Exchangeable Markov multi-state survival processes

Walter Dempsey*

October 26, 2018

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 $♭ \in S$ 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 of the developed models to irregularly sampled and potentially censored multi-state survival data, developing a Markov chain Monte Carlo algorithm for posterior computation.

1 Introduction

In many clinical survival studies, a recruited patient’s health status is monitored on either a regular or intermittent schedule until either (1) an event of interest (e.g., failure) or (2) the end of the study window. Covariates are recorded at the time of recruitment, and treatment protocols per patient are presumed known at baseline. In the simplest survival study, health status

*Department of Statistics, Harvard University, One Oxford Street, Cambridge, MA 02138, U.S.A. E-mail: wdempsey@fas.harvard.edu
$Y(t)$ 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 \cite{Diggle2008, Farewell2010, Henderson2010, Kurland2009}.

In this paper, we consider the setting in which the health process takes values in some pre-specified “state-space”. For example, in the illness-death model, we summarize the current state of the participant as taking one of three possible values \{Healthy, Ill, Dead\}. Such a process can be thought of as a coarse view of the state of health for a patient over time. The continuing importance of multi-state processes in applications cannot be overstated \cite{Jepsen2015, van_den_Hout2016}.

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 \ref{sec:exchangeable}) lead to the concept of exchangeable Markov multi-state survival processes. The purpose of this paper is to characterize the set of multi-state survival processes and show how this 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.

\subsection{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” \cite[p. 457]{Aalen2008}. Martingales and counting processes form the basis of this nonparametric approach. The fundamental concept of the product-integral unifies discrete and continuous-time processes. Nonparametric methods, however, do not adequately handle intermittent observations. For example, \cite{Aalen2015} 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 three-month intervals and subsequently at roughly twelve-month intervals. To deal with the intermit-
tency 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]. Jepsen et al. [2015] discuss the importance of multi-state and competing risks models for complex models of cirrhosis progression. Here again, the nonparametric approach assumes observation times correspond to transition times of the multi-state process.

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

\[
T_u = \inf \{ t \geq 0 : Y(u, t) = \flat \}.
\]
For all \( u \in \mathbb{N} \), we assume \( Y(u, 0) \neq \flat \) at recruitment, so \( T_u > 0 \). Multiple absorbing states \( \{\flat_c\} \) representing different terminal events may occur, such as for competing risk processes.

Without loss of generality, we assume \( \mathcal{S} = \{1, \ldots, s\} =: [s] \). For example, if the state-space is \( \mathcal{S} = \{\text{Alive, Dead}\} \), we recode this to \( [2] = \{1, 2\} \). In this case, \( \flat = 2 \) is the flatlining state. At each time \( t \), the population-level process is given by \( Y(t) = \{Y(u, t) \mid u \in \mathbb{N}\} \in [s]^\mathbb{N} \). We write \( y \) to denote a generic element of \( [s]^\mathbb{N} \). We write \( Y_A \) to denote the restriction of the state space process to \( u \in A \subset \mathbb{N} \). We call \( Y_{[n]} \) the \( n \)-restricted state-space process for \( [n] := \{1, \ldots, n\} \). We write \( y_{[n]} \) to denote a generic element of \( [s]^n \).

### 2.2 Transition graphs for multi-state survival process

The transition graph represents the set of potential transitions between elements of the multi-state survival process out of the set of \( s^2 \) possible transitions. The transition graph is a directed graph \( G = (V, E) \). The vertex set \( V = [s] \) is all potential states; the directed edge set \( E \) contains all edges \( (i, j) \) such that at jump times the Markov process can jump from \( i \) to \( j \). In the illness-death model, for example, a patient can jump from \( \text{Alive} \) to \( \text{Ill} \) but not back; therefore \( (\text{Alive, Ill}) \) is in the edge set but not \( (\text{Ill, Alive}) \). Example 2 in supplementary section C provides additional details. In the bi-directional illness-death model, both edges are present in the transition graph. In continuous-time, jumps can only occur between distinct states, in which case the number of possible edges is \( s(s - 1) \). In this case, \( (i, i) \notin E \) for all \( i \in V \). Any absorbing state \( i \in [s] \) satisfies \( (i, j) \notin E \) for all \( j \neq i \in [s] \). We write \( P_G \) to denote the set of \( |S| \) by \( |S| \) transition matrices \( P \) satisfying \( \sum_{j \in V} P_{i,j} = 1, P_{i,j} \geq 0 \) for all \( i, j \in V \), and \( P_{i,j} = 0 \) for all \( (i, j) \notin E \). In the continuous-time setting, define \( P_{i,i} = 1 - \sum_{j,(i,j) \in E} P_{i,j} \).

### 2.3 Consistency under subsampling

We first note that sample size is often an arbitrary choice based on power considerations and/or patient recruitment constraints. Statistical models should be agnostic to such considerations. That is, 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 \( Y_{[m]} \) for \( m > n \). Define the restriction operator \( R_{m,n} \) to be the restriction of \( Y_{[m]} \) to the first \( n \) individuals. Then the process is consistent under subsampling if \( R_{m,n}(Y_{[m]}) \) is a version of \( Y_{[n]} \) for all \( [m] \supseteq [n] \). Under the consistency assumption, the process \( Y_{[n]} \)
satisfies lack of interference; mathematically,

$$\text{pr}(Y_{[n]} \in A \mid H_{[n]}(t)) = \text{pr}(Y_{[n]} \in A \mid H_{[n]}(t))$$

where $H_{[m]}(t)$ is the $\sigma$-field generated by the variables $Y(i, s)$ for $i \in [l]$ and $s \leq t$. 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. Without it, one is forced into the awkward situation of only being interested in the current collected dataset. Consistency ensures statistical inference to be a special case of “reasoning by induction” [De Finetti, 1972].

### 2.4 Exchangeability

We next note that the patient labels, $u \in \mathbb{N}$, are also arbitrary. Therefore, any suitable multi-state survival process must be agnostic to patient relabeling. We define a multi-state survival process $Y$ to be [partially] exchangeable if for any permutation $\sigma : [n] \rightarrow [n]$, the relabeled process $Y^\sigma_{[n]} = \{Y(\sigma(1), t), \ldots, Y(\sigma(n), t) \mid t \in T\}$ is a version of $Y_{[n]}$.

### 2.5 Time-homogeneous Markov process

We say $Y_{[n]}$ is a time-homogeneous Markov process if, for every $t, t' \geq 0$, the conditional distribution of $Y_{[n]}(t + t')$ given the multi-state survival process history up to time $t$, $H_{[n]}(t)$, only depends on $Y_{[n]}(t)$ and $t'$. This Markovian assumption is a simplifying assumption which leads to mathematically tractable conclusions. In this paper, we restrict our attention to time-homogeneous processes; therefore, we simply say $Y_{[n]}$ is Markovian.

### 3 Markov, exchangeable multi-state survival processes

We define a multi-state survival process that is Markovian, exchangeable, and consistent under sampling as a Markov, exchangeable multi-state survival process. Below, we characterize these processes in both discrete and continuous time. The behavior is markedly different in each setting with continuous-time Markov processes exhibiting much more complex behavior – allowing for both single-unit changes at time $t$ as well as positive fraction changes – showing why choice of time-scale matters in applied settings.
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 $\Sigma$ on $\mathcal{P}_G$. That is, the initial state $Y(0)$ is drawn from an exchangeable distribution on $[s]$. Then at time $t$, the transition distributions for each $u \in \mathbb{N}$ are given by

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

i.e., the $(i, i')$ entry of $P_t$, which is a random transition matrix drawn from $\Sigma$. Let $Y^*_\Sigma$ denote a discrete-time process constructed by this procedure with probability measure $\Sigma$. 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. The proof is left to supplementary section A.

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

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 $\Sigma$, and a set of constants associated with the edge set, denoted $c = \{c_{i,i'} | (i,i') \in E\}$. Unlike discrete-time, however, the jumps occur at random times.

Consider the $n$-restricted state space process. If $y[n]$ is the current state then the holding time in this state is exponentially distributed with a rate parameter depending on the current configuration (see Section 5.3). At a jump time $t$, one of two potential events can occur: (a) a single unit $u \in [n]$ experiences a transition from $Y(u,t-) = i$ to a state $i' \neq i$, or (b) a subset of $[n]$ (potentially a singleton) experience a simultaneous transition. Given the transition matrix $P(t)$, these simultaneous transitions are independent and identically distributed transitions according to $P(t) \in \mathcal{P}_G$. The transition matrix $P(t)$ is obtained from a measure $\Sigma$ on $\mathcal{P}_G$ and a set of constants $c$; unlike the discrete-time setting, the measure $\Sigma$ need not be integrable. Let $Y^*_\Sigma,c$ denote a continuous-time process constructed by this procedure with probability measure $\Sigma$.

**Theorem 2** (Continuous-time characterization). Let $Y = (Y(t), t \in \mathbb{R}^+)$ be a Markov, exchangeable multi-state survival process; and $I_s$ be the $s \times s$ identity
matrix. Then there exists a probability measure $\Sigma$ on $\mathcal{P}_G$ satisfying
\[ \Sigma(\{I_s\}) = 0 \quad \text{and} \quad \int_{\mathcal{P}_G} (1 - P_{\min}) \Sigma(dP) < \infty, \quad P_{\min} = \min_{i \in [s]} P_{i,i} \quad (1) \]
and constants $c = \{c_{i,i'} \geq 0 \mid (i, i') \in E\}$ such that $Y_{\Sigma,c}$ is a version of $Y$.

Theorem 2 generalizes Proposition 4.3 in [Dempsey and McCullagh, 2017] from the simple survival process setting. The multi-state survival process contains many more well-known examples from health and epidemiology – survival, illness-death, competing risk, and co-morbidity processes. We highlight these examples in supplementary section [C]. Within the discussion of examples, we extend Theorems 1 and 2 to the setting of recurrent events (see Corollaries 2 and 3). Here, Theorem 2 is used to characterize the continuous-time Markov chain $Y_{[n]}$ in terms of (1) exponential holding rates and (2) transition matrix at jump times. We start by defining the characteristic index – a set of functions $\zeta : [s]^n \to \mathbb{R}^+$.  

Definition 1 (Configuration vector). For $y_{[n]} \in [s]^n$, define $x_{[n]} \in [n]^s$ as the configuration vector – an $s$-vector summary of the number of units in each state. For example, if $s = 2$, $n = 4$, and $y_{[4]} = (1, 2, 2, 1)$, then $x_{[4]} = (2, 2)$; for $y_{[4]} = (1, 1, 2, 1)$ then $x_{[4]} = (3, 1)$.

Definition 2 (Characteristic index). The characteristic index $\zeta_n(y_{[n]})$ is defined as the normalizing constant to the integral representation obtained from Theorem 2
\[ \zeta_n(y_{[n]}) = \int_{\mathcal{P}_G} \left(1 - \prod_{j=1}^s P_{j,j}^{x_j} \right) \Sigma(dP) + \sum_{i \in [s]} x_i \sum_{i' : (i,i') \in E} c_{i,i'} \]
where $x_j$ is the $j$th component of the configuration vector associated with $y_{[n]}$, and the sum is set to 0 when $\{i' \in V s.t. (i, i') \in E\} = \emptyset$. Condition [1] implies the characteristic index is finite for any $y_{[n]} \in [s]^n$.

For simple survival processes, the characteristic index defined here simplifies to the characteristic index as defined in [Dempsey and McCullagh, 2017]. At a jump-time $t$, the probability of transition from $Y_{[n]}(t-)$ to $Y_{[n]}(t)$ is
\[ q(Y_{[n]}(t-), Y_{[n]}(t)) = \frac{1}{\zeta_n(Y_{[n]}(t-))} \left[ \int_{\mathcal{P}_G} \prod_{a \in [n]} P[Y(u,t-), Y(u,t)] \Sigma(dP) \right. \]
\[ + \left. \delta(\#\{u \in [n] \mid Y(u,t-) \neq Y(u,t)\} = 1) \sum_{i',(i,i') \in E} c_{i,i'} \delta(Y(u',t-) = i, Y(u',t) = i') \right] \]
\[ = \frac{\lambda(Y_{[n]}(t-), Y_{[n]}(t))}{\zeta_n(Y_{[n]}(t-))} \]

7
where $P[i, i'] = P_{i,i'}$, $\delta(\cdot)$ is the indicator function, $u'$ is the single unit to experience a transition, and $\lambda(\cdot, \cdot)$ is the non-normalized transition function.

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 data). Theorems 1 and 2 reinforce this distinction 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 $Y_{[n]}$ transitions into state $y_{[n]} \in [s]^n$. To each such state $y_{[n]}$, 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 particles.

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 $\tau_n(y) \propto \zeta_n(y)$.

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 $\nu$ which depends on choice of time-scale.

5.2 Density function

Since the evolution of the process $Y_{[n]}$ 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 $Y_{[n]}(t_1-)$ to $Y_{[n]}(t_1)$ is

$$
\nu \zeta_n (Y_{[n]}(t_1-)) \exp \left( -\nu \zeta_n (Y_{[n]}(t_1-)) t_1 \right) \int_{0}^{t_1} q(Y_{[n]}(t_1-), Y_{[n]}(t_1)) dt_1 \times \lambda(Y_{[n]}(t_1-), Y_{[n]}(t_1)) \exp \left( -\nu \zeta_n (Y_{[n]}(t_1-)) t_1 \right) dt_1 \times q(Y_{[n]}(t_1-), Y_{[n]}(t_1)) = \exp \left( -\nu \zeta_n (Y_{[n]}(t_1-)) t_1 \right) \prod_{j=1}^{k} \lambda(Y_{[n]}(t_j-), Y_{[n]}(t_j)).
$$

(2)

The number of transitions $k$ is a random variable whose distribution is determined by (2), and hence by $\zeta_n$. Note that with probability one $T_i < \infty$ for each $i \in [n]$; therefore, the transition at time $t_k$ is such that $Y_{[n]}(t_k) = (\flat, \ldots, \flat) := \bar{\flat}$ (i.e., all units have failed by time $t_k$). By definition $\zeta_n(\bar{\flat}) = 0$ and so the integral in equation (2) is finite with probability one.

The $n$-dimensional joint distribution is continuous in the sense that it has no fixed atoms. For $n \geq 2$, it is not continuous with respect to Lebesgue measure in $\mathbb{R}^n$. The one-dimensional marginal process is a Markov multi-state survival process with holding rates $\{\zeta_1(i) \mid i \in [s]\}$, assuming a valid starting state. For example, in a survival process the only valid starting state is “Alive”.

Although the argument leading to (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 $Y_{[n]}$, 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 (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 $Y_{[n]} = y_{[n]}$. The conditional distribution is best described via a set of paired conditional hazard measures, $\{\Lambda^{(c)}_{i,i'}, \Lambda^{(a)}_{i,i'}\}_{(i,i') \in E}$. For $(i,i') \in E$, the pair consists of a continuous component $\Lambda^{(c)}_{i,i'}$ in addition to an atomic measure $\Lambda^{(a)}_{i,i'}$ with positive mass only at the observed transition times of $y_{[n]}$. 
For a time \( t \), not a transition time of \( y_{[n]} \), consider the new unit transitioning from state \( i \) to \( i' \). Then the continuous component \( \Lambda_{i,i'}^{(c)} \) has hazard and cumulative hazard

\[
h_{i,i'}(t) = \zeta_n(y_{[n+1]}(t^-)) - \zeta_n(y_{[n+1]}(t)) := (\Delta \zeta)(y_{[n+1]}(t))
\]

\[
H_{i,i'}(t) = \int_0^t h_{i,i'}(s)ds
\]

Note that \( \zeta_{n+1}(y_{[n+1]}(t)) \) is piecewise constant, so the integral is trivial to compute, but censoring implies that it is not necessarily constant between successive failures.

Now let \( t \) be an observed transition time (i.e., \( y_{[n]}(t^-) \neq y_{[n]}(t) \)) 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

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

or, on the probability scale,

\[
\exp(-\Lambda_{i,i'}(\{t\})) = \frac{\lambda(y_{[n+1]}(t^-), y_{[n+1]}(t))}{\lambda(y_{[n]}(t^-), y_{[n]}(t))}.
\]

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 \( Y_{[n]} = y_{[n]} \). For \( s > 0 \), let \( \{t_j\}_{j=1}^L \) denote the observed transition times of \( y_{[n]} \) within the time-window \((t, t + s)\). Then, the probability that the unit stays in state \( i \) for at least \( s > 0 \) time points is

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

This conditional probability is used to construct the Markov Chain Monte Carlo sampling procedure in Section 7. For any exchangeable, Markov multisate survival process with absorbing state \( \{\gamma_c\} \) such that \( Y_u(0) \not\in \{\gamma_c\} \) with probability one, \( \zeta_1(i) > 0 \) for all \( i \in [s] \setminus \{\gamma_c\} \) implies that the continuous components \( \Lambda_{i,i'}^{(c)} \) for \( i \not\in \{\gamma_c\} \) have infinite total mass, so the time until reaching an absorbing state is finite with probability one. This implies, for instance, that in the Aalen comorbidities study (see Example 3 in Section C) that the process will terminate in death with probability one.
Although the above conditional hazards look complex, it is not difficult to generate the transition times sequentially for processes whose characteristic index admits a simple expression for the above expressions. Right censoring is automatically accommodated by the integral in the continuous component, so the observed trajectory $Y_{[n]}$ may be incomplete. Below we introduce the self-similar harmonic process – a multi-state survival process which admits such simple expressions. We use this process as a building block for more complex models in further sections.

5.4 Self-similar harmonic process

Theorem 2 proves tied failures are an intrinsic aspect of Markov multi-state survival processes. As stated in Section 4, grouped data usually are the result of rounding of intrinsically continuous data. For these processes to be useful in biomedical applications, it is essential that the model should not be sensitive to rounding. Sensitivity to rounding is addressed by restricting attention to processes whose conditional distributions are weakly continuous, i.e., a small perturbation of the transition times gives rise to a small perturbation of the conditional distribution.

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 $(i, j) \in E$ an independent harmonic process with parameters $(\nu_{i,j}, \rho_{i,j})$. For $(i, j) \in E$ let $t_{1}^{(i,j)} < \ldots < t_{k(i,j)}^{(i,j)}$ denote the unique observed transition times from $i$ to $j$ for $Y_{[n]}$ and let $Y_{[n]}^{z}(t; i) = \#\{u \in [n] \text{ s.t. } Y_{u}(t) = i\}$; then the continuous component of the hazard is given by:

$$H_{i,j}(t) = \sum_{l: t_{l}^{(i,j)} \leq t} \nu_{i,j} \frac{t_{l}^{(i,j)} - t_{l-1}^{(i,j)}}{Y_{[n]}^{z}(t_{l-1}^{(i,j)}; i) + \rho_{i,j}} + \nu_{i,j} \frac{t - t_{m}^{(i,j)}}{Y_{[n]}^{z}(t_{m}^{(i,j)}; i) + \rho_{i,j}}.$$  

where the sum runs over transition times $t_{l}^{(i,j)} \leq t$, and $t_{m}^{(i,j)}$ is the last such event. The discrete component is a product over transition times

$$\prod_{l: t_{l}^{(i,j)} \leq t} \frac{Y_{[n]}^{z}(t; i) + \rho_{i,j}}{Y_{[n]}^{z}(t-; i) + \rho_{i,j}}. \quad (3)$$

For small $\{\rho_{i,j}\}_{(i,j) \in E}$, the combined discrete components above 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 $\Sigma$ on $P_G$ is

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

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

While the self-similar harmonic process has strong appeal for use in applied work, we show below that it is not a universally optimal choice. The strong assumption embedded in the self-similar harmonic process is that the transition processes are independent across edges $(i,i') \in E$. This implies that at each transition time only transitions along a single edge $(i,i') \in E$ are possible. We argue below that while this may make sense in specific instances, additional care is needed in writing down appropriate models for multi-state survival processes 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 use the bidirectional illness-death process as an illustrative example. Recall this process has three states, \{Healthy, Ill, Dead\} (i.e, \{1, 2, 3\}); see Example 2 for further details. The state “Dead” is unique and distinct from the states “Healthy” and “Ill”. Indeed, the latter two states require the individual to be categorized more broadly as alive, and are thus refinements of this more general state (i.e. “Alive”). Suppose the labels “Healthy” and “Ill” 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.

The above discussion leads to two natural constraints: (1) the state “Dead” (State 3) is unique and distinct from the other states, and (2) the states “Healthy” (State 1) and “Ill” (State 2) should be considered partially exchangeable [De Finetti, 1972]. To satisfy this, we require particular constraints on the measure $\Sigma$ on $3 \times 3$ transition matrices. First, the measure must only take positive mass on one of two sets of transition matrices: (a) $P$ with off-diagonal positive mass only in entries $(1,3)$ and/or $(2,3)$, or (b) $P$ with off-diagonal positive mass only in entries $(1,2)$ and/or $(2,1)$. The first set are transition matrices representing transitions from “Healthy” or “Ill” to “Dead”. The second are transitions between “Healthy” and “Ill” or vice versa. The partition $B = \{B_1, B_2\} = \{\{1,2\}, \{3\}\}$ splits the space into two disjoint sets. Within each block $B_i$ the states are partially exchangeable. We say that the
bidirectional illness-death process is thus relatively partially exchangeable with respect to the partition \( B \).

Let \((n_1, n_2)\) denote the number of individuals in states “Healthy” and “Ill” directly preceding a transition of type (a). Then the probability that \(d_1 \leq n_1\) and \(d_2 \leq n_2\) individuals respectively transition to the state “Dead” is proportional to 
\[
\int p_{1,1}^{n_1-d_1} p_{1,3}^{d_1} p_{2,2}^{n_2-d_2} p_{2,3}^{d_2} \tilde{\Sigma}(dP),
\]
where \(\tilde{\Sigma}(dP)\) is the measure restricted to type (a) transition matrices. That is, \(\tilde{\Sigma}\) puts positive mass on transition matrices \(P\) such that \(P_{1,2} = P_{2,1} = 0\) so \(P_{k,3} = 1 - P_{k,3}\) for \(k = 1, 2,\) and \(P_{3,3} = 1\) (i.e., “Dead” is an absorbing state). In this paper, we focus on the following choice of the restricted measure:
\[
\hat{\Sigma}(dP) = \nu_{1,1} \cdot P_{1,1}^{\nu_{1,1}-1} (1 - P_{1,1})^{-1} \delta(P_{2,2} = P_{1,1}) dP_{1,1} dP_{2,2}.
\]
A similar formula exists for transitions of type (b). The choice corresponds to a proportional model on the logarithmic scale. It strongly links \(P_{1,1}\) and \(P_{2,2}\) with baseline measure equivalent to that for a harmonic process. In supplementary section D, we provide more details on the connection to the proportional conditional hazards model. We now consider extending this approach to construction of the measure \(\Sigma\) for any multi-state survival process.

### 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 \(G\).

In other words, a \(B\)-composable process is any process that is relatively partially exchangeable with respect to \(B\). For the bi-directional illness-death process (ex. 2), \(B = (\{1, 2\}, \{3\})\). For comorbidities (ex. 4), \(B\) partitions the risk processes. Two risk processes in the same partition may represent the same underlying process. For the competing risks (ex. 4), \(B = (\{1\}, B_2, \ldots, B_k)\) where \((B_2, \ldots, B_k)\) partition the absorbing states and the single state, “Alive”, is distinct which implies \(B_1 = \{1\}\). Every process is composable via the degenerate partition where each \(B_j\) is a singleton and \(k = s\). The partition accounts for absorbing states, \(i \in \{\varnothing\},\) by satisfying \(B(i') \cap B(i) = \emptyset\) for \(i' \not\in \{\varnothing\},\) where \(B(j)\) is the component of \(B\) to which state \(j \in [s]\) belongs.

If \(Y\) is \(B'\)-composable and \(B'\) is a refinement of the partition \(B\), then \(Y\) is also \(B\)-composable. To avoid confusion, from here on, when we say \(Y\) is \(B\)-composable, we assume there does not exist a refinement \(B'\) such that \(Y\) is
also $B'$-composable. Definition 3 is similar in spirit to that of Schweder 2007—both aim to formalize the notion that state changes in the process $Y$ are due to changes in different components.

6.2 Choice of measure $\Sigma$ for a $B$-composable process

Here, we construct an appropriate measure $\Sigma$ 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 in $B_j$ to states in $B_{j'}$, $\tilde{\Sigma}_{jj'}(dP)$, by

$$
\nu_{j,j'} P_{i(j),i(j')}^{\rho_{j,j'}-1} \left(1 - P_{i(j),i(j')}\right)^{-1}
\prod_{l \in B_j \setminus i(j)} \delta \left[ P_{l,l'}^{\gamma_{l,j,j'}} = P_{i(j),i(j')} \right] dP_{i(j),i(j)}
\prod_{l \in B_j} \prod_{m \in B_{j'} : (l,m) \in E} \delta \left[y_{l,m} = \alpha_{l,m}\right] dy_{l,m}
$$

where $\gamma_{l,j,j'} > 0$, $\gamma_{i(j),j'} = 1$, $\alpha_{l,m} \in [0, 1]$, $P_{l,m} \in [0, 1]$ 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 $P_{l,l} = P_{i(j),i(j')}$. Lines (4) and (5) 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) addresses the fact that a single state $l \in B_j$ can transition to multiple states in $B_{j'}$. Here, we choose an atomic measure located at $\alpha_{l,m}$, where the parameters satisfy $\sum_{m \in B_{j'} : (l,m) \in E} \alpha_{l,m} = 1$. These parameters lead to a multinomial distribution across transitions. We take this approach because it leads to simple Gibbs updates of the underlying parameters. In the example presented in section 9, Mild CAV status can transition to either No CAV or Severe CAV status. This is captured by presupposing a binomial distribution over these transitions.

Note, for $B = \{\{1\}, \ldots, \{s\}\}$, the above construction yields the self-similar harmonic process. For $B = \{\{1,2\}, \{3\}\}$, the above construction yields the bi-directional illness-death process from Section 6.
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 \((t, Y[t], V, \Delta)\) consists of the appointment schedule \(t\), multi-state process measurements \(Y[t]\), and a failure/censoring time \(V\), and a censoring indicator \(\Delta\). 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 \(t_{k-1} = (t_1, \ldots, t_{k-1})\) and observation values \(Y[t_{k-1}]\), the next appointment time \(t_k\) satisfies

\[
t_k \perp Y \mid (t_{k-1}, Y[t_{k-1}]). \tag{7}
\]

In other words, 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. This assumption combined with variational independence implies the component of the likelihood associated with the appointment schedule can be ignored for maximum likelihood estimation of parameters associated with the multi-state survival process. Under \(7\), we propose a Markov Chain Metropolis Hastings algorithm for posterior computation.

7.1 Uniformization

Uniformization \cite{Jensen1953, Hobolth2009} is a well-known technique for generating sample paths for a given Markov state-space process. Take a time-homogeneous, continuous-time Markov process with \(s \times s\) transition matrix \(Q\) (i.e., a \(s \times s\) matrix satisfying \(Q_{ii} = -\sum_{j \neq i} Q_{ij}\) and \(Q_{ij} \geq 0\)). A sample path can be generated via Gillespie’s algorithm: (1) given in state \(i \in [s]\), generate a random exponentially distributed holding time \(\tilde{T}\) with rate \(-Q_{ii}\); (2) after \(\tilde{T}\) time steps, randomly transition to a new state \(i' \neq i \in [s]\) with probability \(-Q_{ij}/Q_{ii}\). Alternatively, a sample path can be generated by uniformization with parameter \(\Omega \geq \max_i |Q_{ii}|\): (1) generate sequence of ‘potential’ transition times \(\tilde{w}\) from a homogeneous Poisson process with intensity \(\Omega\); (2) run a discrete-time Markov chain with transition matrix \(B = (I + \frac{Q}{\Omega})\) on the times \(\tilde{w}\) to generate sequence of states \(\tilde{s}\); (3) construct \(\hat{w}\) by removing elements of \(\tilde{w}\) where the process does not transition to a new state, and construct \(\hat{s}\) to be the sequence of states at times \(\hat{w}\). The sample path is represented by \((\hat{w}, \hat{s})\). Both algorithms generate sample paths from the same Markov state-space process; however, uniformization is highly adaptable to Markov Chain
Monte Carlo techniques. See Rao and Teh [2013] for an excellent discussion of uniformization-based MCMC for Markov processes.

7.2 The MCMC Algorithm

Direct application of existing Gibbs samplers based on uniformization to our 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. In this section, we derive an MCMC algorithm for posterior computations given irregularly sampled multi-state survival data.

7.2.1 Prior specification and MCMC updates

We start by specifying priors on the parameters $\{\nu_{j,j',\gamma_{l,j'}}\}_{l \in [s], j,j' \in [k]}$, and $\{\alpha_{l,m}\}_{l,m \in [s]}$. We write $\Phi$ to denote the complete set of parameters. We write $\bar{\alpha}, \bar{\gamma}, \text{et cetera}$ to denote each subset of parameters. Recall for identifiability reasons $\gamma_{i(j),j'} = 1$ for $j,j' \in [k]$. For all other pairs $(l,j')$, we take the prior to be a log-normal mean-zero distribution $\log(\gamma_{l,j'}) \sim N(0, 1)$. Weakly informative default priors are an alternative Gelman et al. [2008]; this corresponds to a log-Cauchy prior with center 0 and scale parameter $s_{l,m}$. However, we saw minimal differences in simulation studies for our current setting. The complete-data likelihood is non-conjugate, so we perform Metropolis-Hastings updates, conditional on both the complete trajectory $Y_{[n]}$ and all other parameters.

For $\rho_{j,j'}$, we follow Dempsey and McCullagh [2017] and take $\rho_{j,j'} := \rho$ as a fixed tuning parameter. Next, define $\lambda_{j,j'} = \nu_{j,j'} \cdot \rho$. Scaling by $\rho$ allows direct comparison of $\lambda_{j,j'}$ across various choices of the tuning parameter. We choose a Gamma prior, $\lambda_{j,j'} \sim \text{Gamma}(\alpha, \beta)$, which is conjugate to the complete-data likelihood. The posterior distribution conditional on the complete trajectory and other parameters is

$$
\lambda_{j,j'} \mid Y_{[n]}, \bar{\gamma}, \bar{\alpha}, \rho \sim \text{Gamma} \left( \alpha + k_{j,j'}, \beta + \rho \int_0^\infty \zeta_n(Y_{[n]}(s); \bar{\gamma}, \bar{\alpha}, \rho) ds \right) \quad (8)
$$

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 $\bar{\lambda}_{j'} = (\nu_{l,1,j'}, \ldots, \nu_{l,m_{l,j'},j'})$. Then the posterior is conjugate and

$$
\bar{\alpha}_{l,j'} \mid Y_{[n]}, \bar{\gamma}, \bar{\lambda}, \rho \sim \text{Dir} \left( \nu_{l,1,j'} + k_{l,1,j'}, \ldots, \nu_{l,m_{l,j'},j'} + k_{l,m_{l,j'}} \right) \quad (9)
$$
where $k_{l,m',j'}$ counts the number of transitions from state $l$ to state $m'$ in $Y_{[n]}$.  

7.2.2 Conditional sampling patient trajectories

We now adapt uniformization to construct a Gibbs sampling of the patient trajectory given all other trajectories $Y_{-i} = y_{-i}$, parameters $\Phi$, and the previous trajectory for this patient $\tilde{y}_i$. For patient $i$, we have observation $(t_i, Y_i[t_i], V_i, \Delta_i)$. The appointment schedule $t_i$ is an ordered sequence $0 \leq t_{i,0} < \ldots < t_{i,k_i} \leq V_i$ where $t_{i,k_i} = V_i$ if and only if $\Delta_i = 1$. By the Markov property, we can focus on generating the patient trajectory for fixed parameters for each interval $[t_{i,j}, t_{i,j+1}]$ separately. Let $t_{i,j} < \hat{t}_{1,j} < \ldots < \hat{t}_{L,j} \leq t_{i,j+1}$ be the unique transition times for all other patients within the interval $[t_{i,j}, t_{i,j+1}]$. Denote this set of transition times $\tilde{t}_j$. At each time $t \in [t_{i,j}, t_{i,j+1}]$, define $\Omega_t = C \cdot \max_{(i,i') \in E} |\Lambda_{i,i'}^{(c)}(t)|$ for some constant $C > 1$. By definition, this is a piecewise constant function that changes at unique transition times or censoring times. Next, sample a Poisson process $w \subset [t_{i,j}, t_{i,j+1}]$ with piecewise-constant rate $R_t = \Omega_t - \Lambda_{\tilde{y}_i(t), \tilde{y}_i(t)}^{(c)}(t)$, where $\Lambda_{\tilde{y}_i(t), \tilde{y}_i(t)}^{(c)}(t)$ is the continuous component of the conditional distribution. Finally, let $u_i$ denote the transition times from the previous trajectory $\tilde{y}_i$. Then let $T = w \cup \tilde{t} \cup u_i$ denote the union of these times. We can then apply the forward-filtering, backward sampling algorithm with transition matrix $B_t = (I + \Lambda^{(c)}/\Omega_t)$ at times $t \in W \cup (V \setminus \tilde{t})$ and transition matrix $\Lambda^{(a)}$ at times $\tilde{t}$. This combines the standard Gibbs sampling based on uniformization with the added complexity of the atomic component associated with the conditional distributions for exchangeable, Markov multi-state survival processes.

7.2.3 MCMC procedure

In each iteration, we proceed sequentially through patients, sampling a latent multi-state path for patient $Y_i$ given all other latent processes $Y_{-i} := Y_{[n]\setminus i}$ as discussed in Section 7.2.2. Then, conditional on the multi-state process $Y_{[n]}$, we perform Metropolis-Hastings updates for $\tilde{\gamma}$ as the complete-data likelihoods are easy to compute via equation (2). We end each iteration by using equation (8) and (9) to sample from the posterior for $\tilde{\lambda}$ and $\tilde{\alpha}$ respectively. One issue with this procedure is sequential sampling of the latent process is computationally expensive and can significantly slow down the MCMC procedure. To address this issue, we also propose an approximate MCMC algorithm in which the latent processes are only updated every few iterations. We see performance is not significantly altered, and run time drops significantly. In the illustrative example below, we show 1000 iterations of the MCMC procedure where the
latent process updates runs every 25 iterations has similar performance as the standard MCMC procedure with significant decrease in overall runtime.

8 MCMC procedure: a simulation example

We illustrate the sampling procedure on a bi-directional illness-death model (example 2). We set parameters as follows: first, we assume that marginally a healthy participant transitions to ill and dead after 2 and 5 years respectively (on average); ill participants to both healthy and dead on average every 3 years respectively. Both healthy and ill participants took on average 3 years to transition to failure. We assume a sample of \( N = 250 \) individuals, with 150 initially healthy and 100 initially unhealth, were generated.

First, assume all transitions are observed. Maximum likelihood estimation is performed. Next, assume the state of each individual is observed annually, with the transition time to failure observed. Traceplots in Figure 1a suggest convergence of the MCMC procedure after the first 100 iterations. The MCMC sampler gives posteriors for the parameters. Table 1 contains these estimates. We see good performance for \( \nu_{12} \), \( \gamma_{21} \), and \( \gamma_{22} \). The posterior for \( \nu_{11} \) reflects the observation schedule; indeed, increasing the frequency of observation significantly improves these posteriors. In particular, under complete observations, the posterior distributions are approximately equal in distribution to the asymptotically normal confidence intervals.

| Parameter | True Value | Maximum Likelihood Estimate | Lower CI | Upper CI | Mean | 5% Quantile | 95% Quantile |
|-----------|------------|------------------------------|----------|----------|------|-------------|-------------|
| \( \nu_{11} \) | 0.50 | 0.53 | 0.43 | 0.64 | 0.15 | 0.09 | 0.23 |
| \( \nu_{12} \) | 0.20 | 0.19 | 0.13 | 0.25 | 0.20 | 0.15 | 0.25 |
| \( \gamma_{21} \) | 0.70 | 0.75 | 0.64 | 0.86 | 0.85 | 0.67 | 1.07 |
| \( \gamma_{22} \) | 1.71 | 1.71 | 1.28 | 2.13 | 1.23 | 1.62 | 2.09 |

Table 1: Parameter estimation

Removing the first 100 iterations as burn-in, posterior distributions are presented in Figure 1b. Black curves are the MCMC sampler procedure; red curves are the approximate MCMC sampling procedure with latent process updates every 25 iterations. We see distributions are approximately equal in all cases, with largest errors for \( \nu_{11} \). This supports the aforementioned difficulty in estimating \( \nu_{11} \) due intermittent observations. This is a consequence of the observation schedule being infrequent compared to the underlying stochastic
dynamics. Complete observations (i.e., more frequent observations) significantly improves estimation of $\nu_{11}$.

![Traceplots for Gibbs sampler](image1)

![Approximate posterior density](image2)

(a) Traceplots for Gibbs sampler  (b) Approximate posterior density

Figure 1: MCMC traceplots and densities for simulation example

Beyond posterior distributions for parameters, one is typically interested in posterior distributions of the survival functions. Note, there are two distinct sources of variation – (1) intermittent observations and (2) parameter uncertainty. The MCMC sampling procedure accounts for both, allowing for survival functions to be constructed for each iteration of the MCMC sampler using each iterations’ latent process and parameters. Figure 2 presents the pointwise median, 5%, and 95% survival at every time since recruitment when the individual is healthy and ill at baseline respectively. The red curves are the true survival function given healthy/ill at baseline. We see that the posteriors for survival functions are almost exactly equal, suggesting intermittent observations did not significantly impact our ability to predict survival of future patients.

9 A worked example: Cardiac allograft vasculopathy (CAV) data

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](http://cran.r-project.org/web/packages/msm) maintained by Christopher Jackson. Cardiac allograft vas-
culopathy (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 3. Yearly examinations occurred for up to 18 years following the transplant. Mean follow-up time, however, is 5.9 years. Survival times are observed and/or censored, but CAV state was only observed at appointment times prior to death/censoring. For censored records, the censoring time is assumed to be 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. There was no overlap in the these two patient subsets.

We set \( B = (\{1, 2, 3\}, \{4\}) \), and assume the underlying process is a \( B \)-composable, exchangeable, Markov multi-state survival process. Parameters are \( \{p_{ij,j'}, \lambda_{ij,j'}\}_{i,j,j' = 1,2} \), \( \{\gamma_{l,1}, \gamma_{l,2}\}_{l \in B_1} \), 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, the prior on \( \alpha = \alpha_{2,1} \) is set to a Beta distribution with parameters 2 and 8 respectively. For parameters \( \nu_{11} \) and \( \nu_{12} \), the Gamma prior has hyperparameters 1 and 1. Parameters \( (\gamma_{21}, \gamma_{22}, \gamma_{31}, \gamma_{32}) \) have independent, standard log-normal priors. We use the approximate MCMC sampler to perform inference.

Traceplots in Figure 4a suggest convergence after the first 100 iterations. The MCMC sampler gives posteriors for the parameters. The posterior mean of \( \nu_{11} \) is 0.75 (i.e., marginal time until transition from state 1 to state 2 is 1.33 years). Parameters \( (\gamma_{21}, \gamma_{31}) \) have posterior means (2.43, 0.65), trans-
lating into marginal holding times of 0.59 and 2.02 years respectively. The posterior mean for $\alpha$ 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 $\nu_{12}$ is 0.52 (i.e., marginally the holding time in state 1 until a transition to state 4 is 1.92 years). This suggests that in state 1, disease progression is slightly more likely than failure. Parameters ($\gamma_{22}, \gamma_{32}$) have posterior means (1.10, 1.81) respectively. This translates marginally into holding times of 1.76 and 1.10 years respectively. For state 2, the distribution of $\gamma_{22}$ does overlap 1.0 suggesting that failure transitions from states 1 and 2 may occur at similar rates.

Figure 3: Cardiac allograft vasculopathy (CAV) transition diagram.

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 as well as the Kaplan-Meier survival function estimator. Note at baseline, all patients are in state 1; therefore, the Kaplan-Meier curve should

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

Figure 4: MCMC traceplots and densities for CAV study

We next consider the posterior distributions for the survival functions. Figure 5a plots median survival at each time $t$ over all iterations of the MCMC sampler as well as the Kaplan-Meier survival function estimator. Note at baseline, all patients are in state 1; therefore, the Kaplan-Meier curve should
be compared with the median survival curve given the new patient is in state 1. We see that the posterior survival curve is significantly lower. This reflects expected disease progression since baseline; that is, a new patient in state 1 at baseline who is alive at time \( t \) is likely to have seen a progression in their disease state, leading to an increase in their current risk of failure. Under the multi-state survival process model, the expected survival time from baseline given new patient is in state 1, 2, and 3 is 8.75, 8.32, and 7.32 respectively. Under the Kaplan-Meier estimator, the expected survival time from baseline is 9.66. We have included the 5% and 95% quantiles for the survival function at each time \( t \) when the patient is in state 3 at baseline. Figure 5b plots median survival over all iterations of the MCMC sampler given the user is alive at time \( t = 5 \). Under the multi-state survival process model, the expected survival time from time 5 given new patient is in state 1, 2, and 3 is 5.97, 5.55, and 5.04 respectively. We again include the 5% and 95% quantiles for the survival function at each time \( t \) when the patient is in state 3 at baseline.

![Figure 5b: Survival functions at baseline](image)

![Figure 5b: Survival functions at \( t = 5 \)](image)

Figure 5: Survival functions given “No CAV” (black), “Mild/Moderate CAV” (red), and “Severe CAV” (blue); Kaplan-Meier estimator (dotted black).

10 Concluding remarks

The preceding pages lay a theoretical and methodological foundation for the development of exchangeable, Markov multi-state survival processes. The approach put forward sits between the prior parametric and non-parametric approaches for multi-state survival data. A representation theorem characterizes the entire space of exchangeable, Markov multi-state processes via a measure \( \Sigma \) and constants \( \mathbf{c} \). Restrictions on choice of measure were driven by prior work on weak continuity and the notion of composable systems developed in this paper. To fit the models to intermittently observed multi-state survival
data, extensions of existing MCMC samplers for Markov jump processes were required; an approximate MCMC sampler version was derived that achieved good empirical performance in simulations and led to significant decreases in runtime. Application to cardiac allograft vasculopathy (CAV) data showed how to interpret posterior parameter distributions. Comparison to Kaplan-Meier estimators showed how the model adjusts the non-parametric survival function estimator to account for disease progression.

References

O. Aalen, O. Borgan, R.D. Gill, and N. Keiding. Interaction between life history events: nonparametric analysis of prospective and retrospective data in the present of censoring. *Scand. J. Statist.*, 7:161–171, 1980.

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.

David Aldous. *Probability Distributions on Cladograms*, pages 1–18. Springer, 1996.

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.

D.G. Clayton. A monte carlo method for bayesian inference in frailty models. *Biometrics*, 47:467–485, 1991.

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 measurements 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.

N.L. Hjort. Nonparametric bayes estimators based on beta processes in models for life history data. *Annals of Statistics*, 18:1259–1294, 1990.

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. Aktuarietidskr.*, 36:87–91, 1953.

P. Jepsen, H. Vilstrup, and P.K. Andersen. The clinical course of cirrhosis: The importance of multistate models and competing risks analysis. *Hepatology*, 62(1):292–302, 2015.

J.D. Kalbfleisch. Nonparametric bayesian analysis of survival time data. *J. Roy. Statist. Soc. B*, 40:214–221, 1978.

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. Handbook of missing data methodology. Technical report, National Research Council of the National Academies, Washington, D.C., 2010.
A Proof of discrete-time characterization

We start with a discussion of an equivalent matrix representation used in proofs of both discrete and continuous-time characterizations.

A.1 Matrix equivalent representation

For every \( y \in [s]^N \), there exists an equivalent representation as a matrix with an infinite number of rows and \( k \) columns \( M \in [s]^{N \otimes k} \) where the first row \( M_1 = [y_1, \ldots, y_k] \). Let \( M_{i, i} \in [s]^N \) denote the \( i \)th column of \( M \) (i.e., \( M = [M_{i, 1} | \ldots | M_{i, k}] \)). Exchangeability of \( Y \) implies column exchangeability of \( M \). That is, for a set of permutations \( (\sigma_1, \ldots, \sigma_k) \) such that \( \sigma_i : \mathbb{N} \to \mathbb{N} \) for \( i = 1, \ldots, k \), we have

\[
M = [M_{1, 1} | \ldots | M_{i, k}] \overset{D}{=} [M_{1, 1}^{\sigma_1} | \ldots | M_{i, k}^{\sigma_k}] = M^\sigma
\]

where \( \overset{D}{=} \) stands for equivalent in distribution. We define this property as column-wise exchangeable. Note exchangeability implies column-wise exchangeability but not vice versa. Restriction acts column-wise

\[
M^{[n]} = [M_{1, 1}^{[n]} | \ldots | M_{i, k}^{[n]}]
\]
with such matrices in one-to-one correspondence with elements in \([s]^{[n,k]}\).

We define an action of the matrix representation \(A\) on \(y \in [s]^N\) by \(A(y) = (A_1y_1, A_2y_2, \ldots)\). In other words, the \(i\)th row of \(A\), \(A_i\), acts on \(y_i\) by sending it to \(A_iy_i\). The identity map \(I\) is defined by each row \(I_i\) being equal to \([12\ldots k]\); then \(I(y) = y\) for all \(y \in [s]^N\). The equivalent vector representation of \(I\) is defined as \(\text{id} \in [s]^N\).

We express the asymptotic frequency of \(A\) by \(|A|_k = (|A_1|, \ldots, |A_k|)\) assuming \(|A_i|\) exists.

The proofs below are for the complete graph case. As \(G\) is simply a restriction of the measure to a particular subset of transition matrices \(P_G\), the proofs below yield the desired results.

**Proof of Theorem 1.** By Kolmogorov consistency, \(Y_{[n]}\) is a Markov chain governed by transition probability rules \(\Pr(Y(t) = y' | Y(t-1) = y)\). Restriction to \([n]\) yields a transition rule for \(y, y' \in [s]^{[n]}\):

\[
\Pr_n(Y_{[n]}(t) = y' | Y_{[n]}(t-1) = y) = \Pr(Y(t) = R_{n}^{-1}(y') | Y(t-1) = y^*)
\]

where \(R_{n}\) is the restriction operation so \(R_{n}^{-1}(y') = \{y \in [s]^N \text{ s.t. } R_{n}(y) = y'\}\) and \(y^* \in R_{n}^{-1}(y)\). Without loss of generality, we focus on time \(t = 1\).

We define a measure \(\eta\) by

\[
\eta(\cdot) := \Pr(\cdot | \text{id})
\]

Via the matrix representation, we can think of \(\eta\) as a measure on matrices \(A \in [s]^{N \otimes k}\). Restriction to \([n]\) yields \(A^{[n]} \sim \eta^{[n]}(\cdot) = \Pr_{nk}(\cdot | \text{id}_{nk})\). The action of \(A^{[n]}\) on \(x \in [s]^{[n]}\) is then given by

\[
A^{[n]}(x) = \eta^{[n]}(x) = \Pr_n(\cdot | I_n(x)) = \Pr_n(\cdot | x)
\]

as we require.

The above argument shows that there exists a measure \(\eta\) such that \(Y^*\) defined by

\[
Y^*(t) = (A_t \circ A_{t-1} \circ \ldots A_1)(Y^*_0)
\]

is equivalent in distribution to \(Y\). Here \(A_t\) are independent, identical distributed draws from \(\eta\) for each time \(t \in \mathbb{N}\).

**Proof of Corollary 2.** Consider the recurrent event process \(Y\) up time \(\tau < \infty\). Then \(Y^* = \tau \wedge Y\) is a version of \(Y\) on \(t \in 0, 1, \ldots, \tau\). Let \(\eta_{\tau}\) to denote the
measure associated with $Y^*$. For $P \in \mathcal{P}_\tau$, let $R_{\tau',\tau}(A)$ be the restriction of this $\mathcal{P}_{\tau'}$. Then for $\tau' < \tau$

$$
\eta_{\tau}(\{P \in \mathcal{P}_\tau \mid R_{\tau',\tau}(P) = P^*\}) = \eta_{\tau'}(\{P^*\})
$$

So we have consistency across $\tau > 0$. We define the measure $\eta$ on $\mathcal{P}_\infty$ by

$$
\eta(\cdot) = \lim_{\tau \uparrow \infty} \eta_{\tau}(\cdot)
$$

is the unique measure such that $Y^*$ is a version of $Y$. \hfill \Box

B Proof of continuous-time characterization

Again the proof below is for the complete graph case. As $G$ is simply a restriction of the measure to a particular subset of transition matrices $\mathcal{P}_G$, the proof below yields the desired result.

Proof of Theorem \[\square\] Like in the discrete-case, we construct the measure $\eta$ from the transition rule which governs $Y$. This will connect $Y^*$ to $Y$ such that they are equal in law.

Since $Y_{[n]}$ is a Markov process on $[s]^{[n]}$, it is governed by a transition rate function

$$
Q_n(y, y') = \lim_{t \downarrow 0} \frac{1}{t} \Pr(Y_{[n]}(t) = y' \mid Y_{[n]}(0) = y).
$$

We start by describing the key characteristics of the transition rate function

1. The transition rate function exhibits finite activity:

$$
\sum_{y' \neq y} Q_n(y, y') < \infty
$$

2. The transition rate function is exchangeable. That is, for any $\sigma : \mathbb{N} \to \mathbb{N}$ and $y \neq y'$:

$$
Q_n(y, y') = Q_n(y^\sigma, (y')^\sigma).
$$

3. The transition rate functions are consistent. That is, for $y, y' \in [s]^{[m]}$ and $m \leq n$,

$$
Q_m(y, y') = Q_n(Y^*, R_{m,n}^{-1}(y'))
$$

where $R_{m,n}^{-1}$ is the inverse of the restriction operator from $[n]$ to $[m]$ and $y^* \in R_{m,n}^{-1}(y)$. 

27
We then define the measure for $A \in [s]^{[n] \times s} \setminus \{id_{k,n}\}$ as

$$\eta_n(A) = Q_n(id_{k,n}, A)$$

This measure is column-wise exchangeable by exchangeability of $Q$ and satisfies

$$\eta_n(A) = \eta(A^* : [s]^{[n] \times s} \mid (A^*)^{[n]} = A)$$

(10)

for all $m \leq n$ and $A \in [s]^{N \times s}$ by consistency of $Q$.

The measure $\eta(A) = Q_n(id_k, A)$ is also column-wise exchangeable and satisfies

\begin{itemize}
  \item $\eta(\{id_k\}) = 0$ (i.e., a transition must occur) and
  \item $\eta(A \mid A^{[n]} \neq id_{k,n}) < \infty$ (i.e., finite, restricted activity)
\end{itemize}

Following [Pitman, 2003], we construct a process $Y^* = (Y^*(t), t \geq 0)$ via its finite restrictions $Y^*_{[n]} = (Y^*_{[n]}(t), t \geq 0)$. Let $A = \{(t, A) \subset \mathbb{R}^+ \times [s]^{N \otimes k}\}$ be a Poisson point process with intensity $dt \otimes \eta$. Given an initial state $Y^*(0)$, for each $t > 0$ if $t$ is an atom of $A$ then

\begin{itemize}
  \item if $A^*_{[n]} \neq id_{k,n}$, then set $Y^*_{[n]}(t) = A^*_{[n]}(Y^*_{[n]}(t-))$
  \item otherwise $Y^*_{[n]}(t) = Y^*_{[n]}(t-)$
\end{itemize}

The difference between the continuous and discrete-time setting is the random time between jumps and that the jumps (1) occur for an infinite fraction of the units as $n \to \infty$, or (2) occur for a single unit $u \in \mathbb{N}$. By construction for $m \geq n$, the restriction of $Y^*_{[n]}$ to $[n]$ is consistent with $Y^*_{[n]}$ so we have $Y^*$ is a unique $[s]^N$-valued process.

**Lemma 1.** The process $Y^*$ is a Markov, exchangeable state-space process.

*Proof.* Consistency is given by the above argument; the process is Markovian by construction and the assumptions on $Y$. Exchangeability is due to $\eta$ being a column-wise exchangeable measure since $\eta_{[n]}$ are finite, column-wise exchangeable measures. \qed

The final concern before showing that $Y^*$ is stochastically equivalent to $Y$ is the uniqueness of the measure $\eta$ related to the restricted measures $\eta_n$.

**Lemma 2.** There exists unique measure $\eta$ on $[s]^{N \otimes s}$ which satisfies (1) $\eta(\{id_k\}) = 0$, (2) $\eta(A \in [s]^{N \otimes s} \mid A^{[n]} \neq id_{k,n}) < \infty$, and

$$\eta(A^* \in [s]^{N \otimes s} \mid (A^*)^{[n]} = A) = \eta_n(A)$$

(11)

for all $n > 0$ and $A \in [s]^{[n] \otimes s}$.  

28
Proof. The sets
\[ \{ A^* \in [s]^{N \otimes s} \mid (A^*)^{[n]} = A \} \]
are a \( \pi \)-system generating the \( \sigma \)-field on \([s]^{N \otimes s}\). The above discussion proves equation [11] and the measure is additive. Therefore, uniqueness is a consequence of any measure extended to a \( \sigma \)-algebra being unique if the measure is \( \sigma \)-finite.

Lemma 3. \( Y^* \) is a version of \( Y \).

Proof. First, the finite restrictions \( \eta_n \) satisfies
\[
\eta_n(\{ A \in [s]^{[n] \otimes s} \mid A(y) = y' \}) = \sum_{A : A(y) = y'} Q_n(I_{k,n}, A) = Q_n(I_{k,n}(y), A(y)) = Q_n(y, y').
\]
And has finite activity:
\[
\eta_n(\{ A \in [s]^{[n] \otimes s} \mid A(y) = y \}) = \sum_{y' \neq y} Q_n(y, y') < \infty
\]
Therefore \((Y^*)^{[n]}\) is an Markov, exchangeable process with jump rates \( Q_n(\cdot, \cdot) \). By Kolmogorov’s extension theorem, the unique process \( Y^* \) is a version \( Y \).

We still need to show \( \eta \) can be decomposed into the respective components:

- **Dislocation measure**: measure on \( s \times s \) transition matrices\( \sim \Sigma \) which satisfies
  \[
  \Sigma(\{I_k\}) = 0
  \]
  \[
  \int_{P_k} (1 - P_{\min}) \Sigma(dP) < \infty
  \]
  where \( P_{\min} = \min_i P_{i,i} \).

- **Erosion measures**: Let \( A \in [s]^{N \times s} \) then we call this \( A = \text{id} \) and we flip a single unite \( u \in N \) (i.e., \( A_{u,i} = i' \)). Let \( \mu_{i,i'}^u \) be this point mass measure. Define
  \[
  \mu_{i,i'} = \sum_{u \in \mathbb{N}} \mu_{i,i'}^u
  \]
- The combined measure is given by:
  \[
  \eta_{\Sigma,c}(\cdot) = \mu_{\Sigma}(\cdot) + \sum_{i \neq i' \in [s]} c_{i,i'} \mu_{i,i'}(\cdot)
  \]
Lemma 4. The measure \( \eta_{\Sigma,c} \) is a column-wise exchangeable measure satisfying the necessary constraints.

Proof. We prove this for each component of \( \eta_{\Sigma,c} \). First, \( \mu_{\Sigma}({\{id_k\}}) = 0 \) by construction. Moreover, for \( P \in \mathcal{P}_k \)

\[
\mu_P({\{A | A^{[n]} \neq id_{s,n}\}}) \leq \sum_{j=1}^s \mu_P({\{A | A^{[n]} \neq j \text{ for all } u \in [n]\}})
\]

\[
\leq k(1 - p_{\min}^n) \leq n \cdot k(1 - p_{\min}^n)
\]

which implies

\[
\mu_{\Sigma}({\{A | A^{[n]} \neq id_{s,n}\}}) \leq n \cdot k \int_{\mathcal{P}_k} (1 - p_{\min}) \Sigma(dP) < \infty
\]

by the above assumptions.

Second, \( \mu_{i,i'}({\{id_k\}}) = 0 \) by construction. Moreover,

\[
\sum_{i \neq i' \in [s]} c_{ii'} \mu_{i,i'}({\{A | A^{[n]} \neq id_{s,n}\}}) \leq c_{\max} \sum_{i \neq i' \in [s]} \sum_{u \in [n]} \mu_{i,i'}^{[u]}({\{A | A^{[n]} \neq id_{s,n}\}})
\]

\[
\leq c_{\max} \left( \frac{s}{2} \right) n < \infty.
\]

\[\square\]

So we have that \( Y^* \) is a version of \( Y \) and for any \( \Sigma \) and \( c \) the measure is \( \mu_{\Sigma,c} \) is column-wise exchangeable satisfying necessary constraints. It rests to connect show that the measure \( \eta \) can be decomposed such that there exists \( \Sigma \) and \( c \) such that \( \eta = \mu_{\Sigma,c} \).

Lemma 5. For \( \eta \) constructed from \( Q \), \( \eta \)-almost every \( A \in [s]^{\mathbb{N} \otimes s} \) possesses asymptotic frequencies \( |A|_s \in \mathcal{P}_s \).

Proof. The \( \eta \) by construction satisfies the necessary conditions. We set \( \eta_n = \eta \) on \( \{A | A^{[n]} \neq id_{s,n}\} \). Then \( \eta_n \) is column-wise exchangeable for \( (\sigma_1, \ldots, \sigma_s) \) such that \( \sigma_i : \mathbb{N} \to \mathbb{N} \) fixes \([n]\).

We can find a column-wise exchangeable measure by simply considering ignoring the first \( n \) rows. Let \( \eta'_n \) be measure obtained from \( \eta \) after applying the \( n \)-shift function \( \phi_n : A \rightarrow A' \). Then \( \eta'_n \) is column-wise exchangeable and therefore has asymptotic frequencies. But asymptotic frequencies only depend
on such an n-shift for every fixed n > 0 (i.e., |A|_s = |φ_n(A)|_s); therefore, η_n-almost every A has asymptotic frequencies.

To prove η-almost every A has asymptotic frequencies we simply note that η_n ↑ η and therefore the monotone convergence theorem completes proof.

Lemma 6. There exists a measure Σ such that the restriction of η to \{ |A|_s \neq I_s \} is equivalent to μ_Σ.

Proof. Let φ_n(A)^[m] denote the restriction of φ_n(A) to [m]. Then

\[
\eta_n(\phi_n(A)^[2] \neq id_{s,2} \mid |A|_s = P) = \eta_n^\phi(\{\phi_n(A)^[2] \neq id_{s,2} \mid |A|_s = P) \\
= \eta_n^\phi(\{A^2 \neq id_{s,2} \mid |A|_s = P) \\
= \mu_P(\{A^2 \neq id_{s,2}\}) \\
\geq 1 - p_{\min}^2 \geq 1 - p
\]

\[ \Rightarrow \eta_n(\{A^2 \neq id_{s,2}\}) \geq \int_{P_k} (1 - p_{\min}) \Sigma_n(dP) \]

with \( \Sigma_n = \eta_n 1[|A|_k \neq I_k] \). As \( n \to \infty \) this yields,

\[ \infty > \eta(\{\phi(A)^[2] \neq id_{s,2}\}) \geq \int_{P_s} (1 - p_{\min}) \Sigma(dP) \]

and \( \Sigma(\{id_s\}) = 0 \) by construction.

It rests to show that \( \mu_\Sigma = 1[|A|_s \neq id_s] \eta \). We have

\[ \eta(\{A[n] = A^*, |A|_s \neq id_s\}) = \lim_{m \to \infty} \eta_m(\{A[n] = A^*, A[m] \neq id_{s,m}, |A|_s \neq id_s\}) \]

The right hand side is equivalent to

\[
\eta_m^\phi(\{A[n] = A^*, |A|_s \neq id_s\}) = \int_{P_s} \mu_P(\{A[n] = A^*\}) |\eta_m^\phi(\mid|A|_s \in dP) \\
= \int_{P_k} \mu_P(\{A[n] = A^*\}) \Sigma(dP) \\
= \mu_\Sigma(\{A[n] = A^*\})
\]

Lemma 7. There exists a set of constants \( \{c_{i,i'}\}_{i \neq i' \in [s]} \) such that the restriction of η to \{ |A|_s = I_s \} is equivalent to \( \mu_\Sigma(\cdot) = \sum_{i \neq i' \in [s]} c_{i,i'} \mu_{i,i'}(\cdot) \).
Proof. We restrict our attention to the set of $A$ where $A^{[2]} \neq \text{id}_{s,2}$ but $\phi_3(A) = \text{id}_s$. This set $B$ contains all single unit transition. As the measure $\eta_3^{\phi}$ is proportional to the point mass at $\text{id}_s$, then $\eta$ restricted to the event $\{A^{[2]} \neq \text{id}_{s,2}, \phi_3(A) = \text{id}_s, |A|_s = \text{id}_s\}$ is the sum
\[ \sum_{A \in B} c_A \delta_A (\cdot). \]

If $A$ contains more than a single unit transition, exchangeability forces $c_A = 0$ since $\eta(\{A|A^{[2]} \neq \text{id}_{s,2}\}) < \infty$. The same argument shows $A \in [s]^{\mathbb{N} \otimes s}$ such that $|A|_s = I_s$ and $c_A > 0$ implies $A$ is a single unit transition. \hfill $\square$

This concludes the proof. \hfill $\square$

C Examples

Here, we describe several important examples that motivate the current study of multi-state survival processes.

Example 1 (Survival process). A survival process has state space $\{\text{Alive, Dead}\}$ with transitions governed by the simple graph shown in Figure 6. In this case, $s = 2$ and the edge-set is the singleton $\{(1, 2)\}$; because the state “Dead” is absorbing, the space $\mathcal{P}_G$ is equivalent to the one-dimensional space $p \in [0, 1)$. Restricting to $[n]$, suppose that all individuals at time $t$ are still at risk. In discrete-time, the probability of $d$ individuals passing away between times $t$ and $t + 1$ is equal to

\[ \int_0^1 p^{n-d}(1 - p)^d \Sigma(dp) \]

where $\Sigma$ is a probability measure on $(0, 1]$. The marginal distribution of the survival time for each patient is geometric. Letting $\Sigma(dp)$ be the conjugate prior $\nu \cdot p^{\alpha-1}(1 - p)^{\beta-1} dp$ with $\alpha, \beta > 0$ yields a discrete-version of the “beta-splitting” process [Aldous, 1996]. The marginal geometric distribution in this case has parameter $\beta/(\alpha + \beta)$.

In continuous-time, the probability of $d$ individuals passing away between times $t$ and $t + 1$ is proportional to

\[ \int_0^1 p^{n-d}(1 - p)^d \Sigma(dp) + \delta(d = 1)c_{1,2} \]
Figure 6: Graph representation of survival process

where $\Sigma$ is a measure on $(0, 1]$ satisfying $\int (1 - p)\Sigma(dp) < \infty$. In continuous-time, the marginal distribution is exponential. The conjugate prior now relaxes the constraints to $\beta > -1$. Considering choice of measure, Dempsey and McCullagh [2017] suggest choosing measure with $\beta = 0$ – called the harmonic process. The harmonic process is the only family of Markov survival processes with weakly continuous predictive distributions – a key property in applied work. The chance of singleton events is set to zero (i.e., $c_{1,2} = 0$).

**Example 2** (Illness-death process). The illness-death process has state space $\{\text{Healthy}, \text{Unhealthy}, \text{Dead}\}$ with transitions governed by the simple graph shown in Figure 7. The state “Dead” (i.e., $s = 3$) is absorbing, the space $P_G$ is equivalent to a three-dimensional space. The bi-directional illness-death process includes the additional edge (Unhealthy,Healthy), allowing the patient to recover. Both processes can be viewed as refinements of the survival process.

Figure 7: Graph representation of the illness-death process

An issue arises for the bi-directional illness-death process when the definitions of “Healthy” and “Unhealthy” are arbitrary (i.e., have no scientific value). Potentially the labels are exchangeable and, if so, the process is a Markov exchangeable survival process. Such considerations lead to natural constraints on the choice of measure – see section 6 for a discussion.

**Example 3** (Comorbidities). Comorbidities are multiple stochastic processes experienced simultaneously by the same patient. Figure 8 for example, represents $L$ binary risk processes each with an absorbing state. In general, $Y(i, t) = (Y_1(i,t), \ldots, Y_L(i,t))$ is an $L$-vector state-space process.

An example of comorbidities is provided by Aalen et al. [1980] where two events (onset of menopause, and occurrence of chronic skin disease) were studied. Patients could also experience a third event, death. In this case, we
have $L = 2$ binary risk processes each with absorbing states and then a final absorbing state of death.

**Example 4** (Competing risks). A patient may experience failure for a multitude of reasons. Figure 9 shows a setting where failure can be caused by $L$ risks. Unlike comorbidities, a patient may only experience one of the competing risks.

**Example 5** (Recurrent events). Recurrent events are events that occur more than once per patient. Examples include recurring hospital admissions, tumor recurrence, and repeated heart attacks. For recurrent events, the state-space is countably infinite; however, the state-space is structured. We can assume all patients initial values are zero (i.e. $Y(u, 0) = 0$) and given $Y(u, t) = k$ then the transition at the next jump time must be to $k + 1$. This structure allows us to provide the following discrete and continuous-time characterizations of the exchangeable, Markov recurrent event processes – extending Theorems 1 and 2.

**Corollary 2** (Discrete-time characterization). Let $Y = (Y(t), t \in \mathbb{N})$ be a discrete-time Markov, exchangeable recurrent event process. Then there exists a probability measure $\Sigma$ on $[0, 1]^\mathbb{N}$ such that $Y^*_\Sigma$ is a version of $Y$.

**Proof of Corollary 2.** Consider the recurrent event process $Y$ up time $\tau < \infty$. Then $Y^* = \tau \wedge Y$ is a version of $Y$ on $t \in 0, 1, \ldots, \tau$. Let $\eta_r$ to denote the
measure associated with \( Y^* \). For \( P \in \mathcal{P}_\tau \), let \( R_{\tau',\tau}(A) \) be the restriction of this \( P_{\tau'} \). Then for \( \tau' < \tau \)

\[
\eta_{\tau'}(\{P \in \mathcal{P}_\tau \mid R_{\tau',\tau}(P) = P^*\}) = \eta_{\tau'}(\{P^*\})
\]

So we have consistency across \( \tau > 0 \). We define the measure \( \eta \) on \( \mathcal{P}_\infty \) by

\[
\eta(\cdot) = \lim_{\tau \uparrow \infty} \eta_{\tau}(\cdot)
\]

is the unique measure such that \( Y^* \) is a version of \( Y \).

\[ \square \]

**Corollary 3 (Continuous-time characterization)**. Let \( Y = (Y(t), t \in \mathbb{R}^+) \) be a continuous-time Markov, exchangeable recurrent event process such that \( Y_u(0) = 0 \) for all \( u \in \mathbb{N} \). Let \( I \) denote the infinite identity matrix. Then there exists a probability measure \( \Sigma \) on \( [0, 1]^\mathbb{N} \) satisfying

\[
\Sigma(\{I\}) = 0 \quad \text{and} \quad \int_{[0,1]^\mathbb{N}} (1 - P_{\min}) \Sigma(dp) < \infty \quad \text{where} \quad P_{\min} = \min_{i \in \mathbb{N}} P_i
\]

and a set of constants \( c = \{c_{i,i+1} \mid i \in \mathbb{N}\} \) such that \( Y^{\Sigma,c}_t \) is a version of \( Y \).

A similar proof can be constructed for the continuous-time setting and is therefore omitted.

\section{D Details on choice of measure}

Let \( Z \) be a positive, stationary Lévy process on \( \mathbb{R}_+ \). As these processes are positive, it is natural to work with the cumulant function

\[
K(t) = \log \left( \mathbb{E} \left[ e^{-Z(t)} \right] \right) = \log \left( \mathbb{E} \left[ e^{-tX} \right] \right)
\]

for \( t \geq 1 \) and \( X = Z(1) \) is an infinitely divisible distribution. The Lévy-Khintchine characterization for positive, stationary, Lévy processes implies

\[
K(t) = - \left[ \gamma t + \int_0^\infty (1 - e^{-ty})w(dy) \right]
\]

for some \( \gamma \geq 0 \) and measure \( w(\cdot) \) on \( \mathbb{R}_+ \), called the Lévy measure, such that the integral is finite for all \( t > 0 \).

\cite{DempseyMcCullagh2017} showed that every exchangeable, Markov survival process can be generated via a Lévy process construction. The proof stems from connecting \( \gamma \) to the erosion measures (i.e., \( c \geq 0 \) in Theorem (2))
and the Lévy measure to the dislocation measures (i.e., $\Sigma(\cdot)$ in Theorem (2)). For instance, the harmonic process can be constructed via a Lévy process with $\gamma = 0$ and $w(dy) = \nu e^{-\rho y} dy/(1 - e^{-y})$.

Now consider the proportional conditional hazards model as described by Kalbfleisch [1978], Hjort [1990], and Clayton [1991]. In the proportional conditional hazards model, the hazard for individual $i$ is $w_i Z(t)$ for some $w_i > 0$ typically $w_i = \exp(x_i' \beta)$ depending on a set of baseline covariates $x_i$. Then the conditional survival density for particle $i$ is $\exp \left( -w_i \int_0^t Z(t) \right) \left( 1 - e^{-w_i Z(t)} \right)$.

Assume there is a single covariate that is a factor with a finite number of levels (i.e., $x_i \in \{1, \ldots, k\} := [k]$). Then $w_i = w_{x_i}$; that is, there are a finite set of weights. The joint marginal density can be derived in a similar way as before. Here, however, the non-normalized transition rules are

$$
\lambda(R, D) = \mathbb{E} \left( e^{-Z(t) \sum_{i \in R} w_{x_i} \prod_{i \in D} (1 - e^{-w_{x_i} Z(t)})} \right)
= \mathbb{E} \left( \prod_{j=1}^k \left( \exp(-Z(t))^{w_j} \right)^{r_j} \left( 1 - \exp(-Z(t))^{w_j} \right)^{d_j} \right)
= \int_0^1 \prod_{j=1}^k (p^{w_j})^{r_j} \left( 1 - p^{w_j} \right)^{d_j} \Sigma(dp)
$$

where $r_j = \# \{ i \in R \text{ s.t. } X_i = j \}$, $d_j = \# \{ i \in D \text{ s.t. } X_i = j \}$, and $\Sigma(\cdot)$ is the dislocation measure. The final equality is due to the connection between the Lévy measure and the dislocation measure. It is clear from above that the proportional conditional hazards model corresponds to a particular choice of the dislocation measure. Namely, the proportional conditional hazards model corresponds to $p_i \rightarrow p^{w_{x_i}}$. So on the $[0, 1]$-scale, the model is conditionally proportional on the log-scale. That is, $\log(p_i) = w_{x_i} \log(p)$. Alternative choices exist. For example, the model may be conditionally proportional on the logistic scale; that is, $\log(p_i/(1 - p_i)) = w_{x_i} \log(p/(1 - p))$. We do not pursue such alternatives in this paper.