is the notion of recurrent events used to study recurring episodes of infection or hospitalization. This is often modeled by introducing multi-state models, described by Hougaard [Hou00], which are mostly based on modeling with Markov assumptions. In these models, estimators often target cumulative intensity processes which provide insight into the expected number of upcoming transitions. These estimators are now known as the Nelson-Aalen estimator for transition rates and the Aalen-Johansen estimator of transition probabilities, see Aalen [Aal78] and Borgan [Bor05]. Datta and Satten [DS01] discovered that the Nelson-Aalen and Aalen-Johansen estimators remain consistent on data that does not follow the Markov assumption. The use case for multi-state models without the Markov assumption, referred to here as non-Markov models, has been showcased by Putter and Spitoni in [PS18]. Their simulation studies demonstrate that, in some instances, Markov estimation introduces a bias when the data does not follow the Markov assumption. In this case, there is a favorable bias-variance trade-off when using the landmark Nelson-Aalen and Aalen-Johansen estimators, which target transition rates and probabilities of a non-Markov model.
However, transition rates and transition probabilities do not hold full information about the distribution of these state processes without the Markov assumption, as they do not capture the temporal inter-dependencies between consecutive jump events, necessary, for example, for calculating variances or higher moments. To fill this gap, Bathke and Christiansen [BC24] have recently proposed a novel framework involving bivariate transition rates but without suitable non-parametric estimators. This paper aims to address the lack of estimation procedures for general bivariate intensity rates. We demonstrate the relationship to the estimation of bivariate survival functions, particularly the works of Dabrowska [Dab88] and Prentice and Cai [PC92]. Our aim is to construct a landmark estimator, demonstrate its consistency through almost sure uniform convergence, and show applicability. Uniform convergence is a widely accepted method for understanding the asymptotic properties of estimators, as demonstrated by Shorack and Wellner [SW09]. Asymptotic normality of the estimators is out of the scope of this paper, see the work of Bathke and Furrer [BF24a] for further work on that problem.
Bivariate rates have another important application apart from being used for classical survival analysis applications. They are also key for the calculation of life insurance premiums or reserves of insurance contracts with special payment functions in non-Markov models. This is important because life insurance calculation is generally done in Markov or semi-Markov models, which introduce systematic model risk. Insurance cash flows typically include transition payments from one state to another, as well as sojourn payments for remaining in the corresponding state. Transition payments can be modeled using jump processes, while sojourn payments can be modeled using indicator functions. Guibert and Planchet [GP18] were able to estimate reserves in a general non-Markov multi-state model for sojourn payments. Christiansen [Chr21] extended this to reserves of insurance cash flows with both transition and sojourn payments, using path-independent payment functions. He showed that the consistency properties of the estimators can be transferred to reserve estimation. Our contribution is to show that our bivariate estimators of transition rates and probabilities can be used to estimate reserves for specific path-dependent cash

flows and second moments of future liabilities, which are important for many types of risk analysis. This is illustrated by a numerical example, presenting the estimation technique. The paper is structured as follows. In Section 3.2 the general multi-state modeling framework is introduced, here we connect jump processes and indicator processes. In Section 3.3 we introduce our bivariate notation. Section 3.4 introduces the definition of bivariate conditional rates and the corresponding integral equations, which will be connected to bivariate Peano-series. Section 3.5 focuses on estimating the bivariate conditional rates using a bivariate landmark Nelson-Aalen estimator, as well as the bivariate transition probabilities using the bivariate landmark Aalen-Johansen estimator. In this Section, we introduce Theorem 3.5.10 which is the main contribution of this paper. Section 3.6 demonstrates how these estimators can be utilized to form estimators for cumulative expected values of general cash flows with inter-dependencies between two time points in the context of life insurance. Section 3.7 introduces a numerical example, where the set up we introduce in Section 3.6 is used to form estimators for expected future cash flows in the case of scaled payments in life insurance.

3.2 Multi-state modeling framework

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$. We consider an individual with the state-process

$$Z = (Z(t))_{t\geq 0},$$

which describes the current state that the individual is in. This process is modeled as an adapted càdlàg jump process on a finite state space $\mathcal{Z}$. When modeling these objects, we want to be able to describe past, present, and future developments. Thus, we assume that we are currently at time $s \geq 0$. So the time interval $[0, s]$ represents the past and the present, and the time interval (s, ∞) represents the future. This time point s is arbitrary but fixed throughout this paper. Based on Z we define additional processes, first the state indicator processes $(I_i)_{i\in\mathcal{Z}}$ by

$$I_i(t) := \mathbb{1}_{\{Z(t)=i\}}, \quad t \geq 0,$$

and second the transition counting processes $(N_{ij})_{i,j\in\mathcal{Z}:i\neq j}$ by

$$N_{ij}(t) := \#\{u \in (0, t] : Z(u^-) = i, Z(u) = j\}, \quad t \geq 0.$$

We generally assume that

$$\mathbb{E}[N_{ij}(t)^2] < \infty, \quad t \geq 0\ , i, j \in \mathcal{Z},\ i \neq j, \tag{3.2.1}$$

which in particular implies that Z has almost surely no explosions. Let

$$N_{ii}(t) := -\sum_{j:j\neq i} (N_{ij}(t) - N_{ij}(s)), \quad t > s \quad i \in \mathcal{Z}. \tag{3.2.2}$$

This construction is done so that we can use the abbreviation

$$\mathrm{d}N_{ii}(t) = -\sum_{j:j\neq i} \mathrm{d}N_{ij}(t), \quad t > s, \quad i \in \mathcal{Z}.$$

Definition (3.2.2) and many other following definitions are only defined for $t > s$. For retrospective modeling, this restriction can be removed if necessary, see Bathke and Christiansen [BC24] for these extensions of the basic theory. Since time s is fixed, we omit it in the notation, but one should keep in mind the dependence of many of our definitions on the parameter s. The following equation shows a useful direct link between the processes $(N_{ij})_{i,j\in\mathcal{Z}}$ and $(I_i)_{i\in\mathcal{Z}}$:

$$I_i(t) = I_i(s) + \sum_{j\in\mathcal{Z}} \int_{(s,t]} N_{ji}(\mathrm{d}u), \quad t \geq s, i \in \mathcal{Z}. \tag{3.2.3}$$

The latter integral and all following integrals in this paper are meant as path-wise Lebesgue-Stieltjes integrals, which in the case of (3.2.3) only exists almost surely.

The sigma algebra $\mathcal{F}_s$ represents the information available at time s. In practice, we often use a reduced information set $\mathcal{G}_s = \sigma(\xi)$ for evaluations, which is generated by a discrete random variable ξ. For general two-dimensional computations we assume

$$\sigma(Z(s)) \subseteq \mathcal{G}_s \subseteq \mathcal{F}_s. \tag{3.2.4}$$

This assumption is needed for calculating bivariate conditional transition rates and bivariate transition probabilities in the general non-Markov model, see Bathke and Christiansen [BC24].
The special case $\mathcal{G}_s = \sigma(Z(s))$ is known as the as-if Markov model, since we use only data that would have been used in the Markov model as well. The choice of $\mathcal{G}_s$ can be influenced by many factors. Some of these are listed below:

- Numerical complexity,
- Lack of data,
- Regulatory requirements.

Currently, Markov modeling is widely used in practice. Therefore, most of the available data is of Markov type. Thus, the natural extension would be the as-if Markov model, as it combines classical Markov information with a non-Markov model. In addition, the numerical complexity increases whenever we work with a broader information model. We therefore focus on an estimation procedure that conditions on discrete random variables, which includes the as-if Markov model because $\#\mathcal{Z} < \infty$.

3.3 Notation

For two-dimensional variables, definitions of processes and integrals, we use $\boldsymbol{t} = (t_1, t_2)$, $f(\boldsymbol{u}^-) = f(u_1^-, u_2^-)$ for a bivariate upper continuous function f that has uniform lower limits, and $f(\mathrm{d}\boldsymbol{u}) = f(\mathrm{d}u_1, \mathrm{d}u_2)$ for a two-variable function f with finite two-dimensional total variation, see the Appendix Definition 3.A.5 for the definition of the total variation used in this paper. In addition, we use partial ordering on $\mathbb{R}^2$, meaning that $\boldsymbol{x} \leq \boldsymbol{y}$ if and only if $x_1 \leq y_1$ and $x_2 \leq y_2$. We extend this notation so that $\boldsymbol{x}$ can be set into relation with a one-dimensional constant. For example, $\boldsymbol{x} \leq y$ iff $x_1 \leq y$ and $x_2 \leq y$. Consequently, $(s, \boldsymbol{t}]$ is a two-dimensional rectangle that is open at the bottom and the left boundaries and closed at the top and right boundaries with the corner points (s, s) and (t_1, t_2). Apart from two-dimensional variables, we use various indices to model different jumps at different times. For this we introduce the two-dimensional indices $\boldsymbol{i} = (i_1, i_2)$.

3.4 Conditional transition rates

This section introduces the conditional transition rates of Christiansen [Chr21] and Bathke and Christiansen [BC24]. We assume that we are currently at time s and have the information $\mathcal{G}_s = \sigma(\xi)$ available, where ξ is a discrete random variable called the landmark. Since the parameter s is fixed, we omit it in the notation

Let $P_{z,i} = (P_{z,i}(t))_{t>s}$ and $Q_{z,ij} = (Q_{z,ij}(t))_{t>s}$ for $i, j \in \mathcal{Z}$ be the almost surely unique càdlàg paths that satisfy

$$\begin{aligned} P_{z,i}(t) &= \mathbb{E}[I_i(t)|\xi = z], \\ Q_{z,ij}(t) &= \mathbb{E}[N_{ij}(t)|\xi = z]. \end{aligned}$$

Let $\boldsymbol{P}_{z,\boldsymbol{i}} = (\boldsymbol{P}_{z,\boldsymbol{i}}(\boldsymbol{t}))_{t_1,t_2>s}$ and $\boldsymbol{Q}_{z,\boldsymbol{ij}} = (\boldsymbol{Q}_{z,\boldsymbol{ij}}(\boldsymbol{t}))_{t_1,t_2>s}$ for $\boldsymbol{i}, \boldsymbol{j} \in \mathcal{Z}^2$ be the almost surely unique surfaces that are càdlàg in each variable and satisfy

$$\begin{aligned} \boldsymbol{P}_{z,\boldsymbol{i}}(\boldsymbol{t}) &= \mathbb{E}[I_{i_1}(t_1)I_{i_2}(t_2)|\xi = z], \\ \boldsymbol{Q}_{z,\boldsymbol{ij}}(\boldsymbol{t}) &= \mathbb{E}[N_{i_1j_1}(t_1)N_{i_2j_2}(t_2)|\xi = z], \end{aligned}$$

for all these definitions refer to a regular conditional probability $\mathbb{P}(\cdot|\xi = z)$. These functions are almost surely unique because the càdlàg property ensures a almost surely unique definition by their values at rational times, which are countably many.

Remark 3.4.1.
Following Section 3.2, $\xi = Z(s)$ is a natural example of a landmark, as seen for example in the as-if Markov model. In this case, $z \in \mathcal{Z}$. But the landmark can also include discrete external landmarks such as gender. So, in general, z need not to be in $\mathcal{Z}$. The fact that the landmark ξ is discrete is important for the landmarking ideas we use in this paper. See Bladt and Furrer [BF24b] for an approach to conditioning on continuous information.

Definition 3.4.2.
For $i, j \in \mathcal{Z}$ and $t \in (s, \infty)$ let the function $(\Lambda_{z,ij}(t))_{t \geq s}$ be defined by

$$\Lambda_{z,ij}(t) = \int\limits_{(s,t]} \frac{\mathbb{1}_{\{P_{z,i}(u^-)>0\}}}{P_{z,i}(u^-)} Q_{z,ij}(\mathrm{d}u), \tag{3.4.1}$$

and for $\boldsymbol{i}, \boldsymbol{j} \in \mathcal{Z}^2$ and $\boldsymbol{t} \in (s, \infty)^2$ let the function $(\boldsymbol{\Lambda}_{z,\boldsymbol{ij}}(\boldsymbol{t}))_{\boldsymbol{t} \geq s}$ be defined by

$$\boldsymbol{\Lambda}_{z,\boldsymbol{ij}}(\boldsymbol{t}) = \int\limits_{(s,\boldsymbol{t}]} \frac{\mathbb{1}_{\{\boldsymbol{P}_{z,\boldsymbol{i}}(\boldsymbol{u}^-)>0\}}}{\boldsymbol{P}_{z,\boldsymbol{i}}(\boldsymbol{u}^-)} \boldsymbol{Q}_{z,\boldsymbol{ij}}(\mathrm{d}\boldsymbol{u}), \tag{3.4.2}$$

where we use the convention $\frac{0}{0} := 0$.

In case of $t > s$ and $\boldsymbol{t} > s$ we denote $\Lambda_{z,ij}(\mathrm{d}t)$ as (univariate) conditional transition rate and $\Lambda_{z,\boldsymbol{ij}}(\mathrm{d}\boldsymbol{t})$ as bivariate conditional transition rate. In order to ensure existence of (3.4.1) and (3.4.2) we generally assume that

$$\begin{aligned} \Lambda_{z,ij}(t) &< \infty, \quad t \geq s, \\ \boldsymbol{\Lambda}_{z,\boldsymbol{ij}}(\boldsymbol{t}) &< \infty, \quad t \geq s, \end{aligned} \tag{3.4.3}$$

almost surely for all $i, j \in \mathcal{Z}, \boldsymbol{i}, \boldsymbol{j} \in \mathcal{Z}^2$. According to Christiansen [Chr21] it holds that

$$P_{z,i}(t) = P_{z,i}(s) + \sum_{j \in \mathcal{Z}} \int\limits_{(s,t]} P_{z,j}(u^-) \Lambda_{z,ji}(\mathrm{d}u), \quad t \geq s \tag{3.4.4}$$

for all $i \in \mathcal{Z}$. This is equivalent to the product integral

$$P_z(t) = P_z(s) \prod_{(s,t]} (Id + \Lambda_z(\mathrm{d}u)), \quad t > s.$$

This is a generalization of Kolmogorov's forward equation to non-Markov models. Bathke and Christiansen [BC24] extended this result to bivariate conditional transition rates with the integral equation

$$\begin{aligned} \boldsymbol{P}_{z,\boldsymbol{i}}(\boldsymbol{t}) = \boldsymbol{P}_{z,\boldsymbol{i}}(\boldsymbol{s}) &+ P_{z,i_2}(s) \left(P_{z,i_1}(t_1) - P_{z,i_1}(s)\right) + P_{z,i_1}(s) \left(P_{z,i_2}(t_2) - P_{z,i_2}(s)\right) \\ &+ \int\limits_{(s,\boldsymbol{t}]} \sum_{\boldsymbol{j} \in \mathcal{Z}^2} \boldsymbol{P}_{z,\boldsymbol{j}}(\boldsymbol{u}^-) \boldsymbol{\Lambda}_{z,\boldsymbol{ji}}(\mathrm{d}\boldsymbol{u}), \end{aligned} \tag{3.4.5}$$

for $\boldsymbol{t} \geq s$ and $\boldsymbol{i} = (i_1, i_2) \in \mathcal{Z}^2$.

These integral equations are similar to the inhomogeneous Volterra equations found in the survival setting, see Prentice and Cai [PC92].
It should be noted here that these processes depend on the current time s, which is omitted only for the sake of conciseness.

3.5 Bivariate conditional Nelson-Aalen and Aalen-Johansen estimation

We begin this section with the general non-parametric estimation setup, which is similar to the non-Markov setup in Christiansen [Chr21]. The non-parametric approach of this paper is deeply rooted in the fact that many practitioners, for example in the German actuarial community, use purely non-parametric estimators. Moreover, the non-parametric approach can be used as a preview to better calibrate parametric models. First, we present the general estimation setup.
For the rest of this paper we assume that we are observing $n \in \mathbb{N}$ individuals:

$$(Z^1(t))_{t\in[0,R^1]}, \dots, (Z^n(t))_{t\in[0,R^n]},$$

where R^m, $m \in \mathbb{N}_{\leq n}$ is the right censoring time of the data. The additional process N^m_{ij} and the random variable ξ^m define the counting process and the information of the m-th observed individual. For the following estimators and the convergence results of these estimators, we make some key assumptions about our data.

Assumption 3.5.1.

a) All observations $(Z^i, R^i)_{i\leq n}$ are independent and identically distributed.

b) R^1 is stochastically independent of $(Z^1(t))_{t\geq s}$ and ξ^1 and almost surely greater than s.

It means that we assume that ξ^m is not affected by censoring. That is, it is observable in the data at time s for all individuals.

Remark 3.5.2.
The estimation model could be defined exactly as in Christiansen [Chr21], but we think that left censoring and weaker conditions on right censoring are not worth the notational overhead. One main application of censoring or truncation in the actuarial context in Section 3.6 is that the portfolio of insured, and thus the data we use for estimation, is not fully closed, so we have data that may be right censored but is rarely left censored.

Remark 3.5.3.
There are many other approaches to the modelling of censoring than the one we present in this paper. There is the example of inverse probability of censoring weighting methods, which allow for additional dependency between the jump process and the censoring mechanism. This in turn requires a parametric censoring model, which distinguishes their approach from our fully non-parametric one, as already mentioned by Bladt and Furrer in [BF24b]. In addition, current work on IPCW methods in multi-state models uses a simpler setup, see the work of Mostajabi and Datta [MD13] and Siriwardhana, Kulasekera and Datta [SKD18].

We continue with a short introduction to the univariate landmark Nelson-Aalen and Aalen-Johansen estimators, see Aalen [Aal78], Christiansen [Chr21], and Bladt and Furrer [BF24b] for reference.
Let R be a random variable that describes the right censoring and ξ the random variable that generates the information $\mathcal{G}_s$. This allows us to define the expected value of the observable censored jump processes and indicator processes.

$$\begin{aligned} P^c_{z,j}(t) &:= \mathbb{E}[\mathbb{1}_{\{Z_t=j\}}\mathbb{1}_{\{t<R\}}\mathbb{1}_{\{\xi=z\}}], \\ Q^c_{z,jk}(t) &:= \mathbb{E}[N_{jk}(t\wedge R)\mathbb{1}_{\{\xi=z\}}], \end{aligned}$$

for $j,k \in \mathcal{Z}$ and $t > s$. This in turn allows us to define censored transition rates as

$$\Lambda^c_{z,jk}(t) := \int\limits_{(s,t]} \frac{\mathbb{1}_{\{P^c_{z,j}(u^-)>0\}}}{P^c_{z,j}(u^-)} Q^c_{z,jk}(\mathrm{d}u),$$

for $j,k \in \mathcal{Z}$ and $t > s$.
All of these processes depend on the index z which takes values in the image of ξ. For this approach to give comprehensible information, the random variable ξ must only take a finite number of values. This is the reason why this approach only works for a finite information space.
One of the reasons for using transition rates in Markov and non-Markov modeling is the following equality. If we take a time τ such that $\mathbb{P}(\tau \leq R) > 0$, we get

$$\Lambda_{z,jk}(t) = \Lambda^c_{z,jk}(t),$$

for $j,k \in \mathcal{Z}$ and $t \in (s,\tau]$, see for example the ideas used by Glidden [Gli02] or Christiansen [Chr21] in the proof of Theorem 7.3. This means that the censoring does not affect the conditional transition rates as long as $t < \tau$.
Estimating processes such as $\Lambda_{z,jk}(t)$ is generally challenging, since one might divide by zero. Bladt and Furrer [BF24b] use what they call a perturbation of $\Lambda_{z,jk}(t)$. Applying their idea onto the univariate conditional transition rates leaves us with

$$\begin{aligned} \Lambda^{(\epsilon)}_{z,jk}(t) &:= \int\limits_{(s,t]} \frac{1}{P^c_{z,j}(u^-)\vee\epsilon} Q^c_{z,jk}(\mathrm{d}u), \\ P^{(\epsilon)}_z(t) &:= P_z(s)\prod_{(s,t]}(Id + \Lambda^{(\epsilon)}_z(\mathrm{d}u)), \end{aligned}$$

for $j,k \in \mathcal{Z}$, $t > s$ and $\epsilon > 0$. The parameter ϵ will be used as the perturbation factor for the rest of this paper. The next step is to define the corresponding estimators for the univariate conditional transition rates and probabilities.

$$\begin{aligned} I^{(n)}_{z,j}(t) &:= \frac{1}{n}\sum_{m=1}^{n} \mathbb{1}_{\{\xi^m=z\}}\mathbb{1}_{\{t\leq R^m\}}\mathbb{1}_{\{Z^m_t=j\}}, \\ N^{(n)}_{z,jk}(t) &:= \frac{1}{n}\sum_{m=1}^{n} \mathbb{1}_{\{\xi^m=z\}}N^m_{jk}(t\wedge R^m) \end{aligned}$$

$$\Lambda^{(n,\epsilon)}_{z,jk}(t) := \int_{(s,t]} \frac{1}{I^{(n)}_{z,j}(u^-) \vee \epsilon} N^{(n)}_{z,jk}(\mathrm{d}u),$$
$$P^{(n,\epsilon)}_z(t) := P^{(n)}_z(s) \prod_{(s,t]} (Id + \Lambda^{(n,\epsilon)}_z(\mathrm{d}u)),$$

where $P^{(n)}_z(s) := \left(\frac{I^{(n)}_{z,j}(s)}{\sum_i I^{(n)}_{z,i}(s)} \right)_{j \in \mathcal{Z}}$, $j,k \in \mathcal{Z}$, and $\epsilon > 0$.
The next step is to use these ideas for the estimation of bivariate conditional transition rates.

Definition 3.5.4.
For $\boldsymbol{t} = (t_1, t_2) \in (s,\infty)^2$, and $\boldsymbol{i} = (i_1, i_2), \boldsymbol{j} = (j_1, j_2) \in \mathcal{Z}^2$ we want to estimate expected values of the censored jump processes and indicator processes:

$$\boldsymbol{P}^c_{z,\boldsymbol{i}}(\boldsymbol{t}) := \mathbb{E}[I_{i_1}(t_1) I_{i_2}(t_2) \mathbb{1}_{\{\boldsymbol{t} < R\}} \mathbb{1}_{\{\xi = z\}}],$$
$$\boldsymbol{Q}^c_{z,\boldsymbol{ij}}(\boldsymbol{t}) := \mathbb{E}[N_{i_1 j_1}(t_1 \wedge R) N_{i_2 j_2}(t_2 \wedge R) \mathbb{1}_{\{\xi = z\}}],$$

and the corresponding bivariate conditional transition rates:

$$\boldsymbol{\Lambda}^c_{z,\boldsymbol{ij}}(\boldsymbol{t}) := \int_{(s,\boldsymbol{t}]} \frac{\mathbb{1}_{\{\boldsymbol{P}^c_{z,\boldsymbol{i}}(\boldsymbol{u}^-) > 0\}}}{\boldsymbol{P}^c_{z,\boldsymbol{i}}(\boldsymbol{u}^-)} \boldsymbol{Q}^c_{z,\boldsymbol{ij}}(\mathrm{d}\boldsymbol{u}).$$

Again, R is a random variable which describes right censoring for the state process.

One of the reasons for using conditional transition rates in the univariate setting is the fact that they are effective even on censored data. This is one of the main reasons for using transition rates in both Markov and non-Markov models. The following lemma captures this feature in the bivariate situation.

Lemma 3.5.5.
Bivariate censored conditional transition rates satisfy the equation

$$\boldsymbol{\Lambda}_{z,\boldsymbol{ij}}(\boldsymbol{t}) = \boldsymbol{\Lambda}^c_{z,\boldsymbol{ij}}(\boldsymbol{t}),$$

for $\boldsymbol{t} \in (s,T]^2$ and $\boldsymbol{i}, \boldsymbol{j} \in \mathcal{Z}^2$, if $\mathbb{P}(\boldsymbol{t} \leq R) > 0$.

Proof. For $\boldsymbol{i}, \boldsymbol{j} \in \mathcal{Z}^2$, we use Campbell's theorem (3.A.3), the law of iterated expectation, and Assumption 3.5.1:

$$\begin{aligned}
\boldsymbol{Q}^c_{z,\boldsymbol{ij}}(\mathrm{d}\boldsymbol{u}) &= \mathbb{E}[\mathbb{1}_{\{\xi = z\}} \mathbb{1}_{\{\boldsymbol{u} \leq R\}} N_{i_1 j_1}(\mathrm{d}u_1) N_{i_2 j_2}(\mathrm{d}u_2)] \\
&= \mathbb{E}[\mathbb{1}_{\{\xi = z\}} \mathbb{E}[\mathbb{1}_{\{\boldsymbol{u} \leq R\}} | (Z_t)_{t \geq s}] N_{i_1 j_1}(\mathrm{d}u_1) N_{i_2 j_2}(\mathrm{d}u_2)] \\
&= \mathbb{P}(\xi = z) \mathbb{E}[\mathbb{1}_{\{\boldsymbol{u} \leq R\}}] \mathbb{E}[N_{i_1 j_1}(\mathrm{d}u_1) N_{i_2 j_2}(\mathrm{d}u_2) | \xi = z] \\
&= \mathbb{P}(\xi = z) \mathbb{P}(\boldsymbol{u} \leq R) \boldsymbol{Q}_{z,\boldsymbol{ij}}(\mathrm{d}\boldsymbol{u}),
\end{aligned}$$

and with similar calculations

$$\boldsymbol{P}^c_{z,\boldsymbol{i}}(\boldsymbol{u}^-) = \boldsymbol{P}_{z,\boldsymbol{i}}(\boldsymbol{u}^-)\mathbb{P}(\xi = z)\mathbb{P}(\boldsymbol{u} \leq R).$$

If $\mathbb{P}(\boldsymbol{u} \leq R) > 0$ and $\boldsymbol{u} > s$, we can plug these equations into the definition of $\boldsymbol{\Lambda}^c_{\boldsymbol{ij}}(\boldsymbol{t})$ and the factor $\mathbb{P}(\xi = z)\mathbb{P}(\boldsymbol{u} \leq R)$ cancels out. In the case that $\mathbb{P}(\xi = z) = 0$ this statement is still true, because then $\boldsymbol{\Lambda}^c_{\boldsymbol{ij}}(\boldsymbol{t})$ is equal to 0 due to the previous calculations and $\boldsymbol{\Lambda}_{\boldsymbol{ij}}(\boldsymbol{t})$ is equal to 0 due to the convention $\frac{0}{0} = 0$. Thus, we get the desired equation. □

We can conclude that for the estimation of $\boldsymbol{\Lambda}_{z,\boldsymbol{ij}}(\boldsymbol{t})$, an estimator for $\boldsymbol{\Lambda}^c_{z,\boldsymbol{ij}}(\boldsymbol{t})$ works well, if $\boldsymbol{t} < \boldsymbol{\tau}$ where we define the endpoint $\boldsymbol{\tau}$ to be a time-point that satisfies

$$\mathbb{P}(\boldsymbol{\tau} \leq R) > 0.$$

In the rest of this paper, we always assume that $\boldsymbol{\tau}$ satisfies this condition and is greater than s. In practice, τ can be understood as the right endpoint of the data window, we are able to observe and is mostly predetermined by the application. We have difficulties estimating a process like $\boldsymbol{\Lambda}^c_{z,\boldsymbol{ij}}(\boldsymbol{t})$ in the case that $\boldsymbol{P}^c_{z,\boldsymbol{i}}(\boldsymbol{u}^-)$ is close to 0. There are different strategies to circumvent this problem. See for example IV.1.2 from Andersen, Borgan and Gill [And+93] where they use assumptions on the behaviour of jump processes to circumvent this problem. For the estimation of bivariate survival functions this gets solved by the fact that $P^c(u-)$ goes to zero monotonically as u goes to infinity, which does not generally happen in a multi-state model. We use the perturbation idea of Bladt and Furrer [BF24b], which we adapt from the univariate case. This corresponds to the new process $\boldsymbol{\Lambda}^{(\epsilon)}_{z,\boldsymbol{ij}}(\boldsymbol{t})$ with

$$\boldsymbol{\Lambda}^{(\epsilon)}_{z,\boldsymbol{ij}}(\boldsymbol{t}) := \int_{(s,\boldsymbol{t}]} \frac{1}{\boldsymbol{P}^c_{z,\boldsymbol{i}}(\boldsymbol{u}-) \vee \epsilon} \boldsymbol{Q}^c_{z,\boldsymbol{ij}}(\mathrm{d}\boldsymbol{u}),$$

for $\epsilon > 0$, $\boldsymbol{i}, \boldsymbol{j} \in \mathcal{Z}^2$ and $\boldsymbol{t} > s$. This $\boldsymbol{\Lambda}^{(\epsilon)}_{z,\boldsymbol{ij}}(\boldsymbol{t})$ is estimated as a replacement for $\boldsymbol{\Lambda}_{z,\boldsymbol{ij}}(\boldsymbol{t})$. In practice, the use of $\boldsymbol{\Lambda}^{(\epsilon)}_{z,\boldsymbol{ij}}$ is hardly a problem for two reasons. First, in the proof of Theorem 3.5.10, we see that

$$\boldsymbol{\Lambda}_{z,\boldsymbol{ij}}(\boldsymbol{t}) = \lim_{\epsilon \to \infty} \boldsymbol{\Lambda}^{(\epsilon)}_{z,\boldsymbol{ij}}(\boldsymbol{t}),$$

which means that we get arbitrarily close to the unperturbed conditional transition rates. Second, for practical calculations and estimates, we generally use discretisation of our conditional transition rates and probabilities, i.e. a discretisation in the time and value axis. This means that the only way to get infinitely close to zero is to make a jump to zero. These jumps to zero are not a problem if ϵ is taken to be smaller than the smallest jump. See Bladt and Furrer [BF24b] example after Definition 2.4 for similar reasoning in the case of perturbation of the univariate conditional transition rates. Furthermore, in practice a complete convergence of transition probabilities to zero can be stopped anyway by a forced transition of states. Examples of this are transitions to annuity at certain ages. In implementation, this leads to forced transitions, for example, into the state "dead"

or "retired". These forced transitions prevent the different transition probabilities from converging to zero. So in practice we usually get

$$\boldsymbol{\Lambda}_{z,\boldsymbol{ij}}(\boldsymbol{t}) = \boldsymbol{\Lambda}^{(\epsilon)}_{z,\boldsymbol{ij}}(\boldsymbol{t})$$

for $t \in (s, \boldsymbol{\tau}]$, $\boldsymbol{i}, \boldsymbol{j} \in \mathcal{Z}^2$, and a fixed $\epsilon > 0$. We start by defining the estimators for the expected value of censored jump processes and indicator processes.

$$\begin{aligned}
\boldsymbol{I}^{(n)}_{z,\boldsymbol{i}}(\boldsymbol{t}) &:= \frac{1}{n}\sum_{m=1}^{n} \mathbb{1}_{\{\xi^m = z\}} \mathbb{1}_{\{\boldsymbol{t} < R^m\}} \mathbb{1}_{\{Z^m_{t_1} = i_1\}} \mathbb{1}_{\{Z^m_{t_2} = i_2\}}, \\
\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{t}) &:= \frac{1}{n}\sum_{m=1}^{n} \mathbb{1}_{\{\xi^m = z\}} N^m_{i_1 j_1}(t_1 \wedge R^m) N^m_{i_2 j_2}(t_2 \wedge R^m),
\end{aligned}$$

for $\boldsymbol{i}, \boldsymbol{j} \in \mathcal{Z}^2$ and $\boldsymbol{t} \in [s, \infty)$. These can be used to define the bivariate landmark Nelson-Aalen estimator

$$\boldsymbol{\Lambda}^{(n,\epsilon)}_{z,\boldsymbol{ij}}(\boldsymbol{t}) := \int_{(s,\boldsymbol{t}]} \frac{1}{\boldsymbol{I}^{(n)}_{z,\boldsymbol{i}}(u^-) \vee \epsilon} \boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\mathrm{d}u),$$

and bivariate landmark Aalen-Johansen estimators as $P^{(n,\epsilon)}_{z,ik}$, the unique solution to the integral equation

$$\begin{aligned}
\boldsymbol{P}^{(n,\epsilon)}_{z,\boldsymbol{i}}(\boldsymbol{t}) = \boldsymbol{P}_{z,\boldsymbol{i}}(s) &+ P_{z,i_2}(s)\left(P^{(n,\epsilon)}_{z,i_1}(t_1) - P_{z,i_1}(s)\right) + P_{z,i_1}(s)\left(P^{(n,\epsilon)}_{z,i_2}(t_2) - P_{z,i_2}(s)\right) \\
&+ \int_{(s,\boldsymbol{t}]} \sum_{\boldsymbol{j} \in \mathcal{Z}^2} \boldsymbol{P}^{(n,\epsilon)}_{z,\boldsymbol{j}}(\boldsymbol{u}^-) \boldsymbol{\Lambda}^{(n,\epsilon)}_{z,\boldsymbol{ji}}(\mathrm{d}\boldsymbol{u}),
\end{aligned} \tag{3.5.1}$$

for $\boldsymbol{i}, \boldsymbol{j} \in \mathcal{Z}^2$ and $\boldsymbol{t} > s$, where the uniqueness comes from Theorem 4.1 from Bathke and Christiansen [BC24].

Remark 3.5.6.
If there is no censoring, all the state processes of the insured follow the equation

$$\mathbb{1}_{\{Z^n_t = j\}} = \mathbb{1}_{\{j = Z^n_s\}} + \sum_{k \in \mathcal{Z}} \int_{(s,t]} N^n_{kj}(\mathrm{d}u).$$

This indicates that solving equation (3.5.1) yields the same estimator as simple averages. Hence, this method of estimation is only useful when censoring occurs before the relevant time points.

Remark 3.5.7.
In the construction of the estimation model, we assume that we observe individuals until they are censored. The construction of the landmark Nelson-Aalen estimator does not

depend on this knowledge. We only need to observe the individuals until they reach an absorption time, i.e. they jump into an absorption state. This is the case because there can be no jumps after the absorption time, so $\boldsymbol{\Lambda}^{(n,\epsilon)}_{z,\boldsymbol{ij}}(\boldsymbol{t})$ does not depend on this information. And since the bivariate landmark Aalen-Johansen estimator only depends on the landmark Nelson-Aalen estimator, this information is not really needed.

To achieve convergence of these estimators, it is necessary to ensure convergence of all the univariate and bivariate estimators. The process begins with the almost certain uniform convergence in n of the univariate processes.
Referring to Bladt and Furrer [BF24b], we get almost sure uniform convergence of the estimators for the univariate conditional transition rates. If s and τ are selected, so that $\tau > s$, $\mathbb{P}(\tau \leq R) > 0$, $i, j \in \mathcal{Z}$ and Assumptions 3.5.1 are met, we get

$$\begin{aligned}
\sup_{t\in(s,\tau]} \left| I^{(n)}_{z,j}(t) - P^{c}_{z,j}(t) \right| &\xrightarrow[n\to\infty]{a.s.} 0, \\
\sup_{t\in(s,\tau]} \left| N^{(n)}_{z,jk}(t) - Q^{c}_{z,jk}(t) \right| &\xrightarrow[n\to\infty]{a.s.} 0, \\
\sup_{t\in(s,\tau]} \left| P^{(n,\epsilon)}_{z,j}(t) - P^{(\epsilon)}_{z,j}(t) \right| &\xrightarrow[n\to\infty]{a.s.} 0, \\
\sup_{t\in(s,\tau]} \left| \Lambda^{(n,\epsilon)}_{z,jk}(t) - \Lambda^{(\epsilon)}_{z,jk}(t) \right| &\xrightarrow[n\to\infty]{a.s.} 0.
\end{aligned} \tag{3.5.2}$$

The idea of the estimation procedure is to estimate censored conditional transition rates and censored transition probabilities, use these to estimate conditional transition rates because they do not depend on uncensored data, and finally to use these uncensored conditional transition rates to estimate the uncensored transition probabilities via the product integral.

Theorem 3.5.8.
If the Assumption 3.5.1 holds, we get uniform convergence of the estimators. For $i, j \in \mathcal{Z}$

$$\begin{aligned}
\lim_{\epsilon\to 0}\lim_{n\to\infty} \sup_{t\in(s,\tau]} \left| \Lambda^{(n,\epsilon)}_{z,jk}(t) - \Lambda_{z,jk}(t) \right| &\overset{a.s}{=} 0, \\
\lim_{\epsilon\to 0}\lim_{n\to\infty} \sup_{t\in(s,\tau]} \left| P^{(n,\epsilon)}_{z,j}(t) - P_{z,j}(t) \right| &\overset{a.s}{=} 0.
\end{aligned}$$

The proof for this theorem follows at the end of this section.
The setup we use for the estimation of bivariate conditional transition rates is based on the same idea. The aim of the following statements is to assert uniform convergence for the bivariate estimators.

Lemma 3.5.9.
If $\boldsymbol{\tau} \in [s, \infty)$ so that $\mathbb{P}(\boldsymbol{\tau} \leq R) > 0$ and Assumption 3.5.1 holds, we get uniform convergence of the estimators $\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}$ and $\boldsymbol{I}^{(n)}_{z,\boldsymbol{i}}$:

$$\sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]} \left| \boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{t}) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{t}) \right| \xrightarrow[n\to\infty]{a.s.} 0,$$

$$\sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]}\left|\boldsymbol{I}^{(n)}_{z,\boldsymbol{i}}(\boldsymbol{t})-\boldsymbol{P}^{c}_{z,\boldsymbol{i}}(\boldsymbol{t})\right|\xrightarrow[n\to\infty]{a.s.}0,$$

for $\boldsymbol{i},\boldsymbol{j}\in\mathcal{Z}^2$.

The proof for this lemma follows at the end of this section.
These convergences can be transferred to the convergence of conditional transition rates and probabilities. Again, the convergence in ϵ is taken care of.

Theorem 3.5.10.
If the Assumption 3.5.1 holds, we get uniform convergence of the estimators. For $\boldsymbol{i},\boldsymbol{j}\in\mathcal{Z}^2$

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]}\left|\boldsymbol{\Lambda}^{(n,\epsilon)}_{z,\boldsymbol{ij}}(\boldsymbol{t})-\boldsymbol{\Lambda}_{z,\boldsymbol{ij}}(\boldsymbol{t})\right|\overset{a.s}{=}0,$$

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]}\left|\boldsymbol{P}^{(n,\epsilon)}_{z}(\boldsymbol{t})-\boldsymbol{P}_{z}(\boldsymbol{t})\right|\overset{a.s}{=}0.$$

The proof for this theorem follows at the end of this section.
This theorem somewhat describes consistency of the estimators $\boldsymbol{\Lambda}^{(n,\epsilon)}_{z,\boldsymbol{ij}}$ and $\boldsymbol{P}^{(n,\epsilon)}_{z}$. In practice one can fix an ϵ and let n grow to get a reasonably close estimation of $\boldsymbol{\Lambda}_{z,\boldsymbol{ij}}$ and $\boldsymbol{P}_z$. In the proof of this theorem we show, among other things, that the solution of the equation (3.A.1) as a functional of $\boldsymbol{\Lambda}_{z,\boldsymbol{ij}}$ and $\Lambda_{z,ij}$ is Hadamard continuous. This statement is already known, see Gill, Van Der Laan and Wellner [GLW95]. They did not add a complete proof, so we include this part in our proof. Additionally, they state that this functional is Hadamard differentiable.

Remark 3.5.11.
For the convergence of these landmark estimators we use $\boldsymbol{\tau}$, meaning $\mathbb{P}(\boldsymbol{\tau}\leq R)$. In a basic multi-state model with one state process, this can be simplified to $\boldsymbol{\tau}=(\tau,\tau)$ with $\mathbb{P}(\tau\leq R)$. This is because both jump processes or indicator processes come from the same state-process and therefore generally fall under the same censoring time. We have introduced this freedom of the censoring variable to allow for natural generalization to two jump processes corresponding to two different subjects of examination.

Remark 3.5.12.
In the context of actuarial estimation, so called retrospective estimation is important because past developments can be non-measurable regarding the current information $\sigma(\mathcal{G}_s)$, see Section 3.2. The definition of the retrospective estimators and the proof of convergence works similar to the prospective estimations, the main difference is the consideration of left truncation. For these reasons, it is out of the scope of this paper.

Remark 3.5.13.
We introduce the idea that one would follow to prove asymptotic normality, i.e. weak convergence to a tight Gaussian process of the introduced estimators. To prove the asymptotic normality of $\boldsymbol{\Lambda}^{(n,\epsilon)}_{z,\boldsymbol{ij}}(\boldsymbol{t})$ and $\boldsymbol{P}^{(n,\epsilon)}_{z}(\boldsymbol{t})$ one uses the Hadamard differentiability

of the equation (3.5.1) interpreted as a functional and ideas introduced in the proof of Theorem 4.5 in Bladt and Furrer [BF24b], which are easily generalised to two dimensions, for the demonstration of the asymptotic normality of $\boldsymbol{\Lambda}^{(n,\epsilon)}_{z,\boldsymbol{ij}}(\boldsymbol{t})$. Both of these ideas are based on the asymptotic normality of $\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{t})$ and $\boldsymbol{I}^{(n)}_{z,\boldsymbol{i}}(\boldsymbol{t})$, where the second also bases on the first with the help of equation (3.2.3). The asymptotic normality of $\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{t})$ follows with the bracketing central limit theorem, Theorem 2.11.9 in Van Der Laan [Laa96]. The convergence via the bracketing central limit theorem relies on the bracketing number, which is the minimum number of sets P_i in the partition of $[s,\tau]^2$ such that

$$\mathbb{E}[\sup_{\boldsymbol{t},\boldsymbol{s}\in P_i} |N_{j_1k_1}(t_1\wedge R)N_{j_2k_2}(t_2\wedge R) - N_{j_1k_1}(s_1\wedge R)N_{j_2k_2}(s_1\wedge R)|^2] < \mu^2$$

for μ close to zero. This can be reduced to a problem of one-dimensional partitions, by applying the equality $ab-dc = a(b-c)+c(a-d)$, which allows us to work with differences in one dimension rather than two. This reduces the problem of the bracketing number significantly, since the jump processes are all monotonically increasing and bounded in expectation. From here, one uses the classical arguments for one-dimensional jump processes to follow the asymptotic normality. This is the idea behind the asymptotic normality of the estimators, the full proof can be found in the work of Bathke and Furrer [BF24a]. This paper focuses on the uniform convergence of the estimators and provides a proof of concept for the estimation of bivariate conditional transition rates and their application in the actuarial context, which will be illustrated through the numerical example in Section 3.7.

Proof of Lemma 3.5.9. From the strong law of large numbers, we immediately get

$$\begin{aligned}\left|\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{t}) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{t})\right| &\xrightarrow[n\to\infty]{a.s.} 0,\\ \left|\boldsymbol{I}^{(n)}_{z,\boldsymbol{i}}(\boldsymbol{t}) - \boldsymbol{P}^{c}_{z,\boldsymbol{i}}(\boldsymbol{t})\right| &\xrightarrow[n\to\infty]{a.s.} 0,\end{aligned} \tag{3.5.3}$$

for all $\boldsymbol{t}\in(s,\boldsymbol{\tau}]$. The first step is to generalize this to uniform convergence for $\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{t})$. We use an idea that is similar to the idea of the univariate Glivenko-Cantelli theorem. We define the sets

$$C_{m,n} := \left\{ \sum_{\boldsymbol{i},\boldsymbol{j}\in\mathcal{Z}^2} |\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{\tau})| \le m \right\}$$

and $C_m := (C_{m,n}$ eventually $)$. This means that $\mathbb{1}_{C_m}\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{t})$ is eventually bounded on $(s,\boldsymbol{\tau}]$. Now because of equation (3.5.3) we have $\mathbb{P}(C_m)=1$. Thus, we can assume $\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(t)$ to be bounded eventually for all $\boldsymbol{i}\in\mathcal{Z}^2$.
We now take a partition of $(s,\boldsymbol{\tau}]$ into finitely many pairwise disjoint rectangles $R_i = (\boldsymbol{u}_i,\boldsymbol{t}_i)\subset(s,\boldsymbol{\tau}]$ and $[s,\tau]=\cup_{i\in I}\overline{R}_i$. With $\mathrm{d}\boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(R_i)\le\epsilon$ and $\mathrm{d}\boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\Delta R_i)\le\epsilon$, where $\mathrm{d}\boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}$ is the one or two-dimensional Lebesgue-Stieltjes measure generated by $\boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}$. This works because for a monotonous function in two-variables, discontinuities can only appear on a countable set of parallels to the x- and y-axis, see for example the work of Young and

Young [YY24]. Thus, for every $\boldsymbol{x} \in (s, \boldsymbol{\tau}]$, we get an $a \in I$ with $\boldsymbol{x} \in R_a = (u_a, t_a)$ or $\boldsymbol{x}$ is on the edge of one of these rectangles. In the first case, we get

$$\begin{aligned}\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{u}_a) \leq \boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{x}) \leq \boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{t}_a^-)\\ \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{u}_a) \leq \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{x}) \leq \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{t}_a^-)\end{aligned}$$

and

$$\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{u}_a) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{t}_a^-) \leq \boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{x}) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{x}) \leq \boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{t}_a^-) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{u}_a).$$

Adding 0 we get

$$\begin{aligned}\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{u}_a) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{u}_a) + \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{u}_a) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{t}_a^-) \leq \boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{x}) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{x})\\ \boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{t}_a^-) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{t}_a^-) + \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{t}_a^-) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{u}_a) \geq \boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{x}) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{x}).\end{aligned}$$

The last step is, to evaluate $\boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{t}_a^-) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(s)$ on every rectangle R_i. We get

$$\begin{aligned}&\left|\boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(t_{a,1}^-, t_{a,2}^-) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(u_{a,1}, u_{a,2})\right|\\ &\qquad\leq \left|\boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(u_{a,1}, t_{a,2}^-) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(u_{a,1}, u_{a,2})\right|\\ &\qquad\quad + \left|\boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(t_{a,1}^-, u_{a,2}) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(u_{a,1}, u_{a,2})\right|\\ &\qquad\quad + \left|\boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{t}_a^-) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(t_{a,1}^-, u_{a,2}) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(u_{a,1}, t_{a,2}^-) + \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{u}_a)\right|\\ &\qquad\leq 3\epsilon.\end{aligned}$$

All in all we get the following assertion

$$\begin{aligned}\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{u}_a) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{u}_a) - 3\epsilon \leq \boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{x}) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{x})\\ \boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{t}_a^-) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{t}_a^-) + 3\epsilon \geq \boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{x}) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{x}).\end{aligned}$$

Now $\forall \epsilon > 0$ take the previously defined partition $(R_i)_{i\in\mathcal{Z}}$ and define $N(\epsilon)$ such that $\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{u}_i) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{u}_i) > -\epsilon$ and $\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{t}_i-) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{t}_i-) < \epsilon$ for all R_i. This is possible due to the finite nature of the partition and (3.5.3). Then we get $\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{x}) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{x}) < 4\epsilon$. For the second case, where $\boldsymbol{x}$ is on the edge of one of the rectangles, we can apply the one-dimensional Glivenko-Cantelli idea on a parallel to the x- or y-axis and get to the same outcome. Thus,

$$\sup_{\boldsymbol{x}\in[s,\boldsymbol{\tau}]} \left|\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{x}) - \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{x})\right| \xrightarrow[n\to\infty]{} 0$$

on C_m. Referring to the fact that $\boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\tau)$ is bounded and that $\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{\tau}) \xrightarrow[n\to\infty]{a.s.} \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\boldsymbol{\tau})$, $\mathbb{P}(C_m) = 1$ for m large enough. All in all, we get the almost sure uniform convergence of $\boldsymbol{N}^{(n)}_{z,\boldsymbol{ij}}(\boldsymbol{t})$.

For the uniform convergence of $\boldsymbol{I}^{(n)}_{z,\boldsymbol{i}}(\boldsymbol{t})$ we use the fact that we can calculate the state-indicator process from the full information over all the jump processes and the information at the current time s and then transfer the uniform convergence of the conditional transition

rates to the transition probabilities. This is a common idea for the convergence of transition probabilities, see for example the work of Bladt and Furrer [BF24b]. In our case we have the formula

$$\begin{aligned} I_{i_1}(t_1)I_{i_2}(t_2) = I_{i_1}(s)I_{i_2}(s) + I_{i_2}(s)\sum_{k\in\mathcal{Z}} N_{ki_1}(t_1) + I_{i_1}(s)\sum_{k\in\mathcal{Z}} N_{ki_2}(t_2) \\ + \sum_{\boldsymbol{k}\in\mathcal{Z}^2} N_{k_1i_1}(t_1)N_{k_2i_2}(t_2). \end{aligned}$$

Because of the censoring we need to adjust this formula to

$$\begin{aligned} I_{i_1}(t_1)I_{i_2}(t_2)\mathbb{1}_{\{\boldsymbol{t}<R\}} = I_{i_1}(s)I_{i_2}(s) + I_{i_2}(s)\sum_{k\in\mathcal{Z}} N_{ki_1}(t_1\wedge R) \\ + I_{i_1}(s)\sum_{k\in\mathcal{Z}} N_{ki_2}(t_2\wedge R) + \sum_{\boldsymbol{k}\in\mathcal{Z}^2} N_{k_1i_1}(t_1\wedge R)N_{k_2i_2}(t_2\wedge R) \\ - \mathbb{1}_{\{t_1\geq R\vee t_2\geq R\}} I_{i_1}(t_1\wedge R)I_{i_2}(t_2\wedge R). \end{aligned}$$

This means that the estimator $I^{(n)}_{z,ij}(\boldsymbol{t})$ satisfies the following equation:

$$\begin{aligned} \boldsymbol{I}^{(n)}_{z,\boldsymbol{i}}(\boldsymbol{t}) = I^{(n)}_{z,\boldsymbol{i}}(s) + \boldsymbol{I}^{(n)}_{z,i_1}(s)\sum_{k\in\mathcal{Z}} N^{(n)}_{z,ki_2}(t_2) + I^{(n)}_{z,i_2}(s)\sum_{k\in\mathcal{Z}} N^{(n)}_{z,ki_1}(t_1) \\ + \sum_{\boldsymbol{k}\in\mathcal{Z}^2} \boldsymbol{N}^{(n)}_{z,\boldsymbol{ki}}(\boldsymbol{t}) - C^{(n)}_{z,\boldsymbol{i}}(\boldsymbol{t}), \end{aligned} \tag{3.5.4}$$

where

$$C^{(n)}_{z,\boldsymbol{i}}(\boldsymbol{t}) := \frac{1}{n}\sum_{m=1}^{n} \mathbb{1}_{\{R^m\leq t_1\vee R^m\leq t_2\}}\mathbb{1}_{\{Z^m_{t_1\wedge R^m}=i_1, Z^m_{t_2\wedge R^m}=i_2\}}\mathbb{1}_{\{\xi^m=z\}},$$

for $\boldsymbol{i}\in\mathcal{Z}^2$. Analogously, we get an equality for the occupation probabilities:

$$\begin{aligned} \boldsymbol{P}^c_{z,\boldsymbol{i}}(\boldsymbol{t}) = \boldsymbol{P}^c_{z,\boldsymbol{i}}(s) + P^c_{z,i_1}(s)\sum_{k\in\mathcal{Z}} Q^c_{z,ki_2}(t_2) + P^c_{z,i_2}(s)\sum_{k\in\mathcal{Z}} Q^c_{z,ki_1}(t_1) \\ + \sum_{\boldsymbol{k}\in\mathcal{Z}^2} \boldsymbol{Q}^c_{z,\boldsymbol{ki}}(\boldsymbol{t}) - C_{z,\boldsymbol{i}}(\boldsymbol{t}), \end{aligned} \tag{3.5.5}$$

where $C_{z,\boldsymbol{i}}(\boldsymbol{t}) := \mathbb{P}(t_1\geq R\vee t_2\geq R, Z_{t_1\wedge R}=i_1, Z_{t_2\wedge R}=i_2, \xi=z)$ for $\boldsymbol{i}\in\mathcal{Z}^2$. The next step is to show $\sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]}\left|C_{z,\boldsymbol{i}}(\boldsymbol{t}) - C^{(n)}_{z,\boldsymbol{i}}(\boldsymbol{t})\right| \xrightarrow[n\to\infty]{a.s.} 0$. We have

$$\begin{aligned} &I_{i_1}(t_1\wedge R)I_{i_2}(t_2\wedge R)\mathbb{1}_{\{R\leq t_1\vee R\leq t_2\}} \\ &= \Big[I_{i_1}(s)I_{i_2}(s) + I_{i_2}(s)\sum_{k\in\mathcal{Z}} N_{ki_1}(t_1\wedge R) + I_{i_1}(s)\sum_{k\in\mathcal{Z}} N_{ki_2}(t_2\wedge R) \\ &\quad + \sum_{\boldsymbol{k}\in\mathcal{Z}^2} N_{k_1i_1}(t_1\wedge R)N_{k_2i_2}(t_2\wedge R)\Big]\mathbb{1}_{\{R\leq t_1\vee R\leq t_2\}}. \end{aligned}$$

Similar to the approach with $\boldsymbol{P}_{z,\boldsymbol{i}}(\boldsymbol{t})$ and $\boldsymbol{I}^{(n)}_{z,\boldsymbol{i}}(\boldsymbol{t})$, we can now separate $C_{z,\boldsymbol{i}}(\boldsymbol{t})$ and $\boldsymbol{C}^{(n)}_{z,\boldsymbol{i}}(\boldsymbol{t})$ into four summands that are monotonically increasing and bounded on a set with probability

1. Thus, with the same two-dimensional Glivenko-Cantelli arguments we used before and the strong law of large numbers, we get the almost sure uniform convergence of $\boldsymbol{C}_{z,\boldsymbol{i}}^{(n)}(\boldsymbol{t})$ to $\boldsymbol{C}_{z,\boldsymbol{i}}(\boldsymbol{t})$. With equations (3.5.5), (3.5.4), the uniform convergence of the univariate estimators and the law of large numbers, we get the almost sure uniform convergence of $\boldsymbol{I}_{z,\boldsymbol{i}}^{(n)}$ to $\boldsymbol{P}_{z,\boldsymbol{i}}^{c}$ on $(s, \boldsymbol{\tau}]$. □

Proof of Theorem 3.5.10. For the proof of this theorem, we want to apply the continuity of the functional

$$L : (F, G) \mapsto \int\limits_{(s,\boldsymbol{t}]} \frac{1}{F(u)} \mathrm{d}G(u),$$

see Lemma 3.A.3 for the precise definition of continuity and the contributing function spaces. For this lemma, we need that $\frac{1}{\boldsymbol{I}_{\boldsymbol{i}}^{(n)} \vee \epsilon}$, $\boldsymbol{N}_{z,\boldsymbol{ij}}^{(n)}$, $\frac{1}{\boldsymbol{P}_{z,\boldsymbol{i}}^{c} \vee \epsilon}$, and $\boldsymbol{Q}_{z,\boldsymbol{ij}}^{c}$ are all of almost surely uniform bounded variation in n, see the Appendix for the definition of the variation 3.A.5.
For $\boldsymbol{N}_{z,\boldsymbol{ij}}^{(n)}$ and $\boldsymbol{Q}_{z,\boldsymbol{ij}}^{c}$, the almost sure uniform bounded variation follows directly from the monotonicity and the fact that $\boldsymbol{N}_{z,\boldsymbol{ij}}^{(n)}(\tau)$ is almost surely bounded, see the proof of Lemma 3.5.9. For the uniform bounded variation of $\frac{1}{\boldsymbol{I}_{\boldsymbol{i}}^{(n)} \vee \epsilon}$ and $\frac{1}{\boldsymbol{P}_{z,\boldsymbol{i}}^{c} \vee \epsilon}$, we use the fact that

$$\Big\| \frac{1}{f(\boldsymbol{x}) \vee \epsilon} \Big\|_{v_1}^{(2)} \leq \frac{1}{\epsilon} + \frac{1}{\epsilon^4} \| f(\boldsymbol{x}) \|_{v_1}^{(2)}.$$

Now $\| \boldsymbol{I}_{\boldsymbol{i}}^{(n)} \|_{v_1}^{(2)}$ and $\| \boldsymbol{P}_{z,\boldsymbol{i}}^{c} \|_{v_1}^{(2)}$ are both of almost surely bounded variation by equation (3.2.3) and the same techniques as before. With Lemma 3.A.3 the statement follows immediately.
The next step is the convergence of $\boldsymbol{P}_{\boldsymbol{i}}^{(n,\epsilon)}(\boldsymbol{t})$. We first need to take care of the problem that $\boldsymbol{P}_{\boldsymbol{i}}^{(n,\epsilon)}(\boldsymbol{t})$ is defined as a solution to an integral equation. For that we rewrite the integral equation. We start by defining $\tilde{P}^{(\epsilon)}(\boldsymbol{t})$:

$$\tilde{P}_{l(i-1)+k}^{(\epsilon)}(\boldsymbol{t}) := P_{ki}^{(\epsilon)}(\boldsymbol{t}).$$

This $\tilde{P}^{(\epsilon)}(\boldsymbol{t})$ corresponds to a column by column translation of the matrix $\boldsymbol{P}^{(\epsilon)}(\boldsymbol{t})$. Now rearrange $\boldsymbol{\Lambda}_{z,\boldsymbol{ij}}(\boldsymbol{t})$ into the matrix $\Lambda^{(\epsilon)}$, such that $\tilde{P}^{(\epsilon)}(\boldsymbol{t})$ is a solution to the inhomogeneous Volterra equation

$$Y(\boldsymbol{t}) = \phi^{(\epsilon)}(\boldsymbol{t}) + \int\limits_{(s,\boldsymbol{t}]} Y(\boldsymbol{u}^{-}) \Lambda^{(\epsilon)}(\mathrm{d}\boldsymbol{u}), \tag{3.5.6}$$

where $\phi^{(\epsilon)}$ corresponds to the one-dimensional part of the two-dimensional integral equation

$$\phi^{(\epsilon)}(\boldsymbol{t})_{l(i_2-1)+i_1} := -P_{i_2}^{(\epsilon)}(s) P_{i_1}^{(\epsilon)}(s) + P_{i_2}^{(\epsilon)}(t_1) P_{i_1}^{(\epsilon)}(s) + P_{i_2}^{(\epsilon)}(s) P_{i_1}^{(\epsilon)}(t_2),$$

with $l = \#\mathcal{Z}$. This works because for $\boldsymbol{t} > s$, the calculation of $P_{\boldsymbol{i}}(\boldsymbol{t})$ only relies on $\boldsymbol{\Lambda}_{\boldsymbol{ji}}(\boldsymbol{t})_{\boldsymbol{j} \in \mathcal{Z}^2}$, thus, we can rearrange $\boldsymbol{P}_{\boldsymbol{i}}$ and $\boldsymbol{\Lambda}_{\boldsymbol{ji}}(\boldsymbol{t})_{\boldsymbol{j} \in \mathcal{Z}^2}$ into a vector and a matrix. This transformed the rather complicated integral equation (3.4.5) into the well known and

understood inhomogeneous Volterra equation.
From Lemma 3.A.1 we know that the solution to this inhomogeneous Volterra equation is

$$\tilde{\boldsymbol{P}}^{(\epsilon)}(\boldsymbol{t}) = \phi^{(\epsilon)}(\boldsymbol{t}) + \int\limits_{(s,\boldsymbol{t}]} \phi^{(\epsilon)}(\boldsymbol{u}^-)\boldsymbol{\Lambda}^{(\epsilon)}(\mathrm{d}\boldsymbol{u})\mathcal{P}((\boldsymbol{u},\boldsymbol{t}],\boldsymbol{\Lambda}^{(\epsilon)}),$$

with

$$\mathcal{P}((s,\boldsymbol{t}],\Lambda) := Id + \sum_{n=1}^{\infty} \int\limits_{s<\boldsymbol{u}^{(1)}<\ldots<\boldsymbol{u}^{(n)}\leq t} \cdots \int \boldsymbol{\Lambda}(\mathrm{d}\boldsymbol{u}^{(1)}) \cdots \boldsymbol{\Lambda}(\mathrm{d}\boldsymbol{u}^{(n)}).$$

Using this solution, we can calculate the difference between the solution of the estimated integral equation and the solution to the theoretical integral equation, where we use $\tilde{\boldsymbol{P}}^{(n,\epsilon)}(\boldsymbol{t})$ as the solution to the inhomogeneous Volterra equation with estimated $\boldsymbol{\Lambda}^{(n,\epsilon)}$ and $\phi^{(n,\epsilon)}$. We divide the estimation error into the estimation error of the integrand and the integrator.

$$\begin{aligned}
&\sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]} \left\| \tilde{\boldsymbol{P}}^{(\epsilon)}(\boldsymbol{t}) - \tilde{\boldsymbol{P}}^{(n,\epsilon)}(\boldsymbol{t}) \right\|_\infty \\
&\qquad \leq \sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]} \left\| \phi^{(\epsilon)}(\boldsymbol{t}) - \phi^{(n,\epsilon)}(\boldsymbol{t}) \right\| \\
&\qquad\quad + \sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]} \left\| \int\limits_{(s,\boldsymbol{t}]} \left(\phi^{(\epsilon)}(\boldsymbol{u}^-) - \phi^{(n,\epsilon)}(\boldsymbol{u}^-) \right) \boldsymbol{\Lambda}^{(n,\epsilon)}(\mathrm{d}\boldsymbol{u})\mathcal{P}((\boldsymbol{u},\boldsymbol{t}],\boldsymbol{\Lambda}^{(n,\epsilon)}) \right\|_\infty \qquad (3.5.7) \\
&\qquad\quad + \sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]} \left\| \int\limits_{(s,\boldsymbol{t}]} \phi^{(\epsilon)}(\boldsymbol{u}^-) \Big[\Lambda^{(\epsilon)}(\mathrm{d}\boldsymbol{u})\mathcal{P}((\boldsymbol{u},\boldsymbol{t}],\boldsymbol{\Lambda}^{(\epsilon)}) - \right. \\
&\qquad\qquad\qquad \left. \boldsymbol{\Lambda}^{(\epsilon,n)}(\mathrm{d}\boldsymbol{u})\mathcal{P}((\boldsymbol{u},\boldsymbol{t}],\boldsymbol{\Lambda}^{(\epsilon,n)}) \Big] \right\|_\infty \qquad (3.5.8)
\end{aligned}$$

The first two summands (3.5.7) can be taken care of by simple monotonicity arguments:

$$\begin{aligned}
&\sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]} \left\| \phi^{(\epsilon)}(\boldsymbol{t}) - \phi^{(n,\epsilon)}(\boldsymbol{t}) \right\|_\infty \\
&\quad + \sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]} \left\| \int\limits_{(s,\boldsymbol{t}]} \phi^{(\epsilon)}(\boldsymbol{u}^-) - \phi^{(n,\epsilon)}(\boldsymbol{u}^-)\boldsymbol{\Lambda}^{(n,\epsilon)}(\mathrm{d}\boldsymbol{u})\mathcal{P}((\boldsymbol{u},\boldsymbol{t}],\boldsymbol{\Lambda}^{(n,\epsilon)}) \right\|_\infty \\
&\qquad\qquad \leq \sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]} \left\| \phi^{(\epsilon)}(\boldsymbol{t}) - \phi^{(n,\epsilon)}(\boldsymbol{t}) \right\|_\infty \left(2 + \left\| \mathcal{P}((s,\boldsymbol{t}],\boldsymbol{\Lambda}^{(n,\epsilon)}) \right\|_\infty \right).
\end{aligned}$$

For the second summand (3.5.8) we again segment the statistical error for the different parts of the integral. We get

$$\begin{aligned}
&\sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]} \left\| \int\limits_{(s,\boldsymbol{t}]} \phi^{(\epsilon)}(\boldsymbol{u}^-) \left[\boldsymbol{\Lambda}^{(\epsilon)}(\mathrm{d}\boldsymbol{u})\mathcal{P}((\boldsymbol{u},\boldsymbol{t}],\boldsymbol{\Lambda}^{(\epsilon)}) - \boldsymbol{\Lambda}^{(\epsilon,n)}(\mathrm{d}\boldsymbol{u})\mathcal{P}((\boldsymbol{u},\boldsymbol{t}],\boldsymbol{\Lambda}^{(\epsilon,n)}) \right] \right\|_\infty \\
&\qquad \leq \sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]} \left\| \int\limits_{(s,\boldsymbol{t}]} \phi^{(\epsilon)}(\boldsymbol{u}^-)\boldsymbol{\Lambda}^{(\epsilon)}(\mathrm{d}\boldsymbol{u}) \left[\mathcal{P}((\boldsymbol{u},\boldsymbol{t}],\boldsymbol{\Lambda}^{(\epsilon)}) - \mathcal{P}((\boldsymbol{u},\boldsymbol{t}],\boldsymbol{\Lambda}^{(\epsilon,n)}) \right] \right\|_\infty \qquad (3.5.9)
\end{aligned}$$

$$+ \sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]} \Big\| \int_{(s,\boldsymbol{t}]} \phi^{(\epsilon)}(\boldsymbol{u}^-) \left[\boldsymbol{\Lambda}^{(\epsilon)}(\mathrm{d}\boldsymbol{u}) - \boldsymbol{\Lambda}^{(\epsilon,n)}(\mathrm{d}\boldsymbol{u})\right] \mathcal{P}((\boldsymbol{u},\boldsymbol{t}], \boldsymbol{\Lambda}^{(\epsilon,n)}) \Big\|_\infty. \quad (3.5.10)$$

For the second summand (3.5.10), we use two-dimensional integration by parts and the fact that $\mathcal{P}((\boldsymbol{u},\boldsymbol{t}], \boldsymbol{\Lambda}^{(\epsilon,n)})$ is monotone in $\boldsymbol{u}$ and $\boldsymbol{t}$ and thus, has a finite two-dimensional variation, and that with the definition of $P_i^{(\epsilon)}$, the definition of $\|\cdot\|_\infty$, and (37) from Christiansen and Furrer [CF22], we have

$$\Big\| \|\phi^{(\epsilon)}\|_{v_1}^{(2)} \Big\|_\infty \leq \left[\Big\| \sum_{i\in\mathcal{Z}} \|P_i^{(\epsilon)}\|_{v_1}^{(1)} \Big\|_\infty \right]^2 \leq \left[\Big\| \|P(s) \prod_{(s,t]} (Id + \Lambda^{(\epsilon)}(\mathrm{d}u))\|_{v_1}^{(1)} \Big\|_\infty \right]^2 < \infty.$$

All in all, we find an upper bound of the form

$$C(n) \sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]} \Big\| \boldsymbol{\Lambda}^{(\epsilon)}(\boldsymbol{t}) - \boldsymbol{\Lambda}^{(n,\epsilon)}(\boldsymbol{t}) \Big\|_\infty,$$

where $C(n)$ contains, the variation of $\phi^{(\epsilon)}$, and $\mathcal{P}((s,\boldsymbol{\tau}], \Lambda^{(\epsilon,n)})$, which is all almost surely eventually bounded in n.
For the first summand (3.5.9) we get with the monotonicity of $\Lambda^{(\epsilon)}$:

$$\begin{aligned}
&\sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]} \Big\| \int_{(s,\boldsymbol{t}]} \phi^{(n,\epsilon)}(\boldsymbol{u}^-) \Lambda^{(\epsilon)}(\mathrm{d}\boldsymbol{u}) \left[\mathcal{P}((\boldsymbol{u},\boldsymbol{t}], \Lambda^{(\epsilon)}) - \mathcal{P}((\boldsymbol{u},\boldsymbol{t}], \Lambda^{(\epsilon,n)})\right] \Big\|_\infty \\
&\leq \sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]} \sup_{\boldsymbol{u}\in(s,\boldsymbol{\tau}]} \Big\| \mathcal{P}((\boldsymbol{u},\boldsymbol{t}], \Lambda^{(\epsilon)}) - \mathcal{P}((\boldsymbol{u},\boldsymbol{t}], \Lambda^{(\epsilon,n)}) \Big\|_\infty \sup_{\boldsymbol{u}\in(s,\boldsymbol{\tau}]} \Big\| \phi^{(n,\epsilon)}(\boldsymbol{u}) \Big\|_\infty \sup_{\boldsymbol{u}\in(s,\boldsymbol{\tau}]} \Big\| \Lambda^{(\epsilon)}(\boldsymbol{u}) \Big\|_\infty.
\end{aligned}$$

Thus, we need to evaluate $\Big\| \mathcal{P}((\boldsymbol{u},\boldsymbol{t}], \Lambda^{(\epsilon)}) - \mathcal{P}((\boldsymbol{u},\boldsymbol{t}], \Lambda^{(\epsilon,n)}) \Big\|_\infty$. By applying the Duhamel equality, see Lemma 3.A.2, we get:

$$\begin{aligned}
&\Big\| \mathcal{P}((\boldsymbol{u},\boldsymbol{t}], \Lambda^{(\epsilon)}) - \mathcal{P}((\boldsymbol{u},\boldsymbol{t}], \Lambda^{(\epsilon,n)}) \Big\|_\infty \\
&\quad = \Big\| \int_{(\boldsymbol{u},\boldsymbol{t}]} \mathcal{P}((\boldsymbol{u},\boldsymbol{t}] \cap (0,s), \Lambda^{(\epsilon)}) (\Lambda^{(\epsilon)}(\mathrm{d}s) - \Lambda^{(\epsilon,n)}(\mathrm{d}s)) \mathcal{P}((\boldsymbol{u},\boldsymbol{t}] \cap (s,\infty), \Lambda^{(\epsilon,n)}) \Big\|_\infty.
\end{aligned}$$

Again, we can apply two-dimensional integration by parts and find that there is a constant $C(n)$ which is almost surely bounded eventually and satisfies

$$\begin{aligned}
&\Big\| \int_{(\boldsymbol{u},\boldsymbol{t}]} \mathcal{P}((\boldsymbol{u},\boldsymbol{t}] \cap (0,s), \Lambda^{(\epsilon)}) (\Lambda^{(\epsilon)}(\mathrm{d}s) - \Lambda^{(\epsilon,n)}(\mathrm{d}s)) \mathcal{P}((\boldsymbol{u},\boldsymbol{t}] \cap (s,\infty), \Lambda^{(\epsilon,n)}) \Big\|_\infty \\
&\qquad\qquad \leq C(n) \sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]} \Big\| \Lambda^{(\epsilon)}(\boldsymbol{t}) - \Lambda^{(n,\epsilon)}(\boldsymbol{t}) \Big\|_\infty.
\end{aligned}$$

This $C(n)$ contains the variation of $\mathcal{P}((\boldsymbol{u},\boldsymbol{t}] \cap (0,s), \Lambda^{(\epsilon)})$ and of $\mathcal{P}((\boldsymbol{u},\boldsymbol{t}] \cap (s,\infty), \Lambda^{(\epsilon,n)})$, which is again almost surely bounded eventually, because they are both monotone in $\boldsymbol{u}$ and $\boldsymbol{t}$. Combining all of this, we get the desired assertion:

$$\sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]} \Big\| \tilde{P}^{(\epsilon)}(\boldsymbol{t}) - \tilde{P}^{(n,\epsilon)}(\boldsymbol{t}) \Big\|_\infty \xrightarrow[n\to\infty]{a.s.} 0.$$

Now we have the convergence in n of the estimators to $\boldsymbol{\Lambda}^{(\epsilon)}_{z,\boldsymbol{ij}}(\boldsymbol{t})$ and $\boldsymbol{P}^{(\epsilon)}_{z,\boldsymbol{i}}(\boldsymbol{t})$. The next step is the convergence in ϵ.
For the convergence of $\boldsymbol{\Lambda}^{(\epsilon)}_{z,\boldsymbol{ij}}$ we start by showing point wise convergence. We apply the monotone convergence theorem on $\boldsymbol{\Lambda}^{(\epsilon)}_{z,\boldsymbol{ij}}$. This works because of (3.4.3) and equation (35) of Christiansen [Chr21]. We get

$$\int\limits_{(s,\boldsymbol{t}]} \frac{1}{\boldsymbol{P}^{c}_{z,\boldsymbol{i}}(\boldsymbol{u}^-) \vee \epsilon} \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\mathrm{d}\boldsymbol{u}) \xrightarrow[\epsilon\to 0]{a.s.} \int\limits_{(s,\boldsymbol{t}]} \frac{1}{\boldsymbol{P}^{c}_{z,\boldsymbol{i}}(\boldsymbol{u}^-)} \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\mathrm{d}\boldsymbol{u}) = \int\limits_{(s,\boldsymbol{t}]} \frac{\mathbb{1}_{\{\boldsymbol{P}^{c}_{z,\boldsymbol{i}}(\boldsymbol{u}^-)>0\}}}{\boldsymbol{P}^{c}_{\boldsymbol{z,i}}(\boldsymbol{u}^-)} \boldsymbol{Q}^{c}_{z,\boldsymbol{ij}}(\mathrm{d}\boldsymbol{u}),$$

for $\boldsymbol{t} \in (s,\boldsymbol{\tau}]$. Thus, $\boldsymbol{\Lambda}^{(\epsilon)}_{z,\boldsymbol{ij}}(\boldsymbol{t}) \xrightarrow[\epsilon\to 0]{a.s.} \boldsymbol{\Lambda}_{z,\boldsymbol{ij}}(\boldsymbol{t})$, for all $\boldsymbol{t} \in (s,\boldsymbol{\tau}]$. Now $\boldsymbol{\Lambda}^{(\epsilon)}_{z,\boldsymbol{ij}}$ and $\boldsymbol{\Lambda}_{z,\boldsymbol{ij}}$ are both monotonically increasing and bounded on $(s,\boldsymbol{\tau}]$. With arguments similar to the two-dimensional proof of the Glivenko-Cantelli theorem, the point wise convergence can be extended to

$$\sup_{\boldsymbol{t}\in(s,\boldsymbol{\tau}]} \left|\boldsymbol{\Lambda}^{(\epsilon)}_{z,\boldsymbol{ij}}(\boldsymbol{t}) - \boldsymbol{\Lambda}_{z,ij}(\boldsymbol{t})\right| \xrightarrow[\epsilon\to 0]{a.s.} 0.$$

The same argumentation works for the univariate conditional transition rates. The application of the monotone convergence theorem proceeds in the same way as the bivariate case.
For the convergence of transition probabilities, we use the same argumentation we used in the convergence in n, namely integration by parts and Duhamel equality.
Now we have

$$\sup_{\boldsymbol{t}\in(s,\tau]} \left|\boldsymbol{\Lambda}^{(n,\epsilon)}_{z,\boldsymbol{ij}}(\boldsymbol{t}) - \boldsymbol{\Lambda}_{z,\boldsymbol{ij}}(\boldsymbol{t})\right| \leq \sup_{\boldsymbol{t}\in(s,\tau]} \left|\boldsymbol{\Lambda}^{(n,\epsilon)}_{z,\boldsymbol{ij}}(\boldsymbol{t}) - \boldsymbol{\Lambda}^{(\epsilon)}_{z,\boldsymbol{ij}}(\boldsymbol{t})\right| + \sup_{\boldsymbol{t}\in(s,\tau]} \left|\boldsymbol{\Lambda}^{(\epsilon)}_{z,\boldsymbol{ij}}(\boldsymbol{t}) - \boldsymbol{\Lambda}_{z,\boldsymbol{ij}}(\boldsymbol{t})\right|.$$

This directly concludes the proof of the statement. □

Sketch of the proof for Theorem 3.5.8. The proof uses convergences in (3.5.2) and ideas that are analogous to the ideas that were used in the proof of the ϵ convergence of Theorem 3.5.10. □

3.6 Uniform convergence in actuarial estimation

Similar to the general multi-state framework, we define $Z = (Z(t))_{t\geq s}$, as the state process of the insured with the state indicator process $(I_i)_{i\in\mathcal{Z}}$ and the transition counting process $(N_{ij})_{i,j\in\mathcal{Z}}$. The landmark $\xi = Z(s)$, similar to the information model used in Markov modeling in the actuarial context, is realistic in application, but different conditioning is also feasible.
Let B be the insurance cash flow of an individual life insurance contract. We assume B to be an adapted càdlàg process with paths of finite variation. For modeling purposes we use a maximum contract horizon of T, which means that

$$B(\mathrm{d}t) = 0, \quad t > T.$$

Several models have been proposed to estimate the expected value of these future cash flows. The Markov model only allows for restrictive payment functions in the future cash flow $B(t)$ and requires the Markov assumption. The more general semi-Markov model allows for the estimation of cash flows with options as policyholder behavior, see Ahmad et al. in [ABF22]. The disadvantage of this model is that it still introduces systematic model risk because we cannot verify the required Markov assumption and that the estimation is computationally expensive. Therefore, modeling without the Markov assumption in a non-Markov model is the more general way to model future cash flows. This has been done for scaled cash flows by Christiansen and Furrer in [CF22]. There they use an approach that uses contract-dependent estimates to account for policyholder behavior. The disadvantage is that, in practice, re-estimating for different insurance contracts is time-consuming and introduces other problems, particularly in terms of data usage and storage.
The conditional bivariate transition rates and transition probabilities introduced in this paper, allow for an estimation procedure that is independent of the design of the insurance contract and allows for a wide variety of payment functions. In practice, this variability allows a service provider to estimate these rates and probabilities once with its combined data and then distribute them to the insurance companies without distributing the data. This is already good practice, for example, with the German Actuarial Association and its semi-Markov tables.
We now introduce the different cash flows $B(t)$ that we use to model the cash flow of insurance contracts.
Classical Markov modeling uses payment functions that allow payments to depend on the current state of the insured and on current jumps from one state to another. A classical cash flow that includes both of these option is a cash flow with a so-called one-dimensional canonical representation.

Definition 3.6.1.
A stochastic process B is said to have a *one-dimensional canonical representation* if there exist real-valued functions $(B_i)_i$ on $[0,\infty)$ with finite variation on finite intervals and measurable functions $(b_{ij})_{ij:i\neq j}$, which also have finite variation on finite intervals such that

$$B(t) = \sum_{i\in\mathcal{Z}} \int_{(s,t]} I_i(u^-)B_i(\mathrm{d}u) + \sum_{\substack{i,j\in\mathcal{Z}\\ i\neq j}} \int_{(s,t]} b_{ij}(u^-)N_{ij}(\mathrm{d}u), \quad t \geq s. \tag{3.6.1}$$

In practice there often arise cash flows that do not have such a representation. Reasons for that can be contract modifications or general analysis of higher-dimensional cash flows. For this reason we allow for a more complex structure.

Definition 3.6.2.
A stochastic process B is said to have a *two-dimensional canonical representation* if there exist real-valued functions $(B_i)_i$ on $[0,\infty)$ which are a difference of two non-decreasing upper continuous functions, real-valued functions $(B_{ij})_{ij}$ on $[0,\infty)^2$ which are a difference of two non-decreasing upper-continuous functions, and measurable and bounded real-valued

functions $(b_{ikl})_{ikl}$, $(b_{ijkl})_{ijkl}$ on $[0,\infty)^2$ which are also the difference of two non-decreasing upper-continuous functions, such that

$$\begin{aligned}\boldsymbol{B}(\boldsymbol{t}) = &\sum_{i,j\in\mathcal{Z}} \int\limits_{(s,\boldsymbol{t}]} I_i(u_1^-)I_j(u_2^-)B_{ij}(\mathrm{d}u_1,\mathrm{d}u_2)\\ &+ \sum_{\substack{i,k,l\in\mathcal{Z}\\ k\neq l}} \int\limits_{(s,\boldsymbol{t}]} I_i(u_1^-)b_{ikl}(u_1^-,u_2^-)B_i(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2)\\ &+ \sum_{\substack{i,j,k,l\in\mathcal{Z}\\ i\neq j,k\neq l}} \int\limits_{(s,\boldsymbol{t}]} b_{ijkl}(u_1^-,u_2^-)N_{ij}(\mathrm{d}u_1)N_{kl}(\mathrm{d}u_2), \quad \boldsymbol{t}\geq s.\end{aligned} \tag{3.6.2}$$

These processes with one- or two-dimensional canonical representations have been used by Bathke and Christiansen [BC24] to model the conditional variance of future liabilities in an as-if Markov model and cash flows with special types of path-dependent payout functions. They do this by computing conditional expected values of stochastic processes with a one- or two-dimensional canonical representation. We first define the conditional expected values of these cash flows

$$\begin{aligned}\boldsymbol{A}_z(t) &:= \mathbb{E}\Big[\int\limits_{(s,t]^2} \boldsymbol{B}(\mathrm{d}\boldsymbol{u})\Big|\xi = z\Big],\\ A_z(t) &:= \mathbb{E}\Big[\int\limits_{(s,t]} B(\mathrm{d}u)\Big|\xi = z\Big].\end{aligned}$$

If discounting is assumed to be deterministic, the discounting factor can be included into the payment functions. Then the conditional expected value of these future cash flows is a state wise prospective reserve.
In practice, a positive probability landmark should be used to condition the cash flows. These cash flows with a two-dimensional canonical representation have already been studied by Bathke and Christiansen in [BC24]. They showed

$$\begin{aligned}\boldsymbol{A}_z(t) = &\sum_{i,j\in\mathcal{Z}} \int\limits_{(s,t]^2} \boldsymbol{P}_{z,ij}(\boldsymbol{u}^-)B_{ij}(\mathrm{d}\boldsymbol{u})\\ &+ \sum_{\substack{i,k,l\in\mathcal{Z}\\ k\neq l}} I_i(s) \int\limits_{(s,t]^2} b_{ikl}(u_1^-,u_2^-)B_i(\mathrm{d}u_1)P_{z,k}(u_2^-)\Lambda_{z,kl}(\mathrm{d}u_2)\\ &+ \sum_{\substack{\boldsymbol{i}\in\mathcal{Z},\boldsymbol{j}\in\mathcal{Z}\\ i_1\neq j_1}} \int\limits_{(s,t]} \int\limits_{(s,t]\times(s,u_1)} b_{j_2i_1j_1}(u_1^-,u_2^-)\boldsymbol{P}_{z,\boldsymbol{i}}(u_2^-,u_3^-)\boldsymbol{\Lambda}_{z,\boldsymbol{ij}}(\mathrm{d}u_2,\mathrm{d}u_3)B_{j_2}(\mathrm{d}u_1)\\ &+ \sum_{\boldsymbol{i},\boldsymbol{j}\in\mathcal{Z}_{\neq}} \int\limits_{(s,t]^2} b_{i_1j_1i_2j_2}(u_1^-,u_2^-)\boldsymbol{P}_{z,\boldsymbol{i}}(\boldsymbol{u}^-)\boldsymbol{\Lambda}_{z,\boldsymbol{ij}}(\mathrm{d}\boldsymbol{u}^-).\end{aligned} \tag{3.6.3}$$

In the one-dimensional case, Christiansen [Chr21] showed

$$A_z(t) = \sum_{i \in \mathcal{Z}} \int_{(s,t]} P_{z,i}(u^-) B_i(\mathrm{d}u) + \sum_{\substack{i,j \in \mathcal{Z} \\ i \neq j}} \int_{(s,t]} b_{ij}(u^-) P_{z,i}(u^-) \Lambda_{z,ij}(\mathrm{d}u).$$

These $\boldsymbol{A}_z$ and A_z can be understood as a functionals $\boldsymbol{A}_z(P_z, \boldsymbol{P}_z, \Lambda_z, \boldsymbol{\Lambda}_z)$ and $A_z(P_z, \Lambda_z)$ of the uni- and bivariate conditional transition rates and transition probabilities. For the purposes of reserving in the insurer's balance sheet, we now need to estimate these conditional expected values. In practice, we first estimate the conditional transition rates and probabilities as seen in the previous section and then plug them into the functionals $\boldsymbol{A}_z$ and A_z. This works because of the following theorems.

Theorem 3.6.3.
For a stochastic process that has a one-dimensional canonical cash flow representation, the functional A_z is continuous in sup-norm in all arguments, meaning that the convergence in sup-norm in the domain results in convergence in the codomain. Additionally,

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \left| A_z(P_z^{(n,\epsilon)}, \Lambda_z^{(n,\epsilon)})(t) - A_z(P_z, \Lambda_z)(t) \right| \overset{a.s.}{=} 0, \quad \forall t > s.$$

The proof of this theorem follows at the end of this section.

Theorem 3.6.4.
For a stochastic process that has a two-dimensional canonical cash flow representation, the functional $\boldsymbol{A}_z$ is continuous in sup-norm in all 4 arguments, meaning that the convergence in sup-norm in the domain results in convergence in the codomain.
Additionally,

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \left| \boldsymbol{A}_z(P_z^{(n,\epsilon)}, \boldsymbol{P}_z^{(n,\epsilon)}, \Lambda_z^{(n,\epsilon)}, \boldsymbol{\Lambda}_z^{(n,\epsilon)})(t) - \boldsymbol{A}_z(P_z, \boldsymbol{P}_z, \Lambda_z, \boldsymbol{\Lambda}_z)(t) \right| \overset{a.s.}{=} 0, \quad \forall t > s.$$

The proof of this theorem follows at the end of this section.

Remark 3.6.5.
All in all, the convergence we get is an almost sure uniform convergence, including the convergence in ϵ. This type of convergence is different from the already established convergence for the one-dimensional cash flows by Christiansen and Furrer [CF22]. They used $\mathcal{L}^1$ convergence in p variation and manage to circumvent the use of the ϵ perturbation. The $\mathcal{L}^1$ convergence in p-variation is different from almost sure convergence in supremum norm, see Vitalis convergence theorem.

The previous two theorems can now be used to estimate the conditional expected values of one-dimensional canonical cash flows, to estimate the conditional second moment of one-dimensional canonical cash flows that turn out to follow a stochastic process with a two-dimensional canonical representation, or to estimate the conditional expected values of path-dependent cash flows that can be rewritten as a stochastic process with a two-dimensional canonical representation. The general idea follows three steps:

i) Estimate the conditional transition rates from data.

ii) Plug in the estimated conditional transition rates and solve equation (3.5.1) and equation (3.4.4).

iii) Plug these estimators in the functionals $\boldsymbol{A}_z$ and A_z.

Proof of Theorem 3.6.4. We show the theorem for $t = T$, for all $t < T$ the proof works the same. We start by showing the continuity of $\boldsymbol{A}_z$ with the help of Lemma 3.A.3 and especially the fact that $B(F, G) \mapsto \int F(s)\mathrm{d}G(s)$ is Hadamard differentiable. For this differentiability to apply, we need bounded variation of all the integrators and integrands in equation (3.6.3). For the first integral, we have bounded variation in n of $\boldsymbol{P}_z^{(n,\epsilon)}$ and $\boldsymbol{P}_z$ as we have seen in the proof of Theorem 3.5.10 with the use of equation (3.2.3), additionally we have that B_{ij} is of bounded variation because it was defined as the difference of two non-decreasing upper continuous functions.
For the second summand, we need $\int_{(s,T]} b_{ikl}(u_1^-, u_2^-)B_i(\mathrm{d}u_1)$ to be of bounded variation in u_2. For that we use the fact that b_{ikl} and B_i are both functions that can be written as the difference of two non-decreasing functions b_{ikl}^+, b_{ikl}^-, B_i^+ and B_i^-. Plugging all of these in, we can bound the variation by

$$\begin{aligned}\Big\| \int_{(s,T]} b_{ikl}(u_1^-, u_2^-)B_i(\mathrm{d}u_1)\Big\|_{v_1}^{(1)} \leq & \int_{(s,T]} b_{ikl}^+(u_1^-, T) - b_{ikl}^+(u_1^-, s)B_i^+(\mathrm{d}u_1) \\ & + \int_{(s,T]} b_{ikl}^-(u_1^-, T) - b_{ikl}^-(u_1^-, s)B_i^+(\mathrm{d}u_1) \\ & + \int_{(s,T]} b_{ikl}^+(u_1^-, T) - b_{ikl}^+(u_1^-, s)B_i^-\mathrm{d}u_1) \\ & + \int_{(s,T]} b_{ikl}^-(u_1^-, T) - b_{ikl}^-(u_1^-, s)B_i^-(\mathrm{d}u_1).\end{aligned}$$

Additionally, $P_z^{(n,\epsilon)}$ and P_z are of uniform bounded variation, which can be seen similarly to $\boldsymbol{P}_z^{(n,\epsilon)}$ and $\boldsymbol{P}_z$ with equation 3.2.3. For the last two summands, we use the same ideas and the fact, that B_j, b_{ijk}, and b_{ijkl} are all the difference of two non-decreasing functions. Thus, we have proved the continuity of A_z and with that found a way to estimate the conditional expected value of a stochastic process with a two-dimensional canonical cash flow representation. □

Sketch of the proof of Theorem 3.6.3. The proof of this theorem uses the same ideas already used for the continuity of the two-dimensional cash flow. We use the fact that A_i and a_{ij} both have finite variation and the fact, that integrals of these forms are Hadamard differentiable, see Lemma 3.A.3. □

3.7 Numerical example

We consider a numerical example, where the bivariate conditional transition rates and transition probabilities arise naturally in the conditional expected values of a future cash flow. The focus of the numerical study is to illustrate how our two-dimensional cash flows from Section 3.6 can be applied in practice and how censoring affects the estimations. This is intended to be an exploratory example to illustrate an actuarial application of the estimation of bivariate conditional transition rates.
The basic setup is similar to that used for the numerical study in the paper by Buchardt and Møller [BM15]. As a technical basis, we consider a survival model, i.e. a Markov model with the two states alive (1) and dead (4). The modeled insurance contract starts at the age of 40 and the retirement age is 65. The insured pays a lump sum of 100,000 at the beginning of the contract and an annual premium of 10,000 until the insured reaches the age of 65 and receives annual pension payments. All payments cease when the insured dies or reaches the age of 100. We get the following sojourn payment function:

$$B_1(\mathrm{dt}) := (-10,000\mathbb{1}_{\{t<65\}} + p\mathbb{1}_{\{100\geq t\geq 65\}})\,\mathrm{dt},$$

where p is the annuity calculated on the technical basis according to the equivalence principle. The technical basis is generally used to price insurance contracts, while the extended model of the market basis is used for the insurers' balance sheets. The technical basis uses the transition intensity

$$\mu^{\star}_{14}(t) := 0.005 + 10^{(5.728-10+0.038t)},$$

and an interest rate of 0%. The market basis consists of four states, with states (1) and (4) denoting alive and dead states, mirroring the technical basis. In addition, state (2) denotes the free policy option and state (3) denotes surrender, where free policy means that premiums are waived and benefits are reduced accordingly. Surrender means a complete cessation of both benefits and payments with a final transitional payment, usually in the form of the current reserve. We therefore introduce two options that model policyholder behavior for the market basis. The additional payment functions are modeled as follows:

$$\begin{aligned} b_{13}(t) &= V_1(t), \\ b_{23}(t,u) &= V_1^{+}(t)\rho(u), \end{aligned}$$

for the transition payments, where $V_1^{+}(t)$ is the technical reserve of the pension payments, and

$$B_2(\mathrm{dt},u) := \rho(u)p\mathbb{1}_{\{100\geq t\geq 65\}}\,\mathrm{dt},$$

for the sojourn payments, where p is the premium that is paid in state (1). The factor $\rho(u)$ defines the reduction of the benefits in the free policy state (2). It is calculated by

$$\rho(u) = \frac{A_1^{+}(u) - A_1^{-}(u)}{A_1^{+}(u)},$$

where $A_1^-(t)$ is the technical reserve of premium payments. The dependence of the payments functions on u models the dependence on the transition time to state (2). For this reason, bivariate transition rates are natural in this application. See the work of Ahmad et al. in [ABF22] for the theoretical background of free policy option calculations. We model the market basis as a semi-Markov model with transition intensities

$$\begin{aligned}
\mu_{14}(t) &:= \mu_{14}^{\star}(t),\\
\mu_{24}(t) &:= \mu_{14}^{\star}(t),\\
\mu_{12}(t) &:= 0.1\mathbb{1}_{\{t\leq 65\}}\\
\mu_{13}(t) &:= 0.05\mathbb{1}_{\{t\leq 65\}}\\
\mu_{23}(t,u) &:= (0.05 + 0.2\mathbb{1}_{\{u\in[\frac{1}{2},\frac{5}{2})\}})\mathbb{1}_{\{t\leq 65\}},
\end{aligned}$$

and an interest rate of 3% annually. We estimate the state wise prospective reserve of the market basis in an as-if Markov model, i.e. the landmark is $\xi = Z(40)$ and since 40 is the starting point of the insurance contract, we assume that each insured is in state (1) at $s = 40$. For an as-if Markov model and a cash flow with scaled payments such as the free policy option, Bathke and Christiansen in [BC24] found a two-dimensional canonical representation for the state wise prospective reserve. Adapting it to this example, we get

$$\begin{aligned}
\boldsymbol{A}_1(t) = &\int_{(40,t]} \frac{\kappa(40)}{\kappa(u)} p_{1,1}(u^-) B_1(\mathrm{d}u) + \int_{(40,t]} \frac{\kappa(40)}{\kappa(u)} b_{13}(u) p_{1,1}(u^-) \Lambda_{1,(1,3)}(\mathrm{d}u)\\
&+ \int_{(40,t]} \int_{(40,65]\times(40,t)} \frac{\kappa(40)}{\kappa(t)} \rho(u_1) \Big(\boldsymbol{p}_{1,(1,1)}(\boldsymbol{u}^-) \boldsymbol{\Lambda}_{1,(1,1)(2,2)}(\mathrm{d}\boldsymbol{u})\\
&\qquad - \sum_{k\in\{3,4\}} \boldsymbol{p}_{1,(1,2)}(\boldsymbol{u}^-) \boldsymbol{\Lambda}_{1,(1,2)(2,k)}(\mathrm{d}\boldsymbol{u}) \Big) p \mathbb{1}_{\{t\geq 65\}} \mathrm{d}t\\
&+ \int_{(40,t]\times(40,t]} \frac{\kappa(s)}{\kappa(u_1)} \rho(u_1) b_{23}(u_2, u_1) \boldsymbol{p}_{1,(1,2)}(\boldsymbol{u}^-) \boldsymbol{\Lambda}_{1,(1,2)(2,3)}(\mathrm{d}\boldsymbol{u}),
\end{aligned} \tag{3.7.1}$$

where $t \leq 100$ and $\kappa(\cdot)$ is the value of a deterministic cash account. The prospective reserve is an example where the estimation using the as-if Markov approach is natural, because we have jumps from state (2) to state (3) or (4) in the cash flow that are dependent on the jump from state (1) to state (2). Such dependencies are the strength of the conditional bivariate estimation.

We introduce some technical background to the estimation of the prospective reserve in the as-if Markov model. The estimation of the conditional bivariate transition rates and transition probabilities for the market basis can be simplified. We can reduce the required transition rates to $\Lambda_{(1,2),(2,4)}$,$\Lambda_{(1,2),(2,3)}$, $\Lambda_{(1,2),(1,2)}$, $\Lambda_{(1,4),(1,4)}$ and $\Lambda_{(1,3),(1,3)}$ where the last two are only needed to estimate the transition probability p_{11}. This reduction originates from the reduced number of possible jumps in the market basis. The required transition rates can also be reduced. We only need p_{11}, p_{12} and p_{13}. This greatly reduces computation

time and space.
Counting the number of steps required to compute the transition rates and probabilities, we get a time complexity of order $\mathcal{O}(n^2)$ and a space complexity of order $\mathcal{O}(n^2)$. Intuitively, we get this complexity because there are $\mathcal{O}(n^2)$ time points on the time grid given by the data, and we need to recursively compute the estimator on this grid.
To compute the plug-in estimator, we can use a type of differential equation, in this case a simple recursive formula due to the finiteness of the grid points. This reduces the computational complexity for the computation of the state wise prospective reserves from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$. We use the decomposition

$$\boldsymbol{A}_1(P_1^{(n,\epsilon)}, \boldsymbol{P}_1^{(n,\epsilon)}, \Lambda_1^{(n,\epsilon)}, \boldsymbol{\Lambda}_1^{(n,\epsilon)})(t) - \boldsymbol{A}_1(0) =: C_1(t) + C_2(t) + C_3(t), \quad t \geq 40,$$

where C_1 contains the first two, C_2 the third and C_3 the fourth summand of the equation (3.7.1), with the estimated transition rates and transition probabilities plugged in. Now all the three summands can be calculated recursively. Let $\{t_n\}_{n\in I}$ be the the set of all jump points defining the time grid imposed by the data. We start with $t_0 = 40$ and $C_1(40) = C_2(40) = C_3(40) = 0$. We get

$$C_1(t_n) = C_1(t_{n-1}) + \int_{t_{n-1}}^{t_n} p_{1,j}^{(n,\epsilon)}(s-)\Big(B_1(\mathrm{d}s) + b_{13}(s)\Lambda_{1,13}^{(n,\epsilon)}(\mathrm{d}s)\Big).$$

For the second one, we get

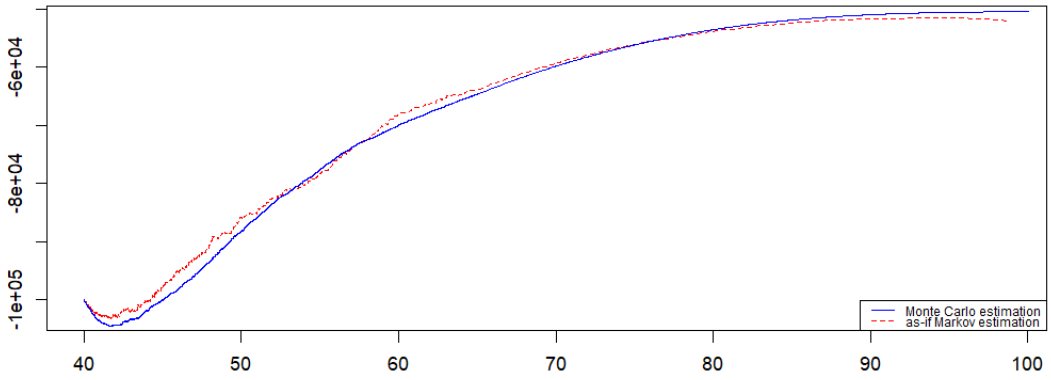
$$\begin{aligned} C_2(t_n) = C_2(t_{n-1}) + \int_{(40,t_{n-1}-)}^{(65,t_n-)} \int_{u_2}^{t_n} B_2(\mathrm{d}s, u_1)\Big(&\boldsymbol{P}_{1,(1,1)}^{(n,\epsilon)}(\boldsymbol{u}-)\boldsymbol{\Lambda}_{1,(1,1)(2,2)}^{(n,\epsilon)}(\mathrm{d}\boldsymbol{u}) \\ &- \sum_{k\in\{3,4\}} \boldsymbol{P}_{1,(1,2)}^{(n,\epsilon)}(\boldsymbol{u}-)\boldsymbol{\Lambda}_{1,(1,2)(2,k)}^{(n,\epsilon)}(\mathrm{d}\boldsymbol{u})\Big) \\ + \int_{t_{n-1}}^{t_n} p\mathbb{1}_{\{t\geq 65\}}\mathrm{d}(s) \int_{(40,40)}^{(65,t_{n-1}-)} \rho(u_1)\Big(&\boldsymbol{P}_{1,(1,1)}^{(n,\epsilon)}(\boldsymbol{u}-)\boldsymbol{\Lambda}_{1,(1,1)(2,2)}^{(n,\epsilon)}(\mathrm{d}\boldsymbol{u}) \\ &- \sum_{k\in\{3,4\}} \boldsymbol{P}_{1,\boldsymbol{j}}^{(n,\epsilon)}(\boldsymbol{u}-)\boldsymbol{\Lambda}_{1,(1,2)(2,k)}^{(n,\epsilon)}(\mathrm{d}\boldsymbol{u})\Big). \end{aligned}$$

The last integral in the formula for C_2 can again be calculated recursively in every step. For C_3, we get

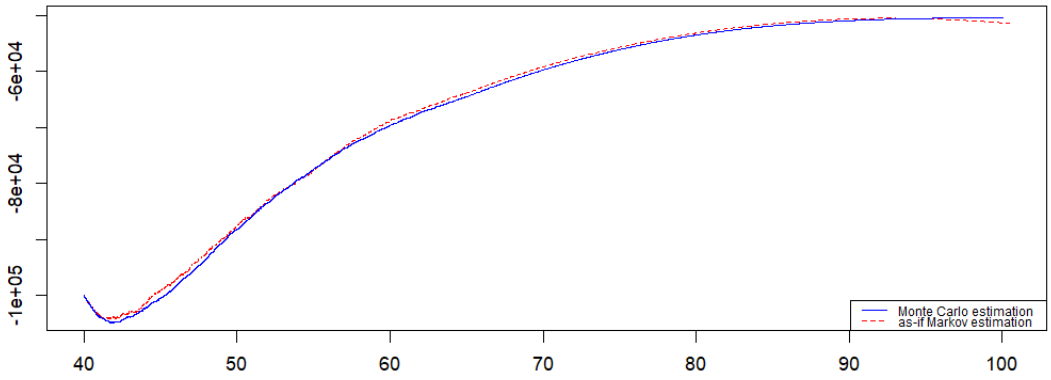
$$C_3(t_n) = C_3(t_{n-1}) + \int_{(40,t_{n-1})}^{(65,t_n)} b_{23}(u_2, u_1)\boldsymbol{P}_{1,(12)}^{(n,\epsilon)}(\boldsymbol{u}-)\boldsymbol{\Lambda}_{1,(1,2)(2,3)}^{(n,\epsilon)}(\mathrm{d}\boldsymbol{u}).$$

We estimate in R [R C23] by generating artificial data following the semi-Markov model described as the market basis using the package *AalenJohansen* by Bladt and Furrer [BF23].
We simulate $5,000$ independent and identically distributed realizations for the defined insurance contract with scaled payments. We use $\mathrm{Unif}(65, 120)$ distributed random variables as right censoring.

To illustrate the numerical feasibility and strength of the plug-in estimator, we plot in figures 3.1a and 3.1b the estimate for the state wise prospective reserve of the insurance



(a) $n = 1000$



(b) $n = 5000$

Figure 3.1: Prospective reserve $\boldsymbol{A}_1(t)$ with time horizon $T = 100$ in the as-if Markov model estimated with $n = 1000$ and $n = 5000$ censored individuals, the plug-in estimator and the landmark Nelson-Aalen and Aalen-Johansen estimators versus Monte Carlo estimation with $n = 10000$ uncensored individuals.

contract in an as-if Markov model on the market basis for $n = 1000$ and $n = 5000$ individuals, together with a Monte Carlo simulation of the same state wise prospective reserve without censoring and $10,000$ individuals for comparison. We see that the bivariate estimator, with $n = 5000$, estimates the state wise prospective of the particular insurance product similarly to the Monte Carlo estimator, which is using uncensored data and more individuals. The convergence of the estimator for a reasonable sample size n is illustrated in this numerical example. However, making further quantitative statements about the

speed of convergence or confidence bounds would require either four-dimensional covariance estimation or bootstrapping, both of which are numerically challenging in our case.
A challenge for the interpretation of the estimator $P^{(n,\epsilon)}_{z,(jk)}$ is that in the censored case the bivariate estimator may not satisfy the expected monotonicity and positivity properties. This property is very similar to the findings in [Dab88] where they found the Kaplan-Meier estimator on the plane to be non-monotonic. This explains the higher volatility observed in the plug-in estimator for $n = 1000$ but does not change the fact that the estimator consistently estimates the state wise prospective reserve, even when censoring is involved. Moreover, the estimators are easy to apply, since we do not need to solve any continuous differential equations, but were able to obtain recursive formulas for the estimators at each level.
We conclude that, despite numerical challenges, the bivariate estimator can be successfully applied to the estimation of state wise prospective reserves in non-Markov models. The bivariate landmark Nelson-Aalen and Aalen-Johansen estimators, combined with the plug-in estimators for state wise prospective reserve, provide a versatile tool because the theory applies not only to insurance contracts with contract modifications, but also to other cash flows that have a two-dimensional representation.

3.8 Concluding remarks

In this paper, we generalized the landmark Nelson-Aalen and the landmark Aalen-Johansen estimators to the estimation of bivariate conditional transition rates and bivariate transition probabilities in general non-Markov models as introduced by Bathke and Christiansen [BC24]. The analysis of intertemporal dependencies via bivariate conditional transition rates is necessary in non-Markov multi-state modeling, in contrast to classical Markov modeling, because the absence of the Markov assumption leads to incomplete distribution knowledge gained from univariate conditional transition rates.
The use of analytical techniques well-known from bivariate Kaplan-Meier estimation and conditional Aalen-Johansen estimation leads to uniform convergence of both estimators. We obtain a blueprint for the estimation of bivariate conditional transition rates and transition probabilities from right censored data.
Future research endeavors could further address the question of asymptotic normality and bootstrapping. Additionally, an extension of the uniform convergence to the entire plane, possibly via martingale techniques, is of interest.
Finally, this study closes a gap between techniques in survival analysis and premium and reserve calculation for insurance contracts. In the case of an insurance contract with scaled payments, we construct estimators for the first moment of all future payments and present a proof of concept with a numerical example using simulated data. This is the basis of any type of risk assessment, where the state process of the insured follows the non-restrictive non-Markov model. Furthermore, estimating bivariate conditional transition rates and transition probabilities allows the estimation of second moments of future liabilities with a canonical one-dimensional cash flow in the non-Markov model. Altogether, this can be used to evaluate the widespread use of Markov assumptions in the reserving and premium calculation for insurance portfolios.

Acknowledgments

The author would like to express gratitude to Marcus Christiansen and Christian Furrer for their invaluable discussions regarding the direction and theoretical background of this paper.

3.A Appendix

The following first two lemmas are two-dimensional generalizations of well-known statements in the field of product integrals. These generalizations have already been mentioned in Gill & Johansen [GJ90] and Gill, Van Der Laan & Wellner [GLW95].

Lemma 3.A.1.
For a monotonically increasing real-matrix valued function Λ and a function $\phi(\boldsymbol{t})$, the equation

$$Y(\boldsymbol{t}) = \phi(\boldsymbol{t}) + \int_{(s,\boldsymbol{t}]} Y(\boldsymbol{u}^-)\Lambda(\mathrm{d}\boldsymbol{u}), \tag{3.A.1}$$

has the solution

$$\tilde{P}(\boldsymbol{t}) = \phi(\boldsymbol{t}) + \int_{(s,\boldsymbol{t}]} \phi(\boldsymbol{u}^-)\Lambda(\mathrm{d}\boldsymbol{u})\mathcal{P}((\boldsymbol{u},\boldsymbol{t}],\Lambda), \tag{3.A.2}$$

where we use the Peano-series definition

$$\mathcal{P}((s,\boldsymbol{t}],\Lambda) := Id + \sum_{n=1}^{\infty} \int_{s<\boldsymbol{u}^{(1)}<\ldots<\boldsymbol{u}^{(n)}\leq\boldsymbol{t}} \cdots \int \Lambda(\mathrm{d}\boldsymbol{u}^{(1)}) \cdots \Lambda(\mathrm{d}\boldsymbol{u}^{(n)}).$$

In the definition of the Peano-series we use the usual partial ordering of $\mathbb{R}^2$ for the definition of the domain.

Proof. See the proof of Proposition 3.1 from Gill, Van Der Laan & Wellner [GLW95]. □

The next Lemma is a generalization of the well-known Duhamel equality.

Lemma 3.A.2.
For monotonically increasing functions $A, B : \mathbb{R}^2 \to \mathbb{R}^{n\times n}$ we have the following equation

$$\mathcal{P}(R,A) - \mathcal{P}(R,B) = \int_R \mathcal{P}(R\cap(\mathbf{0},s),A)\,(A(\mathrm{d}s) - B(\mathrm{d}s))\,\mathcal{P}(R\cap(s,\infty),B),$$

for a rectangle $R \subset \mathbb{R}^2$.

Proof. See the proof of Proposition 3.3 from Gill, Van Der Laan & Wellner [GLW95]. □

Lemma 3.A.3.
The functional

$$A : (F, G) \mapsto \int \frac{1}{F(s)} \mathrm{d}G(s)$$

is continuous at any point (F, G) on the space of càdlàg functions with bounded variation endowed with the supremum norm, where $\frac{1}{F}$ is of bounded variation and bounded away from 0 and G is of bounded variations for sequences $\frac{1}{F_n}$ and G_n that are uniformly of bounded variation as well.

Proof. This follows immediately by the weak continuous Hadamard differentiability of the functional $B : (F, G) \mapsto \int F(s)\mathrm{d}G(s)$ and the fact that $F \mapsto \frac{1}{F}$ is also weak continuous Hadamard differentiable as long as $F \geq \delta > 0$. The statement can be found in Gill, Van Der Laan & Wellner[GLW95] Lemma 5.1. □

Additionally, we need a two-dimensional generalization of the well known Campbell theorem, see for example Milbrodt and Helbig [MH99].

Theorem 3.A.4. *Campbell Theorem*
Let η be a point process on $(\mathbb{R}^d, \mathbb{B}(\mathbb{R}^d))$ with intensity measure λ and let $u : \mathbb{R}^d \to \mathbb{R}$ be a measurable function. Then

$$\int u(x)\eta(\mathrm{d}x)$$

is a random variable and

$$\mathbb{E}\left[\int u(x)\eta(\mathrm{d}x)\right] = \int u(x)\lambda(\mathrm{d}x), \tag{3.A.3}$$

whenever $u \geq 0$ or $\int |u(x)|\lambda(\mathrm{d}x) < \infty$.

Proof. The proof of this can be easily adapted from the proof of the one-dimensional Campbell Theorem, see for example Milbrodt and Helbig [MH99]. □

For the estimation of integrals, we use the so called one and two-dimensional variation of a process. We do a short introduction here.

Definition 3.A.5.
We define the one-dimensional variation of a function $f : [a, b] \to \mathbb{R}$ as

$$\|f\|_{v_1}^{(1)} := \|f\|_\infty + \sup_{\tau=\{t_i\}_{i\leq n}} \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|,$$

where the supremum goes over all finite partitions τ of $[a, b]$. The two-dimensional variation of a function $f : [a, b] \times [c, d] \to \mathbb{R}$ is defined as

$$\begin{aligned}\|f\|_{v_1}^{(2)} := \sup_{\tau=\{t_i\}\times\{\tilde{t}_j\}_{i,j\leq n}} \sum_{i=1}^{n}\sum_{j=1}^{n} |f(t_{i-1}, \tilde{t}_{j-1}) - f(t_{i-1}, \tilde{t}_j) - f(t_i, \tilde{t}_{j-1}) + f(t_i, \tilde{t}_j)| \\ + \|f(a, \cdot)\|_{v_1}^{(1)} + \|f(\cdot, c)\|_{v_1}^{(1)} - \|f\|_\infty,\end{aligned}$$

where the supremum goes over all finite partitions of $[a, b] \times [c, d]$.

This variation definition is mainly used for the following integral inequality

Lemma 3.A.6.

$$\left| \int f(\boldsymbol{u}) g(\mathrm{d}\boldsymbol{u}) \right| \leq \|f\|_\infty \|g\|_{v_1}^{(2)},$$

when f is bounded and g has finite two-dimensional variation.

The proof bases on work from Clarkson and Adams [CA33] and Adams and Clarkson [AC34], see Theorem 5.

Chapter 4

Non-parametric estimators of scaled cash flows

Abstract

In multi-state life insurance, incidental policyholder behavior gives rise to expected cash flows that are not easily targeted by classic non-parametric estimators if data is subject to sampling effects. We introduce a scaled version of the classic Aalen–Johansen estimator that overcomes this challenge. Strong uniform consistency and asymptotic normality are established under entirely random right-censoring, subject to lax moment conditions on the multivariate counting process. In a simulation study, the estimator outperforms earlier proposals from the literature. Finally, we showcase the potential of the presented method to other areas of actuarial science.

Keywords: Aalen–Johansen; forward rates; incidental policyholder behavior; life insurance; non-Markov model.

4.1 Introduction

In the mathematics of life insurance, multi-state modeling has been a canonical approach for more than half a century, owing back to the unification based on Markov chains in [Hoe69]. Today, popular models in both theory and practice include Markov chain models [Nor91; MS97] and semi-Markov models [Hoe72; JD84; Hel08; Chr12; BMS15], but also more sophisticated alternatives such as doubly stochastic and aggregate models [Buc17; ABF23] have found some use. Until recently, the associated actuarial literature focused primarily on valuation techniques and computational aspects, while statistical methodology played a somewhat lesser role.

Estimation in multi-state models is, generally speaking, a challenging undertaking, not least due to sampling effects such as right-censoring. In Markov models, the celebrated Nelson–Aalen and Aalen–Johansen estimators [Aal78; AJ78] may be applied directly for non-parametric predictive purposes or instead for model diagnostics and performance assessment of (semi-)parametric alternatives. Recently, the idea of fully non-Markov modeling has gained traction – not only in actuarial science [Chr21; CF22], but also in

biostatistics and other related fields [PS18; Mal+21; Nie+23]. By imposing no assumption whatsoever on the intertemporal dependence structure of the state process, one seeks to reduce systematic risk and increase robustness. Simulation studies underpin that, if there is a substantial departure from Markovianity, a robust approach is clearly to be preferred [PS18].

Importantly, the Aalen–Johansen estimator remains a consistent and asymptotically normal estimator for occupation probabilities even when the Markov assumption is relaxed [DS01; Gli02]. The landmark Aalen–Johansen estimator, which just invokes sub-sampling, extends this directly to transition probabilities [PS18] and, if the insurance contract admits a so-called deterministic canonical cash flow representation, also to state-wise prospective reserves [Chr21] based on the plug-in principle.

An insurance contract admits a deterministic canonical cash flow representation if the associated sojourn and transition payments are deterministic, see Definition 5.1 in [Chr21] for the specifics. This assumption is often violated in practice, a key example being so-called scaled payments [CF22]. Such payments arise naturally in the context of implicit policyholder options, especially in connection with free policy conversion and stochastic retirement, see [Hen+14; BM15; BMS15; GN16; Fur22]. Here, the contractual payments are scaled by a factor depending on the exercise time(s) of the option(s), as to account for the fact that free policy conversion or early retirement should lead to reduced future benefits.

The actuarial literature currently offers two non-parametric approaches that are able to handle scaled payments in the non-Markov regime under entirely random right-censoring. The first is the one-dimensional estimator of [CF22], which is of usual Nelson–Aalen and Aalen–Johansen type, but uses auxiliary data based on a change of measure argument lifted from [Fur22]. The second is the two-dimensional estimator of [Bat24], which is an extension of the Volterra estimator of P.J. Bickel for bivariate survival functions [Dab88; GLW95] to multi-state models. Both estimators possess certain deficiencies, however. The auxiliary data used by the estimator of [CF22] comes with extra noise not originally present in the sample. The estimator of [Bat24] is computationally demanding and inherits the inferior tail behavior of the Volterra estimator.

This paper introduces a novel scaled (or weighted) one-dimensional Aalen–Johansen estimator that resolves the aforementioned issues. We establish strong uniform consistency and asymptotic normality of the estimator. En passant, we also establish asymptotic normality of the two-dimensional estimator from [Bat24]. Finally, we compare the finite sample performances of the estimators in a simulation study. Overall, the scaled estimator is found to have the best performance.

Throughout, primary focus lies on scaled payments arising from the inclusion of policyholder options, which imposes a specific structure on the scaling. However, the theoretical insights do not really rely on this structure, and our results are therefore more broadly applicable, including to situations where scaling is due to discounting. In particular, this leads to an extension of Theorem 3.1 of [AB23] to the non-Markov regime. Further, it allows for a constructive and simultaneous definition of forward interest and forward transition rates,

addressing lasting concerns in the actuarial literature [MP05; Nor10; Buc14; CN15; Buc17; BFS19].

The paper is structured as follows. In Section 4.2, we introduce the basic multi-state setup, including the scaled payment process. Section 4.3 focuses on representations of the expected cash flow and the resulting targets for estimation. The estimators of interest are defined in Section 4.4, which also contains the main results on consistency and asymptotic normality. In Section 4.5, we present the comparative simulation study. Finally, Section 4.6 showcases the broader actuarial applicability of our approach.

4.2 Preliminaries

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be an underlying filtered probability space. The state of the insured is modeled by a non-explosive jump process $Z = (Z_t)_{t\geq 0}$ on a finite state space $\mathcal{J}$ with $Z_0 \equiv z_0 \in \mathcal{J}$. Note that Z is not assumed to be say Markov or semi-Markov. We associate with Z a multivariate counting process N, which components $N_{jk} = (N_{jk}(t))_{t\geq 0}$ are given by

$$N_{jk}(t) = \#\{s \in (0,t] : Z_{s-} = j, Z_s = j\}.$$

Throughout, we assume that $\mathbb{E}[N_{jk}(t)] < \infty$ for all components, which in particular implies the aforementioned non-explosion of Z.

Incidental modeling of policyholder behavior such as free policy conversion and stochastic retirement proceeds by considering a state space of the form $\mathcal{J} = \mathcal{J}_0 \cup \mathcal{J}_1$, where $\mathcal{J}_0$ and $\mathcal{J}_1$ are the states pre- and post-exercise of the policyholder option, respectively. This interpretation is meaningful as long as we require $z_0 \in \mathcal{J}_0$ and $\mathcal{J}_1$ to be absorbing for Z, that is

$$\mathbb{P}(Z_t \in \mathcal{J}_1, Z_s \in \mathcal{J}_0) = 0 \text{ for } t \leq s.$$

These assumptions are thus made throughout.

The first (and only) hitting time of $\mathcal{J}_1$ by Z is denoted τ:

$$\tau = \inf\{t > 0 : Z_t \in \mathcal{J}_1\}.$$

We can interpret τ as the exercise time of the policyholder option.

Insurance contracts may be described by payment streams consisting of benefits less premiums. We are interested in contracts of the scaled form

$$\begin{aligned}
B(\mathrm{d}t) &= H(t)B^\circ(\mathrm{d}t), & B(0) &= B^\circ(0),\\
H(t) &= \rho(\tau, Z_{\tau-}, Z_\tau)^{\mathbb{1}_{\{\tau\leq t\}}},\\
B^\circ(\mathrm{d}t) &= \sum_{j\in\mathcal{J}} \mathbb{1}_{\{Z_{t-}=j\}} B_j(\mathrm{d}t) + \sum_{\substack{j,k\in\mathcal{J}\\ j\neq k}} b_{jk}(t) N_{jk}(\mathrm{d}t), & B^\circ(0) &\equiv b_0 \in \mathbb{R},
\end{aligned}$$

where each B_j is a right-continuous real function of finite variation, describing the deterministic sojourn payments in the respective state; each b_{jk} is a measurable real function which is bounded on compacts, describing the deterministic transition payments between the respective states; and each scaling factor $[0,\infty) \ni t \mapsto \rho(t,j,k)$ is non-negative and predictable with respect to the information generated by Z, the filtration $\mathbb{F}^Z$. In practice, it is common to determine the scaling factors as to maintain actuarial equivalence on some safe-side basis and, due to the typically simple nature of that basis, each scaling factor would be deterministic rather than just predictable.

In these contracts, all payments are scaled by a common factor upon exercise of the policyholder option. If the factor is below one, the payments are decreased, while if it is above one, the payments are increased. The idea is that, for instance, early retirement should lead to reduced retirement benefits, while delayed retirement should lead to increased retirement benefits. Throughout, we assume that each $\rho(\cdot,j,k)$ is uniformly bounded on compacts. However, the specific form of B in terms of B° and ρ has a redundancy and thus allows one to reparametrize the model such that each scaling factor $\rho(\cdot,j,k)$ is uniformly bounded by one.

Solely to decrease the notational burden, we assume that $\mathbb{P}(\tau = t)B_j(\mathrm{d}t) = 0$. In other words, lump sum sojourn payments may not occur simultaneously with the exercise of the policyholder option. If such lump sum sojourn payments are possible, they might always be included as transition payments from $\mathcal{J}_0$ to $\mathcal{J}_1$ instead.

In case of a finite maximal contract time $T \in (0,\infty)$, the prospective reserve at contract inception reads

$$V(0-) = b_0 + V(0), \qquad V(0) = \int_0^T \frac{1}{\kappa(t)} A(\mathrm{d}t),$$

for a deterministic savings account κ assumed right-continuous, of finite variation, and bounded away from zero, and where the expected accumulated cash flow A is defined according to $A(t) = \mathbb{E}[B(t) - B(0)]$.

In this paper, we focus on non-parametric estimation of $V(0-)$ through A. This might at first appear as a rather narrow approach. For instance, the as-if-Markov approach of [Chr21; CF22] considers expected accumulated cash flows and prospective reserves on the form

$$A(t_0,t) = \mathbb{E}[B(t) - B(t_0)\,|Z_{t_0}],$$
$$V(t_0) = \int_{t_0}^T \frac{\kappa(t_0)}{\kappa(t)} A(t_0,\mathrm{d}t).$$

However, our approach may be adapted to also deal with this situation. The main idea is to use the landmark Z_{t_0}, meaning one should condition on or sub-sample with respect to events of the type $\{Z_{t_0} = i\}$, which is also the approach adopted in [Chr21; CF22]. We merely focus on the situation at contract inception to reduce the notational burden. If conditioning with respect to Z_{t_0} is replaced by more sophisticated information such as (Z_{t_0}, U_{t_0}), where U is the duration process associated with Z, kernel methods could be used, confer with [BF24b]. Kernel methods are outside the scope of this paper.

4.3 Valuation

If data consisted of complete observations of n independent insured, with the ℓ'th insured represented by Z^ℓ, we could use the estimator

$$[0,\infty) \ni t \mapsto \frac{1}{n}\sum_{\ell=1}^{n} B^\ell(t),$$

where B^ℓ refers to B, but with Z replaced by Z^ℓ and similarly for the associated multivariate counting processes.

However, complete observation is unrealistic. Unless the insurance portfolio is closed, insured enter and exit it at random times. We therefore seek a representation of the expected accumulated cash flow with components that may naturally be targeted for non-parametric estimation even when observation is incomplete. This typically involves hazards.

4.3.1 One-dimensional representations

The first representation is inspired by results in [CF22]. The approach of [CF22] utilizes the general change of measure techniques developed in [Fur22]. We offer a slight adjustment of this approach that is both more direct and avoids the condition that the scaling factors $\rho(\cdot, j, k)$ be uniformly bounded by one.

Define p_j^ρ, p_{jk}^ρ, p_j, and p_{jk} according to

$$\begin{aligned} p_j^\rho(t) &= \mathbb{E}[H(t)\mathbb{1}_{\{Z_t=j\}}], & p_j(t) &= \mathbb{E}[\mathbb{1}_{\{Z_t=j\}}], & j &\in \mathcal{J}, \\ p_{jk}^\rho(t) &= \mathbb{E}\Big[\int_0^t H(s)N_{jk}(\mathrm{d}s)\Big], & p_{jk}(t) &= \mathbb{E}[N_{jk}(t)], & j,k &\in \mathcal{J}, j \neq k. \end{aligned}$$

Further, define Λ_{jk}^ρ and Λ_{jk} according to

$$\Lambda_{jk}^\rho(t) = \int_0^t \frac{1}{p_j^\rho(s-)} p_{jk}^\rho(\mathrm{d}s), \qquad \Lambda_{jk}(t) = \int_0^t \frac{1}{p_j(s-)} p_{jk}(\mathrm{d}s), \qquad j,k \in \mathcal{J}, j \neq k.$$

Remark 4.3.1.
Note that for $j,k \in \mathcal{J}_0$, $j \neq k$, it holds that $p_j^\rho = p_j$ and $p_{jk}^\rho = p_{jk}$ and therefore $\Lambda_{jk}^\rho = \Lambda_{jk}$. Furthermore, for $j \in \mathcal{J}_1, k \in \mathcal{J}_0$ it holds that $\Lambda_{jk}^\rho = \Lambda_{jk} = 0$.

Throughout, we shall assume that $\Lambda_{jk}(t) < \infty$ and $\Lambda_{jk}^\rho(t) < \infty$. Since each $\rho(\cdot, j, k)$ is assumed uniformly bounded on compacts, the former implies the latter if also $\rho(\cdot, j, k)$ is uniformly bounded away from zero on the support of τ.

Remark 4.3.2.
That each $\rho(\cdot, j, k)$ is uniformly bounded away from zero on the support of τ is a sufficient but not necessary condition for $\Lambda_{jk}(t) < \infty$ to imply $\Lambda^{\rho}_{jk}(t) < \infty$. Let $j, k \in \mathcal{J}_1$, $j \neq k$, which is the interesting case. If for instance the compensator C_{jk} of N_{jk} is given by

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

compare with the case of Z being Markov, and $\rho(\cdot, j, k)$ is strictly positive, then

$$\begin{aligned} p^{\rho}_{jk}(\mathrm{d}t) &= \mathbb{E}[\rho(\tau, Z_{\tau-}, Z_\tau)\mathbb{P}(Z_{t-} = j \mid Z_\tau)]\alpha_{jk}(\mathrm{d}t) \\ &= \mathbb{E}[\rho(\tau, Z_{\tau-}, Z_\tau)\mathbb{1}_{\{Z_{t-}=j\}}]\alpha_{jk}(\mathrm{d}t) = p^{\rho}_j(t-)\alpha_{jk}(\mathrm{d}t), \end{aligned}$$

so that $\Lambda^{\rho}_{jk}(\mathrm{d}t) = \alpha_{jk}(\mathrm{d}t)$. In similar fashion, it holds that $\Lambda_{jk}(\mathrm{d}t) = \alpha_{jk}(\mathrm{d}t)$, so that $\Lambda^{\rho}_{jk} = \Lambda_{jk}$. Therefore, whether each $\rho(\cdot, j, k)$ is uniformly bounded away from zero on the support of τ or not, it holds that $\Lambda_{jk}(t) < \infty$ implies $\Lambda^{\rho}_{jk}(t) < \infty$.

In the following, let Λ_{jj}, $j \in \mathcal{J}$, be such that all row sums are constantly equal to zero, that is $-\Lambda_{jj} = \sum_{k \neq j} \Lambda_{jk}$. To the contrary, let Λ^{ρ}_{jj} be given by

$$\begin{aligned} -\Lambda^{\rho}_{jj} &= -\Lambda_{jj}, && j \in \mathcal{J}_0, \\ -\Lambda^{\rho}_{jj} &= \sum_{\substack{k \in \mathcal{J} \\ k \neq j}} \Lambda^{\rho}_{jk}, && j \in \mathcal{J}_1. \end{aligned}$$

Note that all row sums $\sum_{k \in \mathcal{J}} \Lambda^{\rho}_{jk}$ are not necessarily constantly equal to zero, unless each $\rho(\cdot, j, k)$ is constantly equal to one.

Proposition 4.3.3.
It holds that

$$A(\mathrm{d}t) = \sum_{j \in \mathcal{J}} p^{\rho}_j(t-)\Big(B_j(\mathrm{d}t) + \sum_{\substack{k \in \mathcal{J} \\ k \neq j}} b_{jk}(t)\Lambda^{\rho}_{jk}(\mathrm{d}t)\Big)$$

and that

$$p^{\rho}_j(t) = \Big[\prod_0^t \big(Id + \Lambda^{\rho}(\mathrm{d}s)\big)\Big]_{z_0 j}, \qquad j \in \mathcal{J}.$$

Remark 4.3.4.
Another way to phrase the second part of the theorem, which is also utilized in the proof, is that $(p^{\rho}_j)_j$ uniquely solves the forward integral equations

$$p^{\rho}_j(\mathrm{d}t) = \sum_{k \in \mathcal{J}} p^{\rho}_k(t-)\Lambda^{\rho}_{kj}(\mathrm{d}t), \qquad j \in \mathcal{J},$$

with initial conditions $p^{\rho}_j(0) = \mathbb{1}_{\{j=z_0\}}$.

Proof. The first result is trivial upon noticing that

$$p_j^\rho(t-) = \mathbb{E}[H(t)\mathbb{1}_{\{Z_{t-}=j\}}] + \mathbb{E}[(1-\rho(\tau, Z_{\tau-}, Z_\tau))\mathbb{1}_{\{\tau=t\}}\mathbb{1}_{\{Z_{t-}=j\}}],$$

where the latter term vanishes since we assumed that $\mathbb{P}(\tau = t)B_j(\mathrm{d}t) = 0$.

The second result is more involved. Introduce

$$q_j(t) = \Big[\prod_0^t \big(\mathrm{Id} + \Lambda^\rho(\mathrm{d}s)\big)\Big]_{z_0 j}.$$

These are the unique solution to a set of forward integral equations, which due to the specific choice of the diagonal of Λ^ρ read

$$\begin{aligned} q_j(\mathrm{d}t) &= \sum_{k\in\mathcal{J}_0} q_k(t-)\Lambda_{kj}(\mathrm{d}t), && j \in \mathcal{J}_0, \\ q_j(\mathrm{d}t) &= \sum_{k\in\mathcal{J}} q_k(t-)\Lambda^\rho_{kj}(\mathrm{d}t), && j \in \mathcal{J}_1, \end{aligned}$$

with boundary conditions $q_j(0) = \mathbb{1}_{\{j=z_0\}}$ for $j \in \mathcal{J}_0$ and $q_j(0) = 0$ for $j \in \mathcal{J}_1$.

Recall, that

$$\mathbb{1}_{\{Z_t=j\}} = \mathbb{1}_{\{j=z_0\}} + \sum_{\substack{k\in\mathcal{J}\\k\neq j}} \big(N_{kj}(t) - N_{jk}(t)\big). \tag{4.3.1}$$

Let $j \in \mathcal{J}_0$. Then (4.3.1) reads

$$\mathbb{1}_{\{Z_t=j\}} = \mathbb{1}_{\{j=z_0\}} + \sum_{\substack{k\in\mathcal{J}_0\\k\neq j}} N_{kj}(t) - \sum_{\substack{k\in\mathcal{J}\\k\neq j}} N_{jk}(t), \tag{4.3.2}$$

and since $p_j^\rho(t) = p_j(t)$, by taking expectations on both sides of (4.3.2) we obtain

$$p_j^\rho(\mathrm{d}t) = \sum_{k\in\mathcal{J}_0} p_k(t-)\Lambda_{kj}(\mathrm{d}t), \qquad p_j^\rho(0) = \mathbb{1}_{\{j=z_0\}}.$$

Let instead $j \in \mathcal{J}_1$. Then (4.3.1) implies

$$H(t)\mathbb{1}_{\{Z_t=j\}} = \sum_{\substack{k\in\mathcal{J}\\k\neq j}} \int_0^t H(s)N_{kj}(\mathrm{d}s) - \sum_{\substack{k\in\mathcal{J}_1\\k\neq j}} \int_0^t H(s)N_{jk}(\mathrm{d}s). \tag{4.3.3}$$

Upon taking expectations on both sides of (4.3.3), we finally obtain

$$p_j^\rho(\mathrm{d}t) = \sum_{k\in\mathcal{J}} p_k^\rho(t-)\Lambda^\rho_{kj}(\mathrm{d}t), \qquad p_j^\rho(0) = 0.$$

Collecting results and referring to the uniqueness of the solution to the forward integral equations, we conclude that $p_j^\rho = q_j$ for all $j \in \mathcal{J}$, as desired. □

Remark 4.3.5.
The identification of p^ρ with the product integral of Λ^ρ is the non-trivial part of Proposition 4.3.3. As a special case, we obtain that

$$p_j(t) = \Big[\prod_0^t \big(\mathrm{Id} + \Lambda(\mathrm{d}s)\big)\Big]_{z_0 j}. \tag{4.3.4}$$

Such identities were the focal point of [Ove19], who associated the left-hand side with a limit of products of transition probabilities and showed that if these transition probabilities are suitably regular, then $\Lambda_{jk}(t) < \infty$ and (4.3.4) holds. By pertaining to the key identity (4.3.1) and properties of the product integral, the proof becomes much shorter and simpler.

The representation of A in Proposition 4.3.3 is pleasing, since A is essentially cast as a well-behaved map of Λ^ρ. However, the matrix function Λ^ρ is – strictly speaking – not a hazard matrix. The presence of H has entailed a choice of diagonal that does not ensure the row sums to be constantly zero. This differs from the approach in [CF22], where the introduction of an auxiliary jump process allows for a hazard interpretation. We briefly summarize this approach and compare it to our approach.

Suppose each $\rho(\cdot, j, k)$ is uniformly bounded by one, and define $\widetilde{Z} = (\widetilde{Z}_t)_{t\geq 0}$ according to

$$\tilde{Z}_t = \mathbb{1}_{\{Z_t \in \mathcal{J}_0\}} Z_t + \mathbb{1}_{\{Z_t \in \mathcal{J}_1\}} \big(\mathbb{1}_{\{U \leq \rho(\tau, Z_{\tau-}, Z_\tau)\}} Z_t + \mathbb{1}_{\{U > \rho(\tau, Z_{\tau-}, Z_\tau)\}} \nabla\big),$$

where ∇ is an artificial cemetery state and $U \sim \mathrm{Unif}(0,1)$ is independent of Z. Set $\widetilde{\mathcal{J}} = \mathcal{J} \cup \{\nabla\}$, denote by $\widetilde{N}$ the multivariate counting process corresponding to $\widetilde{Z}$, and define

$$\widetilde{p}_j(t) = \mathbb{E}[\mathbb{1}_{\{\tilde{Z}_t = j\}}], \qquad \widetilde{p}_{jk}(t) = \mathbb{E}[\widetilde{N}_{jk}(t)], \qquad \widetilde{\Lambda}_{jk}(t) = \int_0^t \frac{1}{\widetilde{p}_j(s-)} \widetilde{p}_{jk}(\mathrm{d}s)$$

for $j, k \in \widetilde{\mathcal{J}}$, $j \neq k$. Further, set

$$-\widetilde{\Lambda}_{jj} = \sum_{\substack{k \in \widetilde{\mathcal{J}} \\ k \neq j}} \widetilde{\Lambda}_{jk}, \qquad j \in \widetilde{\mathcal{J}},$$

so that all row sums are constantly equal to zero. In Subsection 4.1 of [CF22] it is, among other things, shown that

$$A(\mathrm{d}t) = \sum_{j \in \mathcal{J}} \widetilde{p}_j(t-) \Big(B_j(\mathrm{d}t) + \sum_{\substack{k \in \mathcal{J} \\ k \neq j}} b_{jk}(t) \widetilde{\Lambda}_{jk}(\mathrm{d}t) \Big)$$

and that

$$\widetilde{p}_j(t) = \Big[\prod_0^t \big(\mathrm{Id} + \widetilde{\Lambda}(\mathrm{d}s)\big)\Big]_{z_0 j}, \qquad j \in \mathcal{J}.$$

From the latter identity it is rather obvious that Λ^ρ corresponds exactly to $\widetilde{\Lambda}$ with the row and column corresponding to the artificial cemetery state removed. What is then the advantage of the latter representation? This representation is convenient insofar that it retains the interpretation of $\widetilde{\Lambda}$ as an actual hazard matrix and thus of $\widetilde{p}$ as an actual probability vector, from which estimation procedures and their asymptotic properties almost suggest themselves. The representation is inconvenient, on the other hand, insofar that it requires an auxiliary added source of randomness, namely U, as well as the condition that the scaling factors be uniformly bounded by one.

4.3.2 Two-dimensional representation

Two-dimensional representations were proposed and studied in [BC24]. We here give a concise summary adapted to the case of scaled payments. In comparison to the one-dimensional representation, two further assumptions are required, namely the strengthened moment condition $\mathbb{E}[N_{jk}(t)^2] < \infty$ and that $\rho(\cdot, j, k)$ is deterministic, where $j, k \in \mathcal{J}$, $j \neq k$.

Let $\boldsymbol{t} = (t_1, t_2)$ and $\boldsymbol{j} = (j_1, j_2)$. Define $p_{\boldsymbol{j}}$ and $p_{\boldsymbol{jk}}$ according to

$$\begin{aligned} p_{\boldsymbol{j}}(\boldsymbol{t}) &= \mathbb{E}[\mathbb{1}_{\{Z_{t_1}=j_1\}}\mathbb{1}_{\{Z_{t_2}=j_2\}}], && j_1, j_2 \in \mathcal{J}, \\ p_{\boldsymbol{jk}}(\boldsymbol{t}) &= \mathbb{E}[N_{j_1k_1}(t_1)N_{j_2k_2}(t_2)], && j_1, k_1, j_2, k_2 \in \mathcal{J}, j_1 \neq k_1, j_2 \neq k_2. \end{aligned}$$

Further, define $\Lambda_{\boldsymbol{jk}}$ according to

$$\Lambda_{\boldsymbol{jk}}(\boldsymbol{t}) = \int_{\boldsymbol{0}}^{\boldsymbol{t}} \frac{1}{p_{\boldsymbol{j}}(\boldsymbol{s}-)} p_{\boldsymbol{jk}}(\mathrm{d}\boldsymbol{s}), \qquad j_1, k_1, j_2, k_2 \in \mathcal{J}, j_1 \neq k_1, j_2 \neq k_2.$$

Throughout, we shall assume that $\Lambda_{\boldsymbol{jk}}(\boldsymbol{t}) < \infty$. The following result summarizes and adapts Example 2.4, Section 7, Theorem 3.3, and Theorem 4.3 from [BC24] to the current setting.

Proposition 4.3.6.
Suppose that $\mathbb{E}[N_{jk}(t)^2] < \infty$ and that $\rho(\cdot, j, k)$ is deterministic for $j, k \in \mathcal{J}$, $j \neq k$. It then holds that

$$\begin{aligned} &A(t) - b_0 \\ &= \sum_{j \in \mathcal{J}_0} \int_0^t p_j(s-) \Big(B_j(\mathrm{d}s) + \sum_{\substack{k \in \mathcal{J}_0 \\ k \neq j}} b_{jk}(s) \Lambda_{jk}(\mathrm{d}s) + \sum_{k \in \mathcal{J}_1} \rho(s, j, k) b_{jk}(s) \Lambda_{jk}(\mathrm{d}s) \Big) \\ &\quad + \sum_{j_2 \in \mathcal{J}_1} \int_0^t \sum_{\substack{j_1 \in \mathcal{J}_0 \\ k_1 \in \mathcal{J}_1}} \int_{\boldsymbol{0}}^{(t,s-)} \rho(u_1, j_1, k_1) \Big(\sum_{\substack{k_2 \in \mathcal{J} \\ k_2 \neq j_2}} p_{(j_1,k_2)}(\boldsymbol{u}-) \Lambda_{(j_1,k_2)(k_1,j_2)}(\mathrm{d}\boldsymbol{u}) \\ &\qquad\qquad - \sum_{\substack{k_2 \in \mathcal{J}_1 \\ k_2 \neq j_2}} p_{\boldsymbol{j}}(\boldsymbol{u}-) \Lambda_{\boldsymbol{jk}}(\mathrm{d}\boldsymbol{u}) \Big) B_{j_2}(\mathrm{d}s) \end{aligned}$$

$$+ \sum_{\substack{j_2,k_2\in\mathcal{J}_1\\ j_2\neq k_2}} \sum_{\substack{j_1\in\mathcal{J}_0\\ k_1\in\mathcal{J}_1}} \int_{\mathbf{0}}^{(t,t)} \rho(u_1,j_1,k_1) b_{j_2k_2}(u_2) p_{\boldsymbol{j}}(\boldsymbol{u}-) \Lambda_{\boldsymbol{jk}}(\mathrm{d}\boldsymbol{u})$$

and that $(p_{\boldsymbol{j}})_{\boldsymbol{j}}$ uniquely solves the following system of inhomogeneous Volterra integral equations:

$$\begin{aligned} p_{\boldsymbol{j}}(\boldsymbol{t}) &= \mathbb{1}_{\{z_0=j_1=j_2\}} + \mathbb{1}_{\{z_0=j_1\}}\big(p_{j_2}(t_2) - \mathbb{1}_{\{z_0=j_2\}}\big) + \mathbb{1}_{\{z_0=j_2\}}\big(p_{j_1}(t_1) - \mathbb{1}_{\{z_0=j_1\}}\big) \\ &\quad + \sum_{\substack{k_1,k_2\in\mathcal{J}\\ k_1\neq j_1, k_2\neq j_2}} \int_{\mathbf{0}}^{\boldsymbol{t}} \Big(p_{\boldsymbol{k}}(\boldsymbol{s}-)\Lambda_{\boldsymbol{kj}}(\mathrm{d}\boldsymbol{s}) - p_{(k_1,j_2)}(\boldsymbol{s}-)\Lambda_{(k_1,j_2)(j_1,k_2)}(\mathrm{d}\boldsymbol{s}) \\ &\qquad\qquad - p_{(j_1,k_2)}(\boldsymbol{s}-)\Lambda_{(j_1,k_2)(j_2,k_1)}(\mathrm{d}\boldsymbol{s}) + p_{\boldsymbol{j}}(\boldsymbol{s}-)\Lambda_{\boldsymbol{jk}}(\mathrm{d}\boldsymbol{s})\Big), \end{aligned}$$

where $\boldsymbol{j} = (j_1, j_2) \in \mathcal{J} \times \mathcal{J}$.

In comparing the one- and two-dimensional representations, we note that the two-dimensional representation needs further technical assumptions and is more involved, in particular from a numerical point of view. On the other hand, it achieves true separation between the probabilistic components and the contract elements, that is the payments and the scaling factors. This is convenient if one needs to calculate the expected accumulated cash flow across many different contract specifications. It also allows for a more clear separation between valuation and estimation in accordance with common actuarial workflows.

4.4 Non-parametric estimation

In this paper, we are interested in the situation where observation of Z is right-censored. To this end, we introduce a strictly positive random variable R describing right-censoring. Letting η denote the (possible infinite) absorption time of Z, we thus observe the pair

$$\big((Z_t)_{0\leq t\leq R}, \eta \wedge R\big).$$

Under such an incomplete observation scheme, the behavior of Z beyond the right endpoint of the support of R of course remains elusive. In the following, we thus fix a θ strictly below this right endpoint and focus on the time interval $[0,\theta]$. The extension to the maximal interval is not pursued.

Throughout, we shall assume that right-censoring is entirely random, meaning that Z is independent of R. This is a central assumption which can, essentially, only be weakened in case Z is Markov. It implies that

$$\Lambda^{\rho}_{jk}(t) = \int_0^t \frac{1}{p_j^{\rho,\mathsf{c}}(s-)} p_{jk}^{\rho,\mathsf{c}}(\mathrm{d}s), \qquad j,k\in\mathcal{J}, j\neq k, \tag{4.4.1}$$

$$\Lambda_{\boldsymbol{jk}}(\boldsymbol{t}) = \int_{\mathbf{0}}^{\boldsymbol{t}} \frac{1}{p_{\boldsymbol{j}}^{\mathsf{c}}(\boldsymbol{s}-)} p_{\boldsymbol{jk}}^{\mathsf{c}}(\mathrm{d}\boldsymbol{s}), \qquad j_1,k_1,j_2,k_2\in\mathcal{J}, j_1\neq k_1, j_2\neq k_2, \tag{4.4.2}$$

for t, t_1, and t_2 below θ and with

$$p_j^{\rho,\mathsf{c}}(t) = \mathbb{E}[H(t)\mathbb{1}_{\{Z_t=j\}}\mathbb{1}_{\{t<R\}}], \qquad p_{jk}^{\rho,\mathsf{c}}(t) = \mathbb{E}\Big[\int_0^t H(s)\mathbb{1}_{\{s\le R\}}N_{jk}(\mathrm{d}s)\Big],$$
$$p_{\boldsymbol{j}}^{\mathsf{c}}(\boldsymbol{t}) = \mathbb{E}[\mathbb{1}_{\{Z_{t_1}=j_1\}}\mathbb{1}_{\{Z_{t_2}=j_2\}}\mathbb{1}_{\{\boldsymbol{t}<R\}}], \qquad p_{\boldsymbol{jk}}^{\mathsf{c}}(\boldsymbol{t}) = \mathbb{E}[N_{j_1k_1}(t_1\wedge R)N_{j_2k_2}(t_2\wedge R)].$$

In the following, the non-scaled case of each $\rho(\cdot,j,k)$ being constantly equal to one is indicated by dropping the superscript ρ, which is also in accordance with previous definitions and notation.

The components in (4.4.1) and (4.4.2) are natural targets for estimation. However, the fact that the denominators might become zero is somewhat of a technical menace. We follow the suggestion of [BF24b] to change the target of estimation by means of perturbation. That is, we instead consider

$$\Lambda_{jk}^{\rho,\varepsilon}(t) = \int_0^t \frac{1}{p_j^{\rho,\mathsf{c}}(s-)\vee\varepsilon}p_{jk}^{\rho,\mathsf{c}}(\mathrm{d}s), \qquad j,k\in\mathcal{J}, j\neq k,$$
$$\Lambda_{\boldsymbol{jk}}^{\varepsilon}(\boldsymbol{t}) = \int_{\boldsymbol{0}}^{\boldsymbol{t}} \frac{1}{p_{\boldsymbol{j}}^{\mathsf{c}}(\boldsymbol{s}-)\vee\varepsilon}p_{\boldsymbol{jk}}^{\mathsf{c}}(\mathrm{d}\boldsymbol{s}), \qquad j_1,k_1,j_2,k_2\in\mathcal{J}, j_1\neq k_1, j_2\neq k_2,$$

for some $\varepsilon > 0$. We retain the choice of diagonal, that is we let $\Lambda_{jj}^{\rho,\varepsilon}$ be given by

$$-\Lambda_{jj}^{\rho,\varepsilon} = \sum_{\substack{k\in\mathcal{J}\\k\neq j}}\Lambda_{jk}^{\varepsilon}, \qquad j\in\mathcal{J}_0,$$
$$-\Lambda_{jj}^{\rho,\varepsilon} = \sum_{\substack{k\in\mathcal{J}\\k\neq j}}\Lambda_{jk}^{\rho,\varepsilon}, \qquad j\in\mathcal{J}_1.$$

Furthermore, we set

$$p_j^{\rho,\varepsilon}(t) = \Big[\prod_0^t\big(\mathrm{Id}+\Lambda^{\rho,\varepsilon}(\mathrm{d}s)\big)\Big]_{z_0j}, \qquad j\in\mathcal{J}.$$

Finally, we define $(p_{\boldsymbol{j}}^{\varepsilon})_{\boldsymbol{j}}$ as the unique solution to the following system of inhomogeneous Volterra integral equations:

$$\begin{aligned} p_{\boldsymbol{j}}^{\varepsilon}(\boldsymbol{t}) = {} & \mathbb{1}_{\{z_0=j_1=j_2\}} \\ & + \mathbb{1}_{\{z_0=j_1\}}\big(p_{j_2}^{\varepsilon}(t_2)-\mathbb{1}_{\{z_0=j_2\}}\big) + \mathbb{1}_{\{z_0=j_2\}}\big(p_{j_1}^{\varepsilon}(t_1)-\mathbb{1}_{\{z_0=j_1\}}\big) \\ & + \sum_{\substack{k_1,k_2\in\mathcal{J}\\k_1\neq j_1,k_2\neq j_2}}\int_{\boldsymbol{0}}^{(t,t)}\Big(p_{\boldsymbol{k}}^{\varepsilon}(\boldsymbol{s}-)\Lambda_{\boldsymbol{kj}}^{\varepsilon}(\mathrm{d}\boldsymbol{s}) - p_{(k_1,j_2)}^{\varepsilon}(\boldsymbol{s}-)\Lambda_{(k_1,j_2)(j_1,k_2)}^{\varepsilon}(\mathrm{d}\boldsymbol{s}) \\ & \qquad - p_{(j_1,k_2)}^{\varepsilon}(\boldsymbol{s}-)\Lambda_{(j_1,k_2)(j_2,k_1)}^{\varepsilon}(\mathrm{d}\boldsymbol{s}) + p_{\boldsymbol{j}}^{\varepsilon}(\boldsymbol{s}-)\Lambda_{\boldsymbol{jk}}^{\varepsilon}(\mathrm{d}\boldsymbol{s})\Big), \end{aligned} \tag{4.4.3}$$

where $\boldsymbol{j}=(j_1,j_2)\in\mathcal{J}\times\mathcal{J}$.

In the limit $\varepsilon\to 0$, we recover the original expressions, see also Theorem 5.9 in [Bat24].

The classic identity (4.3.1) allows us to establish the following link between the one-dimensional components:

$$p_j^{\rho,\mathsf{c}}(t) = \mathbb{1}_{\{j=z_0\}} - C_j^{\rho}(t) + \sum_{\substack{k\in\mathcal{J}_0\\k\neq j}} \big(p_{kj}^{\rho,\mathsf{c}}(t) - p_{jk}^{\rho,\mathsf{c}}(t)\big) - \sum_{k\in\mathcal{J}_1} p_{jk}^{\mathsf{c}}(t), \qquad j\in\mathcal{J}_0, \tag{4.4.4}$$

$$p_j^{\rho,\mathsf{c}}(t) = -C_j^{\rho}(t) + \sum_{k\in\mathcal{J}_0} p_{kj}^{\rho,\mathsf{c}}(t) + \sum_{\substack{k\in\mathcal{J}_1\\k\neq j}} \big(p_{kj}^{\rho,\mathsf{c}}(t) - p_{jk}^{\rho,\mathsf{c}}(t)\big), \qquad j\in\mathcal{J}_1, \tag{4.4.5}$$

where C_j^{ρ} is defined according to

$$C_j^{\rho}(t) = \mathbb{E}[H(R)\mathbb{1}_{\{Z_R=j\}}\mathbb{1}_{\{R\leq t\}}], \qquad j\in\mathcal{J}.$$

Similarly, in regards to the two-dimensional case we have

$$\begin{aligned}
p_{\boldsymbol{j}}^{\mathsf{c}}(\boldsymbol{t}) &= \mathbb{1}_{\{z_0=j_1=j_2\}} - \mathbb{1}_{\{z_0=j_1=j_2\}}\mathbb{E}[\mathbb{1}_{\{R\leq t_1\vee R\leq t_2\}}]\\
&\quad + \mathbb{1}_{\{z_0=j_2\}}\Big(\sum_{\substack{k\in\mathcal{J}\\k\neq j_1}} p_{kj_1}^{\mathsf{c}}(t_1) - p_{j_1k}^{\mathsf{c}}(t_1)\Big) + \mathbb{1}_{\{z_0=j_1\}}\Big(\sum_{\substack{k\in\mathcal{J}\\k\neq j_2}} p_{kj_2}^{\mathsf{c}}(t_2) - p_{j_2k}^{\mathsf{c}}(t_2)\Big)\\
&\quad - \mathbb{1}_{\{z_0=j_2\}}\Big(\sum_{\substack{k\in\mathcal{J}\\k\neq j_1}} p_{kj_1}^{\mathsf{c},1}(\boldsymbol{t}) - p_{j_1k}^{\mathsf{c},1}(\boldsymbol{t})\Big) - \mathbb{1}_{\{z_0=j_1\}}\Big(\sum_{\substack{k\in\mathcal{J}\\k\neq j_2}} p_{kj_2}^{\mathsf{c},2}(\boldsymbol{t}) - p_{j_2k}^{\mathsf{c},2}(\boldsymbol{t})\Big)\\
&\quad + \sum_{\substack{k_1,k_2\in\mathcal{J}\\k_1\neq j_1,k_2\neq j_2}} \Big(p_{\boldsymbol{kj}}^{\mathsf{c}}(\boldsymbol{t}) + p_{\boldsymbol{jk}}^{\mathsf{c}}(\boldsymbol{t}) - p_{(j_1k_2)(k_1j_2)}^{\mathsf{c}}(\boldsymbol{t}) - p_{(k_1j_2)(j_1k_2)}^{\mathsf{c}}(\boldsymbol{t})\Big)\\
&\quad - \sum_{\substack{k_1,k_2\in\mathcal{J}\\k_1\neq j_1,k_2\neq j_2}} \Big(p_{\boldsymbol{kj}}^{\mathsf{c},3}(\boldsymbol{t}) + p_{\boldsymbol{jk}}^{\mathsf{c},3}(\boldsymbol{t}) - p_{(j_1k_2)(k_1j_2)}^{\mathsf{c},3}(\boldsymbol{t}) - p_{(k_1j_2)(j_1k_2)}^{\mathsf{c},3}(\boldsymbol{t})\Big),
\end{aligned} \tag{4.4.6}$$

where $p_{kj}^{\mathsf{c},1}(\boldsymbol{t}) = \mathbb{E}[N_{kj}(t_1\wedge R)\mathbb{1}_{\{R\leq t_1\vee R\leq t_2\}}]$, $p_{kj}^{\mathsf{c},2}(\boldsymbol{t}) = \mathbb{E}[N_{kj}(t_2\wedge R)\mathbb{1}_{\{R\leq t_1\vee R\leq t_2\}}]$ and $p_{\boldsymbol{kj}}^{\mathsf{c},3}(\boldsymbol{t}) = \mathbb{E}[N_{k_1j_1}(t_1\wedge R)N_{k_2j_2}(t_2\wedge R)\mathbb{1}_{\{R\leq t_1\vee R\leq t_2\}}]$.

4.4.1 Estimators

Consider i.i.d. replicates

$$\big((Z_t^{\ell})_{0\leq t\leq R^{\ell}}, \eta^{\ell}\wedge R^{\ell}\big)_{\ell=1}^{n}$$

and form the following empirical processes

$$\begin{aligned}
\mathbb{I}_j^{\rho}(t) &= \frac{1}{n}\sum_{\ell=1}^{n} H^{\ell}(t)\mathbb{1}_{\{Z_t^{\ell}=j\}}\mathbb{1}_{\{t<R^{\ell}\}},\\
\mathbb{N}_{jk}^{\rho}(t) &= \frac{1}{n}\sum_{\ell=1}^{n}\int_0^t H^{\ell}(s)\mathbb{1}_{\{s\leq R^{\ell}\}}N_{jk}^{\ell}(\mathrm{d}s),\\
\mathbb{I}_{\boldsymbol{j}}(\boldsymbol{t}) &= \frac{1}{n}\sum_{\ell=1}^{n}\mathbb{1}_{\{Z_{t_1}^{\ell}=j_1\}}\mathbb{1}_{\{Z_{t_2}^{\ell}=j_2\}}\mathbb{1}_{\{\boldsymbol{t}<R^{\ell}\}},
\end{aligned}$$

$$\mathbb{N}_{\boldsymbol{jk}}(\boldsymbol{t}) = \frac{1}{n}\sum_{\ell=1}^{n} N^{\ell}_{j_1k_1}(t_1 \wedge R^{\ell})N^{\ell}_{j_2k_2}(t_2 \wedge R^{\ell}),$$

as non-parametric estimators of $p^{\rho,\mathsf{c}}_{j}$, $p^{\rho,\mathsf{c}}_{jk}$, $p^{\mathsf{c}}_{\boldsymbol{j}}$, and $p^{\mathsf{c}}_{\boldsymbol{jk}}$, respectively. Similar to (4.4.4) and (4.4.5), we find the following link for the one-dimensional case:

$$\mathbb{I}^{\rho}_{j}(t) = \mathbb{1}_{\{j=z_0\}} - \mathbb{C}^{\rho}_{j}(t) + \sum_{\substack{k\in\mathcal{J}_0\\k\neq j}} \big(\mathbb{N}^{\rho}_{kj}(t) - \mathbb{N}^{\rho}_{jk}(t)\big) - \sum_{k\in\mathcal{J}_1}\mathbb{N}_{jk}(t), \qquad j\in\mathcal{J}_0, \tag{4.4.7}$$

$$\mathbb{I}^{\rho}_{j}(t) = -\mathbb{C}^{\rho}_{j}(t) + \sum_{k\in\mathcal{J}_0}\mathbb{N}^{\rho}_{kj}(t) + \sum_{\substack{k\in\mathcal{J}_1\\k\neq j}} \big(\mathbb{N}^{\rho}_{kj}(t) - \mathbb{N}^{\rho}_{jk}(t)\big), \qquad j\in\mathcal{J}_1, \tag{4.4.8}$$

where $\mathbb{C}^{\rho}_{j}$ is defined according to

$$\mathbb{C}^{\rho}_{j}(t) = \frac{1}{n}\sum_{\ell=1}^{n} H^{\ell}(R^{\ell})\mathbb{1}_{\{Z^{\ell}_{R^{\ell}}=j\}}\mathbb{1}_{\{R^{\ell}\leq t\}}, \qquad j\in\mathcal{J}.$$

Similarly, in regards to the two-dimensional case we have

$$\begin{aligned}
\mathbb{I}_{\boldsymbol{j}}(\boldsymbol{t}) = {} & \mathbb{1}_{\{z_0=j_1=j_2\}} \\
& + \mathbb{1}_{\{z_0=j_2\}}\Big(\sum_{\substack{k\in\mathcal{J}\\k\neq j_1}} \mathbb{N}_{kj_1}(t_1) - \mathbb{N}_{j_1k}(t_1)\Big) \\
& + \mathbb{1}_{\{z_0=j_1\}}\Big(\sum_{\substack{k\in\mathcal{J}\\k\neq j_2}} \mathbb{N}_{kj_2}(t_2) - \mathbb{N}_{j_2k}(t_2)\Big) \\
& - \mathbb{1}_{\{z_0=j_2\}}\Big(\sum_{\substack{k\in\mathcal{J}\\k\neq j_1}} \mathbb{N}^{1}_{kj_1}(\boldsymbol{t}) - \mathbb{N}^{1}_{j_1k}(\boldsymbol{t})\Big) \\
& - \mathbb{1}_{\{z_0=j_1\}}\Big(\sum_{\substack{k\in\mathcal{J}\\k\neq j_2}} \mathbb{N}^{2}_{kj_2}(\boldsymbol{t}) - \mathbb{N}^{2}_{j_2k}(\boldsymbol{t})\Big) \\
& + \sum_{\substack{k_1,k_2\in\mathcal{J}\\k_1\neq j_1,k_2\neq j_2}} \Big(\mathbb{N}_{\boldsymbol{kj}}(\boldsymbol{t}) + \mathbb{N}_{\boldsymbol{jk}}(\boldsymbol{t}) - \mathbb{N}_{(j_1k_2)(k_1j_2)}(\boldsymbol{t}) - \mathbb{N}_{(k_1j_2)(j_1k_2)}(\boldsymbol{t})\Big) \\
& - \sum_{\substack{k_1,k_2\in\mathcal{J}\\k_1\neq j_1,k_2\neq j_2}} \Big(\mathbb{N}^{3}_{\boldsymbol{kj}}(\boldsymbol{t}) + \mathbb{N}^{3}_{\boldsymbol{jk}}(\boldsymbol{t}) - \mathbb{N}^{3}_{(j_1k_2)(k_1j_2)}(\boldsymbol{t}) - \mathbb{N}^{3}_{(k_1j_2)(j_1k_2)}(\boldsymbol{t})\Big),
\end{aligned} \tag{4.4.9}$$

where the estimators are defined according to

$$\begin{aligned}
\mathbb{N}_{jk}(t) &= \frac{1}{n}\sum_{\ell=1}^{n} N^{\ell}_{jk}(t\wedge R^{\ell}), \\
\mathbb{N}^{1}_{jk}(\boldsymbol{t}) &= \frac{1}{n}\sum_{\ell=1}^{n} N^{\ell}_{jk}(t_1\wedge R^{\ell})\mathbb{1}_{\{R^{\ell}\leq t_1\vee R^{\ell}\leq t_2\}}, \\
\mathbb{N}^{2}_{jk}(\boldsymbol{t}) &= \frac{1}{n}\sum_{\ell=1}^{n} N^{\ell}_{jk}(t_2\wedge R^{\ell})\mathbb{1}_{\{R^{\ell}\leq t_1\vee R^{\ell}\leq t_2\}},
\end{aligned}$$

$$\mathbb{N}^3_{\boldsymbol{jk}}(\boldsymbol{t}) = \frac{1}{n}\sum_{\ell=1}^{n} N^\ell_{j_1k_1}(t_1 \wedge R^\ell)N^\ell_{j_2k_2}(t_2 \wedge R^\ell)\mathbb{1}_{\{R^\ell \leq t_1 \vee R^\ell \leq t_2\}}.$$

The above identities are used repeatedly in the proofs, but may also be helpful for the development of efficient implementations.

We are now ready to form the estimators of interest, which will be of Nelson–Aalen and Aalen–Johansen type. The first estimator $\hat{\Lambda}^{\rho,\varepsilon}$ reads

$$\hat{\Lambda}^{\rho,\varepsilon}_{jk}(t) = \int_0^t \frac{1}{\mathbb{I}^\rho_j(s-) \vee \varepsilon}\mathbb{N}^\rho_{jk}(\mathrm{d}s)$$

and targets $\Lambda^{\rho,\varepsilon}_{jk}$. We also define

$$\begin{aligned} -\hat{\Lambda}^{\rho,\varepsilon}_{jj} &= \sum_{\substack{k\in\mathcal{J}\\k\neq j}} \hat{\Lambda}^{\varepsilon}_{jk}, && j \in \mathcal{J}_0,\\ -\hat{\Lambda}^{\rho,\varepsilon}_{jj} &= \sum_{\substack{k\in\mathcal{J}\\k\neq j}} \hat{\Lambda}^{\rho,\varepsilon}_{jk}, && j \in \mathcal{J}_1, \end{aligned}$$

where $\hat{\Lambda}^{\varepsilon}_{jk}$ is the ordinary estimator, corresponding to $\hat{\Lambda}^{\rho,\varepsilon}_{jk}$ without scaling, that is $H \equiv 1$. The associated estimator $\hat{p}^{\rho,\varepsilon}_j$ reads

$$\hat{p}^{\rho,\varepsilon}_j(t) = \Big[\prod_0^t (\mathrm{Id} + \hat{\Lambda}^{\rho,\varepsilon}(\mathrm{d}s))\Big]_{z_0 j}$$

and targets $p^{\rho,\varepsilon}_j$. The next estimator $\hat{\Lambda}^{\varepsilon}_{\boldsymbol{jk}}$ reads

$$\hat{\Lambda}^{\varepsilon}_{\boldsymbol{jk}}(\boldsymbol{t}) = \int_{\boldsymbol{0}}^{\boldsymbol{t}} \frac{1}{\mathbb{I}_{\boldsymbol{j}}(\boldsymbol{s}-) \vee \varepsilon}\mathbb{N}_{\boldsymbol{jk}}(\mathrm{d}\boldsymbol{s})$$

and targets $\Lambda^{\varepsilon}_{\boldsymbol{jk}}$. The associated estimator $\hat{p}^{\varepsilon}_{\boldsymbol{j}}$, which targets $p^{\varepsilon}_{\boldsymbol{j}}$, is defined as the unique solution to (4.4.3), but with all the theoretical quantities replaced by their just introduced estimators.

The perturbation given by ε only plays a role in the theoretical analyses of the estimators. In applications, where n is of course fixed, one takes ε sufficiently close to zero as to not affect the actual estimates, which actually corresponds to $\varepsilon = 0$.

Let $0 = t_0 < t_1 < t_2 < \cdots < t_M$ describe all observed event times in $[0, \theta]$. The estimators are all constant between these time points and, consequently, best described through difference equations. In regards to the one-dimensional case, it holds that $\hat{p}^{\rho,\varepsilon}_j(t_0) = \mathbb{1}_{\{z_0=j\}}$ and

$$\Delta\hat{p}^{\rho,\varepsilon}_j(t_{m+1}) = \sum_{k\in\mathcal{J}} \hat{p}^{\rho,\varepsilon}_k(t_m)\Delta\hat{\Lambda}^{\rho,\varepsilon}_{kj}(t_{m+1}).$$

For the two-dimensional case, it holds that

$$\hat{p}^{\varepsilon}_{\boldsymbol{j}}(t_m, 0) = \hat{p}^{\varepsilon}_{j_1}(t_m)\mathbb{1}_{\{j_2=z_0\}}, \qquad \hat{p}^{\varepsilon}_{\boldsymbol{j}}(0, t_m) = \hat{p}^{\varepsilon}_{j_2}(t_m)\mathbb{1}_{\{j_1=z_0\}},$$

where $\hat{p}_j^{\varepsilon}$ is the ordinary estimator, corresponding to $\hat{p}_j^{\rho,\varepsilon}$ without scaling, that is $H \equiv 1$. Further, if we let

$$\begin{aligned}\Delta\hat{p}_{\boldsymbol{j}}^{\varepsilon}(t_{m_1+1}, t_{m_2+1}) = \hat{p}_{\boldsymbol{j}}^{\varepsilon}(t_{m_1+1}, t_{m_2+1}) \\ - \hat{p}_{\boldsymbol{j}}^{\varepsilon}(t_{m_1+1}, t_{m_2}) - \hat{p}_{\boldsymbol{j}}^{\varepsilon}(t_{m_1}, t_{m_2+1}) + \hat{p}_{\boldsymbol{j}}^{\varepsilon}(t_{m_1}, t_{m_2}),\end{aligned}$$

we obtain

$$\begin{aligned}\Delta\hat{p}_{\boldsymbol{j}}^{\varepsilon}(t_{m_1+1}, t_{m_2+1}) = &\sum_{\substack{k_1,k_2\in\mathcal{J}\\ k_1\neq j_1, k_2\neq j_2}} \hat{p}_{\boldsymbol{k}}^{\varepsilon}(t_{m_1}, t_{m_2})\Delta\hat{\Lambda}_{\boldsymbol{kj}}^{\varepsilon}(t_{m_1+1}, t_{m_2+1}) \\ &- \sum_{\substack{k_1,k_2\in\mathcal{J}\\ k_1\neq j_1, k_2\neq j_2}} \hat{p}_{(k_1,j_2)}^{\varepsilon}(t_{m_1}, t_{m_2})\Delta\hat{\Lambda}_{(k_1,j_2)(j_1,k_2)}^{\varepsilon}(t_{m_1+1}, t_{m_2+1}) \\ &- \sum_{\substack{k_1,k_2\in\mathcal{J}\\ k_1\neq j_1, k_2\neq j_2}} \hat{p}_{(j_1,k_2)}^{\varepsilon}(t_{m_1}, t_{m_2})\Delta\hat{\Lambda}_{(j_1,k_2)(k_1,j_2)}^{\varepsilon}(t_{m_1+1}, t_{m_2+1}) \\ &- \hat{p}_{\boldsymbol{j}}^{\varepsilon}(t_{m_1}, t_{m_2}) \sum_{\substack{k_1,k_2\in\mathcal{J}\\ k_1\neq j_1, k_2\neq j_2}} \Delta\hat{\Lambda}_{\boldsymbol{jk}}^{\varepsilon}(t_{m_1+1}, t_{m_2+1})\end{aligned}$$

with $\hat{\Lambda}_{\boldsymbol{jk}}^{\varepsilon}(t_{m_1}, 0) = 0$, $\hat{\Lambda}_{\boldsymbol{jk}}^{\varepsilon}(0, t_{m_2}) = 0$, and

$$\Delta\hat{\Lambda}_{\boldsymbol{jk}}^{\varepsilon}(t_{m_1+1}, t_{m_2+1}) = \frac{1}{\mathbb{I}_{\boldsymbol{j}}(t_{m_1}, t_{m_2}) \vee \varepsilon}\Delta\mathbb{N}_{\boldsymbol{jk}}(t_{m_1+1}, t_{m_2+1}),$$

where $\Delta\mathbb{N}_{\boldsymbol{jk}}$ reads

$$\Delta\mathbb{N}_{\boldsymbol{jk}}(t_{m_1+1}, t_{m_2+1}) = \frac{1}{n}\sum_{\ell=1}^{n}\Delta N_{j_1k_1}^{\ell}(t_{m_1+1} \wedge R^{\ell})\Delta N_{j_2k_2}^{\ell}(t_{m_2+1} \wedge R^{\ell}).$$

To efficiently calculate $\hat{p}_{\boldsymbol{j}}^{\varepsilon}$ on the relevant two-dimensional grid, one calculates simultaneously across $\boldsymbol{j} \in \mathcal{J} \times \mathcal{J}$ the gnomons

$$\big(\hat{p}_{\boldsymbol{j}}^{\varepsilon}(t_{m+s}, t_{m+1}), \hat{p}_{\boldsymbol{j}}^{\varepsilon}(t_{m+1}, t_{m+s})\big)_{s=1,2,\ldots,M-m},$$

progressing through m from zero to $M-1$ and thus filling in the square. Note that the edges of the gnomons may be calculated in parallel.

4.4.2 Strong uniform consistency

In Section 5 of [Bat24], the approach of Section 3 in [BF24b] was adapted to the two-dimensional case. In particular, using the link (4.4.6) and the corresponding link (4.4.9), the following result was established:

Theorem 4.4.1.
Suppose that $\mathbb{E}[N_{jk}(\theta)^2] < \infty$ for $j, k \in \mathcal{J}$, $j \neq k$. It then holds that

$$\begin{aligned}&\sup_{0\leq\boldsymbol{t}\leq\theta}\big|\hat{\Lambda}_{\boldsymbol{jk}}^{\varepsilon}(\boldsymbol{t}) - \Lambda_{\boldsymbol{jk}}^{\varepsilon}(\boldsymbol{t})\big| \xrightarrow{\text{a.s.}} 0, && n \to \infty, \\ &\sup_{0\leq\boldsymbol{t}\leq\theta}\big|\hat{p}_{\boldsymbol{j}}^{\varepsilon}(\boldsymbol{t}) - p_{\boldsymbol{j}}^{\varepsilon}(\boldsymbol{t})\big| \xrightarrow{\text{a.s.}} 0, && n \to \infty.\end{aligned}$$

The consistency of $\hat{\Lambda}_{jk}^{\rho,\varepsilon}$ and $\hat{p}_{j}^{\rho,\varepsilon}$ may also be established using the approach of Section 3 in [BF24b]. First note that

$$\sup_{0\leq t\leq\theta}\left|\mathbb{C}_{j}^{\rho}(t)-C_{j}^{\rho}(t)\right|\overset{\text{a.s.}}{\to}0,\qquad\qquad n\to\infty. \tag{4.4.10}$$

This follows from the fact that the class of all uniformly bounded, monotone functions on the real line is Glivenko–Cantelli, confer with Theorem 2.4.1 and Theorem 2.7.5 in [VW96]. The assumption that each $\rho(\cdot,j,k)$ is uniformly bounded on $[0,\theta]$ is here of essence.

Theorem 4.4.2.
It holds that

$$\begin{aligned}\sup_{0\leq t\leq\theta}\left|\hat{\Lambda}_{jk}^{\rho,\varepsilon}(t)-\Lambda_{jk}^{\rho,\varepsilon}(t)\right|&\overset{\text{a.s.}}{\to}0,&&n\to\infty,\\ \sup_{0\leq t\leq\theta}\left|\hat{p}_{j}^{\rho,\varepsilon}(t)-p_{j}^{\rho,\varepsilon}(t)\right|&\overset{\text{a.s.}}{\to}0,&&n\to\infty.\end{aligned}$$

Proof. Following for instance the proof of Proposition 3.2 in [BF24b], it suffices to show strong uniform consistency of the components $\mathbb{I}_{j}^{\rho}$ and $\mathbb{N}_{jk}^{\rho}$. However, from the links (4.4.4)-(4.4.5), the links (4.4.7)-(4.4.8), and the strong uniform consistency (4.4.10), it actually even just suffices to establish strong uniform consistency of $\mathbb{N}_{jk}^{\rho}$.

To this end, we use an idea from the proof of Proposition 3.1 in [BF24b]. Let D_m be the event where $\mathbb{N}_{jk}(\theta)\leq m$ occurs eventually (in n). Recall that each $\rho(\cdot,j,k)$ is uniformly bounded on $[0,\theta]$ and note also that $p_{jk}^{\rho,\mathsf{c}}$ is non-decreasing. Since the class of all uniformly bounded, monotone functions on the real line is Glivenko–Cantelli, we may conclude that on D_m it holds that

$$\sup_{0\leq t\leq\theta}\left|\mathbb{N}_{jk}^{\rho}(t)-p_{jk}^{\rho,\mathsf{c}}(t)\right|\overset{\text{a.s.}}{\to}0,\qquad\qquad n\to\infty.$$

In addition, by the strong law of large numbers it holds that

$$\mathbb{N}_{jk}^{\rho}(\theta)\overset{\text{a.s.}}{\to}p_{jk}^{\rho,\mathsf{c}}(\theta),\qquad\qquad n\to\infty,$$

which shows that for m sufficiently large D_m has probability one. Collecting results completes the proof. □

4.4.3 Asymptotic normality

To establish asymptotic normality, here meaning weak convergence to a tight Gaussian process, we use more or less explicitly the so-called bracketing central limit theorem, Theorem 2.11.9 in [VW96]; see however also Appendix D in [BF24b]. In both cases, we first prove weak convergence of the relevant components and then adapt the approach of Section 4 in [BF24b] to obtain weak convergence of the composite quantities. We begin with the one-dimensional representation.

Lemma 4.4.3.
Suppose that $\mathbb{E}[N_{jk}(\theta)^2] < \infty$ for $j, k \in \mathcal{J}$, $j \neq k$. It then holds that

$$\Big(\sqrt{n}(\mathbb{I}_j^\rho(t) - p_j^{\rho,\mathsf{c}}(t))_j, \sqrt{n}(\mathbb{N}_{jk}^\rho(t) - p_{jk}^{\rho,\mathsf{c}}(t))_{j\neq k}\Big)_{t\in[0,\theta]}$$

converges weakly to a centered tight Gaussian process.

Proof. It suffices to show weak convergence of

$$\Big(\sqrt{n}(\mathbb{C}_j^\rho(t) - C_j^\rho(t))_j, \sqrt{n}(\mathbb{N}_{jk}^\rho(t) - p_{jk}^{\rho,\mathsf{c}}(t))_{j\neq k}\Big)_{t\in[0,\theta]}$$

to a tight Gaussian process. This is because

$$\Big(\sqrt{n}(\mathbb{I}_j^\rho(t) - p_j^{\rho,\mathsf{c}}(t))_j, \sqrt{n}(\mathbb{N}_{jk}^\rho(t) - p_{jk}^{\rho,\mathsf{c}}(t))_{j\neq k}\Big)_{t\in[0,\theta]}$$

is a linear transformation hereof, confer with (4.4.4)-(4.4.5) and (4.4.7)-(4.4.8), so it retains the weak convergence to a tight Gaussian process. (The limiting distribution is of course centered.) Convergence of the finite-dimensional distributions follows from the multivariate central limit theorem. According to Problem 1.5.3 in [VW96], it thus suffices to show asymptotic tightness of each component. Note that this is the case for $\sqrt{n}(\mathbb{C}_j^\rho(t) - C_j^\rho(t))_{t\in[0,\theta]}$, given that the class of all uniformly bounded, monotone functions on the real line is Donsker, confer with Subsection 2.7.2 in [VW96]. Further, under the assumptions $\mathbb{E}[N_{jk}(\theta)^2]$ this is also the case for $\sqrt{n}(\mathbb{N}_{jk}^\rho(t) - p_j^{\rho,\mathsf{c}}(t))_{t\in[0,\theta]}$, confer with Example 2.11.16 in [VW96]. Collecting results completes the proof. □

Theorem 4.4.4.
Suppose that $\mathbb{E}[N_{jk}(\theta)^2] < \infty$ for $j, k \in \mathcal{J}$, $j \neq k$. It then holds that

$$\Big(\sqrt{n}(\hat{p}_j^{\rho,\varepsilon}(t) - p_j^{\rho,\varepsilon}(t))_j, \sqrt{n}(\hat{\Lambda}_{jk}^{\rho,\varepsilon}(t) - \Lambda_{jk}^{\rho,\varepsilon}(t))_{j\neq k}\Big)_{t\in[0,\theta]}$$

converges weakly to a centered tight Gaussian process.

Proof. Since the functional $(\Lambda_{jk}^{\rho,\varepsilon})_{j\neq k} \mapsto (p_j^{\rho,\varepsilon})_j$ is Hadamard differentiable, confer with Theorem 8 in [GJ90], it suffices by the functional delta method to establish weak convergence of the former.

Let $t \mapsto X_j^{\rho,\mathsf{c}}(t)$ and $t \mapsto X_{jk}^{\rho,\mathsf{c}}(t)$ be the limit processes of $t \mapsto \sqrt{n}(\mathbb{I}_j^\rho(t) - p_j^{\rho,\mathsf{c}}(t))$ and $t \mapsto \sqrt{n}(\mathbb{N}_{jk}^\rho(t) - p_{jk}^{\rho,\mathsf{c}}(t))$, respectively. According to Lemma 4.4.3, these limit processes are centered tight joint Gaussian. Similar to the proof of Theorem 4.5 in [BF24b], we may show that $t \mapsto \sqrt{n}(\hat{\Lambda}_{jk}^{\rho,\varepsilon}(t) - \Lambda_{jk}^{\rho,\varepsilon}(t))$ converges to the limit process

$$\begin{aligned} t \mapsto &\int_0^t \frac{1}{p_j^{\rho,\mathsf{c}}(s-) \vee \varepsilon} X_{jk}^{\rho,\mathsf{c}}(\mathrm{d}s) - \int_0^t \frac{\mathbb{1}_{\{p_j^{\rho,\mathsf{c}}(s-)>\varepsilon\}} X_j^{\rho,\mathsf{c}}(s-)}{p_j^{\rho,\mathsf{c}}(s-)} \Lambda_{jk}^{\rho,\varepsilon}(\mathrm{d}s) \\ &= \frac{X_{jk}^{\rho,\mathsf{c}}(t)}{p_j^{\rho,\mathsf{c}}(t) \vee \varepsilon} - \int_0^t X_{jk}^{\rho,\mathsf{c}}(s)\,\mathrm{d}\Big(\frac{1}{p_j^{\rho,\mathsf{c}}(s) \vee \varepsilon}\Big) - \int_0^t \frac{\mathbb{1}_{\{p_j^{\rho,\mathsf{c}}(s-)>\varepsilon\}} X_j^{\rho,\mathsf{c}}(s-)}{p_j^{\rho,\mathsf{c}}(s-)} \Lambda_{jk}^{\rho,\varepsilon}(\mathrm{d}s). \end{aligned} \tag{4.4.11}$$

This limit process is again a centered tight Gaussian process. (In principle, one needs to verify that the finite-dimensional distributions are Gaussian, but this follows directly from the above decomposition.) To extend the result to $(\Lambda_{jk}^{\rho,\varepsilon})_{j\neq k}$, it suffices to verify convergence of the finite-dimensional distributions, confer with Problem 1.5.3 in [VW96]. However, this follows immediately by the Cramer–Wold device in combination with (4.4.11) and the joint Gaussianity from Lemma 4.4.3. □

We now turn our attention to the two-dimensional case. We follow the strategy outlined in Remark 5.12 of [Bat24].

Lemma 4.4.5.
Suppose that $\mathbb{E}[N_{jk}(\theta)^4] < \infty$ for $j, k \in \mathcal{J}$, $j \neq k$. It then holds that

$$\Big(\sqrt{n}(\mathbb{I}_{\boldsymbol{j}}(\boldsymbol{t}) - p_{\boldsymbol{j}}^{\mathsf{c}}(\boldsymbol{t}))_{\boldsymbol{j}}, \sqrt{n}(\mathbb{N}_{\boldsymbol{jk}}(\boldsymbol{t}) - p_{\boldsymbol{jk}}^{\mathsf{c}}(\boldsymbol{t}))_{\boldsymbol{j}\neq\boldsymbol{k}}\Big)_{\boldsymbol{t}\in[0,\theta]\times[0,\theta]}$$

converges weakly to a centered tight Gaussian process.

Proof. Similar to the one-dimensional case, confer with the proof of Lemma 4.4.3, the link (4.4.6) and the link (4.4.9) imply that it suffices to establish asymptotic tightness of the components involved. In the following, we argue that the empirical process

$$[0,\theta]\times[0,\theta] \ni \boldsymbol{t} \mapsto \sqrt{n}(\mathbb{N}_{\boldsymbol{jk}}(\boldsymbol{t}) - p_{\boldsymbol{jk}}^{\mathsf{c}}(\boldsymbol{t}))$$

is asymptotically tight. For the other components, the argument is either simpler or quite similar and therefore, for the sake of brevity, omitted. We shall employ the bracketing central limit theorem with the envelope function

$$F = N_{j_1k_1}(\theta)N_{j_2k_2}(\theta),$$

which is square integrable by assumption. To this end, let the μ-bracketing number be denoted by $N_{[]}(\mu, \mathcal{P}, L_2)$. This is the minimum number of sets N_μ in an n-independent partition $\mathcal{P} = \bigcup_{m=1}^{N_\mu} \mathcal{P}_{\mu,m}$ of the index set $[0,\theta]\times[0,\theta]$ such that, for every partitioning set $\mathcal{P}_{\mu,m}$,

$$\mathbb{E}\Big[\sup_{\boldsymbol{t},\boldsymbol{s}\in\mathcal{P}_{\mu,m}} \big|N_{j_1k_1}(t_1\wedge R)N_{j_2k_2}(t_2\wedge R) - N_{j_1k_1}(s_1\wedge R)N_{j_2k_2}(s_2\wedge R)\big|^2\Big] \leq \mu^2.$$

Since the envelope function is square integrable by assumption, it suffices to identify an n-independent partition that confirms the entropy condition

$$\int_0^{\delta_n} \sqrt{\log N_{[]}(\mu, \mathcal{P}, L_2)}\,\mathrm{d}\mu \to 0, \quad \text{for every } \delta_n \downarrow 0.$$

Note that if $\mathcal{P}_{\mu,m}$ is a rectangle $[\underline{t}_m, \bar{t}_m]\times[\underline{s}_m, \bar{s}_m]$, then by applying the equality $ab - cd = a(b-d) + d(a-c)$, we may bound

$$\sup_{\boldsymbol{t},\boldsymbol{s}\in\mathcal{P}_{\mu,m}} \big|N_{j_1k_1}(t_1\wedge R)N_{j_2k_2}(t_2\wedge R) - N_{j_1k_1}(s_1\wedge R)N_{j_2k_2}(s_2\wedge R)\big|^2$$

by twice the sum of

$$\begin{aligned}&\mathbb{E}[N_{j_1k_1}(\theta)^2 N_{j_2k_2}(\theta)(N_{j_2k_2}(\overline{s}_m \wedge R) - N_{j_2k_2}(\underline{s}_m \wedge R))]\\ &+ \mathbb{E}[N_{j_1k_1}(\theta) N_{j_2k_2}(\theta)^2 (N_{j_1k_1}(\overline{t}_m \wedge R) - N_{j_1k_1}(\underline{t}_m \wedge R))]\end{aligned}$$

due to the monotonicity of the processes involved. For each of these terms, the argument of Example 2.11.16 in [VW96] applies. In particular, there exists an n-independent partition $\mathcal{P}$ of at most $4\mathbb{E}[N_{j_1k_1}(\theta)^2 N_{j_2k_2}(\theta)^2]\mu^{-2}$ rectangles such that, for every rectangle $\mathcal{P}_{\mu,m}$,

$$\mathbb{E}\Big[\sup_{\boldsymbol{t},\boldsymbol{s}\in\mathcal{P}_{\mu,m}} \big|N_{j_1k_1}(t_1 \wedge R)N_{j_2k_2}(t_2 \wedge R) - N_{j_1k_1}(s_1 \wedge R)N_{j_2k_2}(s_2 \wedge R)\big|^2\Big] \leq \mu^2.$$

Consequently, the entropy condition is satisfied. To see this, consider $\delta_n \downarrow 0$. For n sufficiently large and an appropriate constant $K > 0$,

$$\begin{aligned}\int_0^{\delta_n} \sqrt{\log N_{[]}(\mu, \mathcal{P}, L_2)}\,\mathrm{d}\mu &\leq \int_0^{\delta_n} \sqrt{\log\{4\mathbb{E}[N_{j_1k_1}(\theta)^2 N_{j_2k_2}(\theta)^2]\mu^{-2}\}}\,\mathrm{d}\mu\\ &\leq K\int_0^{\delta_n} \sqrt{\mu^{-1}}\,\mathrm{d}\mu\\ &= 2K\sqrt{\delta_n} \to 0.\end{aligned}$$

This completes the proof. □

Theorem 4.4.6.
Suppose that $\mathbb{E}[N_{jk}(\theta)^4] < \infty$ for $j, k \in \mathcal{J}$, $j \neq k$. It then holds that

$$\Big(\sqrt{n}(\hat{p}^{\varepsilon}_{\boldsymbol{j}}(\boldsymbol{t}) - p^{\varepsilon}_{\boldsymbol{j}}(\boldsymbol{t}))_{\boldsymbol{j}}, \sqrt{n}(\hat{\Lambda}^{\varepsilon}_{\boldsymbol{jk}}(\boldsymbol{t}) - \Lambda^{\varepsilon}_{\boldsymbol{jk}}(\boldsymbol{t}))_{\boldsymbol{j}\neq\boldsymbol{k}}\Big)_{\boldsymbol{t}\in[0,\theta]\times[0,\theta]}$$

converges weakly to a centered tight Gaussian process.

Proof. The solution to the integral equations (4.4.3) may be expressed with the help of a Peano series, see Lemma A.1 in [Bat24], which is a Hadamard differentiable functional, see Proposition 3.4 in [GLW95]. Therefore, it suffices by the functional delta method to establish weak convergence of

$$\Big(\sqrt{n}(\hat{\Lambda}^{\varepsilon}_{\boldsymbol{jk}}(\boldsymbol{t}) - \Lambda^{\varepsilon}_{\boldsymbol{jk}}(\boldsymbol{t}))_{\boldsymbol{j}\neq\boldsymbol{k}}\Big)_{\boldsymbol{t}\in[0,\theta]\times[0,\theta]}.$$

Let $\boldsymbol{t} \mapsto X^{\mathsf{c}}_{\boldsymbol{j}}(\boldsymbol{t})$ and $\boldsymbol{t} \mapsto X^{\mathsf{c}}_{\boldsymbol{jk}}(\boldsymbol{t})$ be the limit processes of $\boldsymbol{t} \mapsto \sqrt{n}(\mathbb{I}_{\boldsymbol{j}}(\boldsymbol{t}) - p^{\mathsf{c}}_{\boldsymbol{j}}(\boldsymbol{t}))$ and $\boldsymbol{t} \mapsto \sqrt{n}(\mathbb{N}_{\boldsymbol{jk}}(t) - p^{\mathsf{c}}_{\boldsymbol{jk}}(\boldsymbol{t}))$, respectively. Now, similar to the one-dimensional case, confer with the proof of Theorem 4.4.4, one may show that $\boldsymbol{t} \mapsto \sqrt{n}(\hat{\Lambda}^{\varepsilon}_{\boldsymbol{jk}}(\boldsymbol{t}) - \Lambda^{\varepsilon}_{\boldsymbol{jk}}(\boldsymbol{t}))$ converges to the limit process

$$\boldsymbol{t} \mapsto \int_{\boldsymbol{0}}^{\boldsymbol{t}} \frac{1}{p_j^{\rho,\mathsf{c}}(\boldsymbol{s}-) \vee \varepsilon} X^{\mathsf{c}}_{\boldsymbol{jk}}(\mathrm{d}\boldsymbol{s}) - \int_{\boldsymbol{0}}^{\boldsymbol{t}} \frac{\mathbb{1}_{\{p^{\mathsf{c}}_{\boldsymbol{j}}(\boldsymbol{s}-)>\varepsilon\}} X^{\mathsf{c}}_{\boldsymbol{j}}(\boldsymbol{s}-)}{p^{\mathsf{c}}_{\boldsymbol{j}}(\boldsymbol{s}-)} \Lambda^{\varepsilon}_{\boldsymbol{jk}}(\mathrm{d}\boldsymbol{s}),$$

which is again a centered tight Gaussian process. (In principle, one needs to verify that the finite-dimensional distributions are Gaussian, but this follows by standard arguments applied to the above decomposition in combination with two-dimensional integration by parts.) The extension from fixed indices $\boldsymbol{j}$ and $\boldsymbol{k}$ to across indices transpires exactly as in the one-dimensional case. □

4.4.4 Discussion and remarks

Theorems 4.4.4 and 4.4.6 are of qualitative nature. In principle, it is possible to calculate the covariances and, based hereon, propose plug-in estimators. This was done in [Gli02] for the one-dimensional case without scaling, and the scaled case should follow along similar lines. For the two-dimensional case, the calculations become very involved. In general, we suggest instead the use of the bootstrap. Its validity must be verified using, for instance, the (bootstrap) functional delta method and related results, see also Theorem 5.1 in [GLW95] for the validity of the bootstrap for bivariate survival functions.

The results of Subsections 4.4.2 and 4.4.3 concern occupation probabilities and hazards. In practice, we are interested in expected accumulated cash flows and prospective reserves. If, however, we assume the transition payments b_{jk} to be of finite variation, then the convergences immediately carry over to ε-perturbed versions of the quantities of interest by the continuous mapping theorem and functional delta method, respectively, and these quantities converge to the true quantities of interest as $\varepsilon \to 0$, see also Section 6 in [Bat24].

In the next section, we perform a comparative simulation study. In particular, we compare the one-dimensional estimators of [CF22] with the one-dimensional estimators of this paper. Therefore, we briefly introduce the former.

The estimators of [CF22] are based on the results at the end of Subsection 4.3.1. They correspond to the usual Nelson–Aalen and Aalen–Johansen estimators (for occupation probabilities only), but using the data

$$\big((\tilde{Z}_t^\ell)_{0\le t\le R^\ell}, \tilde{\eta}^\ell \wedge R^\ell\big)_{\ell=1}^n$$

with $\tilde{Z}_t^\ell = \mathbb{1}_{\{Z_t^\ell \in \mathcal{J}_0\}} Z_t^\ell + \mathbb{1}_{\{Z_t^\ell \in \mathcal{J}_1\}} (\mathbb{1}_{\{U^\ell \le \rho(\tau^\ell, Z^\ell_{\tau^\ell -}, Z^\ell_{\tau^\ell})\}} Z_t^\ell + \mathbb{1}_{\{U^\ell > \rho(\tau^\ell, Z^\ell_{\tau^\ell -}, Z^\ell_{\tau^\ell})\}} \nabla)$, where $(U_\ell)_{\ell=1}^n$ is an i.i.d. sequence of Unif$(0,1)$-distributed random variables that is independent of $(Z^\ell, R^\ell)_{\ell=1}^n$. The resulting estimators are of the same type as the one-dimensional estimators studied in this paper, but without scaling. Their asymptotic theory and numerical implementation is already well understood, but might then also be derived from the results obtained here. However, the usage of the auxiliary variables $(U_\ell)_{\ell=1}^n$ introduces additional and, as we substantiate in the next section, unnecessary noise.

4.5 Comparative simulation study

The setup of the simulation study is adopted from [Bat24], the emphasize being on simplicity rather than authenticity. We consider a multi-state model with six states as in Figure 4.1, which allows for the modeling of the free policy as well as the surrender option. In this specific example, the jump process can at most be semi-Markov.

The insured has age 40 years at contract inception. The contractual payments consist of an initial premium of $100,000$ monetary units, an annual premium rate while alive of $10,000$ monetary units until retirement at age 65 years, and an annual benefit rate while

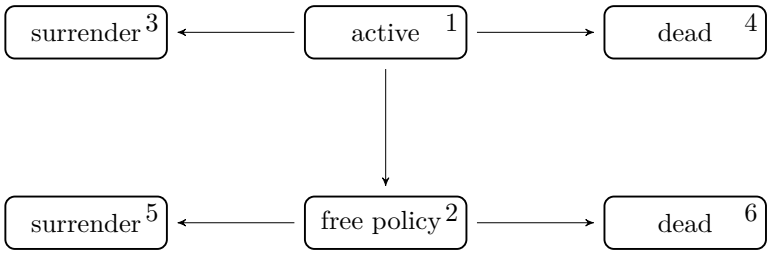


Figure 4.1: Survival model with free policy and surrender options.

alive from onset of retirement. Policyholder options are typically not explicitly included on the technical basis; this is achieved by determining the surrender payments and the free policy factor $t \mapsto \rho(t)$ such that there is no change in technical reserve upon exercise of the option. In particular, the technical reserve may be calculated in a survival model, and only a technical mortality rate and technical interest rate needs to be specified. We suppose that the technical mortality rate is

$$\mu^\star(t) = 0.005 + 10^{(5.728-10+0.038t)},$$

and that the technical interest rate is zero. Finally, we assume that the annual benefit rate has been determined to maintain actuarial equivalence under this technical basis. All in all, we then get

$$\begin{aligned}
&B_1(\mathrm{d}t) = \big(-10,000\mathbb{1}_{[0,25)}(t) + 22,658.671\mathbb{1}_{[25,\infty)}(t)\big)\,\mathrm{d}t,\\
&B_2(\mathrm{d}t) = 22,658.671\mathbb{1}_{[25,\infty)}(t)\,\mathrm{d}t,\\
&b_0 = 100,000, \quad b_{13}(t) = V^\star(t), \quad b_{25}(t) = V^{\star,+}(t), \quad \rho(t) = \frac{V^\star(t)}{V^{\star,+}(t)},
\end{aligned}$$

where $V^\star$ is the technical reserve, and $V^{\star,+}$ is the technical reserve of benefits solely. The resulting scaled payment stream reads

$$\begin{aligned}
&B(\mathrm{d}t)\\
&= \rho(t)^{\mathbb{1}_{\{\tau\le t\}}}\Big(\mathbb{1}_{\{Z_{t-}=1\}}B_1(\mathrm{d}t) + \mathbb{1}_{\{Z_{t-}=2\}}B_2(\mathrm{d}t) + b_{13}(t)N_{13}(\mathrm{d}t) + b_{25}(t)N_{25}(\mathrm{d}t)\Big).
\end{aligned}$$

The one-dimensional representation of the corresponding expected accumulated cash flow reads

$$\begin{aligned}
A(t) = b_0 + \int_0^t \Big(&p_1(s-)B_1(\mathrm{d}s) + p_1(s-)b_{13}(s)\Lambda_{13}(\mathrm{d}s)\\
&+ p_2^\rho(s-)B_2(\mathrm{d}s) + p_2^\rho(s-)b_{13}(s)\Lambda_{25}^\rho(\mathrm{d}s)\Big),
\end{aligned} \tag{4.5.1}$$

while the two-dimensional representation becomes

$$\begin{aligned} A(t) = \int_0^t \int_{\mathbf{0}}^{(t,s-)} \rho(u_1) \Big(& p_{(1,1)}(\boldsymbol{u}-) \Lambda_{(1,1)(2,2)}(\mathrm{d}\boldsymbol{u}) \\ & - p_{(1,2)}(\boldsymbol{u}-) \big(\Lambda_{(1,2)(2,5)}(\mathrm{d}\boldsymbol{u}) + \Lambda_{(1,2)(2,6)}(\mathrm{d}\boldsymbol{u}) \big) \Big) B_2(\mathrm{d}s) \\ & + b_0 + \int_0^t p_1(s-) \Big(B_1(\mathrm{d}s) + b_{13}(s) \Lambda_{13}(\mathrm{d}s) \Big) \\ & + \int_{\mathbf{0}}^{(t,t)} \rho(u_1) p_{(1,2)}(\boldsymbol{u}-) b_{25}(u_2) \Lambda_{(1,2)(2,5)}(\mathrm{d}\boldsymbol{u}). \end{aligned} \tag{4.5.2}$$

Data is simulated from a semi-Markov model with initial state $z_0 = 1$ and the following (duration-dependent) transition rates:

$$\begin{aligned} \mu_{14}(t) &= \mu_{26}(t) = \mu^\star(t), \\ \mu_{12}(t) &= 0.1 \mathbb{1}_{[0,25)}(t), \\ \mu_{13}(t) &= 0.05 \mathbb{1}_{[0,25)}(t), \\ \mu_{25}(t,u) &= \big(0.05 + 0.2 \mathbb{1}_{[\frac{1}{2},\frac{5}{2})}(u)\big) \mathbb{1}_{[0,25)}(t). \end{aligned}$$

The right-censoring times are, solely for illustrative purposes, drawn from the distribution $\mathrm{Unif}(20, 80)$.

4.5.1 Estimates

Recall that the perturbation given by ε only plays a role in the theoretical analyses of the estimators. Correspondingly, throughout the simulation study, we set ε equal to zero and suppress it notationally. True values from the underlying semi-Markov model are calculated using Monte Carlo methods based on an uncensored sample of size $n = 10,000$.

First, we compare our scaled Aalen–Johansen estimator of the scaled occupation probabilities in state 2 with the proposed estimator of [CF22], which relies on change of measure arguments; we use the acronyms SAJ and CMAJ, respectively. The estimators are computationally similar, the complexity being $\mathcal{O}(n)$ for both. In Figure 4.2, we have plotted the resulting estimates for $n = 2,000$ and, in the case of the CMAJ estimator, for three different realizations of its auxiliary uniformly distributed random variables.

One immediate observations is that these uniformly distributed random variables induce additional noise to the estimates; apart from that, the estimators exhibit very similar characteristics. This is further illustrated in Figure 4.3, where we consider 100 realizations of the auxiliary random variables, but take a snapshot at $t = 20$. The immediate and unsurprising conclusion is that the SAJ estimator strictly improves on the CMAJ estimator, since it eliminates the artificial noise introduced by the auxiliary random variables.

We now turn to the comparison of the SAJ estimator with the two-dimensional Aalen–Johansen (2dAJ) estimator. We focus on the expected accumulated cash flow $t \mapsto$

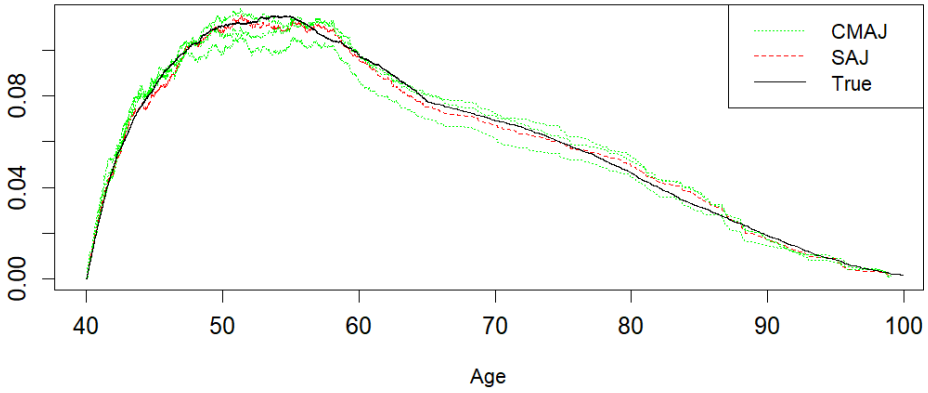


Figure 4.2: Estimates of the scaled occupation probabilities p_2^ρ (black) for $n = 2,000$ using the SAJ (red) and the CMAJ (green) estimators; the latter is illustrated with three different realizations of its auxiliary uniformly distributed random variables.

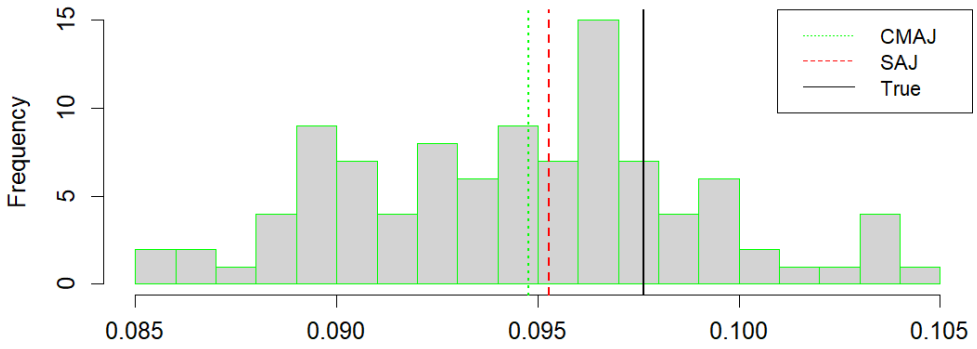


Figure 4.3: Histogram of estimates of the scaled occupation probability $p_2^\rho(20)$ (black) for $n = 2,000$ and 100 realizations of the auxiliary uniformly distributed random variables together with the resulting average estimate (green) and the SAJ estimate (red).

$A(t)$ to avoid comparing apples with pears. From (4.5.1), we obtain the plug-in SAJ estimator

$$t \mapsto b_0 + \int_0^t \Big(\hat{p}_1(s-)B_1(\mathrm{d}s) + \hat{p}_1(s-)b_{13}(s)\hat{\Lambda}_{13}(\mathrm{d}s) + \hat{p}_2^{\rho}(s-)B_2(\mathrm{d}s) + \hat{p}_2^{\rho}(s-)b_{13}(s)\hat{\Lambda}_{25}^{\rho}(\mathrm{d}s)\Big),$$

while (4.5.2) yields the plug-in 2dAJ estimator

$$\begin{aligned} t \mapsto b_0 &+ \int_0^t \hat{p}_1(s-)\Big(B_1(\mathrm{d}s) + b_{13}(s)\hat{\Lambda}_{13}(\mathrm{d}s)\Big) \\ &+ \int_0^t \int_{\mathbf{0}}^{(t,s-)} \rho(u_1)\Big(\hat{p}_{(1,1)}(\boldsymbol{u}-)\hat{\Lambda}_{(1,1)(2,2)}(\mathrm{d}\boldsymbol{u}) \\ &\qquad - \hat{p}_{(1,2)}(\boldsymbol{u}-)\big(\hat{\Lambda}_{(1,2)(2,5)}(\mathrm{d}\boldsymbol{u}) + \hat{\Lambda}_{(1,2)(2,6)}(\mathrm{d}\boldsymbol{u})\big)\Big)B_2(\mathrm{d}s) \\ &+ \int_{\mathbf{0}}^{(t,t)} \rho(u_1)\hat{p}_{(1,2)}(\boldsymbol{u}-)b_{25}(u_2)\hat{\Lambda}_{(1,2)(2,5)}(\mathrm{d}\boldsymbol{u}). \end{aligned}$$

The computation of these estimators requires numerical integration, which reduces to a one- and two-dimensional recursion, respectively. In particular, the 2dAJ estimator is computationally more demanding, the complexity being $\mathcal{O}(n^2)$ rather than $\mathcal{O}(n)$. Comparing the two estimators, one notices another fundamental difference: the latter separates ρ from the estimation procedure. If the contractual payments differ between insured, then so do potentially the scaling factors ρ, which would necessitate up to n SAJ estimates, increasing the complexity of the SAJ estimator to $\mathcal{O}(n^2)$.

In Figures 4.4 and 4.5, we have plotted the resulting estimates for $n = 2,000$ and $n = 5,000$, respectively. In the body, both estimators perform well and essentially identically. However, in the tail, the SAJ estimator outperforms the 2dAJ estimator. This deficiency of the 2dAJ estimator is in line with earlier results for the bivariate survival case reported in, for instance, the last paragraph of Subsection 4.3 in [GJ90]. In light hereof, we now give some further theoretical comments following and extending Section 7 in [Bat24].

Now, in the absence of censoring, the estimators agree and equal the non-parametric maximum likelihood estimator. This follows from the fact that $\mathbb{I}^{\rho} = \hat{p}^{\rho}$ and $\mathbb{I}_{\boldsymbol{j}} = \hat{p}_{\boldsymbol{j}}$ when no censoring is present. Under right-censoring, the 2dAJ estimator may fail to satisfy monotonicity and positivity; this was already noted in [Dab88] for the bivariate survival case. It also introduces additional sub-sampling, which is best illustrated via a simple example. Consider three individuals with the following event times and types:

Individual I:	times $= (\tau^1, \sigma^1)$,	types $= (1 \to 2, 2 \to 4)$,
Individual II:	times $= (\tau^2, R^2)$,	types $= (1 \to 2, \text{right-censoring})$,
Individual III:	times $= (\tau^3, \sigma^3)$,	types $= (1 \to 2, 2 \to 4)$

with $\tau^3 < \tau^1 < \tau^2 < \sigma^1 < R^2 < \sigma^3$. The plug-in SAJ estimate of $A(\sigma^3) - A(R^2)$ becomes $\hat{p}_2^{\rho}(R^2)(B_2(\sigma^3) - B_2(R^2))$, where $\hat{p}_2^{\rho}(R^2)$ is given by

$$\hat{p}_2^{\rho}(R^2) = \rho(\tau^3)\frac{\hat{p}_1^{\rho}(0)}{\mathbb{I}_1^{\rho}(0)}\frac{1}{3} + \rho(\tau^1)\frac{\hat{p}_1^{\rho}(\tau^3)}{\mathbb{I}_1^{\rho}(\tau^3)}\frac{1}{3} + \rho(\tau^2)\frac{\hat{p}_1^{\rho}(\tau^1)}{\mathbb{I}_1^{\rho}(\tau^1)}\frac{1}{3} - \rho(\tau^1)\frac{\hat{p}_2^{\rho}(\sigma^1)}{\mathbb{I}_2^{\rho}(\sigma^1)}\frac{1}{3},$$

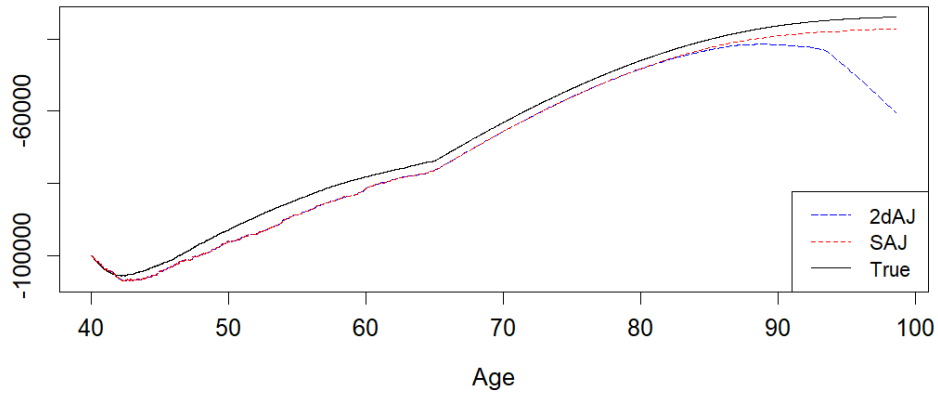


Figure 4.4: Estimates of the expected accumulates cash flow $t \mapsto A(t)$ (black) for $n = 2,000$ using the SAJ (red) and the 2dAJ (blue) estimators.

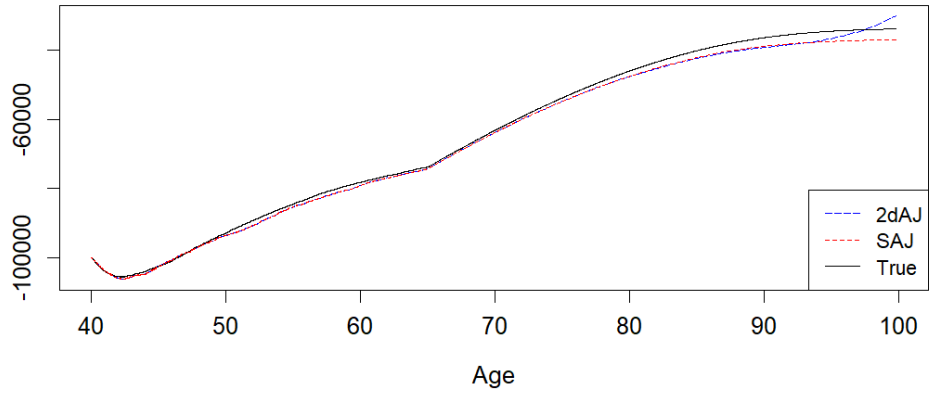


Figure 4.5: Estimates of the expected accumulates cash flow $t \mapsto A(t)$ (black) for $n = 5,000$ using the SAJ (red) and the 2dAJ (blue) estimators.

while the plug-in 2dAJ estimate reads

$$\frac{1}{3}\Big(\rho(\tau^3)\frac{\hat{p}_{(1,1)}(0,0)}{\mathbb{I}_{(1,1)}(0,0)} + \rho(\tau^1)\frac{\hat{p}_{(1,1)}(\tau^3,\tau^3)}{\mathbb{I}_{(1,1)}(\tau^3,\tau^3)} + \rho(\tau^2)\frac{\hat{p}_{11}(\tau^1,\tau^1)}{\mathbb{I}_{(1,1)}(\tau^1,\tau^1)} - \rho(\tau^1)\frac{\hat{p}_{(1,2)}(\tau^1,\sigma^1)}{\mathbb{I}_{(1,2)}(\tau^1,\sigma^1)}\Big)\Big(B_2(\sigma^3) - B_2(R^2)\Big).$$

For these three individuals, right-censoring only takes place from state 2, so the first three summands across estimates agree. The only difference lies in the last summands, that is between

$$\rho(\tau^1)\frac{\hat{p}^{\rho}_2(\sigma^1)}{\mathbb{I}^{\rho}_2(\sigma^1)} \quad \text{and} \quad \rho(\tau^1)\frac{\hat{p}_{(1,2)}(\tau^1,\sigma^1)}{\mathbb{I}_{(1,2)}(\tau^1,\sigma^1)},$$

where the 2dAJ estimator performs additional sub-sampling.

4.5.2 Conclusion

The SAJ estimator strictly outperforms the CMAJ estimator as it foregoes the latter's auxiliary uniformly distributed random variables that add unnecessary noise. The SAJ and the 2dAJ estimators show comparable performance, but have different strengths and weaknesses. The SAJ estimator has much lower variability in the tail, is easier to implement due to its similarity with known estimators for Markov models, and is computationally far less demanding if the scaling factors do not differ between insured. The 2dAJ estimator is computationally demanding and more difficult to implement, however once implemented it achieves true separation between the contractual payments, including the scaling factor, and the statistical estimation of biometric and behavioral risks. This may be particularly important if the estimates are to find application across insurers, as is commonly the case for actuarial tables.

4.6 Extension

The preceding sections focused on scaled payments with H given by

$$H(t) = \rho(\tau, Z_{\tau-}, Z_\tau)^{\mathbb{1}_{\{\tau\leq t\}}}, \tag{4.6.1}$$

where τ was interpreted as the exercise time of some policyholder option. This is not the only actuarial example of some importance. If we instead take

$$H(t) = \exp\Big\{-\int_0^t \delta(u)\,\mathrm{d}u\Big\}, \tag{4.6.2}$$

then B given by

$$B(\mathrm{d}t) = H(t)B^\circ(\mathrm{d}t), \qquad B(0) = B^\circ(0),$$

constitutes the discounted version of B° using the (stochastic) interest rate δ, and $t \mapsto A(t) = \mathbb{E}[B(t)]$ becomes the expected accumulated discounted cash flow.

In the following, we explain how the key results of the preceding sections still apply, even if the stochastic process H is only required to be right-continuous, of finite variation, uniformly bounded on compacts, and – in regards to non-parametric estimation – adapted to the filtration $\mathbb{F}^Z$. In particular, no reference to (4.6.1) is made. This serves to showcase a broader applicability of our contribution and might, especially with (4.6.2) in mind, be of independent interest.

Throughout, we assume that H is adapted to $\mathbb{F}$, right-continuous, of finite variation, and uniformly bounded on compacts. We further define

$$\begin{aligned} \bar{p}_j(t) &= \mathbb{E}[H(t)\mathbb{1}_{\{Z_t=j\}}], \\ \bar{H}_j(\mathrm{d}t) &= \frac{1}{\bar{p}_j(t-)}\mathbb{E}[\mathbb{1}_{\{Z_{t-}=j\}}H(\mathrm{d}t)], \qquad (4.6.3) \\ \bar{\Lambda}_{jk}(\mathrm{d}t) &= \frac{1}{\bar{p}_j(t-)}\mathbb{E}[H(t)N_{jk}(\mathrm{d}t)], \qquad \bar{\Lambda}_{jj} = \bar{H}_j - \sum_{\substack{\ell\in\mathcal{J}\\ \ell\neq j}} \bar{\Lambda}_{j\ell}, \qquad (4.6.4) \end{aligned}$$

for $j,k \in \mathcal{J}$, $j \neq k$. The following result mirrors Proposition 4.3.3.

Proposition 4.6.1.
It holds that

$$A(t) = \sum_{j\in\mathcal{J}} \bar{p}_j(t-)\Big((1+\Delta\bar{H}_j(t))B_j(\mathrm{d}t) + \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} b_{jk}(t)\bar{\Lambda}_{jk}(\mathrm{d}t)\Big)$$

and that

$$\frac{\bar{p}_j(t)}{\mathbb{E}[H(0)]} = \Big[\prod_0^t (Id + \bar{\Lambda}(\mathrm{d}s))\Big]_{z_0 j}, \qquad j \in \mathcal{J}.$$

Proof. Note that

$$\mathbb{E}[H(t)\mathbb{1}_{\{Z_{t-}=j\}}] = \bar{p}_j(t-) + \mathbb{E}\Big[(H(t)-H(t-))\mathbb{1}_{\{Z_{t-}=j\}}] = \bar{p}_j(t-)(1+\Delta\bar{H}_j(t)).$$

It then immediately follows that

$$A(t) = \sum_{j\in\mathcal{J}} \bar{p}_j(t-)\Big((1+\Delta\bar{H}_j(t))B_j(\mathrm{d}t) + \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} b_{jk}(t)\bar{\Lambda}_{jk}(\mathrm{d}t)\Big),$$

which verifies the first statement of the proposition. The proof of the second statement mirrors the proof of Proposition 4.3.3, but requires the following key identity:

$$\begin{aligned} H(t)\mathbb{1}_{\{Z_t=j\}} = \mathbb{1}_{\{j=z_0\}}H(0) &+ \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} \int_0^t H(s)N_{kj}(\mathrm{d}s) \\ &- \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} \int_0^t H(s)N_{jk}(\mathrm{d}s) + \int_0^t \mathbb{1}_{\{Z_{s-}=j\}}H(\mathrm{d}s), \end{aligned} \qquad (4.6.5)$$

which follows from (4.3.1) and an application of integration by parts. □

Remark 4.6.2.
Another way to phrase the second part of the theorem is that $(\bar{p}_j)_j$ uniquely solves the forward integral equations

$$\bar{p}_j(\mathrm{d}t) = \sum_{k\in\mathcal{J}} \bar{p}_k(t-)\bar{\Lambda}_{kj}(\mathrm{d}t), \qquad\qquad j \in \mathcal{J},$$

with initial conditions $\bar{p}_j(0) = \mathbb{1}_{\{j=z_0\}}\mathbb{E}[H(0)]$.

Remark 4.6.3.
The product integral representation of $(\bar{p}_j)_j$, and the equivalent forward integral equations, constitute an extension of previous insights for Markov processes to the non-Markov regime; confer, for instance, with Appendix 1 in [BFM20] as well as Theorem 3.1 in [AB23].

Remark 4.6.4.
If H is taken according to (4.6.2), then the equations (4.6.3)–(4.6.4) yield the correct Ansatz for a proper constructive and simultaneous definition of forward interest and forward transition rates in the spirit of Definition 4.1 in [Buc17], with the forward integral equations given in Remark 4.6.2 offering the appropriate extension of Theorem 4.4 in [Buc17]. This addresses an open question that dates back to at least [MP05]; confer also, for instance, with [Nor10; Buc14; CN15; BFS19].

We now turn to non-parametric estimation under the assumption that H is adapted to $\mathbb{F}^Z$. We shall not carry out all arguments in complete detail, but just report on the key findings. If H is taken according to for instance (4.6.2), one could readily ask why assuming H to be adapted to $\mathbb{F}^Z$ has merit compared to, say, just requiring H (and therefore δ) to be deterministic. However, state-dependent interest rates may appear when dealing with certain reserve-dependent expenses, see Section G in [Nor91] for the basic idea as well as [CDD14] and Section 4 in [BFM20] for further details. Therefore, even under the specification (4.6.2), the case of $\mathbb{F}^Z$-adaptedness carries practical value.

As in Section 4.4, we let the right-censoring time R be independent of Z and suppose $((Z_t)_{0\le t\le R}, \eta\wedge R)$ is observed. In particular, the process H is sufficiently observed. Due to right-censoring being entirely random, we obtain the key identities

$$\bar{H}_j(\mathrm{d}t) = \frac{1}{\bar{p}_j^{\mathsf{c}}(t-)}\mathbb{E}[\mathbb{1}_{\{Z_{t-}=j\}}H(\mathrm{d}t\wedge R)],$$
$$\bar{\Lambda}_{jk}(\mathrm{d}t) = \frac{1}{\bar{p}_j^{\mathsf{c}}(t-)}\mathbb{E}[H(t)N_{jk}(\mathrm{d}t\wedge R)],$$

for t below θ and where we define $\bar{p}_j^{\mathsf{c}}(t) = \mathbb{E}[H(t)\mathbb{1}_{\{Z_t=j\}}\mathbb{1}_{\{t<R\}}]$, $j\in\mathcal{J}$; the identities may be compared to (4.4.1).

Based hereon, we may for an i.i.d. sample directly define estimators of $\bar{\Lambda}$ and $\bar{p}$ based on averages of the above quantities, and just like their counterparts in Section 4.4, these

estimators will be strongly consistent and asymptotically normal. This is due to the fact that the statements and proofs of Section 4.4 only to a very limited degree rely on the specification (4.6.1) and when they do, such as in (4.4.4), the statements and proofs are easily adapted to the general case based on (4.6.5).

Acknowledgments

Significant parts of the research presented in this paper were carried out while Christian Furrer was a Junior Fellow at the Hanse-Wissenschaftskolleg Institute for Advanced Study in Delmenhorst, Germany. The authors are grateful to Marcus C. Christiansen for initiating and facilitating their fruitful research collaboration. Christian Furrer also thanks Martin Bladt for valuable discussions on empirical process theory.

Bibliography

[Aal78] O.O. Aalen. „Nonparametric Inference for a Family of Counting Processes“. In: *Annals of Statistics* 6.4 (1978), pp. 701–726. DOI: 10.1214/aos/1176344247.

[AB23] J. Ahmad and M. Bladt. „Phase-type representations of stochastic interest rates with applications to life insurance“. In: *European Actuarial Journal* 13.2 (2023), pp. 571–606. DOI: 10.1007/s13385-023-00346-4.

[ABF22] J Ahmad, K. Buchardt, and C. Furrer. „Computation of bonus in multi-state life insurance“. In: *ASTIN Bulletin* 52.1 (2022), 291–331. DOI: 10.1017/asb.2021.32.

[ABF23] J. Ahmad, M. Bladt, and C. Furrer. „Aggregate Markov models in life insurance: Properties and valuation“. In: *Insurance: Mathematics and Economics* 113 (2023), pp. 50–69. DOI: 10.1016/j.insmatheco.2023.07.006.

[AC34] C.R Adams and J.A. Clarkson. „Properties of Functions f (x, y) of Bounded Variation“. In: *Transactions of the American Mathematical Society* 36.4 (1934), pp. 711–730. DOI: 10.2307/1989819.

[AJ78] O.O. Aalen and S. Johansen. „An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observations“. In: *Scandinavian Journal of Statistics* 5 (1978), pp. 141–150.

[Ams68] M. H. Amsler. *Les chaines de Markov des assurances vie, invalidité et maladie*. 1968.

[And+93] P.K. Andersen et al. *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer, New York, 1993. DOI: 10.1007/978-1-4612-4348-9.

[Bat24] T. Bathke. „On the Estimation of bivariate Conditional Transition Rates“. In: *European Actuarial Journal* (2024). DOI: 10.1080/03461238.2024.2420067.

[BC24] T. Bathke and M.C. Christiansen. „Two-dimensional forward and backward transition rates“. In: *European Actuarial Journal* 14.2 (2024), pp. 411–436. DOI: 10.1007/s13385-023-00363-3.

[BF23] M. Bladt and C. Furrer. *AalenJohansen: Conditional Aalen-Johansen Estimation*. R package version 1.0. 2023. DOI: https://doi.org/10.32614/CRAN.package.AalenJohansen.

[BF24a] T. Bathke and C. Furrer. *Non-parametric estimators of scaled cash flows*. Preprint, arXiv:2408.13176. 2024. DOI: 10.48550/arXiv.2408.13176.

[BF24b] M. Bladt and C. Furrer. *Conditional Aalen–Johansen estimation*. Preprint, arXiv:2303.02119. 2024. DOI: 10.48550/arXiv.2303.02119.

[BFM20] K. Buchardt, C. Furrer, and T. Møller. „Tax- and expense-modified risk-minimization for insurance payment processes“. In: *Scandinavian Actuarial Journal* 2020.10 (2020), pp. 934–961. DOI: 10.1080/03461238.2020.1790413.

[BFS19] K. Buchardt, C. Furrer, and M. Steffensen. „Forward transition rates“. In: *Finance and Stochastics* 23.4 (2019), pp. 975–999. DOI: 10.1007/s00780-019-00397-0.

[BM15] K. Buchardt and T. Møller. „Life Insurance Cash Flows with Policyholder Behavior“. In: *Risks* 3.3 (2015), pp. 290–317. DOI: 10.3390/risks3030290.

[BMS15] K. Buchardt, T. Møller, and K.B. Schmidt. „Cash flows and policyholder behaviour in the semi-Markov life insurance setup“. In: *Scandinavian Actuarial Journal* 2015.8 (2015), pp. 660–688. DOI: 10.1080/03461238.2013.879919.

[Bor05] O. Borgan. „Nelson–Aalen Estimator“. In: *Encyclopedia of Biostatistics* 2 (2005). DOI: 10.1002/0470011815.b2a11054.

[Buc14] K. Buchardt. „Dependent interest and transition rates in life insurance“. In: *Insurance: Mathematics and Economics* 55 (2014), pp. 167–179. DOI: 10.1016/j.insmatheco.2014.01.004.

[Buc17] K. Buchardt. „Kolmogorov's forward PIDE and forward transition rates in life insurance“. In: *Scandinavian Actuarial Journal* 2017.5 (2017), pp. 377–394. DOI: 10.1080/03461238.2016.1160255.

[CA33] J.A. Clarkson and C.R. Adams. „On Definitions of Bounded Variation for Functions of Two Variables“. In: *Transactions of the American Mathematical Society* 35.4 (1933), pp. 824–854. DOI: 10.2307/1989593.

[CD24] M.C. Christiansen and B. Djehiche. *As-if Markov reserves for reserve-dependen payments*. Preprint, SSRN. 2024. DOI: /10.2139/ssrn.4962272.

[CDD14] M.C. Christiansen, M.M. Denuit, and J. Dhaene. „Reserve-dependent benefits and costs in life and health insurance contracts“. In: *Insurance: Mathematics and Economics* 57 (2014), pp. 132–137. DOI: 10.1016/j.insmatheco.2014.05.009.

[CF22] M.C. Christiansen and C. Furrer. „Extension of as-if-Markov modeling to scaled payments“. In: *Insurance: Mathematics and Economics* 107 (2022), pp. 288–306. DOI: 10.1016/j.insmatheco.2022.09.001.

[Chr10] M.C. Christiansen. „Biometric worst-case scenarios for multi-state life insurance policies“. In: *Insurance: Mathematics and Economics* 47.2 (2010), pp. 190–197. DOI: 10.1016/j.insmatheco.2010.05.002.

[Chr12] M.C. Christiansen. „Multistate models in health insurance“. In: *Advances in Statistical Analysis* 96 (2012), pp. 155–186. DOI: 10.1007/s10182-012-0189-2.

[Chr21] M.C. Christiansen. „On the calculation of prospective and retrospective reserves in non-Markov models“. In: *European Actuarial Journal* 11.2 (2021), pp. 441–462. DOI: 10.1007/s13385-021-00277-y.

[CN15] M. C. Christiansen and A. Niemeyer. „On the forward rate concept in multi-state life insurance“. In: *Finance and Stochastics* 19.2 (2015), pp. 295–327. DOI: 10.1007/s00780-014-0244-9.

[Dab88] D.M. Dabrowska. „Kaplan–Meier Estimate on the Plane“. In: *Annals of Statistics* 16.4 (1988), pp. 1475–1489. DOI: 10.1214/aos/1176351049.

[DS01] S. Datta and G.A. Satten. „Validity of the Aalen–Johansen estimators of stage occupation probabilities and Nelson–Aalen estimators of integrated transition hazards for non-Markov models“. In: *Statistics & Probability Letters* 55.4 (2001), pp. 403–411. DOI: 10.1016/S0167-7152(01)00155-9.

[Fur22] C. Furrer. „Scaled insurance cash flows: representation and computation via change of measure techniques“. In: *Finance and Stochastics* 26.2 (2022), pp. 359–382. DOI: 10.1007/s00780-022-00472-z.

[Gil93] R. D. Gill. „Multivariate Survival Analysis“. In: *Theory of Probability & Its Applications* 37.1 (1993), pp. 18–31. DOI: 10.1137/1137061.

[GJ90] R.D. Gill and S. Johansen. „A Survey of Product-Integration with a View Toward Application in Survival Analysis“. In: *Annals of Statistics* 18.4 (1990), pp. 1501–1555. DOI: 10.1214/aos/1176347865.

[Gli02] D.V. Glidden. „Robust Inference for Event Probabilities with Non-Markov Event Data“. In: *Biometrics* 58.2 (2002), pp. 361–368. DOI: 10.1111/j.0006-341X.2002.00361.x.

[GLW95] R.D. Gill, M.J. Van Der Laan, and J.A. Wellner. „Inefficient estimators of the bivariate survival function for three models“. In: *Annales de l'Institut Henri Poincaré* 31.3 (1995), pp. 545–597.

[GN16] K.S.T. Gad and J.W. Nielsen. „Reserves and cash flows under stochastic retirement“. In: *Scandinavian Actuarial Journal* 2016.10 (2016), pp. 876–904. DOI: 10.1080/03461238.2015.1028432.

[GP18] Q. Guibert and F. Planchet. „Non-parametric inference of transition probabilities based on Aalen–Johansen integral estimators for acyclic multi-state models: application to LTC insurance“. In: *Insurance: Mathematics and Economics* 82 (2018), pp. 21–36. DOI: 10.1016/j.insmatheco.2018.05.004.

[Hel08] M. Helwich. „Durational effects and non-smooth semi-Markov models in life insurance“. PhD thesis. University of Rostock, 2008. DOI: https://doi.org/10.18453/rosdok_id00000324.

[Hen+14] L.F.B. Henriksen et al. „Markov chain modeling of policyholder behavior in life insurance and pension“. In: *European Actuarial Journal* 4 (2014), pp. 1–29. DOI: 10.1007/s13385-014-0091-2.

[HO78] J.M. Hoem and O.O.Aalen. „Actuarial values of payment streams“. In: *Scandinavian Actuarial Journal* 1978.1 (1978), pp. 38–47. DOI: 10.1080/03461238.1978.10414317.

[Hoe69] J.M. Hoem. „Markov Chain Models in Life Insurance“. In: *Blätter der DGVFM* 9 (1969), pp. 91–107. DOI: 10.1007/BF02810082.

[Hoe72] J.M. Hoem. „Inhomogeneous Semi-Markov Processes, Select Actuarial Tables, and Duration-Dependence in Demography“. In: *Population Dynamics*. Ed. by T.N.E. Greville. Academic Press, 1972, pp. 251–296. DOI: 10.1016/B978-1-4832-2868-6.50013-8.

[Hou00] P. Hougaard. *Analysis of Multivariate Survival Data*. Statistics for Biology and Health. Springer New York, 2000. DOI: 10.1007/978-1-4612-1304-8.

[JD84] J. Janssen and R.D. de Dominicis. „Finite non-homogeneous semi-Markov processes: Theoretical and computational aspects“. In: *Insurance: Mathematics and Economics* 3.3 (1984), pp. 157–165. DOI: 10.1016/0167-6687(84)90057-X.

[KGJD02] R. Kaa, M. Goovaerts, and M. Denuit J. Dhaene. *Modern Actuarial Risk Theory*. Springer New York, 2002. DOI: 10.1007/b109818.

[KM58] E.L. Kaplan and P. Meier. „Nonparametric Estimation from Incomplete Observations“. In: *Journal of the American Statistical Association* 53.282 (1958), pp. 457–481. DOI: 10.1080/01621459.1958.10501452.

[Laa96] M.J van der Laan. „Efficient Estimation in the Bivariate Censoring Model and Repairing NPMLE“. In: *The Annals of Statistics* 24.2 (1996), pp. 596–627. DOI: 10.1214/aos/1032894454.

[LP17] G. Last and M. Penrose. *Lectures on the Poisson process*. Vol. 7. Cambridge University Press, 2017.

[Mal+21] N. Maltzahn et al. „A hybrid landmark Aalen–Johansen estimator for transition probabilities in partially non-Markov multi-state models“. In: *Lifetime Data Analysis* 27.4 (2021), pp. 737–760. DOI: 10.1007/s10985-021-09534-4.

[MD13] F. Mostajab and S. Datta. „Nonparametric regression of state occupation, entry, exit, and waiting times with multistate right-censored data“. In: *Statistics in Medicine* 32.17 (2013), pp. 3006–3019. DOI: 10.1002/sim.5703.

[MH99] H. Milbrodt and M. Helbig. *Mathematische Methoden der Personenversicherung*. Walter de Gruyter, 1999. DOI: 10.1515/9783110197952.

[Møl93] C.M. Møller. „A stochastic version of Thiele's differential equation“. In: *Scandinavian Actuarial Journal* 1993 (1993), pp. 1–16. DOI: 10.1080/03461238.1993.10413910.

[MP05] K.R. Miltersen and S. Persson. *Is Mortality Dead? Stochastic Forward Force of Mortality Determined by No Arbitrage*. Working paper, Norwegian School of Economics and Business Administration. 2005.

[MS97] H. Milbrodt and A. Stracke. „Markov models and Thiele's integral equations for the prospective reserve“. In: *Insurance: Mathematics and Economics* 19.3 (1997), pp. 187–235. DOI: 10.1016/S0167-6687(97)00020-6.

[Nel06] R.B. Nelsen. *An Introduction to Copulas*. Springer, 2006. DOI: 10.1007/0-387-28678-0.

[Nie+23] A. Nießl et al. „Statistical inference for state occupation and transition probabilities in non-Markov multi-state models subject to both random left-truncation and right-censoring“. In: *Econometrics and Statistics* 25 (2023), pp. 110–124. DOI: 10.1016/j.ecosta.2021.09.008.

[Nor10] R. Norberg. „Forward mortality and other vital rates – Are they the way forward?“ In: *Insurance: Mathematics and Economics* 47.2 (2010), pp. 105–112. DOI: 10.1016/j.insmatheco.2010.07.002.

[Nor91] R. Norberg. „Reserves in Life and Pension Insurance“. In: *Scandinavian Actuarial Journal* 1991 (1991), pp. 3–24. DOI: 10.1080/03461238.1991.10557357.

[Ove19] M. Overgaard. „State Occupation Probabilities in Non-Markov Models“. In: *Mathematical Methods of Statistics* 28.4 (2019), pp. 279–290. DOI: 10.3103/S1066530719040033.

[PC92] R.L. Prentice and J. Cai. „Covariance and survivor function estimation using censored multivariate failure time data“. In: *Biometrika* 79.3 (1992), pp. 495–512. DOI: 10.1093/biomet/79.3.495.

[Pro05] P.E. Protter. *Stochastic differential equations*. Stochastic Modelling and Applied Probability. Springer Berlin, 2005. DOI: 10.1007/978-3-662-10061-5.

[PS18] H. Putter and C. Spitoni. „Non-parametric estimation of transition probabilities in non-Markov multi-state models: The landmark Aalen–Johansen estimator“. In: *Statistical Methods in Medical Research* 27.7 (2018), pp. 2081–2092. DOI: 10.1177/0962280216674497.

[PZ18] R.L. Prentice and S. Zhao. „Nonparametric estimation of the multivariate survivor function: the multivariate Kaplan-Meier estimator “. In: *Lifetime data analysis* 24 (2018), pp. 3–27. DOI: 10.1007/s10985-016-9383-y.

[R C23] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2023.

[Sho00] G.R. Shorack. *Probability for Statisticians*. Springer Texts in Statistics. Springer, 2000. DOI: 10.1007/978-3-319-52207-4.

[SKD18] C. Siriwardhana, K.B. Kulasekera, and S. Datta. „Flexible semi-parametric regression of state occupational probabilities in a multistate model with right-censored data“. In: *Lifetime data analysis* 24.3 (2018), pp. 464–491. DOI: 10.1007/s10985-017-9403-6.

[SW09] G.R. Shorack and J.A. Wellner. *Empirical Processes with Applications to Statistics*. Classics in Applied Mathematics. unabridged republication of work first published in 1986. Society for Industrial and Applied Mathematics, 2009. DOI: 10.1137/1.9780898719017.

[VW96] A.W. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer New York, 1996. DOI: 10.1007/978-1-4757-2545-2.

[YY24] W.H. Young and G.C. Young. „On the Discontinuities of Monotone Functions of several Variables“. In: *Proceedings of the London Mathematical Society* 2.1 (1924), pp. 124–142. DOI: 10.1112/plms/s2-22.1.124.

Contact

Theis Bathke
Email: theis.bathke@googlemail.com

Affidavit

(English)

I, Theis Bathke, hereby declare that the submitted thesis is my own work. I have only used the sources indicated and have not made unauthorized use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given. I confirm that I followed the guidelines of good scientific practice of the Carl von Ossietzky University Oldenburg.
I further declare that neither the submitted thesis nor parts of it have been presented as part of an examination degree to any other university.

Eidesstattliche Erklärung

(German)

Hiermit versichere ich, Theis Bathke, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Außerdem versichere ich, dass ich die allgemeinen Prinzipien wissenschaftlicher Arbeit und Veröffentlichung, wie sie in den Leitlinien guter wissenschaftlicher Praxis festgelegt sind, befolgt habe.
Ich versichere weiterhin, dass weder diese Arbeit noch Teile davon an einer anderen Universität eingereicht wurden.