Statistica Sinica Preprint No: SS-2018-0403

| Title | Exchangeable Markov multi-state survival processes |
|--|--|
| Manuscript ID | SS-2018-0403 |
| URL | http://www.stat.sinica.edu.tw/statistica/ |
| DOI | 10.5705/ss.202018.0403 |
| Complete List of Authors | Walter Dempsey |
| Corresponding Author | Walter Dempsey |
| E-mail | wdem@umich.edu |
| Notice: Accepted version subject to English editing. | |

Statistica Sinica

Exchangeable Markov multi-state survival processes

Walter Dempsey

University of Michigan

Abstract: We consider exchangeable Markov multi-state survival processes – temporal processes taking values over a state-space S with at least one absorbing failure state $b \in S$ that satisfy natural invariance properties of exchangeability and consistency under subsampling. The set of processes contains many well-known examples from health and epidemiology – survival, illness-death, competing risk, and comorbidity processes; 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: multi-state survival process; exchangeability; Markov process; composable systems; Markov chain Monte Carlo

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, health status Y(t) at time t is a binary variable, dead (0) or alive (1). In clinical trials with health monitoring, Y(t) is a more detailed description of the state of health of the individual, containing relevant patient information, e.g., pulse rate, cholesterol level, cognitive score or CD4 cell count [Diggle et al., 2008, Farewell and Henderson, 2010, Kurland et al., 2009].

In this paper, we study health processes taking 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 the state of health for a patient over 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 paper is to characterize this set of multi-state survival processes and

1. INTRODUCTION

show how the theory of exchangeable stochastic processes fits naturally into the applied framework of event-history analysis. Both the parametric continuous-time Markov process with independent participants and the nonparametric counting process are contained as limiting 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 et al., 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 dynamic path analysis for a liver cirrhosis dataset. In this study, the prothrombin index, a composite blood coagulation index related to liver function, is measured initially at threemonth intervals and subsequently at roughly twelve-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., 2010].

One alternative is to consider parametric models such as continuoustime Markov processes. Prior work [Saeedi and Bouchard-Côté, 2011, Hajiaghayi et al., 2014, Rao and Teh, 2013] has focused on estimation of 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 is that observing sharp changes in health in prior patient trajectories at a particular time since recruitment will not impact the likelihood of a similar sharp change in a future patient at the same timepoint. The proposed approach in this paper balances between 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 prop-

erty. 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 $\flat \in S$ 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 **Y**:

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

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

Without loss of generality, we assume $S = \{1, ..., s\} =: [s]$. For example, if the state-space is $S = \{Alive, Dead\}$, we recode this to $[2] = \{1, 2\}$. At each time t, the population-level process is given by $\mathbf{Y}(t) = \{Y(u, t) \mid u \in V(u, t) \mid u \in V(u, t)\}$

 $\mathbb{N}\} \in [s]^{\mathbb{N}}$. 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 $\mathbf{y}_{[4]} = (1, 1, 2, 1)$ then $\mathbf{x}_{[4]} = (3, 1)$. We write x_i to denote the *i*th entry of $\mathbf{x}_{[n]}$.

Example 1 ((Bidirectional) illness-death process). To make ideas concrete, we take the illness-death process as the running example in this paper. The illness-death process has state space { Healthy, Unhealthy, Dead } with transitions governed by the graph shown in Figure 1. The *bi-directional* illness-death process includes the 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 supplementary section ??.



Figure 1: Graph representation of the illness-death process

2.2 Transition graphs for 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 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 not (Unhealthy, Healthy). 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 $\sum_{i' \in V} P_{i,i'} = 1$, $P_{i,i'} \ge 0$ 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 as it is often an arbitrary choice based on power considerations and/or patient recruitment constraints. Informally, observing n units versus n + 1 units and then restricting to the first n units should be equivalent, i.e., the model should exhibit *consistency under subsampling*.

Consider the multi-state survival process $\mathbf{Y}_{[m]}$ for m > n. Define the restriction operator $R_{m,n}$ to be 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 distribution to $\mathbf{Y}_{[n]}$ for all $[m] \supset [n]$. That is, for any finite subset $\mathbf{t} := (t_1, \ldots, t_k) \subset \mathcal{T}$, $R_{m,n}(\mathbf{Y}_{[m]}(\mathbf{t})) =_{\mathcal{D}} \mathbf{Y}_{[n]}(\mathbf{t})$. Stated another way, $\operatorname{pr}(\mathbf{Y}_{[n]} \in A) = \operatorname{pr}(\mathbf{Y}_{[m]} \in R_{m,n}^{-1}(A))$ for any Borel measurable set A.

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

$$\operatorname{pr}(\mathbf{Y}_{[n]} \in A \mid H_{[m]}(t)) = \operatorname{pr}(\mathbf{Y}_{[n]} \in A \mid H_{[n]}(t))$$

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, ensuring the *n*-restricted multi-state survival process is unaffected by the multi-state survival process for 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} to be *[partially] exchangeable* if for any permutation σ : $[n] \rightarrow [n]$, the relabeled process $\mathbf{Y}_{[n]}^{\sigma} = \{Y(\sigma(1), t), \dots, Y(\sigma(n), t) \mid t \in \mathcal{T}\}$ 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 $t, \mathcal{H}_{[n]}(t)$, only depends on $\mathbf{Y}_{[n]}(t)$ and t'. This Markovian assumption is a simplifying assumption which leads to mathematically tractable conclusions. In this paper, we restrict our attention to timehomogeneous processes; therefore, we simply say $\mathbf{Y}_{[n]}$ is Markovian.

3 Markov, exchangeable multi-state survival

processes

Define a multi-state survival process that is Markovian, exchangeable, and consistent under subsampling as a *Markov, exchangeable 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 choice of time-scale matters in applied settings. All proofs are left to supplementary materials.

3.1 Discrete-time multi-state survival models

In discrete-time, the Markov, exchangeable 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

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

i.e., the (i, i') entry of P_t . Let $\mathbf{Y}_{\Sigma}^{\star}$ denote a discrete-time process constructed by 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 a Markov, exchangeable 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 our running example 1. First, assume all units start in state "Healthy". Next, for each $t \in \mathbb{N}$, define a set $\{Z_{i,i'}^{(t)}\}$ of independent beta random variables with parameters $\gamma \alpha_{i,i'}$ and $\gamma \beta_{i,i'}$ respectively. Set $[P_t]_{1,2} = Z_{1,2}$, $[P_t]_{1,3} = (1 - Z_{1,2}) \times Z_{1,3}$, $[P_t]_{1,1} = 1 - [P_t]_{1,2} - [P_t]_{1,3}$; $[P_t]_{2,3} = Z_{2,3}$ and $[P_t]_{2,2} = 1 - [P_t]_{2,3}$; $[P_t]_{3,3} = 1$. 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

$$P_{1} = \begin{bmatrix} 0.26 & 0.51 & 0.23 \\ 0.00 & 0.54 & 0.46 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}, P_{2} = \begin{bmatrix} 0.25 & 0.50 & 0.26 \\ 0.00 & 0.54 & 0.46 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}, P_{3} = \begin{bmatrix} 0.25 & 0.48 & 0.27 \\ 0.00 & 0.50 & 0.50 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}.$$

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

$$P_{1} = \begin{bmatrix} 0.13 & 0.45 & 0.43 \\ 0.00 & 0.16 & 0.84 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}, P_{2} = \begin{bmatrix} 0.29 & 0.52 & 0.20 \\ 0.00 & 0.19 & 0.81 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}, P_{3} = \begin{bmatrix} 0.01 & 0.69 & 0.30 \\ 0.00 & 0.62 & 0.38 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}.$$

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

3.2 Continuous-time multi-state survival models

In continuous-time, the Markov exchangeable multi-state survival process is governed by a measure on transition matrices, denoted Σ , and a set of constants associated with the edge set, denoted $\mathbf{c} = \{c_{i,i'} | (i,i') \in E\}$. Unlike discrete-time, 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 rate parameter that depends on the number of units in each state, i.e., the current *configuration* $\mathbf{x}_{[n]}$ (see Definition 1). At a 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) experience 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 continuoustime 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 a Markov, exchangeable multi-state survival process; and I_s be the $s \times s$ identity matrix. Then there exists a probability measure Σ on \mathcal{P}_G satisfying

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

and constants $\mathbf{c} = \{c_{i,i'} \ge 0 \mid (i,i') \in E\}$ 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, e.g., 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) transition matrix at jump times. In order to formally define this procedure, we require an additional definition.

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

index, denoted $\zeta(\mathbf{y}_{[n]})$, 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 0 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 $q(\mathbf{Y}_{[n]}(t-), \mathbf{Y}_{[n]}(t))$, equals

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

$$=: \frac{\lambda(\mathbf{Y}_{[n]}(t-),\mathbf{Y}_{[n]}(t))}{\zeta_{n}(\mathbf{Y}_{[n]}(t-))}$$

where $P[i, i'] = P_{i,i'}$, $\delta(\cdot)$ is the indicator function, and $\lambda(\cdot, \cdot)$ is the nonnormalized 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

4. DISCRETIZATION AND ROUNDING

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, as single unit transitions are unrelated to the rest of the population, we recommend setting $\mathbf{c} \equiv 0$ and 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 such a 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 rounding of intrinsically continuous phenomena). Theorems 1 and 2 supplement this scientific distinction with a mathematical one as we see distinct characterizations of discrete and continuous-time processes. One example of the former in survival analysis is time to get pregnant, which should be measured in menstrual cycles. The latter represents 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 a 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 choice of time-scale.

5.2 Density function

Since the evolution of the process $\mathbf{Y}_{[n]}$ is Markovian, it is a straightforward exercise to give an expression for the probability density function for any specific temporal trajectory. 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

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

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

where $\lambda(\cdot, \cdot)$ is the non-normalized transition probabilities. Continuing in this way, it can be seen that 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

$$\exp\left(-\int_0^\infty \nu\zeta_n(\mathbf{Y}_{[n]}(s))\,ds\right)\prod_{j=1}^k \lambda\big(\mathbf{Y}_{[n]}(t_j-),\mathbf{Y}_{[n]}(t_j)\big).\tag{5.2}$$

The number of transitions k is a random variable whose distribution is determined by (5.2), and hence by ζ_n .

Although the argument leading to (5.2) 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. So long as the censoring mechanism is *exchangeability preserving* [Dempsey and McCullagh, 2017], inference based on the joint density given by equation (5.2) is possible. Both simple type I censoring and independent censoring mechanism 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 via a set paired measures consisting of a continuous component $\Lambda_{i,i'}^{(c)}$ and an atomic measure $\Lambda_{i,i'}^{(a)}$ 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', i.e., $\mathbf{y}_{[n+1]}(t-) = (\mathbf{y}_{[n]}(t-), i)$ and $\mathbf{y}_{[n+1]}(t) = (\mathbf{y}_{[n]}(t), i')$. The continuous component has hazard and cumulative hazard

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

The non-normalized transitions $\lambda(\cdot, \cdot)$ are piecewise constant as a function of t, so the integral is trivial to compute, but 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 $\Lambda_{i,i'}^{(a)}$ associated with switching from state i to i'. At each such point, the conditional hazards has an atom with finite mass

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

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

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

and 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 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 previously addressed by restricting attention to processes whose conditional distributions are *weakly continuous*, i.e., small perturbations of transition times imply small perturbations of conditional distributions.

Dempsey and McCullagh [2017] originally studied this question in the context of exchangeable, Markov survival processes. In particular, it is shown 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 $(i, i') \in E$ an independent harmonic process with parameters $(\nu_{i,i'}, \rho_{i,i'})$. For $(i, i') \in E$ let $t_1^{(i,i')} < \ldots < t_{k(i,i')}^{(i,i')}$ 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:

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

where the sum runs over 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 transition times

$$\prod_{l:t_{l}^{(i,i')} \leq t} \frac{\mathbf{Y}_{[n]}^{\sharp}(t;i) + \rho_{i,i'}}{\mathbf{Y}_{[n]}^{\sharp}(t-;i) + \rho_{i,i'}}.$$
(5.3)

For small $\{\rho_{i,i'}\}_{(i,i')\in E}$, the combined discrete components are essentially the same as 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 non-zero, 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

6. COMPOSABLE MULTI-STATE SURVIVAL MODELS

edge $(i, i') \in E$ are possible. While this may make sense in specific cases, additional care is needed in writing down 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" were 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 only take positive mass on one of two sets of transition matrices: (I) P with off-diagonal positive mass only in entries (1,3) and/or (2,3), or (II) P with off-diagonal positive mass only in entries (1,2) and/or (2,1). The first represent transitions from "Healthy" or "Ill" to "Dead". The

6. COMPOSABLE MULTI-STATE SURVIVAL MODELS

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, i.e., $\tilde{\Sigma}$ puts positive mass on transition matrices P such that $P_{1,2} = P_{2,1} = 0$ so $P_{k,k} = 1 - P_{k,3}$ for k = 1, 2, and $P_{3,3} = 1$. Here, we restrict attention to measures of the form:

$$\tilde{\Sigma}(dP) = \nu_{1,1} \cdot P_{1,1}^{\rho_{1,1}-1} (1 - P_{1,1})^{-1} \delta(P_{2,2}^{\gamma} = P_{1,1}) dP_{1,1} dP_{2,2}.$$
(6.4)

Equation (6.4) corresponds to a proportional model on the logarithmic scale, linking $P_{1,1}$ and $P_{2,2}$ via baseline measure for a harmonic process. Details on the connection to the proportional conditional hazards model are provided in supplementary section ??.

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

6. COMPOSABLE MULTI-STATE SURVIVAL MODELS

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 (ex. 1), $B = (\{1, 2\}, \{3\})$. For comorbidities (ex. ??), B partitions the risk processes. For competing risks (ex. ??), $B = (\{1\}, B_2, \ldots, B_k)$ 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, Markov, exchangeable 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

7. PARAMETER ESTIMATION

in B_j to states in $B_{j'}$, $\tilde{\Sigma}_{jj'}(dP)$, by

$$\nu_{j,j'} P_{i(j),i(j)}^{\rho_{j,j'-1}} \left(1 - P_{i(j),i(j)}\right)^{-1} dP_{i(j),i(j)} \tag{6.5}$$

$$\prod_{l \in B_j \setminus i(j)} \delta \left[P_{l,l}^{\gamma_{l,j'}} = P_{i(j),i(j')} \right] dP_{l,l}$$
(6.6)

$$\prod_{l \in B_j} \prod_{m \in B_{j'}: (l,m) \in E} \delta[y_{l,m} = \alpha_{l,m}] dy_{l,m}$$
(6.7)

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 (6.5) and (6.6) 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 $\gamma_{i(j),j'}$ is set to 1 by design, and so the parameters measure risk relative to the chosen *representative state*. Line (6.7) addresses the fact that a single state $l \in B_j$ can transition to multiple states in $B_{j'}$, and leads to a simple Gibbs update procedure.

7 Parameter estimation

In practice, the patient's health status is typically measured at recruitment (t = 0), and 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 $Y(\mathbf{t})$, and 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 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

$$t_k \perp Y \mid (\mathbf{t}_{k-1}, Y(\mathbf{t}_{k-1})). \tag{7.8}$$

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 (7.8).

7.1.1 Prior specification and MCMC updates

We start with 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 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

7. PARAMETER ESTIMATION

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 direct comparison of $\lambda_{j,j'}$ across various choices of the 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

$$\operatorname{Gamma}\left(\alpha + k_{j,j'}, \beta + \rho \cdot \int_0^\infty \zeta_n(Y_{[n]}(s); \bar{\gamma}, \bar{\alpha}, \rho) ds\right)$$
(7.9)

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

Finally, for $l \in [s]$ consider transitions to partition $B_{j'}$. Index 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 $\bar{p}_{l,j'} = (p_{l,1,j'}, \ldots, p_{l,m_{l,j'},j'})$. Then the posterior is conjugate and

$$\bar{\alpha}_{l,j'} \mid \mathbf{Y}_{[n]}, \bar{\gamma}, \bar{\lambda}, \rho \sim \text{Dir}\left(p_{l,1,j'} + k_{l,1,j'}, \dots, p_{l,m_l,j'} + k_{l,m_l,j'}\right)$$
(7.10)

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 which is highly adaptable to MCMC. See Rao and Teh [2013] for an excellent

7. PARAMETER ESTIMATION

discussion of uniformization-based MCMC. Direct application of existing Gibbs samplers based on uniformization to the current setting is problematic due to combinatorial growth in the state-space and time-inhomogeneity of the conditional Markov process. Luckily, this approach can be adjusted using the sequential description from Section 5.3 as guidance.

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 have observe $(\mathbf{t}_i, \mathbf{Y}_i(\mathbf{t}_i), V_i, \Delta_i)$. The appointment schedule \mathbf{t}_i is an ordered sequence $0 \le t_{i,0} < \ldots < t_{i,k_i} \le V_i$. 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 $\tilde{\mathbf{t}}_j, t_{i,j} \leq \tilde{t}_{1,j} < \ldots < \tilde{t}_{L_j,j} \leq t_{i,j+1}$, be 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 $\Omega_t = C \cdot \max_{(i,i') \in E} |\Lambda_{i,i'}^{(c)}(t)|$ 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 transition matrix $\Lambda^{(a)}$ at times $\tilde{\mathbf{t}}_j$.

7.1.3 MCMC procedure

In each MCMC iteration, we proceed sequentially through 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 $\bar{\alpha}$ respectively using equation (7.9) and (7.10). One issue with this procedure is path sampling is computationally expensive. To address this issue, we also propose an approximate MCMC algorithm in which the latent processes are only updated every few iterations. In simulations, posteriors are not significantly altered, while run time drops significantly. For the sake of conciseness, we provide a simulation study of the MCMC procedure in the supplementary materials (see Section 1). We reserve the remainder of this paper for application of the proposed methodology to an irregularly sampled multi-state survival dataset.

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. This data was 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 transplant. Mean follow-up time is 5.9 years. For censored records, the censoring time is set equal to the final appointment time. Out of the 622 patients, only 192 patients were observed in state 2 (Mild CAV) at any point during their follow-up. Out of these 192, 43 of these patients were subsequently observed in state 1. Only 92 patients were observed in state 3 (Severe CAV) at any point during their follow-up. Out of these 92, 12 of these patients were subsequently observed in state 2.



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

We set $B = (B_1 = \{1, 2, 3\}, B_2 = \{4\})$ and $\rho = 10$. 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. As 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 $\lambda_{12} \sim \text{Gamma}(1,1)$. Parameters $(\gamma_{21}, \gamma_{22}, \gamma_{31}, \gamma_{32})$ have independent, standard log-normal priors. The approximate MCMC sampler is used.

Traceplots in Figure 3a suggest convergence after the first 100 iterations. The posterior mean of λ_{11} is 0.80 (i.e., marginal time until transition from state 1 to state 2 is 1.25 years). Parameters (γ_{21}, γ_{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 the patient is a bit more likely to experience progression of the CAV status than 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. Parameters (γ_{22}, γ_{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 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 median survival at each time t over all iterations of the MCMC sampler, the Kaplan-Meier survival function estimator, and point-



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

Figure 3: MCMC traceplots and densities for CAV study wise 5% and 95% quantiles for the posterior survival function given 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 expected disease progression since baseline. The expected restricted-mean survival time is estimated under the restriction to $t \leq 20$ due to study follow-up ending at that time. For states 1, 2, and 3 the expectation is 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 median survival given the user is alive at time t = 5. The expected restricted-mean remaining survival time from time 5 given new patient is in state 1, 2, and 3 is 6.12, 5.67, and 5.11 respectively.



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

8.1 Comparison to alternative life history analyses

We compare our results to 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 non-parametric Aalen-Johansen estimator under the assumption that jump times coincide with 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 in states 2 and 3. The non-parametric estimator suffers from the last-observation carried forward assumption. In particular, disease progression is restricted to occur at observation times. Moreover, it requires state transitions to be observed. This leads to the state 2 having higher survival probability over state 1 later in the study, which is scientifically implausible. The non-parametric survival function more closely matches our current results for t > 5 given s = 0 at baseline. This is due to state 3 having 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 towards the parametric model; indeed, our model assumes the marginal process \mathbf{Y}_u for any $u \in \mathbb{N}$ is a time-homogeneous Markov model. As we have sufficient sample size, the conditional distributions account for model misspecification, yet allow us to deal naturally with intermittent observation times. For the non-parametric (parametric) estimator, the expected restricted-mean survival time in states 1, 2, and 3 is 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



Figure 5: Survival functions from baseline for parametric (red), nonparametric (black), and current (blue) methodology.

from our proposed methodology to that from the non-parametric estimator. Here, the gaps among the non-parametric survival curves given 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 the conditional survival curves are closer than the non-parametric analysis suggests. For the non-parametric estimator, the expected restricted-mean survival time in states 1, 2, and 3 is 8.52, 5.71, and 4.71 respectively.

9 Discussion

This paper 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 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 only consider the time homogeneous Markov setting, which implied the measure Σ is fixed across t and holding times were 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 be proposed to allow for more complex holding time distributions. A hidden Markov structure could account for measurement-error and allow for state misclassification. These are considered important future work.

References

- O. Aalen, O. Borgan, and H. Gjessing. Survival and Event History Analysis: A Process Point of View. Springer-Verlag New York, New York, NY, 1 edition, 2008.
- O. Aalen, T. Pedersen, O. Borgan, R. Hoff, K. Roysland, and S. S. Dynamic path analysis a useful tool to investigate mediation processes in clinical survival trials. *Statistics in Medicine*, 34(29):3866–3887, 2015.
- C. Breto, , D. He, E. Ionides, and A.A. King. Time series analysis via mechanistic models. *The Annals of Applied Statistics*, 3(1):319–348, 2009.
- B. De Finetti. Probability, Induction and Statistics. John Wiley & Sons Ltd, 1 edition, 1972.
- W. Dempsey and P. McCullagh. Weak continuity of predictive distributions for markov survival processes. *Electronic Journal of Statistics*, 11(2): 5406–5451, 2017.
- W. Dempsey and P. McCullagh. Survival models and health sequences. Lifetime Data Analysis (with discussion), 24(4):550–584, 2018.
- P.J. Diggle, I. Sousa, and A. Chetwynd. Joint modeling of repeated mea-

surements and time-to-event outcomes: The fourth armitage lecture. Statistics in Medicine, 27(16):2981–2998, 2008.

- D. Farewell and R. Henderson. Longitudinal perspectives on event history analysis. *Lifetime Data Analysis*, 16(1):102–117, 2010.
- A. Gelman, A. Jakulin, M.G. Pittau, and Y-S Su. A weakly informative default prior distribution for logistic and other regression models. Ann. Appl. Stat., 2(4):1360–1383, 2008.
- M. Hajiaghayi, B. Kirkpatrick, L. Wang, and A. Bouchard-Côté. Efficient continuous-time markov chain estimation. In *International Conference* on Machine Learning (ICML), 2014.
- A. Hobolth and E.A. Stone. Simulation from endpoint-conditioned, continuous-time markov chains on a finite state space, with applications to molecular evolution. Ann. Appl. Stat., 3(3):1225–1310, 2009.
- A. Jensen. Markoff chains as an aid in the study of markoff processes. Skand. Aktuarietiedskr., 36:87–91, 1953.
- B.F. Kurland, L.L. Johnson, B.L. Egleston, and P.H. Diehr. Longitudinal data with follow-up truncated by death: match the analysis method to the research aims. *Statistical Science*, 24(2):211–222, 2009.

- R. Little, R. D'agostino, K. Dickersin, S. Emerson, J. Farrar, C. Frangakis, J. Hogan, G. Molenberghs, S. Murphy, J. Neaton, D. Scharfstein, W. Shih, J. Siegel, H. Stern, M. Cohen, and A. Gaskin. The prevention and treatment of missing data in clinical trials. Technical report, National Research Council of the National Academies, Washington, D.C., 2010.
- E. Peña and M. Hollander. Models for recurrent events in reliability and survival analysis. In *Mathematical reliability: an expository perspective*, pages 105–123, Kluwer, Boston, 2004.
- Edsel A. Peña. Asymptotics for a class of dynamic recurrent event models. Journal of Nonparametric Statistics, 28(4):716–735, 2016.
- V. Rao and Y. W. Teh. Mcmc for continuous-time discrete state systems. In NIPS, 2013.
- A. Saeedi and A. Bouchard-Côté. Priors over recurrent continuous time processes. In *NIPS*, volume 24, pages 2052–2060, 2011.
- V. Schweder. Composable finite markov processes. *Scand. J. Statist.*, 34: 169–185, 2007.