Statistica Sinica **31** (2021), 1807-1828
doi:https://doi.org/10.5705/ss.202018.0403

EXCHANGEABLE MARKOV MULTI-STATE SURVIVAL PROCESSES

Walter Dempsey

University of Michigan

Abstract: We consider *exchangeable Markov multi-state survival processes*, which are temporal processes taking values over a state-space $\mathcal{S}$, with at least one absorbing failure state $\flat \in \mathcal{S}$ that satisfy the natural invariance properties of exchangeability and consistency under subsampling. The set of processes contains many well-known examples from health and epidemiology including survival, illness-death, competing risk, and comorbidity processes. Here, an extension leads to recurrent event processes.

We characterize exchangeable Markov multi-state survival processes in both discrete and continuous time. Statistical considerations impose natural constraints on the space of models appropriate for applied work. In particular, we describe constraints arising from the notion of *composable systems*. We end with an application to irregularly sampled and potentially censored multi-state survival data, developing a Markov chain Monte Carlo algorithm for inference.

Key words and phrases: Composable systems, exchangeability, Markov chain Monte Carlo, Markov process, multi-state survival process.

1. Introduction

In many clinical survival studies, a patient's health status is monitored intermittently until either an event of interest (e.g., failure) or the end of the study window. In a simple survival study, a person's health status $Y(t)$ at time t is a binary variable, namely, dead (0) or alive (1). In clinical trials with health monitoring, $Y(t)$ is a more detailed description of the individual's state of health, containing relevant patient information such as pulse rate, cholesterol level, cognitive score, or CD4 cell count (Diggle, Sousa and Chetwynd (2008); Farewell and Henderson (2010); Kurland et al. (2009)).

In this study, we examine health processes that take values in some prespecified "state-space." For example, in the illness-death model, the participant's current state takes one of three possible values {Healthy, Unhealthy, Dead}. Such a process can be thought of as a coarse view of a patient's state of health over

Corresponding author: Walter Dempsey, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA. E-mail: wdem@umich.edu.

time. When no baseline covariates are measured beyond the initial state $Y(0)$, the model for the set of patient state-space processes should satisfy natural constraints. First, the model should be agnostic to patient labeling. Second, the model should be agnostic to sample size considerations. These natural constraints (mathematically defined in Section 2) lead to *exchangeable Markov multi-state survival processes*. The purpose of this study is to characterize this set of processes and to show how the theory of exchangeable stochastic processes fits naturally into the applied framework of an event-history analysis. Both the parametric continuous-time Markov process with independent participants and the nonparametric counting process are included as special cases. Next, we discuss the notion of "composable systems" and its effect on model specification. A Markov chain Monte Carlo (MCMC) algorithm is then derived for posterior computations, given irregularly sampled multi-state survival data. We end with an application to a cardiac allograft vasculopathy (CAV) multi-state survival study.

1.1. Related work

Odd Aalen was one of the first to recognize the importance of incorporating the "theory of stochastic processes" into an "applied framework of event history analysis" Aalen, Borgan and Gjessing (2008, p.457). Martingales and counting processes form the basis of this nonparametric approach. Nonparametric methods, however, do not adequately handle intermittent observations. For example, Aalen et al. (2015) consider a dynamic path analysis for a liver cirrhosis data set. In this study, the prothrombin index, a composite blood coagulation index related to liver function, is measured initially at three-month intervals, and subsequently at roughly 12-month intervals. To deal with the intermittency of observation times, Aalen et al. (2015) use the "last-observation carried forward" (LOCF) assumption. However, such an assumption is unsatisfactory for highly variable health processes, and can lead to biased estimates (Little et al. (2012)).

One alternative is to consider parametric models such as continuous-time Markov processes. Prior work (Saeedi and Bouchard-Côté (2011); Hajiaghayi et al. (2014); Rao and Teh (2013)) has focused on estimating parametric continuous-time Markov processes under intermittent observations. Most parametric models, however, make strong assumptions about the underlying state-space process; in particular, most models assume independence among patients. One implication of this is that observing sharp changes in health in prior patient trajectories at a particular time since recruitment does not impact the likelihood of a similar sharp change in a future patient at the same timepoint. The proposed approach balances the nonparametric and parametric approaches.

2. Multi-State Survival Models

In this section, we formally define the multi-state survival process and the notions of exchangeability, Kolmogorov-consistency, and the Markov property. We combine these in Section 3 to provide characterization theorems for these processes in discrete and continuous time.

2.1. Multi-state survival process

Formally, the *multi-state survival process*, $\mathbf{Y}$, is a function from the label set $\mathbb{N} \times \mathcal{T}$ into the state space $\mathcal{S}$. For now, we assume the cardinality is finite (i.e., $|\mathcal{S}| < \infty$). If the response is in *discrete time*, then the process is defined on $\mathcal{T} = \mathbb{N}$. If the response is in *continuous time*, then the process is defined on $\mathcal{T} = \mathbb{R}^+$. Each label is a pair (u, t), and the value $\mathbf{Y}(u, t)$ is an element of $\mathcal{S}$ corresponding to the state of patient u at time t.

The distinguishing characteristic of survival processes is *flatlining* (Dempsey and McCullagh (2018)); that is, there exists an absorbing state $\flat \in \mathcal{S}$ such that $Y(u, t) = \flat$ implies $Y(u, t') = \flat$, for all $t' > t$. Thus, the survival time T_u for unit u is a deterministic function of the multi-state survival process $\mathbf{Y}$:

$$T_u = \inf\{t \geq 0 : Y(u, t) = \flat\}.$$

For all $u \in \mathbb{N}$, we assume $Y(u, 0) \neq \flat$ at recruitment; thus, $T_u > 0$. Multiple absorbing states $\{\flat_c\}$ representing different terminal events may occur.

Without loss of generality, we assume $\mathcal{S} = \{1, \ldots, s\} =: [s]$. For example, if the state-space is $\mathcal{S} = \{\text{Alive}, \text{Dead}\}$, we recode this to $[2] = \{1, 2\}$. At each time t, the population-level process is given by $\mathbf{Y}(t) = \{Y(u, t) \,|\, u \in \mathbb{N}\} \in [s]^{\mathbb{N}}$. We write $\mathbf{y}$ to denote a generic element of $[s]^{\mathbb{N}}$. We write $\mathbf{Y}_A$ to denote the restriction of the state-space process to $u \in A \subset \mathbb{N}$. We call $\mathbf{Y}_{[n]}$ the *n-restricted state-space process* for $[n] := \{1, \ldots, n\}$. We write $\mathbf{y}_{[n]}$ to denote a generic element of $[s]^n$. Finally, for $\mathbf{y}_{[n]}$ we define an associated vector $\mathbf{x}_{[n]}$, called the *configuration vector*.

Definition 1 (Configuration vector)**.** For $\mathbf{y}_{[n]} \in [s]^n$, define $\mathbf{x}_{[n]} \in [n]^s$ as the *configuration vector*, an s-vector summary of the number of units in each state. For example, if $s = 2$, $n = 4$, and $\mathbf{y}_{[4]} = (1, 2, 2, 1)$, then $\mathbf{x}_{[4]} = (2, 2)$; for $\mathbf{y}_{[4]} = (1, 1, 2, 1)$, $\mathbf{x}_{[4]} = (3, 1)$. We write x_i to denote the ith entry of $\mathbf{x}_{[n]}$.

Example 1 ((Bidirectional) illness-death process)**.** To make these ideas concrete, we employ the illness-death process as the running example in this paper. The illness-death process has state space {Healthy, Unhealthy, Dead}, with transi-

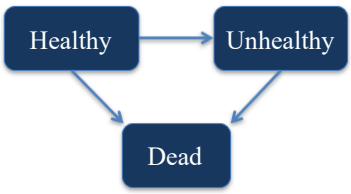


Figure 1. Graph representation of the illness-death process.

tions governed by the graph shown in Figure 1. The *bi-directional* illness-death process includes an additional edge (Unhealthy, Healthy), allowing the patient to recover. The state "Dead" ($s = 3$) is absorbing. Both processes can be viewed as refinements of the survival process. We highlight many other examples in the Supplementary Material, Section S3.

2.2. Transition graph for a multi-state survival process

The *transition graph* associated with a particular multi-state survival process is a directed graph $G = (V, E)$, which represents the set of potential transitions among the states $i \in [s]$. The vertex set $V = [s]$ is the set of all states; the directed edge set E contains all edges (i, i') such that, at jump times, the process can jump from i to i'. In Example 1, a patient can jump from *Healthy* to *Unhealthy*, but not back; therefore, (Healthy, Unhealthy) is in the edge set, but (Unhealthy, Healthy) is not. In the bi-directional case, both edges are present in the transition graph. In continuous-time, jumps can only occur between distinct states, so $(i, i) \notin E$, for all $i \in V$. An absorbing state $i \in [s]$ satisfies $(i, i') \notin E$, for all $i' \neq i \in [s]$. We write $\mathcal{P}_G$ to denote the set of s by s transition matrices P satisfying $\sum_{i' \in V} P_{i,i'} = 1$, $P_{i,i'} \geq 0$ for all $i, i' \in V$, and $P_{i,i'} = 0$ for all $(i, i') \notin E$. In the continuous-time setting, define $P_{i,i} = 1 - \sum_{i',(i,i') \in E} P_{i,i'}$.

2.3. Consistency under subsampling

Statistical models for multi-state survival processes should be agnostic to sample size, because this is often an arbitrary choice based on power considerations and/or patient recruitment constraints. Informally, observing n units should be equivalent to observing $n + 1$ units and then restricting to the first n units; that is, the model should exhibit *consistency under subsampling*.

Consider the multi-state survival process $\mathbf{Y}_{[m]}$ for $m > n$. Define the restriction operator $R_{m,n}$ as the restriction of $\mathbf{Y}_{[m]}$ to the first n individuals. Then, the process is *consistent under subsampling* if $R_{m,n}(\mathbf{Y}_{[m]})$ is equivalent in distribu-

tion to $\mathbf{Y}_{[n]}$, for all $[m] \supset [n]$. That is, for any finite subset $\mathbf{t} := (t_1, \ldots, t_k) \subset \mathcal{T}$, $R_{m,n}\left(\mathbf{Y}_{[m]}(\mathbf{t})\right) =_{\mathcal{D}} \mathbf{Y}_{[n]}(\mathbf{t})$. Stated another way, $\text{pr}(\mathbf{Y}_{[n]} \in A) = \text{pr}(\mathbf{Y}_{[m]} \in R_{m,n}^{-1}(A))$, for any Borel measurable set A.

Under the consistency assumption, the process $\mathbf{Y}_{[n]}$ satisfies *lack of interference*; mathematically,

$$\text{pr}(\mathbf{Y}_{[n]} \in A \,|\, H_{[m]}(t)) = \text{pr}(\mathbf{Y}_{[n]} \in A \,|\, H_{[n]}(t)),$$

where $H_{[l]}(t)$ is the σ-field generated by the variables $Y(u, t')$, for $i \in [l]$ and $t' \leq t$. Lack of interference is essential, because it ensures the n-restricted multi-state survival process is unaffected by the multi-state survival processes of subsequent components. Consistency under subsampling ensures the statistical models are embedded in suitable structures that permit extrapolation.

2.4. Exchangeability

Given no covariates, patient labeling is also an arbitrary choice to which any suitable multi-state survival process must be agnostic. Define a multi-state survival process $\mathbf{Y}$ as *[partially] exchangeable* if for any permutation $\sigma : [n] \to [n]$, the relabeled process $\mathbf{Y}_{[n]}^{\sigma} = \{Y(\sigma(1), t), \ldots, Y(\sigma(n), t) \,|\, t \in \mathcal{T}\}$ is equivalent in distribution to $\mathbf{Y}_{[n]}$. That is, for any finite subset $\mathbf{t} \subset \mathcal{T}$, $\mathbf{Y}_{[n]}^{\sigma}(\mathbf{t}) =_{\mathcal{D}} \mathbf{Y}_{[n]}(\mathbf{t})$.

2.5. Time-homogeneous Markov process

$\mathbf{Y}_{[n]}$ is a *time-homogeneous Markov process* if, for every $t, t' \geq 0$, the conditional distribution of $\mathbf{Y}_{[n]}(t+t')$, given the multi-state survival process history up to time t, $\mathcal{H}_{[n]}(t)$, depends only on $\mathbf{Y}_{[n]}(t)$ and t'. This Markovian assumption is a simplifying assumption that leads to mathematically tractable conclusions. Here, we restrict our attention to time-homogeneous processes; therefore, we simply say $\mathbf{Y}_{[n]}$ is Markovian.

3. Exchangeable Markov Multi-State Survival Processes

Define a multi-state survival process that is Markovian, exchangeable, and consistent under subsampling as an *exchangeable Markov multi-state survival process*. We next characterize these processes in both discrete and continuous time. The behavior is markedly different in each setting, showing why the choice of time scale matters in applied settings. All proofs are left to the Supplementary Materials.

3.1. Discrete-time multi-state survival models

In discrete-time, the exchangeable Markov multi-state survival process is governed by a series of random transition matrices P_t, each drawn independently from a probability measure Σ on $\mathcal{P}_G$. The initial state $\mathbf{Y}(0)$ is drawn from an exchangeable distribution on $[s]$. Then, at time t, the transition distributions for each $u \in \mathbb{N}$ are given by

$$\text{pr}(Y(u,t) = i' \,|\, Y(u,t-1) = i) \sim [P_t]_{i,i'},$$

that is, the (i,i') entry of P_t. Let $\mathbf{Y}^\star_\Sigma$ denote a discrete-time process constructed using this procedure, with probability measure Σ. By construction, the process is an exchangeable, Markov multi-state survival process in discrete time. Theorem 1 states that this procedure describes all such processes.

Theorem 1 (Discrete-time characterization)**.** *Let* $\mathbf{Y} = \{Y(u,t), u \in \mathbb{N}, t \in \mathbb{N}\}$ *be an exchangeable Markov multi-state survival process. Then, there exists a probability measure* Σ *on* $\mathcal{P}_G$ *such that* $\mathbf{Y}^\star_\Sigma$ *is a version of* $\mathbf{Y}$*.*

Example 1 (cont.)**.** We provide an illustrative construction for example 1. First, assume all units start in the state "Healthy". Next, for each $t \in \mathbb{N}$, define a set $\{Z^{(t)}_{i,i'}\}$ of independent beta random variables, with parameters $\gamma\alpha_{i,i'}$ and $\gamma\beta_{i,i'}$, respectively. Set $[P_t]_{1,2} = Z_{1,2}$, $[P_t]_{1,3} = (1 - Z_{1,2}) \times Z_{1,3}$, $[P_t]_{1,1} = 1 - [P_t]_{1,2} - [P_t]_{1,3}$, $[P_t]_{2,3} = Z_{2,3}$, $[P_t]_{2,2} = 1 - [P_t]_{2,3}$, and $[P_t]_{3,3} = 1$. The random matrix P_t governs the illness-death process at time t. We show simulation results for $t = 1, 2, 3$, with $\alpha_{i,i'} = \beta_{i,i'} = 1$, for all i, i'. For $\gamma = 100$, we simulate

$$P_1 = \begin{bmatrix} 0.26 & 0.51 & 0.23 \\ 0.00 & 0.54 & 0.46 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 0.25 & 0.50 & 0.26 \\ 0.00 & 0.54 & 0.46 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}, \quad P_3 = \begin{bmatrix} 0.25 & 0.48 & 0.27 \\ 0.00 & 0.50 & 0.50 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}.$$

For $\gamma = 1$ and the same random seed, we simulate

$$P_1 = \begin{bmatrix} 0.13 & 0.45 & 0.43 \\ 0.00 & 0.16 & 0.84 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 0.29 & 0.52 & 0.20 \\ 0.00 & 0.19 & 0.81 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}, \quad P_3 = \begin{bmatrix} 0.01 & 0.69 & 0.30 \\ 0.00 & 0.62 & 0.38 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}.$$

As $\gamma \to \infty$, P_t converges to a deterministic matrix $P^\star$. In this limiting case, $Y(u,\cdot)$ evolves as an *independent* Markov process with transition matrix $P^\star$.

3.2. Continuous-time multi-state survival models

In continuous time, the exchangeable Markov multi-state survival process is governed by a measure on the transition matrices, denoted by Σ, and a set of constants associated with the edge set, denoted by $\mathbf{c} = \{c_{i,i'} \,|\, (i,i') \in E\}$. Unlike the discrete time case, transitions occur at random times (called jump times).

For the n-restricted process $\mathbf{Y}_{[n]}$, the holding time in any state $\mathbf{y}_{[n]}$ is exponentially distributed, with a rate parameter that depends on the number of units in each state, that is, the current *configuration* $\mathbf{x}_{[n]}$ (see Definition 1). At jump time t, one of two events can occur: (a) a *single unit* $u \in [n]$ experiences a transition; or (b) a *subset of* $[n]$ (potentially a singleton) experiences a simultaneous transition. If the jump time is of type (a), a state $i \in [s]$ is chosen with probability proportional to $x_i \cdot c_{i,\bullet}$, where $c_{i,\bullet} := \sum_{i':(i,i')\in E} c_{i,i'}$. Among the x_i units satisfying $Y(u, t-) = i$, for $u \in [n]$, choose one at random, and transition that unit to state $i' \in [s]$, such that $(i,i') \in E$ with probability proportional $c_{i,i'}/c_{i,\bullet}$. If the jump time is of type (b), a transition matrix $P(t)$ is obtained from a measure Σ on $\mathcal{P}_G$. Given $P(t)$, all units transition according to $P(t) \in \mathcal{P}_G$ under the constraint that at least one unit transitions to a new state. Unlike the discrete-time setting, the measure Σ need not be integrable.

The above procedure provides a high-level construction of a continuous-time process $\mathbf{Y}^{\star}_{\Sigma,c}$. In Section 5, a detailed version of this procedure is given. Theorem 2 states that this procedure describes all such processes.

Theorem 2 (Continuous-time characterization)**.** *Let* $\mathbf{Y} = (\mathbf{Y}(t), t \in \mathbb{R}^+)$ *be an exchangeable Markov multi-state survival process; and let* I_s *be the* $s \times s$ *identity matrix. Then, there exists a probability measure* Σ *on* $\mathcal{P}_G$ *satisfying*

$$\Sigma(\{I_s\}) = 0 \text{ and } \int_{\mathcal{P}_G} (1 - P_{\min})\Sigma(dP) < \infty, \quad P_{\min} = \min_{i\in[s]} P_{i,i} \tag{3.1}$$

and constants $\mathbf{c} = \{c_{i,i'} \geq 0 \,|\, (i,i') \in E\}$*, such that* $\mathbf{Y}^{\star}_{\Sigma,c}$ *is a version of* $\mathbf{Y}$*.*

Theorem 2 generalizes Proposition 4.3 in Dempsey and McCullagh (2017) from the survival setting. The result contains many well-known examples from health and epidemiology including survival, illness-death, competing risk, and comorbidity processes.

The procedure defined in Section 5 characterizes $\mathbf{Y}_{[n]}$ in terms of (1) exponential holding rates, and (2) the transition matrix at the jump times. In order to define this procedure formally, we require an additional definition.

Definition 2 (Characteristic index)**.** For any $\mathbf{y}_{[n]} \in [s]^n$, the *characteristic index*,

denoted by $\zeta(\mathbf{y}_{[n]})$, is defined as

$$\zeta_n(\mathbf{y}_{[n]}) = \int_{\mathcal{P}_G} \left(1 - \prod_{j=1}^{s} P_{j,j}^{x_j}\right) \Sigma(dP) + \sum_{i\in[s]} x_i \sum_{i':(i,i')\in E} c_{i,i'},$$

where the sum is set to zero when $\{i' \in V \text{ s.t. } (i,i') \in E\} = \emptyset$. Condition (3.1) implies the characteristic index is finite, for any $\mathbf{y}_{[n]} \in [s]^n$.

At a jump time t, let $A_{t,[n]} \subseteq [n]$ denote the subset of units that experience a transition. If $|A_{t,[n]}| > 1$, then the jump is a transition of type (b), governed by Σ. If $|A_{t,[n]}| = 1$, then the jump *may* be a transition of type (a) *or* (b) and, therefore, will depend on *both* $\mathbf{c}$ and Σ. The transition probability from $\mathbf{Y}_{[n]}(t-)$ to $\mathbf{Y}_{[n]}(t)$, denoted $q(\mathbf{Y}_{[n]}(t-), \mathbf{Y}_{[n]}(t))$, is equal to

$$\frac{1}{\zeta_n(\mathbf{Y}_{[n]}(t-))} \Bigg[\underbrace{\int_{\mathcal{P}_G} \prod_{u\in[n]} P[Y(u,t-), Y(u,t)] \, \Sigma(dP)}_{\text{Term 1}} + \underbrace{\delta(|A_{t,[n]}| = 1) \sum_{u\in[n]} \delta(u \in A_{t,[n]}) \sum_{i':(i,i')\in E} c_{i,i'} \, \delta(Y(u,t-) = i, Y(u,t) = i')}_{\text{Term 2}} \Bigg]$$
$$=: \frac{\lambda(\mathbf{Y}_{[n]}(t-), \mathbf{Y}_{[n]}(t))}{\zeta_n(\mathbf{Y}_{[n]}(t-))},$$

where $P[i,i'] = P_{i,i'}$, $\delta(\cdot)$ is the indicator function, and $\lambda(\cdot,\cdot)$ is the non-normalized transition function from $[s]^n \times [s]^n \to \mathbb{R}_+$. Term 1 depends on the measure Σ, and is associated with a positive fraction of the *population* transitioning at time t according to $P \sim \Sigma$. Term 2 depends on the constants $\mathbf{c}$, and is associated with the single unit $u \in A_{t,[n]}$ in state i transitioning to state i' with probability proportional to $c_{i,i'}$. A critical question is how to specify Σ and $\mathbf{c}$. In applied work, because single-unit transitions are unrelated to the rest of the population, we recommend setting $\mathbf{c} \equiv 0$; we discuss an appropriate family of parametrized measures Σ_Ψ in Section 5.4.

4. Discretization and Rounding

It has been argued that "there may be no scientific reason to prefer a true continuous time model over a fine discretization" (Breto et al., 2009, p. 325). We tend to disagree with this viewpoint; a basic and very important issue in

multi-state survival analysis is the distinction between inherently discrete data (coming from intrinsically time-discrete phenomena) and grouped data (coming from the rounding of intrinsically continuous phenomena). Theorems 1 and 2 supplement this scientific distinction with a mathematical one, because we see distinct characterizations of discrete- and continuous-time processes. One example of the former in survival analysis is the time taken to get pregnant, which should be measured in menstrual cycles. Continuous-time processes represent the majority of multi-state survival data. For this reason, we focus the remainder of this paper on the continuous-time case.

5. Description of Continuous-Time Process

5.1. Holding times

Let t be the jump time at which the state vector $\mathbf{Y}_{[n]}$ transitions into state $\mathbf{y}_{[n]} \in [s]^n$. To each such state $\mathbf{y}_{[n]}$, we associate an independent exponentially distributed holding time. By choosing the rate functions in an appropriate way, the Markov multi-state survival process can be made both consistent under subsampling and exchangeable under permutation of units.

Corollary 1. *A set of rate functions* $\{\tau_n : [s]^n \to \mathbb{R}^+\}_{n=1}^{\infty}$ *is consistent if it is* proportional to *the characteristic index* $\tau_n(\mathbf{y}_{[n]}) \propto \zeta_n(\mathbf{y}_{[n]})$.

Corollary 1 follows from Theorem 2, and shows how the exponential holding rate relates to the characteristic index; in particular, the difference is a proportionality constant ν, which depends on the choice of time scale.

5.2. Density function

Because the evolution of the process $\mathbf{Y}_{[n]}$ is Markovian, providing an expression for the probability density function for any specific temporal trajectory is straightforward. The probability that the first transition occurs in the interval dt_1, with transition from $\mathbf{Y}_{[n]}(t_1-)$ to $\mathbf{Y}_{[n]}(t_1)$, is

$$\begin{aligned}&\nu\zeta_n(\mathbf{Y}_{[n]}(t_1-)) \exp\left(-\nu\zeta_n(\mathbf{Y}_{[n]}(t_1-))t_1\right) dt_1 \times q(\mathbf{Y}_{[n]}(t_1-), \mathbf{y}_{[n]}(t_1))\\ &= \exp\left(-\nu\zeta_n(\mathbf{Y}_{[n]}(t_1-))t_1\right) dt_1 \times \lambda(\mathbf{Y}_{[n]}(t_1-), \mathbf{Y}_{[n]}(t_1)),\end{aligned}$$

where $\lambda(\cdot,\cdot)$ denotes the non-normalized transition probabilities. Continuing in this way, the joint density for a particular temporal trajectory $\mathbf{Y}_{[n]}$ consisting of k transitions with jump times $0 < t_1 < \cdots < t_k$ is

$$\exp\left(-\int_0^\infty \nu\zeta_n(\mathbf{Y}_{[n]}(s))\,ds\right)\prod_{j=1}^{k}\lambda\big(\mathbf{Y}_{[n]}(t_j-),\mathbf{Y}_{[n]}(t_j)\big). \tag{5.1}$$

The number of transitions k is a random variable, the distribution of which is determined by (5.1), and hence by ζ_n.

Although the argument leading to (5.1) did not explicitly consider censoring, the density function has been expressed in integral form so that censoring is accommodated correctly. The pattern of censoring affects the evolution of $\mathbf{Y}_{[n]}$, and thus affects the integral, but the product involves only transitions and transition times. As long as the censoring mechanism is *exchangeability preserving* (Dempsey and McCullagh (2017)), inference based on the joint density given by equation (5.1) is possible. Simple type I censoring and independent censoring mechanisms both preserve exchangeability.

5.3. Sequential description

Kolmogorov consistency permits ease of computation for the trajectory of a new unit $u' = n+1$, given trajectories for the first n units $\mathbf{Y}_{[n]} = \mathbf{y}_{[n]}$. The conditional distribution is best described using a set of paired measures consisting of a continuous component $\Lambda^{(c)}_{i,i'}$ and an atomic measure $\Lambda^{(a)}_{i,i'}$, with positive mass only at the observed transition times of $\mathbf{y}_{[n]}$, for $(i,i') \in E$.

For a time t, not a transition time of $\mathbf{y}_{[n]}$, consider the new unit transitioning from state i to i'; that is, $\mathbf{y}_{[n+1]}(t-) = (\mathbf{y}_{[n]}(t-), i)$ and $\mathbf{y}_{[n+1]}(t) = (\mathbf{y}_{[n]}(t), i')$. The continuous component has hazard and cumulative hazard

$$h_{i,i'}(t) = \lambda(\mathbf{y}_{[n+1]}(t-), \mathbf{y}_{[n+1]}(t)) \quad \text{and} \quad H_{i,i'}(t) = \int_0^t h_{i,i'}(s)ds,$$

respectively. The non-normalized transitions $\lambda(\cdot,\cdot)$ are piecewise constant as a function of t. Therefore, so the integral is trivial to compute. However, censoring implies it is not necessarily constant between transition times.

Now, let t be an observed transition time (i.e., $\mathbf{y}_{[n]}(t-) \neq \mathbf{y}_{[n]}(t)$) and consider the atomic measure $\Lambda^{(a)}_{i,i'}$ associated with switching from state i to i'. At each such point, the conditional hazard has an atom with finite mass

$$\Lambda^{(a)}_{i,i'}(\{t\}) = \log\frac{\zeta_n(\mathbf{y}_{[n]}(t-))\,q(\mathbf{y}_{[n]}(t-),\mathbf{y}_{[n]}(t))}{\zeta_{n+1}(\mathbf{y}_{[n+1]}(t-))\,q(\mathbf{y}_{[n+1]}(t-),\mathbf{y}_{[n+1]}(t))},$$

or, on the probability scale,

$$\exp(-\Lambda_{i,i'}^{(a)}(\{t\})) = \frac{\lambda(\mathbf{y}_{[n+1]}(t-), \mathbf{y}_{[n+1]}(t))}{\lambda(\mathbf{y}_{[n]}(t-), \mathbf{y}_{[n]}(t))}.$$

The above calculations define the conditional holding time of the new unit after it enters state i at time t (i.e., $Y(n+1, t-) \neq i$ and $Y(n+1, t) = i$), conditional on $\mathbf{Y}_{[n]} = \mathbf{y}_{[n]}$. For $s > 0$, let $\{t_j\}_{j=1}^{L}$ denote the observed transition times of $\mathbf{y}_{[n]}$ within the time window $(t, t+s]$. The probability that the unit stays in state i for *at least* $s > 0$ time points is

$$\exp\left(-\sum_{i':(i,i')\in E} \nu(H_{i,i'}(t+s) - H_{i,i'}(t))\right) \cdot \prod_{j=1}^{L} \exp\left(-\sum_{i':(i,i')\in E} \Lambda_{i,i'}^{(a)}(\{t_j\})\right),$$

which serves as the basis for the proposed MCMC procedure in Section 7.

5.4. Self-similar harmonic process

Theorem 2 implies tied failures are an intrinsic aspect of Markov multi-state survival processes. However, grouped data are often the result of the rounding of intrinsically continuous data. For these models to be useful in biomedical applications, it is essential that they not be sensitive to rounding. This has been addressed previously by restricting attention to processes with conditional distributions that are *weakly continuous*; that is, small perturbations of transition times imply small perturbations of the conditional distributions.

Dempsey and McCullagh (2017) originally studied this question in the context of exchangeable Markov survival processes. In particular, they show that the *harmonic process* is the only Markov survival process with weakly continuous conditional distributions. Here, we extend the harmonic process to a multi-state survival process by associating with each edge $(i, i') \in E$ with an independent harmonic process with parameters $(\nu_{i,i'}, \rho_{i,i'})$. For $(i, i') \in E$, let $t_1^{(i,i')} < \cdots < t_{k(i,i')}^{(i,i')}$ denote the unique observed transition times from i to i' for $\mathbf{Y}_{[n]}$, and let $\mathbf{Y}_{[n]}^{\sharp}(t; i) = \#\{u \in [n] \text{ s.t. } Y_u(t) = i\}$; then, the continuous component of the hazard is given by

$$H_{i,i'}(t) = \sum_{l:t_l^{(i,i')}\leq t} \nu_{i,i'} \frac{t_l^{(i,i')} - t_{l-1}^{(i,i')}}{\mathbf{Y}_{[n]}^{\sharp}(t_{l-1}^{(i,i')}; i) + \rho_{i,i'}} + \nu_{i,i'} \frac{t - t_m^{(i,i')}}{\mathbf{Y}_{[n]}^{\sharp}(t_m^{(i,i')}; i) + \rho_{i,i'}},$$

where the sum runs over the transition times $t_l^{(i,i')} \leq t$, and $t_m^{(i,i')}$ is the last such event. The discrete component is a product over the transition times,

$$\prod_{l:t_l^{(i,i')}\leq t} \frac{\mathbf{Y}^{\sharp}_{[n]}(t;i)+\rho_{i,i'}}{\mathbf{Y}^{\sharp}_{[n]}(t-;i)+\rho_{i,i'}}. \tag{5.2}$$

For small $\{\rho_{i,i'}\}_{(i,i')\in E}$, the combined discrete components are essentially the same as those of the right-continuous version of the Aalen–Johansen estimator.

We call this process the *self-similar harmonic process with transition graph* G. The associated measure Σ on $\mathcal{P}_G$ is

$$\Sigma(dP) = \delta[\#\{p_{i,i'}>0,(i,i')\in E\}=1]\,\nu_\star(1-p_\star)^{-1}p_\star^{\rho_\star-1}dp_\star,$$

where $p_\star$ is the single nonzero, off-diagonal entry, $\delta[\cdot]$ is the indicator function, and $(\nu_\star,\rho_\star)$ are the associated parameters.

While the self-similar harmonic process has strong appeal for use in applied work, we argue it is not universally optimal. The independence assumption implies that at each transition time, only transitions along a single edge $(i,i')\in E$ are possible. While this may make sense in specific cases, additional care is needed when expressing appropriate models in general.

6. Composable Multi-State Survival Models

We now discuss constraints on the multi-state survival models based on decompositions of the state-space $[s]$. We start with the illness-death process (Example 1) as our illustrative example. The state "Dead" is unique, while the states "Healthy" and "Unhealthy" both require the individual to be categorized more broadly as alive. Suppose the labels "Healthy" and "Unhealthy" are uninformative with respect to failure transitions. Then, the refinement is immaterial, and the transition rules should collapse to the transition rule for an exchangeable Markov survival process.

This leads to two natural constraints: (1) state "Dead" is distinct; and (2) states "Healthy" and "Ill" should be considered *partially exchangeable* (De Finetti (1972)). To satisfy this, we constrain the measure Σ to take only positive mass on one of two sets of transition matrices: (I) P, with off-diagonal positive mass in entries $(1,3)$ and/or $(2,3)$ only; and (II) P, with off-diagonal positive mass in entries $(1,2)$ and/or $(2,1)$ only. The first represent transitions from "Healthy" or "Ill" to "Dead." The second represent transitions between "Healthy" and "Ill." The partition $B=\{B_1,B_2\}=\{\{1,2\},\{3\}\}$ splits the state space. We then say the process is partially exchangeable with respect to the partition B.

Let (n_1,n_2) denote the number of individuals in states "Healthy" and "Unhealthy" directly preceding a transition to state "Dead." Then, the probability

that $d_1 \leq n_1$ and $d_2 \leq n_2$ individuals, respectively, transition is proportional to $\int p_{1,1}^{n_1-d_1} p_{1,3}^{d_1} p_{2,2}^{n_1-d_1} p_{2,3}^{d_2} \tilde{\Sigma}(dP)$, where $\tilde{\Sigma}$ is the measure Σ restricted to type-(I) transition matrices; that is, $\tilde{\Sigma}$ puts positive mass on transition matrices P such that $P_{1,2} = P_{2,1} = 0$; thus $P_{k,k} = 1 - P_{k,3}$ for $k = 1, 2$, and $P_{3,3} = 1$. Here, we restrict our attention to measures of the form

$$\tilde{\Sigma}(dP) = \nu_{1,1} \cdot P_{1,1}^{\rho_{1,1}-1}(1 - P_{1,1})^{-1}\delta(P_{2,2}^{\gamma} = P_{1,1})dP_{1,1}dP_{2,2}. \tag{6.1}$$

Equation (6.1) corresponds to a proportional model on the logarithmic scale, linking $P_{1,1}$ and $P_{2,2}$ via a baseline measure for a harmonic process. Details on the connection to the proportional conditional hazards model are provided in the Supplementary Material, Section S4.

6.1. Composable multi-state survival process

We now generalize the above by introducing B-composable processes.

Definition 3. A multi-state survival process is *B-composable* if there exists a partition $B = (B_1, \ldots, B_k)$ of the state-space $[s]$, such that elements within block B_i are *partially exchangeable* with respect to transition graph G.

Definition 3 is similar in spirit to that of Schweder (2007)—both aim to formalize the notion that state changes in the process $\mathbf{Y}$ are due to changes in different components. For the bi-directional illness-death process (example 1), $B = (\{1, 2\}, \{3\})$. For comorbidities (example S4), B partitions the risk processes. For competing risks (example S4), $B = (\{1\}, B_2, \ldots, B_k)$, where $(B_2, \ldots, B_k)$ partition the absorbing states, and the single state "Alive" is distinct, which implies $B_1 = \{1\}$. If $\mathbf{Y}$ is B'-composable and B' is a refinement of the partition B, then $\mathbf{Y}$ is also B-composable. To avoid confusion, from here on, when we say $\mathbf{Y}$ is B-composable, we assume no refinement of B' exists such that $\mathbf{Y}$ is also B'-composable.

6.2. Choice of measure for a composable process

Here, we construct an appropriate measure Σ for a B-composable exchangeable Markovian multi-state survival process. The measure will take positive mass only on transitions from states within B_j to states within $B_{j'}$, for a single choice of $j, j' \in \{1, \ldots, k\} := [k]$ indexing components of the partition B. For each component B_j, let $i(j) \in [s]$ denote a *representative state*. Then, for $j, j' \in [k]$, define the restricted measure on transitions from states in B_j to states in $B_{j'}$, $\tilde{\Sigma}_{jj'}(dP)$, by

$$\nu_{j,j'} P_{i(j),i(j)}^{\rho_{j,j'}-1} \left(1 - P_{i(j),i(j)}\right)^{-1} dP_{i(j),i(j)} \tag{6.2}$$

$$\prod_{l \in B_j \setminus i(j)} \delta\left[P_{l,l}^{\gamma_{l,j'}} = P_{i(j),i(j')}\right] dP_{l,l} \tag{6.3}$$

$$\prod_{l \in B_j} \prod_{m \in B_{j'} : (l,m) \in E} \delta[y_{l,m} = \alpha_{l,m}] dy_{l,m}, \tag{6.4}$$

where $\gamma_{l,j'} > 0$, $\gamma_{i(j),j'} = 1$, $\alpha_{l,m} \in [0,1]$, $P_{l,m} \in [0,1]$, and $y_{l,m} \in [0,1]$, such that $\sum_{m \in B_{j'} : (l,m) \in E} y_{l,m} = 1$. Here, the assumption is $P_{l,m} = (1 - P_{l,l}) \cdot y_{l,m}$, for $l \neq m$, and $P_{l,l} = P_{i(j),i(j)}^{\gamma_{l,j'}}$. Lines (6.2) and (6.3) build the general measure from a baseline harmonic measure and the assumption of proportionality on the logarithmic scale for $P_{l,l}$, where the proportionality constant depends on $l \in B_j$. Note that $\gamma_{i(j),j'}$ is set to one by design. Therefore, the parameters measure the risk relative to the chosen *representative state*. Line (6.4) addresses the fact that a single state $l \in B_j$ can transition to multiple states in $B_{j'}$, leading to a simple Gibbs update procedure.

7. Parameter Estimation

In practice, a patient's health status is typically measured at recruitment ($t = 0$), and then regularly or intermittently thereafter while the patient is under observation. A complete observation on one patient $(\mathbf{t}, Y(\mathbf{t}), V, \Delta)$ consists of the appointment schedule $\mathbf{t}$, multi-state process measurements $Y(\mathbf{t})$, a failure/censoring time V, and a censoring indicator Δ. For censored records, $\Delta = 1$ and the censoring time V is usually, but not necessarily, equal to the date of the most recent appointment or the end of study.

Here, we assume *non-informative observation times*. In particular, given previous appointment times $\mathbf{t}_{k-1} = (t_1, \ldots, t_{k-1})$ and observation values $Y(\mathbf{t}_{k-1})$, the next appointment time t_k satisfies

$$t_k \perp\!\!\!\perp Y \,|\, (\mathbf{t}_{k-1}, Y(\mathbf{t}_{k-1})). \tag{7.1}$$

That is, the conditional distribution of the random interval $t_k - t_{k-1}$ may depend on the observed history, but not on the subsequent health trajectory.

7.1. The MCMC algorithm

In this section, we derive an MCMC algorithm for posterior computations, given irregularly sampled multi-state survival data under assumption (7.1).

7.1.1. Prior specification and MCMC updates

We start with the prior specification. Let Φ denote the complete set of parameters. We use bar notation (e.g., $\bar{\alpha}$, $\bar{\gamma}$) to denote each subset of parameters. Recall that for identifiability reasons, $\gamma_{i(j),j'} = 1$, for $j, j' \in [k]$. For all other pairs (l, j'), the prior is set to $\log(\gamma_{l,j'}) \sim N(0, 1)$. Weakly informative default priors are an alternative (Gelman et al. (2008)). The complete-data likelihood is non-conjugate, so Metropolis–Hastings updates are performed.

We follow Dempsey and McCullagh (2017) and set $\rho_{j,j'} := \rho$ as a fixed tuning parameter. We define $\lambda_{j,j'} = \nu_{j,j'} \cdot \rho$. Scaling by ρ allows for a direct comparison of $\lambda_{j,j'}$ across various choices of ρ. We set a conjugate Gamma prior $\lambda_{j,j'} \sim \text{Gamma}(\alpha, \beta)$. The posterior distribution, given $\mathbf{Y}_{[n]}, \bar{\gamma}, \bar{\alpha}, \rho$, is

$$\text{Gamma}\left(\alpha + k_{j,j'}, \beta + \rho \cdot \int_0^\infty \zeta_n(Y_{[n]}(s); \bar{\gamma}, \bar{\alpha}, \rho) ds\right), \tag{7.2}$$

where $k_{j,j'}$ is the number of transitions between blocks j and j'.

Finally, for $l \in [s]$, consider transitions to partition $B_{j'}$. We index the states in $B_{j'}$ such that a transition from l is possible by $1, \ldots, m_{l,j'}$. Then, the prior for $\bar{\alpha}_{l,j'} = (\alpha_{l,1,j'}, \ldots, \alpha_{l,m_{l,j'},j'})$ is a Dirichlet distribution with parameters $\bar{p}_{l,j'} = (p_{l,1,j'}, \ldots, p_{l,m_{l,j'},j'})$. Furthermore the posterior is conjugate and

$$\bar{\alpha}_{l,j'} \mid \mathbf{Y}_{[n]}, \bar{\gamma}, \bar{\lambda}, \rho \sim \text{Dir}\left(p_{l,1,j'} + k_{l,1,j'}, \ldots, p_{l,m_l,j'} + k_{l,m_l,j'}\right), \tag{7.3}$$

where $k_{l,m',j'}$ counts the number of transitions from state l to m' in $\mathbf{Y}_{[n]}$.

7.1.2. Conditional sampling patient trajectories

Uniformization (Jensen (1953); Hobolth and Stone (2009)) is a well-known technique for generating sample paths for a Markov state-space process, and is highly adaptable to MCMC. See Rao and Teh (2013) for an excellent discussion of a uniformization-based MCMC. A direct application of existing Gibbs samplers based on uniformization to the current setting is problematic, owing to the combinatorial growth in the state-space and the time-inhomogeneity of the conditional Markov process. Luckily, this approach can be adjusted using the sequential description from Section 5.3.

Here, we show how to adapt the uniformization-based Gibbs sampler to sample a patient trajectory $\mathbf{Y}_i$, given all other trajectories $\mathbf{Y}_{-i} = \mathbf{y}_{-i}$, parameters Φ, and the prior iteration's patient trajectory $\tilde{\mathbf{y}}_i$. For patient i, we observe $(\mathbf{t}_i, \mathbf{Y}_i(\mathbf{t}_i), V_i, \Delta_i)$. The appointment schedule $\mathbf{t}_i$ is an ordered sequence $0 \leq t_{i,0} < \cdots < t_{i,k_i} \leq V_i$. By the Markov property, we can focus on sampling

in each interval $[t_{i,j}, t_{i,j+1}]$ separately. Define the set of transition times $\tilde{\mathbf{t}}_j$, for $t_{i,j} \leq \tilde{t}_{1,j} < \cdots < \tilde{t}_{L_j,j} \leq t_{i,j+1}$, as the unique transition times in $\mathbf{y}_{-i}$ within the interval $[t_{i,j}, t_{i,j+1}]$. At each time $t \in [t_{i,j}, t_{i,j+1}]$, define the piecewise constant function $\Omega_t = C \cdot \max_{(i,i')\in E} |\Lambda^{(c)}_{i,i'}(t)|$, with $C > 1$. Sample a Poisson process $\mathbf{w}_j \subset [t_{i,j}, t_{i,j+1}]$ with piecewise constant rate $R_t = \Omega_t - \Lambda^{(c)}_{\tilde{y}_i(t),\tilde{y}_i(t)}(t)$, where $\Lambda^{(c)}_{\tilde{y}_i(t),\tilde{y}_i(t)}(t)$ is the continuous component of the conditional distribution. Let $\mathbf{u}_{i,j}$ denote the transition times of $\tilde{\mathbf{y}}_i$ in the interval $[t_{i,j}, t_{i,j+1}]$. We then apply the forward-filtering, backward sampling algorithm with transition matrix $B_t = (I + \Lambda^{(c)}/\Omega_t)$ at times $t \in \mathbf{w}_j \cup (\mathbf{u}_{i,j} \setminus \tilde{\mathbf{t}}_j)$, and with the transition matrix $\Lambda^{(a)}$ at times $\tilde{\mathbf{t}}_j$.

7.1.3. MCMC procedure

In each MCMC iteration, we proceed sequentially through the patients, sampling a latent multi-state path for patient $\mathbf{Y}_i$, given all other latent processes $\mathbf{Y}_{-i} := \mathbf{Y}_{[n]\setminus i}$, as described in Section 7.1.2. Conditional on $\mathbf{Y}_{[n]}$, we perform Metropolis Hastings updates for $\bar{\gamma}$. We end each iteration by sampling $\bar{\lambda}$ and $\bar{\alpha}$ using equations (7.2) and (7.3), respectively. One issue with this procedure is that the path sampling is computationally expensive. To address this issue, we propose an approximate MCMC algorithm in which the latent processes are only updated every few iterations. In simulations, the posteriors do not change significantly, but the run time drops significantly. For the sake of conciseness, we provide a simulation study of the MCMC procedure in the Supplementary Material (see Section 1). In the remainder of this paper, we apply the proposed methodology to an irregularly sampled multi-state survival data set.

8. Cardiac Allograft Vasculopathy Case Study

To illustrate our methodology, we use data from angiographic examinations of 622 heart transplant recipients at Patworth Hospital in the United Kingdom. These data were downloaded from the R library `http://cran.r-project.org/web/packages/msm`, maintained by Christopher Jackson. Cardiac allograft vasculopathy (CAV) is a deterioration of the arterial walls. Four states were defined for heart transplant recipients: no CAV ($s = 1$), mild/moderate CAV ($s = 2$), severe CAV ($s = 3$), and dead ($s = 4$). The transition graph is given by Figure 2. Yearly examinations occurred for up to 18 years following a transplant. The mean follow-up time is 5.9 years. For censored records, the censoring time is set equal to the final appointment time. Of the 622 patients, only 192 were observed in state 2 (Mild CAV) at any point during their follow-up. Of these

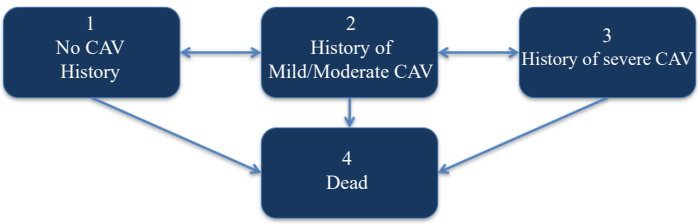


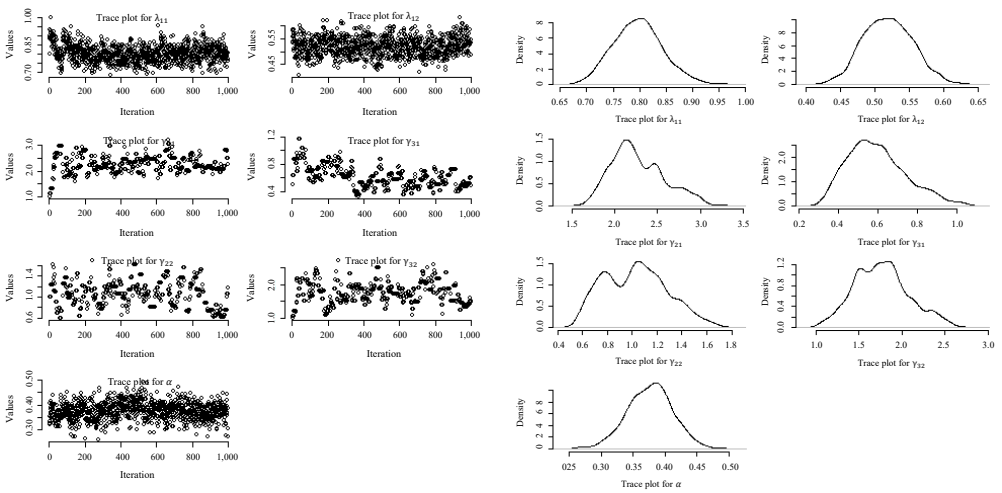
Figure 2. Cardiac allograft vasculopathy (CAV) transition diagram.

192, 43 were subsequently observed in state 1. Only 92 patients were observed in state 3 (Severe CAV) at any point during their follow-up. Of these 92, 12 were subsequently observed in state 2.

We set $B = (B_1 = \{1,2,3\}, B_2 = \{4\})$ and $\rho = 10$. The parameters are $\{\lambda_{(j,j')}\}_{j,j'=1,2}$, $\{\gamma_{l,1}, \gamma_{l,2}\}_{l \in B_1 \setminus 1}$, and $\alpha_{2,1} \in [0,1]$. For identifiability, we set $\gamma_{1,1}$ and $\gamma_{1,2}$ equal to one. Because transitions from state 2 to 1 occur, but should not occur too often, we set the prior on $\alpha = \alpha_{2,1} \sim \text{Beta}(2,8)$. We set λ_{11} and $\lambda_{12} \sim \text{Gamma}(1,1)$. The parameters $(\gamma_{21}, \gamma_{22}, \gamma_{31}, \gamma_{32})$ have independent standard log-normal priors. The approximate MCMC sampler is used.

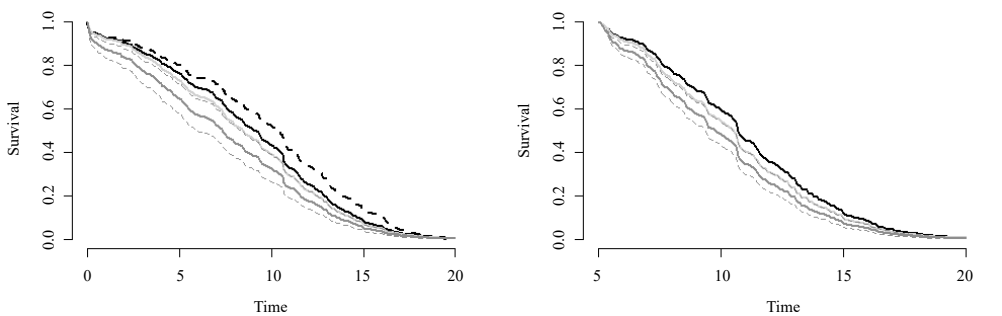
The traceplots in Figure 3a suggest convergence after the first 100 iterations. The posterior mean of λ_{11} is 0.80 (i.e., the marginal time until a transition from state 1 to state 2 is 1.25 years). The parameters $(\gamma_{21}, \gamma_{31})$ have posterior means $(2.27, 0.52)$, translating into marginal holding times of 0.58 and 2.35 years, respectively. The posterior mean for α is 0.38, suggesting that a patient is a bit more likely to experience a progression of his/her CAV status than a regression. The posterior mean for λ_{12} is 0.60 (i.e., marginally, the holding time in state 1 until a transition to state 4 is 1.68 years). This suggests that in state 1, disease progression is slightly more likely than failure. The parameters $(\gamma_{22}, \gamma_{32})$ have posterior means $(1.03, 1.76)$, respectively. This translates marginally into holding times of 1.63 and 0.99 years, respectively. Figure 3b suggests that the failure rate from state 3 is high relative to that from state 1, while the rates from states 1 and 2 are similar.

We next consider the posterior distributions for the survival functions. Figure 4 plots the median survival at each time t over all iterations of the MCMC sampler, Kaplan–Meier survival function estimator, and point-wise 5% and 95% quantiles for the posterior survival function, given the baseline state "Severe CAV" ($s = 3$). We see that the posterior survival curve is significantly lower than the Kaplan-Meier survival function estimator. This reflects the expected disease progression since baseline. The expected restricted-mean survival time



(a) Traceplots for MCMC algorithm (b) Approximate posterior densities

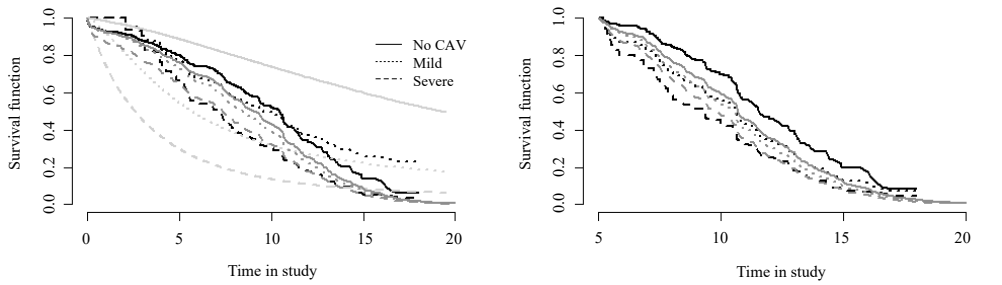
Figure 3. MCMC traceplots and densities for CAV study.



(a) Survival functions at baseline (b) Survival functions at $t = 5$

Figure 4. Survival functions for "No CAV" (black), "Mild/Moderate CAV" (light gray), and "Severe CAV" (dark gray); Kaplan-Meier estimator (dotted black).

is estimated under the restriction $t \leq 20$; because the study follow-up ends at that time. For states 1, 2, and 3, the expectations are 8.84, 8.40, and 7.41, respectively. Recall that all patients are in state 1 at baseline; therefore, the Kaplan-Meier curve should be compared with the median survival curve given the new patient is in state 1. Under the Kaplan–Meier estimator, the expected restricted-mean survival time is 9.66. The 5% and 95% quantiles for the survival function at each time t when the patient is in state 3 at baseline are included. Figure 4b plots the median survival, given that the user is alive at time $t = 5$. The expected restricted-mean remaining survival time from time 5 given that a new patient is in state 1, 2, or 3, is 6.12, 5.67, and 5.11 respectively.



(a) Comparison at baseline.

(b) Comparison at $t = 5$.

Figure 5. Survival functions from baseline for the parametric (light gray), nonparametric (black), and current (dark gray) methodologies.

8.1. Comparison with alternative life history analyses

We compare our results with those obtained from (a) a time-homogeneous parametric Markov model (Saeedi and Bouchard-Côté (2011); Hajiaghayi et al. (2014); Rao and Teh (2013)), and (b) a nonparametric Aalen–Johansen estimator under the assumption that the jump times coincide with the observation times (Aalen et al. (2015)). Figure 5a visualizes the estimated survival curves for each approach. The survival estimator under (a) clearly suffers from model misspecification, likely over-estimating survival in state 1 and under-estimating survival in states 2 and 3. The nonparametric estimator suffers from the last observation carried forward (LOCF) assumption. In particular, disease progression is restricted to occur at observation times. Moreover, it requires that state transitions be observed. This leads to state 2 having a higher survival probability than that of state 1 later in the study, which is scientifically implausible. The nonparametric survival function more closely matches our current results for $t > 5$, given $s = 0$ at baseline. This is because state 3 has a holding time that is much longer than the observation, leading to lower bias under the LOCF assumption. Figure 5a suggests the proposed framework can be viewed as regularizing the nonparametric estimator toward the parametric model; indeed, our model assumes that the marginal process $\mathbf{Y}_u$, for any $u \in \mathbb{N}$, is a time-homogeneous Markov model. Because we have a sufficient sample size, the conditional distributions account for the model misspecification, while allowing us to deal naturally with intermittent observation times. For the nonparametric (parametric) estimator, the expected restricted-mean survival times in states 1, 2, and 3 are 10.85 (24.56), 14.76 (11.62), and 8.36 (5.84), respectively.

Figure 5b compares the estimated conditional survival curves at $t = 5$ from our proposed methodology with that from the nonparametric estimator. Here, the gaps between the nonparametric survival curves, given a state at time $t = 5$, are more spaced as there are few observed transitions among states. The proposed methodology accounts for possible transitions that agree with the data, suggesting that the conditional survival curves are closer than the nonparametric analysis suggests. For the nonparametric estimator, the expected restricted-mean survival times in states 1, 2, and 3 is 8.52, 5.71, and 4.71, respectively.

9. Discussion

This study lays a theoretical and methodological foundation for the development of models based on exchangeable Markov multi-state survival processes. The model class encompasses many examples from health and epidemiology. Section 5.3 demonstrates how the process accommodates dependence, providing a data-generating description of an unobserved trajectory $\mathbf{Y}_{n+1}$, given the observed trajectory $\mathbf{Y}_{[n]}$. The model class, however, is limited in several respects. First, the models are not yet suited to handle dynamicity, an important issue when dealing with recurrent event processes (Peña (2016)). Second, we consider only the time-homogeneous Markov setting, which implies that the measure Σ is fixed across t and the holding times are exponential. Third, in many settings, the observations are noisy and the state-space should be considered latent. We believe these issues can be handled by suitable extensions of the proposed methodology. For recurrent events, a more flexible dependence on the event-history can be introduced (Peña and Hollander (2004)). A semi-Markovian structure could allow for more complex holding-time distributions. A hidden Markov structure could account for measurement errors and allow for state misclassification. These issues are left to future work.

Supplementary Material

In the supplementary materials, we present the proofs for both the discrete-time and continuous-time characterizations of the exchangeable, Markov multi-state survival processes. We then present several important examples that motivate the current study of multi-state survival processes, including survival, illness-death, comorbidities, and competing risk processes. We then discuss the choice of measure and connect our proposal to the proportional conditional hazards model. We end with a simulation example to demonstrate the proposed methods and a link to all code related to the CAV analysis and simulation study.

References

Aalen, O., Borgan, O. and Gjessing, H. (2008). *Survival and Event History Analysis: A Process Point of View*. Springer-Verlag, New York.

Aalen, O., Pedersen, T., Borgan, O., Hoff, R., Roysland, K. and Strohmaier, S. (2015). Dynamic path analysis – A useful tool to investigate mediation processes in clinical survival trials. *Statistics in Medicine* **34**, 3866–3887.

Breto, C., He, D., Ionides, E. and King, A. (2009). Time series analysis via mechanistic models. *The Annals of Applied Statistics* **3**, 319–348.

De Finetti, B. (1972). *Probability, Induction and Statistics*. John Wiley & Sons Ltd.

Dempsey, W. and McCullagh, P. (2017). Weak continuity of predictive distributions for Markov survival processes. *Electronic Journal of Statistics* **11**, 5406–5451.

Dempsey, W. and McCullagh, P. (2018). Survival models and health sequences. *Lifetime Data Analysis (with Discussion)* **24**, 550–584.

Diggle, P., Sousa, I. and Chetwynd, A. (2008). Joint modeling of repeated measurements and time-to-event outcomes: The fourth armitage lecture. *Statistics in Medicine* **27**, 2981–2998.

Farewell, D. and Henderson, R. (2010). Longitudinal perspectives on event history analysis. *Lifetime Data Analysis* **16**, 102–117.

Gelman, A., Jakulin, A., Pittau, M. and Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* **2**, 1360–1383.

Hajiaghayi, M., Kirkpatrick, B., Wang, L. and Bouchard-Côté, A. (2014). Efficient continuous-time Markov chain estimation. In *Proceedings of the 31st International Conference on Machine Learning* **32**, I-638–I-646.

Hobolth, A. and Stone, E. (2009). Simulation from endpoint-conditioned, continuous-time Markov chains on a finite state space, with applications to molecular evolution. *The Annals of Applied Statistics* **3**, 1225–1310.

Jensen, A. (1953). Markoff chains as an aid in the study of Markoff processes. *Skandinavisk Aktuarietidskrift* **36**, 87–91.

Kurland, B., Johnson, L., Egleston, B. and Diehr, P. (2009). Longitudinal data with follow-up truncated by death: Match the analysis method to the research aims. *Statistical Science* **24**, 211–222.

Little, R., D'agostino, R., Dickersin, K., Emerson, S., Farrar, J., Frangakis, C. et al. (2012). The prevention and treatment of missing data in clinical trials. *The New England Journal of Medicine* **367**, 1355–1360.

Peña, E. and Hollander, M. (2004). Models for recurrent events in reliability and survival analysis. In *Mathematical Reliability: An Expository Perspective*, 105–123. Kluwer, Boston.

Peña, E. A. (2016). Asymptotics for a class of dynamic recurrent event models. *Journal of Nonparametric Statistics* **28**, 716–735.

Rao, V. and Teh, Y. W. (2013). MCMC for continuous-time discrete state systems. In *NIPS*.

Saeedi, A. and Bouchard-Côté, A. (2011). Priors over recurrent continuous time processes. In *Proceedings of the 24th International Conference on Neural Information Processing*, 2052–2060.

Schweder, V. (2007). Composable finite Markov processes. *Scandinavian Journal of Statistics* **34**, 169–185.

Walter Dempsey
Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA.
E-mail: wdem@umich.edu

(Received October 2018; accepted January 2020)