Azam Asanjarani, Benoit Lique^{*} and Yoni Nazarathy

Estimation of semi-Markov multi-state models: a comparison of the sojourn times and transition intensities approaches

<https://doi.org/10.1515/ijb-2020-0083>

Received June 3, 2020; accepted December 17, 2020; published online January 6, 2021

Abstract: Semi-Markov models are widely used for survival analysis and reliability analysis. In general, there are two competing parameterizations and each entails its own interpretation and inference properties. On the one hand, a semi-Markov process can be defined based on the distribution of sojourn times, often via hazard rates, together with transition probabilities of an embedded Markov chain. On the other hand, intensity transition functions may be used, often referred to as the hazard rates of the semi-Markov process. We summarize and contrast these two parameterizations both from a probabilistic and an inference perspective, and we highlight relationships between the two approaches. In general, the intensity transition based approach allows the likelihood to be split into likelihoods of two-state models having fewer parameters, allowing efficient computation and usage of many survival analysis tools. Nevertheless, in certain cases the sojourn time based approach is natural and has been exploited extensively in applications. In contrasting the two approaches and contemporary relevant R packages used for inference, we use two real datasets highlighting the probabilistic and inference properties of each approach. This analysis is accompanied by an R vignette.

Keywords: hazard rate; intensity transition; multi-state model; semi-Markov model; sojourn time.

1 Introduction

In biostatistics, many models for survival and reliability analysis are two-state stochastic processes which lead to a particular event such as death, or the outcome of a particular drug treatment. However, applying a multi-state stochastic process allows the modeler to provide a richer and more accurate model by adding more details. These details allow to capture alternative paths to the event of interest, specify all the intermediate events, and also allow to understand partial failure modes in a progressive disease.

In our context, a multi-state stochastic process is a process $X(t)$ for $t \geq 0$, where $X(t)$ can take a finite number of values $1, 2, \dots, \ell$. This process can be considered as a family of random variables $X(t)$ indexed by t . The quantities of interest are often the probability of being in a state at a given time and the distribution of first passage times (the time until the process reaches a given state for the first time from a particular starting state).

In some applications of multi-state stochastic processes, the dependence on the history of the process is negligible. Therefore, for the sake of mathematical tractability, assuming the Markov property (where future transitions between states depend only upon the current state) is convenient. For instance, continuous time Markov chains, which we refer to here as *Markov processes*, are widely used in modeling the movements of

***Corresponding author: Benoit Lique^{*}**, Department of Mathematics and Statistics, Macquarie University, Université de Pau et des Pays de l'Adour, E2S-UPPA, Pau, France; and ACEMS, Queensland University of Technology, Brisbane, Australia, E-mail: benoit.lique^{*}@univ-pau.fr

Azam Asanjarani, The University of Auckland, Auckland, New Zealand, E-mail: azam.asanjarani@auckland.ac.nz.
<https://orcid.org/0000-0001-6115-073X>

Yoni Nazarathy, The University of Queensland, Brisbane, Australia, E-mail: y.nazarathy@uq.edu.au

patients between units of a hospital or between stages of a disease, see for instance [1, 2]; or in the prediction of the incidence rate of diseases, see [3].

However, in certain cases, the Markov assumption is unrealistic. For instance, in the study of human sleep stages, the sleep stages usually do not follow an exponential distribution (constant hazard rate), but can rather have other forms such as a Weibull distribution, see [4]. Further, some aspects of systems' behavior cannot be captured by Markov processes. For instance, the risk of chronic diseases such as AIDS essentially depends on the time since infection, see [5]. For these cases, applying the class of *semi-Markov processes* (SMP), as an extension of Markov processes, is fruitful. Here, future probability transitions depend on the sojourn time (the time spent in the current state), and the clock of each sojourn time is reset to zero after each transition into a new state. SMPs have a variety of applications in healthcare. For instance, for predicting a disease progression [6]; health care manpower supply prediction [7]; and recovery progress of patients with a specific type of disease [8].

For biomedical applications, especially those concerned with characterizing an individual's progression through various stages of a disease, a three-state semi-Markov process known as the *illness-death model* is very popular (see for instance [5, 9]). The illness-death model may also be applied for modeling the trajectory of patients in intensive care units (ICUs) of hospitals (see [10, 11]).

Here, our main focus is on the statistical methodology of semi-Markov processes and more precisely on parametric models of SMPs. We compare and contrast two approaches for defining SMPs which we denote via, **I – sojourn times**, and **II – intensity transition functions**.

Approach I is based on the specification of the *sojourn time distribution*, together with a transition probability matrix of a discrete time Markov chain, which we call the *embedded chain*. Approach II, is based on *intensity transition functions* which when multiplied by an infinitesimal quantity, specify the instantaneous probability of transition between states. Note that in the literature, these are sometimes also called hazard rate functions of the SMP, however they should not be confused with hazard functions of probability distributions (such as for example the sojourn time distributions).

While from a probabilistic perspective, both Approach I and Approach II are equivalent ways for describing an SMP, from a statistical perspective there are differences. In this paper, we highlight that when it comes to parameter estimation, Approach II has significant advantages over Approach I. Intrinsically this is because Approach II can be expressed by using fewer parameters. Further, we can show that the likelihood function of Approach II can be written as the product of likelihoods of two-state models. This is very helpful for reducing the computational effort for likelihood-based parameter estimation. Nevertheless, depending on the application at hand, using either approach may be useful and in certain cases researchers have chosen to use Approach I because it focuses directly on the underlying distributions. We highlight the associated tradeoffs in the paper.

The remainder of this paper is structured as follows. In Section 2, we introduce semi-Markov processes (SMP) via both Approach I and Approach II. We also present the probabilistic relationships between the approaches. We continue with Section 3 focusing on inference and specify the likelihood function for both approaches. In Section 4, we illustrate inference on two available datasets and present how to implement both approaches (for both datasets) using several contemporary R software packages. For both datasets, we compare results of fully parametric models based on both approaches and highlight the implications of each modeling choice. Section 5 presents some concluding remarks. All the numerical results (tables and figures) from the paper are freely reproducible via R code in a comprehensive detailed vignette available in [12].

2 Semi-Markov processes

A multi-state model is a continuous time stochastic process with values in a discrete set that is often applied for longitudinal medical studies, where the patients may experience several events and their related information is collected over time. The complexity of a model greatly depends on the number of states and also on the

possible transitions. Figure 1 demonstrates a general multi-state model. See [13] as a classical reference for multi-state models.

A specific class of examples of multi-state models is the class of Markov processes where the state evolution jumps between levels and the process adheres to the Markov property. However, in many real-world applications, we need a stochastic process that exhibits dependence between jump times. For instance, in biostatistics, where the trajectory of patients in a hospital is considered, using a Markov process as a multi-state model imposes a stringent limitation on the sojourn time distribution in each state. This is because a key consequence of using Markov processes is the fact that state sojourn times follow the exponential distribution and is “non-ageing” with a constant hazard rate. In contrast, in Semi-Markov processes this assumption can be relaxed. Hence with Markov processes one is often lacking the desired degrees of freedom for modeling the dependence between jump times.

Extending the modeling from Markov processes to the class of SMPs removes the restriction of memoryless (exponential) state sojourn times and at the same time, preserves the usefulness of treating the data as jump processes in continuous time. In fact, for semi-Markov processes we consider a relaxation of the Markov property for sojourn times and only the embedded chain of states is required to follow the Markov property. For this reason, SMPs are applied for modeling a variety of phenomena in different areas such as economics, reliability, and health care, see [14].

To define an SMP, consider a homogeneous Markov chain $\{J_n\}_{n \geq 0}$ on states $\{1, 2, \dots, \ell\}$ where the probability of n th ($n \geq 1$) jump from state i to state j for $i \neq j$ is p_{ij} . That is,

$$p_{ij} = \mathbb{P}(J_n = j | J_{n-1} = i). \quad (1)$$

Based on the directed graph associated with these probabilities (where arc $i \rightarrow j$ exists only if $p_{ij} > 0$) states can be classified as either transient or recurrent. A state i is *absorbing* if $p_{ij} = 0$ for all $j \neq i$. Note that under this definition

$$\sum_{j \neq i} p_{ij} = \begin{cases} 1 & \text{for non-absorbing } i, \\ 0 & \text{for absorbing } i. \end{cases}$$

Denote the increasing sequence of jump times by $T_0 = 0 < T_1 < T_2 < T_3 < \dots$ and define $N(t) = \max\{n : T_n \leq t\}$ for $t \geq 0$. This is a count of the number of transitions up to time t . The stochastic process $X(t) = J_{N(t)}$ is said to be a *semi-Markov process (SMP)* if whenever the process enters state i , the next state is j with probability p_{ij} and given that the next state to be entered is j , the time until the transition from i to j is a random variable with cumulative distribution function $F_{ij}(t)$:

$$F_{ij}(t) = \mathbb{P}(\tau_n \leq t | J_{n-1} = i, J_n = j), \quad t \geq 0, \quad (2)$$

where $\tau_n = T_n - T_{n-1}$, see Chapter 4 of [15]. Hence in general, SMPs are not Markov processes as they do not possess the Markov property. Further, a semi-Markov process allows arbitrarily distributed sojourn times in any state but retains the Markov property for the embedded (discrete time) Markov chain, $\{J_n\}_{n \geq 0}$.

Figure 1: An illustration of a multi-state model.

Figure 2: The illness-death model.

In many applications of SMPs in healthcare, a very popular three state semi-Markov process known as the *illness-death model* is applied, see for instance [8].

Illness-death model: The illness-death model is the most common model in epidemiology (see Figure 2), where it is often applied in studying chronic diseases. In this model, we have three states “Health”, “Illness” and “Death”, denoted by 1, 2 and 3, respectively. There are three kinds of transitions: $1 \rightarrow 2$, $2 \rightarrow 3$ and $1 \rightarrow 3$ and state 3 is absorbing. Since this model is often used to describe severe illnesses, there is no possibility of recovery, and therefore the model is irreversible.¹ In cases where treatments may yield remission, it is more appropriate to construct a model with an additional state “Remission” rather than to consider that there is a possibility of moving back to the “Healthy” state. Here, we consider the general multi-state model with all possible transitions as illustrated in Figure 1 which includes both illness-death model and its reversible version (see Figure 3).

We now arrive at our key point dealing with SMPs. Here we contrast the two approaches which we denote via, **I – sojourn times**, and **II – intensity transition functions**. Approach I was already used to define SMPs above. We spell out further details of Approach I and then continue to introduce Approach II.

2.1 Approach I: sojourn times

As defined above, an SMP, $X(t)$, can be constructed via the sequence $\{(J_n, T_n)\}_{n \geq 0}$, of states and jump times respectively. The underlying parameters of this construction involve the transition probabilities of the embedded chain, p_{ij} presented in (1), as well as the distributions of the sojourn times for each transition $i \rightarrow j$, presented in (2).

Figure 3: The reversible illness-death model.

¹ To avoid confusion, note that this term is also used for the different concept of a “reversible” Markov chain appearing in a different setting.

The transition probabilities of the embedded chain are often organized in a matrix, $P = [p_{ij}]$ (where we set $p_{ii} = 0$). Note that P restricted to the non-absorbing states is a stochastic matrix. Further, assuming the sojourn time distributions are continuous, they are often represented in different forms including the probability density function, the survival function or the hazard rate function. We now present the details.

The *probability density function* of the sojourn time is

$$f_{ij}(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \mathbb{P}(\tau_n \in (t, t + \Delta t) | J_{n-1} = i, J_n = j). \quad (3)$$

The corresponding *survival function* is

$$S_{ij}(t) = \mathbb{P}(\tau_n > t | J_{n-1} = i, J_n = j) = 1 - F_{ij}(t). \quad (4)$$

(Note that $S_{ij}(t)$ is a decreasing function, that is $S_{ij}(0) = 1$ and $\lim_{t \rightarrow +\infty} S_{ij}(t) = 0$). The *hazard function* which is often thought of as the probability that a jump occurs in a specified interval $(t, t + \Delta t)$ given no jump before time t , is

$$\alpha_{ij}(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \mathbb{P}(\tau_n \in (t, t + \Delta t) | J_{n-1} = i, J_n = j, \tau_n > t). \quad (5)$$

Here, note that by definition of conditional probability we have

$$\alpha_{ij}(t) = \frac{f_{ij}(t)}{S_{ij}(t)}, \quad \text{and} \quad f_{ij}(t) = \alpha_{ij}(t) e^{-\int_0^t \alpha_{ij}(u) du}. \quad (6)$$

It is also useful to define the probability of staying in a current state i for at least t time units. We call this the survival function of the waiting/holding time in state i , and denote it via

$$S_i(t) = \mathbb{P}(\tau_n > t | J_n = i) = \sum_{j \neq i} p_{ij} S_{ij}(t). \quad (7)$$

2.2 Approach II, intensity transition functions

The first approach required specification of the parameters using two types of objects, transition probabilities of the embedded Markov chain, and the distribution of sojourn times given a transition $i \rightarrow j$. The second approach which we present now is more succinct in that it only requires one type of object: *intensity transition functions*. These functions are defined via

$$\tilde{\alpha}_{ij}(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \mathbb{P}(\tau_n \in (t, t + \Delta t), J_n = j | J_{n-1} = i, \tau_n > t), \quad (8)$$

and are similar in nature to hazard rates. However, they should not be treated as hazard rate functions of a probability distribution. They rather indicate the instantaneous probability of making a transition from state i to state j after spending t time units in state i since the last transition. However summing up over all target states j , we do obtain a hazard rate of a probability distribution which we denote via $\tilde{\alpha}_i(t) = \sum_{j \neq i} \tilde{\alpha}_{ij}(t)$. This is the hazard rate of the waiting/holding time in state i . We can use this approach to obtain an alternative expression to $S_i(t)$ of (7).

$$S_i(t) = e^{-\int_0^t \tilde{\alpha}_i(u) du} = e^{-\int_0^t \sum_{j \neq i} \tilde{\alpha}_{ij}(u) du}. \quad (9)$$

Using Approach II, more formally, the meaning of the intensity transition functions is

$$\lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(X(t + u + \Delta t) = j | X(t) = i, X(t^-) \neq i, \tau_{N(t)+1} > u)}{\Delta t} = \tilde{\alpha}_{ij}(u), \quad (10)$$

where using the notation defined previously, $\tau_{N(t)+1}$ is the time elapsed since the last transition, measured at time t . The conditional probability in (10) is conditional on the fact that the last transition was at time t into state

i and up to time $t + u$ there have not been any further transitions. Given this condition, it indicates the instantaneous probability of: (i) making a transition at time $t + u$. (ii) Making the transition into state j .

It can be shown that specification of the intensity transition functions $\tilde{\alpha}_{ij}(\cdot)$ completely specify the probability law of an SMP. See [16] for a formal description where one can also use the notation $\mathcal{H}(t^-)$ to define the history of the process just before time t , i.e., up to t^- . This can formally be defined as the sigma algebra in a filtration associated with the stochastic process. Related notation often used in the literature is the (conditional on history) transition probability to be in state j at time $t + u$, after being in state i in time t . This is often denoted by $P_{ij}(t, t + u) = \mathbb{P}(X(t + u) = j | X(t) = i, \mathcal{H}(t^-))$.

We note that Approach II can be viewed as a collection of competing risk models where at the instance of arriving to a new state, a new set of cause-specific hazard rates is defined. See for example [17] where a competing risk model is presented as a multi-state model and the cause-specific hazard rates terminology is used.

2.3 Relations between the two approaches

We now wish to show how for the same SMP, one can use either the parameters of Approach I or the parameters of Approach II and interchange between them. A key relationship is the following

$$\tilde{\alpha}_{ij}(t) = \frac{p_{ij}S_{ij}(t)}{S_i(t)}\alpha_{ij}(t) = p_{ij}\frac{f_{ij}(t)}{S_i(t)}. \quad (11)$$

It is established using the conditional probabilities, defined in (1), (4), (7), and (5). This key relationship also yields an interpretation of the *cumulative incidence function* for the $i \rightarrow j$ transition which we denote via $\text{CIF}_{ij}(\cdot)$. This is a common measure used in the field of competing risks, see for example [18]; that determines the probability of transitioning into j before or at time t . By rearranging and integrating both sides of (11) we obtain the following representation of the cumulative incidence function.

$$\text{CIF}_{ij}(t) = \int_0^t S_i(u)\tilde{\alpha}_{ij}(u)du = \int_0^t f_{ij}(u)p_{ij}du = p_{ij}F_{ij}(t). \quad (12)$$

We can now convert between the parameterizations of both approaches as follows.

Approach I $\rightarrow$ Approach II. Given $\alpha_{ij}(\cdot)$ and p_{ij} , obtain $\tilde{\alpha}_{ij}(\cdot)$ as follows.

$$\tilde{\alpha}_{ij}(t) = p_{ij} \frac{e^{-\int_0^t \alpha_{ij}(u)du}}{\sum_{k \neq i} p_{ik} e^{-\int_0^t \alpha_{ik}(u)du}} \alpha_{ij}(t). \quad (13)$$

This follows directly from (11).

Approach II $\rightarrow$ Approach I. Given $\tilde{\alpha}_{ij}(\cdot)$ obtain p_{ij} and $\alpha_{ij}(\cdot)$:

First we have,

$$p_{ij} = \int_0^\infty \tilde{\alpha}_{ij}(t) e^{-\int_0^t \sum_{k \neq i} \tilde{\alpha}_{ik}(u)du} dt. \quad (14)$$

This can also be obtained from (12) by taking $t \rightarrow \infty$. Once p_{ij} values are at hand we again use (11) and isolate $f_{ij}(t)$ to obtain,

$$f_{ij}(t) = \tilde{\alpha}_{ij}(t) \frac{e^{-\int_0^t \sum_{k \neq i} \tilde{\alpha}_{ik}(u)du}}{p_{ij}}, \quad (15)$$

from which we can obtain $\alpha_{ij}(t)$ using (6) in the standard manner.

2.4 Examples

We now present three examples illustrating the relationship between the two approaches.

Example 1: Continuous time Markov chains (CTMC): A (finite state) CTMC is a multi-state stochastic process on states $\{1, \dots, \ell\}$ that is time homogenous and satisfies the Markov property. That is, $\mathbb{P}(X(s+t) = j | X(s) = i, \mathcal{H}(s^-)) = \mathbb{P}(X(t) = j | X(0) = i)$. For simplicity of notation we restrict attention to CTMCs without absorbing states. Such chains can be parameterized in several ways, two of which are reminiscent of Approach I, and Approach II above:

- CTMC Approach I: Define a $\ell \times \ell$ stochastic transition probability matrix P (non-negative entries and rows sum to 1) with 0 entries in the diagonal. Denote the entries via P_{ij} . This is often called the transition probability matrix of the embedded discrete time Markov chain. Then define a vector of rates of length ℓ , λ , where λ_i^{-1} denotes the mean holding time of state i . Then from the theory of CTMCs, see for example [16]; the process evolves as follows. If at time t the process is in state i , an exponential random variable with parameter λ_i is generated to determine a holding duration. After that duration passes a transition from state i to state j occurs with probability P_{ij} . The durations and the transition choice are independent. The process then repeats. It is now evident that such a description of a CTMC is a special case of **Approach I** for SMPs, where

$$\alpha_{ij}(t) = \lambda_i \quad \text{and} \quad p_{ij} = P_{ij}. \quad (16)$$

Observe that $\alpha_{ij}(t)$ is independent of t and independent of the target state j .

- CTMC Approach II: Define a $\ell \times \ell$ generator matrix Q (non-negative entries on the off-diagonal, negative entries on the diagonal, and rows sum to 0). Denote the entries via q_{ij} . Then treat the process as an SMP using **Approach II** with

$$\tilde{\alpha}_{ij}(t) = q_{ij}, \quad (17)$$

Observe that $\alpha_{ij}(t)$ is independent of t .

It is well known from the theory of CTMCs the two parameterizations are equivalent.

CTMC Approach I $\rightarrow$ CTMC Approach II:

$$q_{ij} = \begin{cases} \lambda_i P_{ij} & i \neq j, \\ -\lambda_i & i = j, \end{cases}$$

CTMC Approach II $\rightarrow$ CTMC Approach I:

$$P_{ij} = \frac{q_{ij}}{\sum_{k \neq i} q_{ik}}, \quad \text{and} \quad \lambda_i = \sum_{k \neq i} q_{ik} = -q_{ii}.$$

We now verify that these transformations directly agree with the relationship between Approach I of the SMP and Approach II of the SMP. Using (13) and (16) we have,

$$\tilde{\alpha}_{ij}(t) = P_{ij} \frac{e^{-\int_0^t \lambda_i du}}{\sum_{k \neq i} P_{ik} e^{-\int_0^t \lambda_i du}} \lambda_i = P_{ij} \lambda_i = q_{ij}. \quad (18)$$

Going in the other direction using (14) and remembering that since Q is a generator matrix, $q_{ii} = -\sum_{k \neq i} q_{ik} < 0$, we have,

$$p_{ij} = \int_0^\infty q_{ij} e^{-\int_0^t \sum_{k \neq i} q_{ik} du} dt = q_{ij} \int_0^\infty e^{q_{ii}t} dt = \frac{q_{ij}}{-q_{ii}}.$$

Now using (15) we have,

$$f_{ij}(t) = -q_{ij} \frac{e^{\int_0^t q_{ii} du}}{q_{ij}/q_{ii}} = -q_{ii} e^{q_{ii}t} = \lambda_i e^{-\lambda_i t}.$$

This is an exponential density and hence it has a constant hazard rate $\alpha_{ij}(t) = \lambda_i$.

Example 2: Exponential Sojourn Times: The example above illustrates that an SMP parameterized via Approach I where $\alpha_{ij}(t)$ is constant in time and also independent of j yields constant (Approach II) transition intensity functions, $\tilde{\alpha}_{ij}(t)$. That is, constant (Approach I) hazard rates yield constant transition intensity functions, but under the condition that $\alpha_{ij}(t)$ is the same for all target states j .

We now show that this condition is necessary for having constant transition intensity functions. To see this, use (13) with $\alpha_{ij}(t) = \lambda_{ij}$ where for some j_1 and j_2 , $\lambda_{ij_1} \neq \lambda_{ij_2}$. Now,

$$\tilde{\alpha}_{ij}(t) = p_{ij} \frac{e^{-\lambda_{ij}t}}{\sum_{k \neq i} p_{ik} e^{-\lambda_{ik}t}} \lambda_{ij}. \quad (19)$$

Since $\lambda_{ij_1} \neq \lambda_{ij_2}$ we are not able to cancel out the exponent as was done in (18).

Interestingly, the converse does not hold. The previous example showed that time independent transition intensity functions always yield a CTMC and hence constant (Approach I) hazard rates.

Example 3: Weibull Distributions: The Weibull continuous univariate distribution of a non-negative random variable is parameterized by a shape parameter $\eta > 0$ and a scale parameter $\mu > 0$. It has density, survival function, and hazard rate functions respectively given via,

$$f(t) = \frac{\eta}{\mu} \left(\frac{t}{\mu} \right)^{\eta-1} e^{-(t/\mu)^\eta}, \quad S(t) = e^{-(t/\mu)^\eta}, \quad \alpha(t) = \frac{\eta}{\mu} \left(\frac{t}{\mu} \right)^{\eta-1}.$$

It is appealing in survival analysis and reliability analysis because if $\eta < 1$ the hazard rate is monotonically decreasing; if $\eta = 1$ the hazard rate is constant (an exponential distribution) and if $\eta > 1$ the hazard rate is monotonically increasing.

Say now that we are using Approach I and parameterize all of the sojourn time distributions as Weibull distributions, where for transition $i \rightarrow j$ we have respective scale and shape parameters μ_{ij} and η_{ij} . Now, if we were to consider the Approach II representation, then from (13) the transition intensity functions are

$$\tilde{\alpha}_{ij}(t) = p_{ij} \frac{e^{-(t/\mu_{ij})^{\eta_{ij}}}}{\sum_{k \neq i} p_{ik} e^{-(t/\mu_{ik})^{\eta_{ik}}}} \frac{\eta_{ij}}{\mu_{ij}} \left(\frac{t}{\mu_{ij}} \right)^{\eta_{ij}-1} = K_{ij}(t) t^{\eta_{ij}-1}, \quad (20)$$

where we use the notation $K_{ij}(t)$ to represent all of the components of $\tilde{\alpha}_{ij}(t)$ excluding $t^{\eta_{ij}-1}$. Now the form of $K_{ij}(t)$ can determine if the Approach II representation is of a Weibull type or not. A Weibull type representation will follow if $K_{ij}(t)$ is independent of t , otherwise it is not.

We now see that if for a fixed i and every j_1 and j_2 with $j_1 \neq j_2$ we have that

$$\eta_{ij_1} = \eta_{ij_2}, \quad \text{and} \quad \mu_{ij_1} = \mu_{ij_2}, \quad (21)$$

then, $K_{ij}(t)$ is independent of t and we obtain a Weibull type Approach II representation with,

$$\tilde{\alpha}_{ij}(t) = p_{ij} \frac{\eta_{ij}}{\mu_{ij}} \left(\frac{t}{\mu_{ij}} \right)^{\eta_{ij}-1} = \frac{\eta_{ij}}{p_{ij}^{-1/\eta_{ij}} \mu_{ij}} \left(\frac{t}{p_{ij}^{-1/\eta_{ij}} \mu_{ij}} \right)^{\eta_{ij}-1}, \quad (22)$$

and hence the scale parameter is modified to $p_{ij}^{-1/\eta_{ij}} \mu_{ij}$ and the shape parameter keeps the same form. Otherwise (if there exists j_1 and j_2 such that (21) does not hold), then $\tilde{\alpha}_{ij}(t)$ cannot be of the Weibull type as in (22).

Going the other way, assume we are using Approach II with Weibull type transition intensity functions,

$$\tilde{\alpha}_{ij}(t) = \frac{\eta_{ij}}{\mu_{ij}} \left(\frac{t}{\mu_{ij}} \right)^{\eta_{ij}-1}.$$

Then by integrating (14) we can in principle obtain p_{ij} . It turns out that if for a given state i , the shape parameters of all the transition intensity functions is the same (denote it via η_i) then the integration yields an explicit form,

$$p_{ij} = \frac{\mu_{ij}^{\eta_i}}{\sum_{k \neq i} \mu_{ik}^{\eta_i}}, \quad (23)$$

otherwise, there is not an explicit solution. In such a case where $\eta_{ij} = \eta_i$ for all j , we also have from Equation (15),

$$f_{ij}(t) = \frac{\eta_i}{\mu_{ij}} \left(\frac{t}{\mu_{ij}} \right)^{\eta_i-1} e^{-\sum_{k \neq i} \left(\frac{t}{\mu_{ik}} \right)^{\eta_i}} \frac{\mu_{ij}^{\eta_i}}{\sum_{k \neq i} \mu_{ik}^{\eta_i}}. \quad (24)$$

Further, if $\mu_{ij} = \mu_i$ for all j then p_{ij} of (23) reduces to ℓ_i^{-1} where $\ell_i < \ell$ is the number of possible target states from state i . Note that this form is not specific to the Weibull type intensity transition function, but will hold whenever the intensity transition functions out of state i are the same.

However, with the Weibull case we can continue further and (24) can be represented as,

$$f_{ij}(t) = \ell_i \frac{\eta_i}{\mu_i} \left(\frac{t}{\mu_i} \right)^{\eta_i-1} e^{-\ell_i \left(\frac{t}{\mu_i} \right)^{\eta_i}} = \frac{\eta_i}{\ell_i^{-1/\eta_i} \mu_i} \left(\frac{t}{\ell_i^{-1/\eta_i} \mu_i} \right)^{\eta_i-1} e^{-\left(\frac{t}{\ell_i^{-1/\eta_i} \mu_i} \right)^{\eta_i}},$$

which denotes a Weibull distribution with same shape parameter as the transition intensity function and a scale parameter equal to $\ell_i^{-1/\eta_i} \mu_i$.

3 Likelihood and inference

Here we focus on inference for the fully-parametric semi-Markov model defined by the two above mentioned approaches. Survival analysis usually applies for cohort or clinical studies, hence the data is usually gathered from the same subjects repeatedly during a time interval $[0, T]$. A key issue in survival analysis and event history analysis is the occurrence of incomplete or sparse observations. For instance, in the case of chronic diseases, when the event of study is death, time of occurrence of this event is not observed for the subjects still alive at time T . This type of incomplete observation is called *right-censoring*. There are other kinds of incomplete data like left-censoring and interval censoring, see [19, 20]. Andersen and Keiding [17] present different incomplete data forms which can be handled in multi-state frameworks. Our focus in this paper is right censored data corresponding to when at the end of the observation period, not all individuals under study have reached an absorbing state.

As our focus is on the fully-parametric case we now present the likelihood functions for both Approach I and Approach II. We assume there are n subjects denoted via $h = 1, \dots, n$, for which data is collected and we assume independence between subjects. For each subject, we distinguish between two cases depending on if the subject is in an absorbing state at time T or not. Note that in principle we can set T to be subject specific, but for simplicity we do not do so here. This is recorded via δ_h where $\delta^{(h)} = 1$ implies no right censoring (subject is in an absorbing state at time T), and $\delta^{(h)} = 0$ implies right censoring. Further, we record the state evolution denoted via the sequence $\{j\}$ and the sojourn times denoted via the sequence $\{\tau\}$. For simplicity, we assume that all subjects start at the same fixed and known state, denoted via J_0 . This can be easily generalized.

For subject h during $[0, T]$, we use $N^{(h)}$ to denote the number of state transitions up to time T . The data of the subject is represented via

$$\mathcal{H}^{(h)} = (J_0, J_1^{(h)}, \dots, J_{N^{(h)}}^{(h)}, \tau_1^{(h)}, \dots, \tau_{N^{(h)}}^{(h)}, \delta^{(h)}).$$

Note that if $\delta^{(h)} = 0$ and there is right censoring then we are also interested in the time duration $U^{(h)}$ after the last state is visited. This quantity is represented via,

$$U^{(h)} = T - \sum_{i=1}^{N^{(h)}} \tau_i^{(h)}. \quad (25)$$

As our subjects are independent we can consider the likelihood contribution of each subject h in isolation, and after denoting it via $\mathcal{L}^{(h)}$, the full likelihood is

$$\mathcal{L} = \prod_{h=1}^n \mathcal{L}^{(h)}. \quad (26)$$

The specific form of $\mathcal{L}^{(h)}$ now depends on if we are using Approach I or Approach II and as we show below, in Approach II, it can further be decomposed as in (30).

3.1 Likelihood for Approach I

The form of $\mathcal{L}^{(h)}$ based on $\mathcal{H}^{(h)}$ is,

$$\mathcal{L}^{(h)} = \left(\prod_{k=1}^{N^{(h)}} p_{J_{k-1}J_k} f_{J_{k-1}J_k}(\tau_k) \right) \left(S_{J_{N^{(h)}}}(U^{(h)}) \right)^{1-\delta^{(h)}}, \quad (27)$$

where for brevity we omit the (h) superscripts for the state and sojourn time sequences $\{J\}$ and $\{\tau\}$. To further understand (27) consider a recursive construction where each transition without censoring based on $J_{k-1} \rightarrow J_k$ with a sojourn time of τ_k has a likelihood contribution $p_{J_{k-1}J_k} f_{J_{k-1}J_k}(\tau_k)$. Further, in case of censoring the last censored transition has likelihood contribution $S_{J_{N^{(h)}}}(U^{(h)})$ based on the survival function of the holding time (7), corresponding to the last observed state $J_{N^{(h)}}$.

3.2 Likelihood for Approach II

We refer to the key relationship (11) and replace $p_{ij} f_{ij}(u)$ in (27) with $\tilde{\alpha}_{ij}(u) S_i(u)$ to obtain,

$$\mathcal{L}^{(h)} = \left(\prod_{k=1}^N \tilde{\alpha}_{J_{k-1}J_k}(\tau_k) S_{J_{k-1}}(\tau_k) \right) \left(S_{J_N}(U) \right)^{1-\delta}, \quad (28)$$

where for brevity we omit all superscripts (h) . Further, we manipulate the likelihood contribution of each subject as follows:

$$\begin{aligned} \mathcal{L}^{(h)} &= \left(\prod_{k=1}^N \tilde{\alpha}_{J_{k-1}J_k}(\tau_k) e^{-\int_0^{\tau_k} \tilde{\alpha}_{J_{k-1}}(u) du} \right) \left(e^{-\int_0^U \tilde{\alpha}_{J_N}(u) du} \right)^{1-\delta} \\ &= \prod_{k=1}^{N+1} (\tilde{\alpha}_{J_{k-1}J_k}(\tau_k))^{\mathbb{I}_{\{k \neq N+1\}}} e^{-\int_0^{\tau_k} \tilde{\alpha}_{J_{k-1}}(u) du}. \end{aligned} \quad (29)$$

In the first line we use the representation in (9). For the second line, by increasing the summation from N to $N+1$ we are able to remove δ by defining $\tau_{N+1} = U$, $J_{N+1} \equiv -1$, and using the fact that $\tilde{\alpha}_{ij}(u) \equiv 0$ in cases where i is an absorbing state and $0^0 \equiv 1$.

Now, we are able to separate $\mathcal{L}^{(h)}$ into the form

$$\mathcal{L}^{(h)} = \prod_{i=1}^{\ell} \prod_{j=1}^{\ell} \mathcal{L}_{ij}^{(h)}, \quad (30)$$

following similar ideas as those initially presented in [21]. To achieve such a separation, it is useful to define the indicators

$$\delta_{i \rightarrow j}^{k-1} = \mathbb{1}_{\{J_{k-1}=i, J_k=j\}}, \quad \text{and} \quad \delta_{i \neq j}^{k-1} = \mathbb{1}_{\{J_{k-1}=i, J_k \neq j\}}.$$

We may now manipulate (29) to obtain

$$\mathcal{L}_{ij}^{(h)} = \prod_{k=1}^{N+1} \left(\tilde{\alpha}_{ij}(\tau_k) e^{-\int_0^{\tau_k} \tilde{a}_{ij}(u) du} \right)^{\delta_{i \rightarrow j}^{k-1}} \left(e^{-\int_0^{\tau_k} \tilde{a}_{ij}(u) du} \right)^{\delta_{i \neq j}^{k-1}}. \quad (31)$$

The form of (26), (30), and (31) allows us to treat each transition separately as if each transition intensity transition function $\tilde{\alpha}_{ij}(\cdot)$ has its own set of parameters.

3.3 Parametric forms and covariate information

In carrying out inference, we assume a parametric form for $\alpha_{ij}(\cdot)$ in Approach I or $\tilde{\alpha}_{ij}(\cdot)$ in Approach II. We also allow for covariate information where the natural way to introduce covariates in a multi-state model is a *Cox-like proportional hazard model* which can be defined by using either the hazard function of the sojourn times or by intensity transitions, see [22]. Hence for a vector of covariates Z , we have

$$\alpha_{ij}(t|Z) = \alpha_{ij,0}^{(\theta_{ij})}(t) e^{\beta_{ij}^T Z}, \quad (32)$$

for Approach I. Further we have

$$\tilde{\alpha}_{ij}(t|Z) = \tilde{\alpha}_{ij,0}^{(\tilde{\theta}_{ij})}(t) e^{\tilde{\beta}_{ij}^T Z}, \quad (33)$$

for Approach II.

Here $\alpha_{ij,0}^{(\theta_{ij})}(\cdot)$ and $\tilde{\alpha}_{ij,0}^{(\tilde{\theta}_{ij})}(\cdot)$ are the baseline functions for the hazard rate and transition intensity functions respectively. They each follow a parametric form, determined via θ_{ij} and $\tilde{\theta}_{ij}$ respectively. Further, β_{ij} and $\tilde{\beta}_{ij}$ are the regression parameters for transition $i \rightarrow j$ associated with the covariates of the subject. Observe that there are major differences in the interpretation of the regression coefficients β_{ij} of Approach I and $\tilde{\beta}_{ij}$ of Approach II. The former determine the hazard ratios dealing with sojourn times and bear no effect on the actual transitions of the SMP. The latter, determine the hazard ratios dealing with risks and also affect the instantaneous rate of transitioning.

When estimating the parameters for such a model, with Approach I, we also need to estimate p_{ij} , whereas with Approach II this is not needed as p_{ij} is implicitly determined via (14). Interestingly, via Approach II, we have that the covariates implicitly affect p_{ij} , whereas with Approach I, we may wish to set $p_{ij}(Z)$ using multinomial regression, however it isn't common practice.

3.4 Contrasting inference for Approach I and Approach II

The key difference in the likelihood expressions between Approach I (27) and Approach II (30) and (31) is in computational tractability. As noted by [23] the optimization of the likelihood for Approach I could be challenging when the number of covariates and parameters is large due to a complex form of the sojourn time distribution. This is especially the case when the number of variables is large compared to the size of the data. It

should further be noted that modeling using Approach I requires the embedded chain parameters in addition to the parameters involved in (32). This means that for a model with ℓ states, there may be up to $\ell^2 - \ell$ more parameters to estimate when using Approach I, in comparison to Approach II, where all the parameters appear in (33).

With Approach II, the decoupling in (30) allows to optimize the likelihood for each transition $i \rightarrow j$ separately if each hazard intensity transition function (33) has its own set of parameters $\tilde{\theta}_{ij}$ and $\tilde{\beta}_{ij}$. In such a framework, the full likelihood of the SMP can be maximized separately by maximizing the likelihood of the different separate two-state models, see [22]. Also, the inference of the SMP can be easily handled by using survival models like the Cox model as the separate likelihood arising in survival analysis.

Hence Approach II enables fitting SMP models using survival analysis methods and software, see for example [24–26]. This also applies to the case where extensions to the proportional hazard assumption are needed. The forms in (32) and (33) are clearly based on a proportional hazard assumption with fixed covariates where there is only a multiplicative effect on the baseline hazard functions independent of time. An extension is to consider time-dependent covariates $Z(t)$ as in [17, 27]. Further, in both proportional models the effect of variables is assumed to have a linear (or log-linear) functional form. A more flexible model is to consider non-linear effects using smooth functions of the covariates such as the generalized additive model (GAM) presented by [28]. With such extensions, using Approach II is much more straightforward than using Approach I due to the decoupling and because there is ample statistical software available for survival analysis.

One may also wish to incorporate random effects in the models using frailty models. In this case, both (32) and (33) could also include random effect to take into account the correlation between observations/subjects. This has been handled in [10] for Approach II by exploiting the decoupling of the likelihood. Such random effects may also be incorporated in Approach I, however the computational effort will be greater.

4 Semi-Markov application in practice

In this section, we illustrate the inference of the semi-Markov model on two real datasets. Our purpose with these examples is to help practitioners see how the mathematical details presented above interact with statistical software. In light of this, we chose datasets that have been analyzed previously with a one sided view (either Approach I or Approach II but not both). These datasets are readily available in the R ecosystem. All the numerical results including tables and figures, are freely reproducible through R codes in a comprehensible detail vignette available in [12]. For both datasets, we compare results of full parametric models based on Approach I where we estimate sojourn time distributions and transition probabilities of the embedded chain, and based on Approach II where we estimate transition intensity functions.

To carry out estimation using Approach I we use the *SemiMarkov* package, see [29]. This package implements inference using the likelihood (26) with (27) while allowing to use different parametric forms for the sojourn time distribution including exponential, Weibull, and exponentiated Weibull, see [30]. It also supports covariates with standard inference for the covariate coefficients β_{ij} , including the Wald test. It allows the possibility to include different covariates for each transition. While the focus of the inference with this package is Approach I, as output, the package can automatically provide the intensity transition functions $\tilde{\alpha}_{ij}(\cdot)$, referred to as the “hazard function of the SMP”.

For Approach II, the decoupling in (30) with (31) allows us to use any package that implements inference for survival analysis since for each transition $i \rightarrow j$ we can maximize,

$$\mathcal{L}_{ij} = \prod_{h=1}^n \mathcal{L}_{ij}^{(h)}$$

separately if each hazard intensity transition function (33) is assumed to have its own set of parameters $\tilde{\theta}_{ij}$ and $\tilde{\beta}_{ij}$. For this we can use different R packages including *flexsurv* [26], *survival* [31], and *eha* [32]. In the examples here, we mainly use *flexsurv* which offers an easy way to run different parametric forms for $\tilde{\alpha}_{ij}(\cdot)$

including exponential, Weibull, gamma, and generalized gamma (see [33]) among many other forms. It also supports covariates with standard inference for the covariate coefficients $\tilde{\beta}_{ij}$.

We note that the purpose of this section is not to be an extensive survey of all of the R packages that can be used for inference in multi-state models. For a comprehensive software survey, see the survival task view from R [34]. Our focus here is to illustrate how the aforementioned packages can be used to easily carry out inference, while we also illustrate a few key points.

4.1 Progressive three-state model for the stanford heart transplant data

As an illustrative example, we revisit the analysis of the Stanford Heart Transplant data, freely available through the survival package. A full description of the data can be found in [35]. The dataset presents the survival of patients on the waiting list for the Stanford heart transplant program. We analyze the data similarly to [36] which use their R package `p3state` and propose an illness-death model similar to Figure 2. Their states represent “alive without transplant”, “alive with transplant”, and “dead”.

We consider the same dataset, where patients’ records from October 1967 (start of the program) until April 1974 includes the information of 103 patients where 69 received heart transplants, and 75 died. Three fixed covariates for each patient are available. These include the age at acceptance (**age**), the year of acceptance (**year**), and previous surgery (**surgery**: coded as 1 = yes; 0 = no).

We first use the data without taking into account any covariate effects to show that an exponential distribution on the baseline hazard rate of the sojourn time, $\alpha_{ij}(\cdot)$ does not produce constant intensity transition functions, $\tilde{\alpha}_{ij}(\cdot)$. This point was shown theoretically in Example 2, of Section 2.4 and we now illustrate it using data and the `SemiMarkov` package. Using the package we obtain both $\alpha_{ij}(\cdot)$ and $\tilde{\alpha}_{ij}(\cdot)$ for $(i, j) = (1, 2)$ and $(i, j) = (1, 3)$. The results are in Figure 4 where we see that while $\alpha_{ij}(\cdot)$ are constant, $\tilde{\alpha}_{ij}(\cdot)$ are clearly not constant across time. These curves in fact follow a form like (19).

We now continue to analyze the data fully with Approach I using the `SemiMarkov` package where we compare different parametric forms and covariates in the model. Table 1 presents the inference results of 6 different models investigated depending on the distribution and covariates chosen for each transition. For each model we present the $\hat{\beta}_{ij}$ estimates and their estimated standard error. We also point out that the `SemiMarkov` package yields p -values for the associated Wald tests. We also present the estimates of the transition probabilities of the embedded chain, $\hat{p}_{ij}$. It is evident that these are outputs from likelihood estimation as they don’t fully agree across models (they are not proportion estimates). For the final two models, *Weibull + Select* and *Weibull/Exp + Select* are based on model selection of the significant covariates with more details in the vignette, see [12]. Observe that the final model, *Weibull/Exp + Select*, incorporates two different distributions.

Figure 4: Baseline hazard rate of the sojourn times (Approach-I: $\alpha_{ij}(\cdot)$) and baseline hazard rate of the semi-Markov process (Approach-II: $\tilde{\alpha}_{ij}(\cdot)$) for transition $1 \rightarrow 2$ (top plot) and $1 \rightarrow 3$ (bottom plot).

Table 1: Estimation of the different semi-Markov process modelled by specifying proportional hazard models on the sojourn times-Stanford Heart Transplant data-Approach I.

Model	Transition	Distribution	Surgery		Age		Year		$\hat{p}_{ij}$
			$\hat{\beta}$	$\hat{\sigma}_{\beta}$	$\hat{\beta}$	$\hat{\sigma}_{\beta}$	$\hat{\beta}$	$\hat{\sigma}_{\beta}$	
Exp baseline	1→2	Exp	–	–	–	–	–	–	0.70
	1→3	Exp	–	–	–	–	–	–	0.30
	2→3	Exp	–	–	–	–	–	–	1.00
Exponential	1→2	Exp	–0.18	0.31	–0.01	0.02	–0.18	0.07	0.68
	1→3	Exp	1.70	0.61	0.11	0.02	–0.19	0.11	0.32
	2→3	Exp	–1.40	0.45	0.06	0.02	0.43	0.05	1.00
Weibull	1→2	Weibull	–0.18	0.31	–0.01	0.02	–0.18	0.07	0.68
	1→3	Weibull	1.20	0.63	0.07	0.02	–0.14	0.11	0.32
	2→3	Weibull	–1.10	0.46	0.04	0.02	0.27	0.05	1.00
EWeibull	1→2	EWeibull	–0.21	0.31	–0.01	0.02	–0.16	0.08	0.68
	1→3	EWeibull	1.10	0.62	0.05	0.02	–0.13	0.11	0.32
	2→3	EWeibull	–1.20	0.46	0.02	0.02	0.26	0.10	1.00
Weibull + Select	1→2	Weibull	–	–	–	–	–0.17	0.07	0.68
	1→3	Weibull	1.10	0.63	0.07	0.02	–	–	0.32
	2→3	Weibull	–1.10	0.47	0.05	0.02	0.31	0.06	1.00
Weibull/Exp + Select	1→2	Exp	–	–	–	–	–0.18	0.07	0.68
	1→3	Weibull	1.10	0.63	0.07	0.02	–	–	0.32
	2→3	Weibull	–1.10	0.47	0.05	0.02	0.31	0.06	1.00

For choosing between the different proposed models, we use the *expected Kullbak-Leibler* (EKL) risk as in [10]. This is done via the *Akaike information criterion* (AIC) which is an estimate of the EKL risk in a parametric framework using maximum likelihood method, see [37, 38]. For each of the 6 models described above, we obtain the AIC using output from *SemiMarkov* and present it in the first part of Table 2.

The best model according to the AIC is given by the *Exponentiated Weibull* model. However, due to convergence issues during the optimization (see the vignette [12]), we select the second best result given by the *Weibull/Exp + Select* model as the most appropriate model for this dataset according to AIC using Approach I.

Moving onto inference using Approach II, we investigate and estimate different models by specifying different proportional intensity transition models as defined in Equation (33). We use *flexsurv* to estimate $\hat{\alpha}_{ij}(\cdot)$ for several models which we present in the second part of Table 2. The models *Exponential*, *Weibull*, *Gamma*, and *Generalized Gamma* are all estimated with all three covariates in the model. The additional two

Table 2: Akaike information criterion (AIC) and number of parameters for the models based on the sojourn times (Approach I) and the intensity transition functions (Approach II).

Method	Model	AIC	Number of parameters
Approach I	Exponential baseline	1730.07	4
	Exponential	1653.04	13
	Weibull	1645.55	16
	EWeibull ^a	1632.22	18
	Weibull + Select	1641.67	13
	Weibull/Exp + Select	1639.73	12
Approach II	Exponential	1751.03	12
	Weibull	1716.51	15
	Gamma	1723.97	15
	Generalized Gamma	1708.88	18
	Weibull + Select	1709.44	10
	Generalized Gamma + Select	1701.03	13

^aconvergence issue reported during the estimation.

models, *Weibull + Select*, and *Generalized Gamma + Select*, have a reduced set of covariates with a procedure consistent with that used for Approach I above. More details are in the vignette (see [12]). The AIC based performance of each of these models is presented in the second part of Table 2.

The best model according to the AIC is *Generalized Gamma + select*. In this model only the **age** covariate is used for transition $1 \rightarrow 2$, only the **year** covariate is used for transition $1 \rightarrow 3$, and both the **surgery** and **age** covariates are used for transition $2 \rightarrow 3$. For this dataset, the best model based on Approach I presents a better AIC value (1639.73) in comparison to the best model based on Approach II (1701.03).

As discussed in Section 3.3 the regression coefficients of Approach I and those of Approach II do not have the same interpretation. Indeed for the hazard rate of the sojourn time, the regression coefficients can be interpreted in terms of relative risk on the waiting time (i.e., given an $i \rightarrow j$ transition). For example the positive coefficient for ‘surgery’ for the $1 \rightarrow 3$ transition, as appearing in Table 1, can be interpreted as an indication that death without getting a heart transplant is less favorable for those that had previous surgery. While the hazard rate of the SMP based on transition intensity functions can be interpreted as the subject’s risk of passing from state i to state j . Interpretation of the regression coefficients using both Approach I and Approach II is further discussed in the second example below.

Since a key object of a Semi-Markov process is the transition intensity function, generally called the hazard rate of the SMP, it is also useful to visualize estimates of these functions based on both Approach I and Approach II. Further insight may be gained by plotting these against non-parametric estimators. For this, we used the Breslow estimator [39]; which yields an estimate for the baseline cumulative risk function of a cox-regression type model of each transition. Figure 5 plots the cumulative baseline (all covariates are set to 0) estimated transition intensity functions, estimated via Approach I and Approach II, against the Breslow estimator. These cumulative functions are based on the parameter estimates for the best model of Approach I (*Weibull/Exp + Select*) vs. the best model of Approach II (*Generalized Gamma + Select*). In this case, the plots in Figure 5 hint at a better fit (closer to the non-parametric estimation) of the SMP using Approach I in accordance with the results of the AIC values discussed above.

4.2 Reversible semi-Markov model for the asthma control data

We revisit the analysis of the asthma control data which has been used to illustrate the SemiMarkov R package in [29]. See also the related papers using the same data, [40] and [41]. The data consists of the follow-up of severe asthmatic patients consulting at varied times according to their perceived needs. At each visit, several covariates were recorded and asthma was evaluated. The available data (asthma) from the SemiMarkov package presents a random selection of the 371 patients with at least two visits. In this illustration, we only use the body mass index (BMI) covariate coded: 1 if BMI ≥ 25 and 0 otherwise. Similarly to the analysis in [29]; we use a reversible three-state model as presented in Figure 6. This investigates the evolution of asthma based on, “optimal control” (State 1), “sub-optimal control” (State 2) and “unacceptable control” (State 3). More details about concepts of control scores can be found in [41]. Observe that in this model, none of the states are absorbing and hence the model is said to be *reversible*.

Figure 5: Cumulative of the baseline hazard rate of the SMP (intensity transition function) estimated from best AIC model of Approach-I and Approach-II. A semi-parametric estimation is presented as benchmark.

Figure 6: The reversible three-state model.

We note that this data is based on scheduled visits every three months or according to patients' perceived needs [40], and is thus potentially interval censored since the exact times of state transitions are approximated to be at visit times. This aspect of the analysis was not the focus of [29] and the related papers. While dealing with such interval censoring is important we do not focus on it here further and refer the reader to [42, 43].

For illustration purposes, we concentrate on a Weibull model as proposed in [29]. We perform a first full parametric proportional Weibull model for the sojourn times (Approach I) with the BMI covariate for each transition. This yields significant results of BMI covariate for the transitions $1 \rightarrow 3$ and $3 \rightarrow 1$. Then, we decide to run a sparse model including BMI only for these transitions to facilitate the convergence of the model which can be difficult for a model with too many parameters as reported in [29]. For this new model, BMI regression coefficients remain significant for transitions $1 \rightarrow 3$ and $3 \rightarrow 1$ with $\hat{\beta}_{13} = -0.88$ and $\hat{\beta}_{31} = -0.448$, and respective p -values 0.012, and 0.023. The fact that hazard ratio of the sojourn time associated with these covariates is less than unity (estimated coefficients are negative), indicates that BMI ≥ 25 generally lengthens the duration of the sojourn time in state 1 when making a $1 \rightarrow 3$ transition and generally lengthens the duration of the sojourn time in state 3 when making a $3 \rightarrow 1$ transition. This can also be interpreted as a decrease of the risk of leaving "optimal control" state to "unacceptable control" as well as a decrease of the risk of leaving "unacceptable control" state to "optimal control". However, the magnitude of the estimated coefficients cannot be used to evaluate the change in the hazard ratio on the risk (recall Equations (32) and (33) and the differences between β_{ij} and $\tilde{\beta}_{ij}$). We may also visualize the effect of BMI covariate on the associated risks by plotting the hazard rate of the semi-Markov model (intensity transition functions), deduced from the sojourn times. See Figure 7 where we also plot estimated transition intensity functions using Approach II estimation, which we describe now.

Approach II presented in this paper offer a direct way to model the intensity transition functions and has the advantage of a split likelihood which facilitates an efficient optimization procedure. In the vignette available in [12]; we present the results of the full Weibull proportional semi-Markov model using this Approach. For illustration, we do this in two ways: (i) Optimizing the full likelihood. (ii) Splitting the likelihood for each transition and optimizing individually. As expected, both ways yield the same likelihood estimates. This highlight the advantage of the intensity transition based approach for overcoming potential computational issues in the optimization. Moreover, in some applications, some transitions of interest could then be further explored with any sophisticated survival model of choice as the estimation for each transition is performed separately. See for example [10, 24, 26, 44].

We augment the plots in Figure 7 with the estimated transition intensity functions using Approach II. Note that similarly to Approach I, the BMI effects are in the same direction. However, the interpretation of the regression coefficients is different. For the intensity transition based estimation (Approach II), the BMI regression coefficient estimates are $\hat{\beta}_{13} = -0.5$ and $\hat{\beta}_{31} = -0.67$. Here, in contrast to the Approach I estimates, the exact magnitude of the coefficients may be interpreted as the change in the hazard ratios associated with BMI ≥ 25 .

Figure 7: Hazard rate of the semi-Markov process for each transition from the sojourn time and intensity transition approaches.

5 Concluding remarks

We have surveyed the two main approaches for modeling and inference using semi-Markov processes in a parametric setting. These are Approach I based on sojourn time distributions and the embedded Markov chain, and Approach II based on transition intensity functions. Each approach has its advantages and drawbacks when carrying out inference and analysis and these were described in the paper. We further summarized the formulation of these two approaches and showed relations between them, where we focused on inference for each of the approaches.

In general, the intensity transition based approach (Approach II) allows the likelihood to be split when each transition has its own set of parameters. Such separation into two-state models facilitates more efficient optimization as well as usage of additional modeling tools. The large scale impact of such separation on bigger datasets and/or simulated data as part of a simulation study remains to be carried out in future work. However, even putting the computational aspects aside, there is great value in such separation as it allows to use many survival analysis R packages directly within the context of semi-Markov processes.

Such packages include `flexsurv` [26] which allows to fit various shapes of transition intensity functions including the Royston-Parmar spline model [44]. Further, in a semi-parametric framework the `mstate` package [45] is popular for running multi-state models including the SMP. By exploiting the separation of the likelihood, `mstate` provides the estimation of covariate effects using Cox regression models. In addition, non-linear effects of the covariates could be investigated through smooth functions using the `mgcv` package which enables to run GAM models with survival data, see [46]. Further, with Approach II, we can easily handle random effects using the `frailtypack` package [47]; as has been exploited by [10]. Further, in presence of high dimensional data, elastic-net penalty [48] can be used via the package `glmnet` to fit Cox regression models. Such benefits of using the intensity transition based approach have previously been reported and exploited for general Markov multi-state models as in

[17, 21] and [22]. Our purpose here was to survey these benefits and compare to Approach I as is popularly used with the SemiMarkov package providing a flexible tool for inference of semi-Markov models based on sojourn time hazard rates in a parametric framework [29].

The focus in this paper was purely on time-homogenous SMP in which the absolute time or date is assumed not to affect the probabilistic behavior. In certain situations one should consider time-inhomogeneous models allowing the absolute calendar time to play a role in the probabilistic model and the inference. Such situations can occur if seasonal (periodic) phenomena are present, or in long term longitudinal studies where the nature of the cohort is assumed to vary during the study. In a future study, it may be of interest to contrast Approach I and Approach II in a time-inhomogeneous setting.

Of further interest, we highlight the fact that Phase-Type (PH) distributions may fit naturally in an SMP framework. Such approaches have been previously proposed in [49–51] and [52]. Interestingly, as PH distributions provide a dense semi-parametric approximation of any distribution of a non-negative random variable, it may be of interest to approximate semi-Markov processes with continuous time Markov chains constructed via PH distributions, where there are potentially more than ℓ states in the approximating Markov chain. To the best of our knowledge, such a semi-parametric inference setup has still not been explored. Such a setup may work well with the EM algorithm for parameter estimation of this type of PH distributions, see [53]. We leave this for future work.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: AA and BL are supported by the Australian Research Council (ARC) Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS) under grant number CE140100049. YN is supported by Australian Research Council (ARC) under grant number DP180101602.

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

1. Broyles JR, Cochran JK, Montgomery DC. A statistical Markov chain approximation of transient hospital inpatient inventory. *Eur J Oper Res* 2010;207:1645–57.
2. Taylor GJ, McClean SI, Millard PH. Using a continuous-time Markov model with Poisson arrivals to describe the movements of geriatric patients. *Appl Stoch Model Bus Ind* 1998;14:165–74.
3. Aalen OO, Farewell VT, De Angelis D, Day NE, Nöel Gill O. A Markov model for HIV disease progression including the effect of HIV diagnosis and treatment: application to AIDS prediction in England and Wales. *Stat Med* 1997;16:2191–210.
4. Wang C, Alvarez SA, Ruiz C, Moonis M. Computational modeling of sleep stage dynamics using Weibull semi-Markov chains. In: Sixth international conference on health informatics (HEALTHINF 2013). Barcelona: SciTePress; 2013:122–30 pp.
5. Joly P, Commenges D. A penalized likelihood approach for a progressive three-state model with censored and truncated data: application to AIDS. *Biometrics* 1999;55:887–90.
6. Goshu AT, Dessie ZG. Modelling progression of HIV/AIDS disease stages using semi-Markov processes. *J Data Sci* 2013;11:269–80.
7. Trivedi V, Moscovice I, Bass R, Brooks J. A semi-Markov model for primary health care manpower supply prediction. *Manag Sci* 1987;33:149–60.
8. Polesel R, Romanin-Jacur G. Application of semi-Markov processes to health care planning: a model to study the flow of patients following the dialysis-transplantation plan. In: *Semi-Markov models*. Boston: Springer; 1986:437–50 pp.
9. Boucquemont J, Metzger M, Combe C, Stengel B, Leffondre K, Group NS, et al. Should we use standard survival models or the illness-death model for interval-censored data to investigate risk factors of chronic kidney disease progression? *PLoS One* 2014;9:e114839.
10. Liqueur B, Timsit J-F, Rondeau V. Investigating hospital heterogeneity with a multi-state frailty model: application to nosocomial pneumonia disease in intensive care units. *BMC Med Res Methodol* 2012;12:79.
11. Coeurjolly J-F, Nguele-Makao M, Timsit J-F, Liqueur B. Attributable risk estimation for adjusted disability multistate models: application to nosocomial infections. *Biom J* 2012;54:600–16.
12. Liqueur B, Asanjarani A, Nazarathy Y. Biostatistics-semi-markov-vignette; 2020. Available from: <https://github.com/benoit-liqueur/SEMI-MARKOV-VIGNETTE>.

13. Andersen PK, Borgan O, Gill RD, Keiding N. Statistical models based on counting processes. In: Springer series in statistics. New York: Springer Science & Business Media; 2012.
14. Janssen J. Semi-Markov models: theory and applications. In: Proceedings of the international symposium on semi Markov processes and their applications. Brussels: Springer Science & Business Media; 2013.
15. Ross SM, Kelly JJ, Sullivan RJ, Perry WJ, Mercer D, Davis RM, et al. Stochastic processes. New York: Wiley; 1996, vol 2.
16. Breuer L, Baum D. An introduction to queueing theory: and matrix-analytic methods. Dordrecht: Springer Science & Business Media; 2005.
17. Andersen PK, Keiding N. Multi-state models for event history analysis. *Stat Methods Med Res* 2002;11:91–115.
18. Koller MT, Raatz H, Steyerberg EW, Wolbers M. Competing risks and the clinical community: irrelevance or ignorance? *Stat Med* 2012;31:1089–97.
19. Commenges D, Jacqmin-Gadda H. Dynamical biostatistical models. New York: CRC Press; 2015, 86.
20. Andersen PK. Censored data. Encyclopedia of biostatistics. Hoboken: John Wiley; 2005. <https://doi.org/10.1002/0470011815.b2a11008>.
21. Hougaard P. Multi-state models: a review. *Lifetime Data Anal* 1999;5:239–64.
22. Meira-Machado L, de Uña-Álvarez J, Cadarso-Suárez C, Andersen PK. Multi-state models for the analysis of time-to-event data. *Stat Methods Med Res* 2009;18:195–222.
23. Król A, Saint-Pierre P. SemiMarkov: an R package for parametric estimation in multi-state semi-Markov models. *J Stat Software* 2015;66:1–16.
24. Therneau TM, Grambsch PM. The Cox model. In: Modeling survival data: extending the Cox model. New York: Springer; 2000: 39–77 pp.
25. Cook RJ, Lawless J. The statistical analysis of recurrent events. New York: Springer Science & Business Media; 2007.
26. Jackson CH. Flexsurv: a platform for parametric survival modeling in R. *J Stat Software* 2016;70. <https://doi.org/10.18637/jss.v070.i08>.
27. Andersen PK, Esbjerg S, Sørensen TI. Multi-state models for bleeding episodes and mortality in liver cirrhosis. *Stat Med* 2000; 19:587–99.
28. Hastie T, Tibshirani R. Exploring the nature of covariate effects in the proportional hazards model. *Biometrics* 1990;46: 1005–16.
29. Król A, Saint-Pierre P. SemiMarkov: an R package for parametric estimation in multi-state semi-Markov models. *J Stat Software* 2015;66:784.
30. Foucher Y, Mathieu E, Saint-Pierre P, Durand J-F, Daurès J-P. A semi-Markov model based on generalized Weibull distribution with an illustration for HIV disease. *Biom J: J Math Methods Biosci* 2005;47:825–33.
31. Therneau TM. A package for survival analysis in S; 2015. Available from: <https://CRAN.R-project.org/package=survival>, version 2.38.
32. Broström G. eha: event history analysis; 2019. Available from: <https://CRAN.R-project.org/package=eha>, R package version 2.8.0.
33. Prentice RL. Discrimination among some parametric models. *Biometrika* 1975;62:607–14.
34. Arthur Allignol AL. CRAN task view: survival analysis; 2020. <https://CRAN.R-project.org/view=Survival>, version 2014-07-25.
35. Crowley J, Hu M. Covariance analysis of heart transplant survival data. *J Am Stat Assoc* 1977;72:27–36.
36. Meira-Machado L, Roca-Pardiñas J. p3state. msm: analyzing survival data from an illness-death model. *J Stat Software* 2011; 38:1–18.
37. Liquet B, Sakarovich C, Commenges D. Bootstrap choice of estimators in parametric and semiparametric families: an extension of eic. *Biometrics* 2003;59:172–8.
38. Commenges D, Sayyareh A, Letenneur L, Guedj J, Bar-Hen A. Estimating a difference of Kullback–Leibler risks using a normalized difference of aic. *Ann Appl Stat* 2008;2:1123–42.
39. Breslow NE. Discussion of professor Cox's paper. *J Roy Stat Soc B* 1972;34:216–7.
40. Saint-Pierre P, Bourdin A, Chanez P, Daurès J-P, Godard P. Are overweight asthmatics more difficult to control? *Allergy* 2006;61: 79–84.
41. Saint-Pierre P, Combescure C, Daurès J, Godard P. The analysis of asthma control under a Markov assumption with use of covariates. *Stat Med* 2003;22:3755–70.
42. Commenges D. Inference for multi-state models from interval-censored data. *Stat Methods Med Res* 2002;11:167–82.
43. Van Den Hut A. Multi-state survival models for interval-censored data. CRC Press; 2016.
44. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med* 2002;21:2175–97.
45. de Wreede LC, Fiocco M, Putter H. mstate: an R package for the analysis of competing risks and multi-state models. *J Stat Software* 2011;38:1–30.
46. Wood S, Pya N, Säfken B. Smoothing parameter and model selection for general smooth models (with discussion). *J Am Stat Assoc* 2016;111:1548–75.

47. Rondeau V, Gonzalez JR, Mazroui Y, Mauguen A, Diakite A, Laurent A, et al. Frailtypack: general frailty models: shared, joint and nested frailty models with prediction; evaluation of failure-time surrogate endpoints; 2019. Available from: <https://CRAN.R-project.org/package=frailtypack>, R package version 3.0.3.
48. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J Stat Software* 2011;39:1–13.
49. Latouche G. A phase-type semi-Markov point process. *SIAM J Algebr Discrete Methods* 1982;3:77–90.
50. Malhotra M, Reibman A. Selecting and implementing phase approximations for semi-Markov models. *Stoch Model* 1993;9: 473–506.
51. Titman AC, Sharples LD. Semi-Markov models with phase-type sojourn distributions. *Biometrics* 2010;66:742–52.
52. Aalen OO. Phase type distributions in survival analysis. *Scand J Stat* 1995;22:447–63. <https://www.jstor.org/stable/4616373>.
53. Asmussen S, Nerman O, Olsson M. Fitting phase-type distributions via the EM algorithm. *Scand J Stat* 1996;23:419–41. <http://www.jstor.org/stable/4616418>.