Discrete-time multistate models: estimation, inference, properties

Christian Dudel, Daniel Schneider, Angelo Lorenti, Peng Li, Mikko Myrskylä

October 26, 2023

Abstract

Multi-state models generalize survival analysis to transitions back-and-forth between several states. The majority of the literature conceptualizes multi-state models in continuous time. Discrete-time approaches are rare, despite being well-suited for many practical applications. In this paper, we collect existing and provide new results on discrete-time multi-state models. These models have desirable properties and are easy to apply. Specifically, we set up the model as an absorbing Markov chain, and we discuss estimation and inference. Given certain assumptions on functional form, estimation is straightforward, and can use standard methods widely implemented in statistical software. Moreover, we show that Markov chain multi-state models provide consistent estimates of several estimands even when the underlying data generating process is non-Markovian. We conduct simulations which show that small-sample bias is negligible, and that bias is moderate even when the estimate is not consistent under a non-Markovian data generating process.

Keywords: Multi-state model; discrete-time multi-state model; Markov chain; Markov model; interval-censored data; panel data

1 Introduction

Multi-state models generalize standard survival analysis and competing risk models. They allow multiple transitions between several states. An illness-death model with recovery is a simple example with three states: healthy, ill, and dead. Healthy individuals can become ill; ill individuals can recover; and both healthy and ill individuals might die. Multi-state models like this have a wide range of applications. For instance, Keiding et al. (2001) study health outcomes after bone marrow transplantation to treat leukemia. In their model, after transplantation patients can develop graft versus host disease of different severity, each captured by a separate state; they can relapse; and they can die. Other medical applications include different types of cancer (e.g., Le-Rademacher et al., 2018; Uhry et al., 2010; Putter et al., 2006), heart disease and heart failure (e.g., Ieva et al., 2017; Cannon et al., 2017), and dementia (e.g. Vermunt et al., 2019; Buter et al., 2008). In economics, multi-state models have been extensively used to study labor market trajectories (e.g., Harris et al., 2021; Bijwaard, 2014; Skoog and Ciecka, 2010). Further examples in the social sciences include patterns of family formation (Bonetti et al., 2013) and migration (Vega and Brazil, 2015; Schmertmann, 1999).

A large part of the literature on multi-state models conceptualizes the data generating process as time-continuous. That is, time flows continuously, and transitions between states can happen at any time. Several excellent reviews and textbook treatments of continuous-time multi-state models are available (e.g., Cook and Lawless, 2018; Putter et al., 2007; Andersen and Keiding, 2002). Alternatively, multi-state models can be thought of as evolving in discrete time; i.e., time evolves in discrete steps of fixed length, and the process jumps between states. The literature on discrete-time multi-state models is much sparser than the literature on time-continuous models, spread across several disciplines, and in reviews and textbooks only mentioned in passing, if at all (e.g., Cook and Lawless, 2018; van den Hout, 2017). This gap is surprising, as discrete-time models are useful in many applications. Measurement often is only available on a discrete time scale (Schmid and Berger, 2020). For instance, the health status of patients might only be measured once per week, and panel data which collects repeated measures from respondents at fixed time-intervals is commonly used in the social sciences. Moreover, discrete-time approaches build on transition probabilities which are easy to understand, and estimation is rather straightforward as compared to continuous-time approaches (Tutz and Schmid, 2016).

In this paper, we setup a multi-state model as a discrete-time, homogeneous, and absorbing Markov chain of first order. Such Markov chains describe the transition process through one-step transition probabilities. We discuss estimation for such models, closed-form solutions for some functionals of the process, and inference. We also briefly comment on the implementation of the approach in statistical software. We provide formal results on the asymptotic validity of results if the assumptions underlying the Markov chain are violated, and we conduct simulations to assess finite sample behavior.

A key result of this paper is that neither the Markov assumption nor homogeneity are necessary for our discrete-time approach produce valid results. The Markov assumption implies that the process is memoryless: the next state at time $t+1$ only depends on which

state is occupied at current time t, and not on earlier states at earlier times. This assumption is likely violated in many assumptions. However, we show that several statistics can be consistently estimated with one-step transition probabilities even if the underlying data generating process is non-Markovian. Moreover, one-step transition probabilities for non-Markovian processes can be estimated just like in the Markovian case, even though the derivation/justification is somewhat different. Taken together this means that the Markov chain approach can be used even when one of it's key assumptions does not hold, and our simulations show that finite sample bias is negligible. This does not hold for all statistics which can be derived from a Markov chain. For instance, the expected (residual) lifetime in a state can be consistently estimated, while the variance of the (residual) lifetime in a state cannot. However, in our simulations we show the bias for such statistics is usually modest, at least if the violation of the Markov assumption is moderate. We provide an approach to quantify whether this is the case.

2 Model and estimands

2.1 The Markov chain

Let $(Z_{t\in\mathscr{T}})$ be a time-homogenous, absorbing Markov chain with finite state space $\mathscr{Z} = \mathscr{S}\times\mathscr{T}\cup\mathscr{A}$. Z_t denotes the state occupied by the process at time t. $\mathscr{S}$ is a finite set of transient states, with $|\mathscr{S}| = S$ being the number of transient states. $\mathscr{A}$ is the set of absorbing states and $|\mathscr{A}| = A$. The time scale $t \in \mathscr{T} = (0,1,2,\ldots,T)$ moves in discrete time steps. We use a clock-forward approach, and $t = 0$ refers to the time the process started (Putter et al., 2007). T is the longest possible duration of the process before one of the absorbing states in $\mathscr{A}$ will be reached, although the absorbing state might be reached earlier. Deciding on a value for T depends on the specific application, but should be straightforward for most. For example, in a clinical trial with 5-year follow up, the maximum time will be 5 years. In a more complete example, t could capture age measured in years, $\mathscr{S}$ could consist of the two states "healthy" and "ill", and $\mathscr{A}$ of the absorbing state "dead". Once some maximum age is reached, such as 99 years, individuals die with probability 1, but they might die earlier. Then the state space consists of the states (healthy, age 0), (unhealthy, age 0), (healthy, age 1), ..., (healthy, age 99), (unhealthy, age 99), and (dead). Thus, in our setup, time is absorbed into the state space and combined with the transient states. To keep notation simple, we will assume that the process always starts at time $t = 0$ and evolves in time steps of one unit.

$\Pr[Z_{t+1} = (j,t+1)|Z_t = (i,t)] = p_{ij}^{(t)}$ is the transition probability that the process is in state $j \in \mathscr{S}$ at time $t+1$ after being in state $i \in \mathscr{S}$ at time t. These transition probabilities only depend on the current state at time t and not on the previous history of the process at times $t-1, t-2$, and so on; i.e., $\Pr[Z_{t+1} = (j,t+1)|Z_t = (i,t)] = \Pr[Z_{t+1} = (j,t+1)|Z_t = (i,t), \mathscr{H}_{t-1}]$, where $\mathscr{H}_{t-1} = (Z_{t-1},\ldots,Z_0)$ captures the history of the process up to and including $t-1$. Thus, the process is memoryless and has the Markov property. The distribution of states at any time t is denoted as $\pi_t = (\Pr[Z_t = z_{i,t}])$, which is a vector with $S+A$ entries. π_0 is the distribution at the start of the process. In some cases, π_0 might be fixed, particularly when the process always starts in the same state.

Using this setup, we can collect transition probabilities in a homogeneous transition matrix which is structured by the time of the process t as (Caswell, 2012)

$$\mathbf{P} = \begin{pmatrix} \mathbf{0} & \mathbf{P}_0 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{R}_0 \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_1 & \dots & \mathbf{0} & \mathbf{R}_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{P}_T & \mathbf{R}_Z \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{I}_{\mathscr{A}} \end{pmatrix}, \tag{1}$$

where the $S \times S$ submatrices $\mathbf{P}_t$ contain the transition probabilities $p_{ij}^{(t)}$; the submatrices $\mathbf{R}_t$ are of dimension $S \times A$ and contain the transition probabilities which describe transitions from the transient to the absorbing states; and $\mathbf{I}_{\mathscr{A}}$ is an $A \times A$ identity matrix, as the absorbing states are never left. The transition probabilities could also be arranged in different ways and results do not depend on the specific arrangement used here, but it makes notation easier (Caswell, 2012).

2.2 State expectancies and other estimands

Given the definitions of the process studied here, standard methods for homogeneous, absorbing Markov chains can be used to calculate statistics which describe the process, such as the expected time spent in a state or the lifetime risk of ever entering a state (e.g. Kemeny and Snell, 1971; Iosifescu, 1980). These methods use the transition matrix $\mathbf{P}$ as defined in equation (1).

The expected time spent in any transient state $s \in \mathscr{S} \times \mathscr{T}$ can be calculated using the fundamental matrix $\mathbf{N}$ (Kemeny and Snell, 1971). It is given by

$$\mathbf{N} = (\mathbf{I}_{ST} - \mathbf{U}). \tag{2}$$

$\mathbf{I}_{ST}$ is an identity matrix of size $ST \times ST$ and

$$\mathbf{U} = \begin{pmatrix} \mathbf{0} & \mathbf{P}_0 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_1 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{P}_T \end{pmatrix}, \tag{3}$$

i.e., the transition matrix for the transient states. Each row of $\mathbf{N}$ corresponds to one specific state; the entries in each row give the expected time spent in all other states starting from this specific state. Given the way the model is setup here, each entry $n_{ij}^{(tu)}$ of $\mathbf{N}$ equals $\mathrm{E}[D_{j,t}|Z_u = (i,u)]$, where (i,u) is the starting state corresponding to a specific row and $D_{j,t} = \mathbb{I}[Z_t = z_{j,t}]$ is an indicator variable which is equal to one if the process is in state $j \in \mathscr{Z}$ at time t. The derivation of $\mathbf{N}$ is a classic result in Markov chain theory and essentially based on the repeated application of $\mathbf{U}$; i.e., the repeated application of one-step transition probabilities based on the Markov assumption (for a derivation see Kemeny and Snell, 1971).

Let D_j capture the total time spent in state $j \in \mathscr{S}$. Its expectation is given by

$$\begin{aligned} \mathrm{E}\left[D_j\right] &= \sum_i \Pr[Z_0 = (i,0)] \sum_{t=0}^{T} \mathrm{E}\left(D_{j,t} | Z_0 = z_{i,0}\right) && (4) \\ &= \sum_i \Pr[Z_0 = (i,0)] \left(\sum_{t=0}^{T} n_{ij}^{(t)} \right) - 0.5\mathbb{I}[i=j]. && (5) \end{aligned}$$

The adjustment by half a time unit in case of $i = j$ is based on the assumption that transitions occur mid-interval; i.e., if the process starts in state i at time $t = 0$, then the transition to another state at time $t = 1$ will on average occur at time $t = 0.5$, thus only contributing half a time unit.

Many other quantities can be derived using the fundamental matrix and the transition matrix. For instance, the lifetime risk of ever reaching a state $j \in \mathscr{Z}$ starting from state j is also just given by the entries of $\mathbf{N}$, as $n_{ij}^{(t0)} = \mathrm{E}[D_{j,t} | Z_0 = z_{i,0}] = \Pr[Z_t = z_{j,t} | Z_0 = z_{i,0}]$. Another useful quantity is the probability of being absorbed into a specific absorbing state $a \in \mathscr{A}$, which can be calculated as the row sums of (Kemeny and Snell, 1971)

$$\mathbf{N} \begin{pmatrix} \mathbf{R}_0 \\ \mathbf{R}_1 \\ \vdots \\ \mathbf{R}_Z \end{pmatrix}. \tag{6}$$

More generally, the full distribution of D_j for j in $\mathscr{S}$ can be calculated from the transition matrix given that the Markov assumption holds (Dudel, 2021; Sericola, 2000). This allows to calculate the lifetime risk of reaching j, the variance of D_j, and other characterizations of this distribution.

3 Estimation and inference

3.1 Estimation of transition probabilities

For estimating the transition probabilities the observed transitions are relevant. For instance, if an individual is observed at times $0, 1, 2, \ldots, T$, then this individual will contribute $T - 1$ transitions: the transition from 0 to 1, the transition from 1 to 2, and so on, until the transition from $T + 1$ to T. Let $\mathbf{x}_i$ be a vector of time-constant covariates for individual i; these could capture, for instance, gender, education, or whether in a clinical trial an individual belongs to the treatment or the control group. Time-varying covariates $\mathbf{x}_{it}$ are also easily possible, but we will restrict ourselves to the time-constant case for simplicity.

Given this notation, assuming there is no censoring, and making use of the Markov assumption, the likelihood function is (also see van den Hout, 2017)

$$L = \prod_i \prod_{t=0}^{T-1} \prod_j \prod_k \Pr[Z_{i,t+1} = (k, t+1) | Z_{i,t} = (j,t), \mathbf{x}_i]^{\mathbb{I}[i,j,k,t]}, \tag{7}$$

where $(j,t) \in \mathscr{S} \times t$, $(k,t+1) \in \mathscr{S} \times t+1 \cup \mathscr{A}$, and $\mathbb{I}[i,j,k,t]$ is an indicator variable which equals 1 if individual i transitioned from state j at time t to state k at time $t+1$. To account for right censoring, let δ_{it} equal 0 if individual i is not observed anymore at time t. Assuming that censoring is independent of the other transitions and that censoring is non-informative – i.e., that the censoring process does not depend on similar parameters as the transition process (Kalbfleisch and Prentice, 2002) – the likelihood is proportional to

$$L \propto \prod_i \prod_{t=0}^{T_i-1} \prod_j \prod_k \Pr[Z_{i,t+1} = (k,t+1) | Z_{i,t} = (j,t), \mathbf{x}_i]^{\mathbb{I}[i,j,k,t]}, \tag{8}$$

where T_i is the last time unit i is observed. Left censoring can be dealt with in a similar way.

For parametric estimation, let β_{jkt} be a vector of coefficients which describe the impact of $\mathbf{x}_i$ on the transition from state (j,t) to state $(k,t+1)$. The likelihood is then proportional to

$$L \propto \prod_i \prod_{t=0}^{T_i-1} \prod_j \prod_k \Pr[Z_{i,t+1} = (k,t+1) | Z_{i,t} = (j,t), \mathbf{x}_i, \beta_{jkt}]^{\mathbb{I}[i,j,k,t]}. \tag{9}$$

This setup will have many parameters to estimate, potentially resulting in numerical instabilities and affecting inference (Jóźwiak and Moerbeek, 2012). This is a common issue in multi-state modeling, and a typical solution is to arrive at a reduced model by introducing assumptions on the functional form (e.g., Fiocco et al., 2008). Here, we will include both t and Z_t as covariates; i.e., such that the new vector of covariates is $\mathbf{y}_t = (\mathbf{x}, Z_t, t)$. This leaves us with

$$L \propto \prod_i \prod_{t=0}^{T_i-1} \prod_k \Pr[Z_i = k | \mathbf{y}_{i,t}, \gamma_k]^{\mathbb{I}[i,j,k,t]}, \tag{10}$$

where γ_k is the new set of coefficients, including all coefficients for $\mathbf{x}$ as well as the coefficients for t and for Z_t. Here, Z_i instead of $Z_{i,t+1}$ is used to indicate that the dependence on t is now implicit in the vector of covariates. This means that in the reduced model transitions follow a multinomial distribution conditional on $\mathbf{y}_{i,t}$ and the model parameters; i.e., the reduced model is equivalent to a discrete-time competing risk model.

The reduced model can be estimated by using several approaches for discrete-time competing risks. Here, we will use multinomial logistic regression, which is straightforward to implement and extend. If $M = S + A$ is the total number of transient and absorbing states not accounting for time, then there will be $M-1$ equations, one for each transient and absorbing state minus one as the transition probabilities sum to one. Each of these equations represents the transition probability to one state. The regression equations are given by

$$\log \left[\frac{\Pr[Z_{t+1} = (k,t+1) | \mathbf{y}_t]}{\Pr[Z_{t+1} = (M,t+1) | \mathbf{y}_t]} \right] = \alpha_k + \sum_{i=1}^{M-1} \gamma_{jk} \mathbb{I}(Z_t = z_{k,t}) + f[t] + \mathbf{x}' \gamma_x. \tag{11}$$

On the left hand side of the equation are the log odds of being in state $(k,t+1)$ relative to state $(M,t+1)$. On the right hand side, α_k denotes the intercept; γ_{jk} is the coefficient

for starting in state (j,t); $f[t]$ is some function of the time of the process, such as a simple linear function $\gamma_t t$ or a polynomial of t; and γ_x captures the coefficients for explanatory variables $\mathbf{x}$. Given parameter estimates, the calculation of transition probabilities from equation (11) is straightforward.

3.2 Variance estimation using the block bootstrap

Estimating the variance of functionals of a Markov chain is not straightforward. These functionals are usually non-linear functions of the transition probabilities, making it difficult to find analytical solutions. The block bootstrap is an alternative approach (e.g., Sutradhar and Cook, 2008). It is easily applicable to any functional of the Markov chain, and it has been shown to perform well (Dudel and Myrskylä, 2020). All data belonging to the same unit or individual k is treated as one "block" b_k. B denotes the total number of blocks and $\mathscr{B} = \{b_1, \ldots, b_k, \ldots, b_B\}$ is the set of all blocks. For each of R bootstrap replications, a new sample $\mathscr{B}^*$ is created by sampling B blocks from $\mathscr{B}$ with replacement and the functional of interest is calculated. The variance across bootstrap replications then is used as an estimate of the sampling variance of the functional.

4 Consistency

4.1 Transition probabilities and the state distributions

In this section, we will show that some functionals of the transition probabilities can be consistently estimated using the Markov assumption even when the data generating process (DGP) is non-Markovian, including state expectancies and lifetime risks; e.g., equation (5) provides a consistent estimate of $\mathrm{E}(D_j)$. The proofs require consistent estimates of one-step transition probabilities and of the starting distribution. This is straightforward. When the DGP is non-Markovian, $\Pr[Z_{t+1} = (i,t+1)|Z_t = (j,t)] \neq \Pr[Z_{t+1} = (i,t+1)|Z_t = (j,t), \mathscr{H}_{t-1}]$; i.e., $\Pr[Z_{t+1} = (i,t+1)|Z_t = (j,t)]$ does not provide a complete description of the process. Nevertheless, it can still be estimated using the methods described in the previous section, as standard consistency arguments do not require that $\Pr[Z_{t+1} = (i,t+1)|Z_t = (j,t)]$ is a complete description of the DGP; or put differently, consistent estimation of conditional probabilities does not require that these conditional probabilities are in any sense a complete description of the DGP or that the Markov assumption holds. Moreover, it is still an interesting statistic, as it can be interpreted as the average over all potential histories, $\sum_{\mathscr{H}_{t-1}} \Pr[Z_t = (j,t), \mathscr{H}_{t-1}] \Pr[Z_{t+1} = (i,t+1)|Z_t = (j,t), \mathscr{H}_{t-1}]$ (Putter and Spitoni, 2018). Consistent estimation of the starting distribution, π_0, is also straightforward.

Based on consistent estimates of one-step transition probabilities and the starting distribution, π_t can be estimated consistently when the DGP is non-Markovian. For instance, entry j of π_1 equals

$$\Pr[Z_1 = (j,1)] = \sum_i \Pr[Z_0 = (i,0)] p_{ij}^{(0)}; \tag{12}$$

i.e., it is a linear combination of the starting distribution and one-step transition probabilities and thus consistent by Slutsky's theorem if its components are consistently estimated. This argument can be applied repeatedly to show that the distribution of states at any time t,

$$\Pr[Z_{t+1} = (j, t+1)] = \sum_i \Pr[Z_t = (i,t)] p_{ij}^{(t)}, \tag{13}$$

can be consistently estimated. Thus, for the calculation of the cross-sectional state distribution only the current state is required and the history is irrelevant.

4.2 State expectancies

Given consistent estimates of π_t, consistency of $\mathrm{E}(D_j)$ followsstraightforward. $\mathrm{E}[D_j] = \sum_t \mathrm{E}[D_{j,t}]$ as the expectation is additive. Assuming that individuals spend one full time unit in state j if they occupy it at time t, $\mathrm{E}[D_{j,t}] = \Pr[Z_t = (j,t)]$, which can be taken from π_t. Thus, $\mathrm{E}[D_j] = \sum_t \Pr[Z_t = (j,t)]$, which is equivalent to equation (4):

$$\begin{aligned}
\mathrm{E}\left[D_j\right] &= \sum_t \Pr[Z_t = (j,t)] && (14)\\
&= \sum_t \sum_i \Pr[Z_{t-1} = (i,t-1)] p_{ij}^{(t-1)} && (15)\\
&= \sum_i \Pr[Z_0 = (i,0)] \sum_{t=0}^{T} \Pr[Z_t = (j,t) | Z_0 = (i,0)] && (16)
\end{aligned}$$

The step from (15) to (16) follows from recursively replacing $\Pr[Z_{t-1} = (i,t-1)]$ with its definition from equation (13) and then rearranging terms. The correction by 0.5 time units given in equation (5) is missing but can easily be added.

4.3 Lifetime risk

Another quantity which can in principle be estimated consistently even if the DGP is non-Markovian is the probability of ever reaching a specific state $s \in \mathscr{S}$. To achieve this, s is turned into an absorbing state and moved from $\mathscr{S}$ to $\mathscr{A}$, as was already suggested by Brookmeyer and Abdalla (2018). Then the probability of being absorbed into state s and not other states $a \in \mathscr{A}$ equals the probability of ever reaching s; e.g., the lifetime risk of ever getting disabled. This exploits that while transition probabilities do not have a memory, they are conditional on not having reached an absorbing state before; i.e. $\Pr[Z_{t+1} = (j,t+1) | Z_t = (i,t), s \notin \mathscr{H}_{t-1}] = \Pr[Z_{t+1} = (j,t+1) | Z_t = (i,t)]$.

Formally, the probability of being absorbed in state s can be written as

$$\begin{aligned}
\Pr[s \notin \mathscr{H}_T] &= 1 - \Pr[s \in \mathscr{H}_T] && (17)\\
&= 1 - \sum_i \sum_{t=0}^{T} \sum_{k \in \mathscr{S}, k \neq s} p_{ks}^{(t)} \Pr[Z_t = (k,t) | Z_0 = (i,0)] \Pr[Z_0 = (i,0)]. && (18)
\end{aligned}$$

This is equivalent to equation (6). Consistency follows the same argument as for state expectancies: the calculation does not require the Markov assumption, and its a linear

combination of quantities which can be consistently estimated irrespective of whether the DGP is Markovian or not.

While estimation of the lifetime risk is feasible in this way, it increases the demands regarding the data, as s will in reality not be absorbing. That is, an individual might have been in state s in the past, but this might not be recorded in the data. This could lead to biased transition probabilities, where the bias will depend on the extent to which past visits to state s are missed in the data.

4.4 Other functionals and higher moments

Consistent estimation of higher moments of D_j, and of its complete distribution in general, is not possible if the DGP is non-Markovian. For instance, the variance of D_j requires knowledge of $\mathrm{E}[D_{j,t}D_{k,u}] = \Pr[Z_t = (j,t), Z_k = (k,u)]$ which only can be calculated from memoryless transition probabilities if the Markov assumption holds. In this case detailed knowledge of trajectories of the process is required.

However, the amount of bias when estimating higher moments of D_j using the Markov assumption when the true data generating process is non-Markovian depends on the extent the true process has memory. For instance, the variance of D_j can be written as

$$\begin{aligned} \mathrm{Var}[D_j] &= \sum_{t=0}^{T} \mathrm{Var}[D_{j,t}] + \sum_{0 \le t < k \le T} \mathrm{Cov}[D_{j,t}, D_{j,k}] && (19) \\ &= \sum_{t=0}^{T} \mathrm{Var}[D_{j,t}] + \sum_{t=0}^{T} \mathrm{Cov}[D_{j,t}, D_{j,t+1}] + \sum_{0 \le t \le T, k > t+1} \mathrm{Cov}[D_{j,t}, D_{j,k}]. && (20) \end{aligned}$$

The first and the second sum only require knowledge of $p_{ij}^{(t)}$, while the third sum requires transition probabilities with more memory if the process is non-Markovian. Bias will be proportional to this third sum. If the third sum is small relative to the first and the second sum, then in practice estimates of the variance of D_j using the Markov assumption might provide good approximations. A similar reasoning applies to other moments of the distribution of D_j, and more generally to functionals of transition probabilities: the process might not be memoryless, but it might not have a long-term memory either.

4.5 Testing the Markov assumption

In practical applications, departures from the Markov assumption will not always be clear cut, and might be gradual. For instance, the Markov assumption could strictly speaking be violated, but conditioning transition probabilities on states visited at time $t-1$, $t-2$, etc. might not change them much as compared to one step transition probabilities. This might be particularly true when additionally conditioning on a set of covariates $\mathbf{x}$. Moreover, the Markov assumption could hold for some types of transitions, but not for others. If transitions of the latter type occur with a relatively low probability, our approach also could provide good approximations.

Here, we propose a test of the Markov assumption which is straightforward to implement and easy to interpret. It is defined as follows:

$$\Delta(L) = \frac{1}{N_{Obs}} \sum_i \frac{1}{2} \left| \Pr[Z_{t+1} = (i, t+1) | \mathbf{x}] - \Pr[Z_{t+1} = (i, t+1) | \mathbf{x}, Z_t, Z_{t-1}, \ldots, Z_{t-L}] \right|, \quad (21)$$

where N_{Obs} is the number of observed transitions in the data. For each observed transition in the data, predicted transition probabilities without any lagged state are calculated, and then compared to transition probabilities up to lag $t-L$. Technically, this is equivalent to an average of the dissimilarity index (Kuha and Firth, 2011). $\Delta(L)$ assesses how much individual-level predictions of the state at time $t+1$ change when additional lags of the occupied state are introduced. It always lies between 0 and 1. It equals 1 when all predictions change completely, and it equals 0 if nothing changes. $\Delta(0)$ just conditions on the current state Z_t with no lag. We define $\Delta(-1) = 0$ as a second reference point when neither the current state nor lags are included. Comparing $\Delta(L)$ with $\Delta(0)$ and $\Delta(-1)$ then shows how much prediction can be improved over only using covariates and only using covariates and the Markov assumption. Inference can use the block bootstrap.

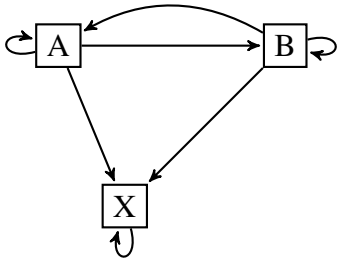
5 Simulations

5.1 Setup

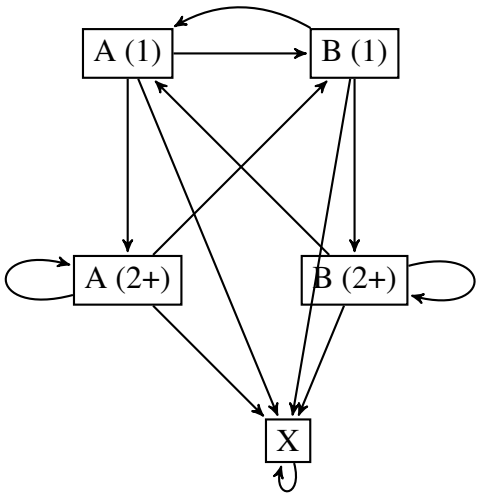
Using simulations we assess the finite sample bias of three Markov chain estimands: state expectancies, lifetime risks, and the variance of the distribution of the time spent in a state. The simulations are set up such that the Markov assumption is violated to varying degrees, ranging from mild violations to processes which strongly depend on the past. In addition, we vary several other parameters, such as the number of states, sample size, duration of the process, and the amount of censoring.

To simulate DGPs which violate the Markov assumption we exploit that in the discrete-time setting, we can introduce duration-dependence of transition probabilities by extending the state space. That is, if m captures the time spent in state i, then the state space is extended to include (i,t,m) instead of just (i,t). This way, it is theoretically possible to include the complete history in the state space $(i,t,\mathscr{H}_{t-1})$, and to capture any discrete-time process with a Markovian model. Practically, however, this will not be possible. For our simulations, estimation will only include (i,t) and not (i,t,m); that is, any state (i,t,m) will be represented by (i,t), irrespective of the value of m. While the true underlying DGP is Markovian, the resulting Markov chain with lumped states is not (Kemeny and Snell, 1971).

For a first set of simulations, the three models in Figure 1 were used to simulate DGPs, for a maximum of up to 10 time steps. Model 1 has two transient states, A and B, and is not duration dependent. In contrast, for Models 2 and 3 the state space is extended to capture duration dependency, and the first time unit in states A and B comes along with different transition probabilities than later time units. In all models the transition probabilities do not depend on t. The transition probabilities were chosen such that Model 2 strongly violates



(a) Model 1



(b) Model 2 and 3

Figure 1: Model structure for the simulations.

the Markov assumption. Model 3 is between Model 1 and Model 2. For all three models, the true values of the estimands of interest can be calculated analytically.

In the simulations, for all three models only a basic multistate model with two transient states is estimated. Based on the estimated transition probabilities, all estimands are calculated using the Markov assumption. These point estimates can be compared to the true values for the underlying process. We conduct our simulations with sample sizes of 100, 250, and 500 individuals. For each sample size, we run 1,000 replications and report the mean relative bias calculated over all 1,000 replications as well as the average value of $\Delta(1)$.

5.2 Results

Results are shown in Table 1 and expressed as mean relative bias (MRB), $\mathrm{E}(S_i - T)/T$, where S_i is the result of replication i of the simulation for a certain statistic and T is the true value. As can be seen, the MRB for $\mathrm{E}(A)$ and $\mathrm{E}(B)$ is always negligible, irrespective of the data generating process (DGP) and sample size. Bias for $\mathrm{Var}(A)$ and $\mathrm{Var}(B)$ can be rather substantial if the DGP is non-Markovian. $\mathrm{R}(A)$ and $\mathrm{R}(B)$ are interesting in that bias is very small at worst; however, the highest values of MRB are not observed for model 2

Table 1: Mean relative bias (MRB) of the state expectancy in states A and B, E(A) and E(B); the variance of the time spend in A and B, Var(A) and Var(B); the lifetime risk of ever reaching state A and B, R(A) and R(B); and average of $\Delta(1)$ for the different simulation setups.

Model	Sample size	E(A)	E(B)	Var(A)	Var(B)	R(A)	R(B)	$E(\Delta(1))$
Model 1	100	0.002	-0.002	-0.005	-0.005	0.000	0.000	0.020
	250	0.004	0.004	-0.001	-0.001	-0.001	0.000	0.013
	500	0.006	-0.006	0.015	0.015	0.000	0.000	0.009
Model 2	100	-0.001	0.001	7.492	7.492	0.001	0.001	0.369
	250	-0.002	0.002	7.946	7.946	0.001	0.001	0.369
	500	0.001	-0.001	7.757	7.757	0.001	0.001	0.368
Model 3	100	-0.006	0.006	0.534	0.534	-0.022	-0.021	0.246
	250	-0.003	0.003	0.540	0.540	-0.020	-0.021	0.249
	500	0.006	-0.006	0.551	0.551	-0.019	-0.022	0.249

but for model 3. This might seem surprising, as model 3 has lower values for $\Delta(1)$ than model 2. However, model 2 is set up in such a way that both states A and B are essentially always reached, leaving little variation, if any.

6 Outlook

Overall, our theoretical findings as well as the results of our simulations show that discrete-time multistate models work extremely well, even if the Markov assumption is violated. Further simulations will emulate more realistic settings and will, for instance, also take into account that transition probabilities might change over time as well as censoring.

References

Andersen, P. K. and Keiding, N. (2002) Multi-state models for event history analysis. *Statistical Methods in Medical Research*, **11**, 91–115.

Bijwaard, G. E. (2014) Multistate event history analysis with frailty. *Demographic Research*, **30**, 1591–1620.

Bonetti, M., Piccaretta, R. and Salford, G. (2013) Parametric and nonparametric analysis of life courses: An application to family formation patterns. *Demography*, **50**, 881–902.

Brookmeyer, R. and Abdalla, N. (2018) Multistate models and lifetime risk estimation: Application to Alzheimer's disease. *Statistics in Medicine*, **38**, 1558–1565.

Buter, T. C., van den Hout, A., Matthews, F. E., Larsen, J. P., Brayne, C. and Aarsland, D. (2008) Dementia and survival in parkinson disease a 12-year population study. *Neurology*, **70**, 1017–1022.

Cannon, J., Roberts, K., Milne, C. and Carapetis, J. R. (2017) Rheumatic heart disease severity, progression and outcomes: A multi-state model. *Journal of the American Heart Association*, **6**, e003498.

Caswell, H. (2012) Matrix models and sensitivity analysis of populations classified by age and stage: a vec-permutation approach. *Theoretical Ecology*, **5**, 403–417.

Cook, R. J. and Lawless, J. F. (2018) *Multistate Models for the Analysis of Life History Data*. CRC Press.

Dudel, C. (2021) Expanding the Markov chain toolbox: Distributions of occupation times and waiting times. *Sociological Methods & Research*, **50**, 401–428.

Dudel, C. and Myrskylä, M. (2020) Estimating the number and length of episodes in disability using a markov chain approach. *Population Health Metrics*, **18**, 15.

Fiocco, M., Putter, H. and van Houwelingen, H. C. (2008) Reduced-rank proportional hazards regression and simulation-based predictions for multi-state models. *Statistics in Medicine*, **27**, 4340–4358.

Harris, M. N., Zhao, X. and Zucchelli, E. (2021) Ageing workforces, ill-health and multi-state labour market transitions. *Oxford Bulletin of Economics and Statistics*, **83**, 199–227.

van den Hout, A. (2017) *Multi-State Survival Models for Interval-Censored Data*. CRC Press.

Ieva, F., Jackson, C. H. and Sharples, L. D. (2017) Multi-state modelling of repeated hospitalisation and death in patients with heart failure: The use of large administrative databases in clinical epidemiology. *Statistical Methods in Medical Research*, **26**, 1350–1372.

Iosifescu, M. (1980) *Finite Markov Processes and Their Applications*. Dover.

Jóźwiak, K. and Moerbeek, M. (2012) Power analysis for trials with discrete-time survival endpoints. *Journal of Educational and Behavioral Statistics*, **37**, 630–654.

Kalbfleisch, J. D. and Prentice, R. L. (2002) *The Statistical Analysis of Failure Time Data*. Wiley.

Keiding, N., Klein, J. P. and Horowitz, M. M. (2001) Multi-state models and outcome prediction in bone marrow transplantation. *Statistics in Medicine*, **20**, 1871–1885.

Kemeny, J. G. and Snell, J. L. (1971) *Finite Markov Chains*. Springer.

Kuha, J. and Firth, D. (2011) On the index of dissimilarity for lack of fit in loglinear and log-multiplicative models. *Computational Statistics & Data Analysis*, **55**, 375–388.

Le-Rademacher, J. G., Peterson, R. G., Thernau, T. M., Sanford, B. L., Stone, R. M. and Mandrekar, S. J. (2018) Application of multi-state models in cancer clinical trials. *Cilincal Trials*, **15**, 489–498.

Putter, H., Fiocco, M. and Geskus, R. B. (2007) Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, **26**, 2389–2430.

Putter, H., van der Hage, J., de Bock, G. H., Elgalta, R. and van de Velde, C. J. H. (2006) Estimation and prediction in a multi-state modelfor breast cancer. *Biometrical Journal*, **48**, 366–380.

Putter, H. and Spitoni, C. (2018) Non-parametric estimation of transition probabilities in non-markov multi-state models: The landmark Aalen–Johansen estimator. *Statistical Methods in Medical Research*, **27**, 2081–2092.

Schmertmann, C. P. (1999) Estimating multistate transition hazards from last-move data. *Journal of the American Statistical Association*, **94**, 53–63.

Schmid, M. and Berger, M. (2020) Competing risks analysis for discrete time-to-event data. *Wiley Interdisciplinary Reviews-Computational Statistics*, **13**, e1529.

Sericola, B. (2000) Occupation times in Markov processes. *Stochastic Models*, **16**, 479–510.

Skoog, G. R. and Ciecka, J. E. (2010) Measuring years of inactivity, years in retirement, time to retirement, and age at retirement within the Markov model. *Demography*, **47**, 609–628.

Sutradhar, R. and Cook, R. J. (2008) Analysis of interval-censored data from clustered multistate processes: application to joint damage in psoriatic arthritis. *Journal of the Royal Statistical Society, Series C*, **57**, 553–566.

Tutz, G. and Schmid, M. (2016) *Modeling Discrete Time-to-Event Data*. Springer.

Uhry, Z., Hedelin, G., Colonna, M., Asselain, B., Arveux, P., Rogel, A., Exbrayat, C., Guldenfels, C., Courtial, I., Soler-Michel, P., Molinie, F., Eilstein, D. and Duffy, S. W. (2010) Multi-state markov models in cancer screening evaluation: a brief review and case study. *Statistical Methods in Medical Research*, **19**, 463–486.

Vega, A. and Brazil, N. (2015) A multistate life table approach to understanding return and reentry migration between mexico and the united states during later life. *Demographic Research*, **33**, 1211–1240.

Vermunt, L., Sikkes, S. A. M., van den Hout, A., Handels, R., Bos, I., van der Flier, W. M., Kern, S., Ousset, P.-J., Maruff, P., Skoog, I., Verhey, F. R. J., Freund-Levi, Y., Tsolaki, M., Wallin, A. K., Rikkert, M. O., Soininen, H., Spiru, L., Zetterberg, H., Blennow, K., Scheltens, P., Muniz-Terrera, G., Visser, P. J., Alzheimer Disease Neuroimaging Initiative, AIBL Research Group and ICTUS/DSA study groups (2019) Duration of preclinical, prodromal, and dementia stages of alzheimer's disease in relation to age, sex, and apoe genotype. *Alzheimer's & Dementia*, **15**, 888–898.