Exchangeable Markov multi-state survival processes

University of Michigan

Supplementary Material

S1 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.

S1.1 Matrix equivalent representation

For every $y \in [s]^{\mathbb{N}}$, there exists an equivalent representation as a matrix with an infinite number of rows and k columns $M \in [s]^{\mathbb{N} \otimes k}$ where the first row $M_{1,\cdot} = [y_1, \ldots, y_k]$. Let $M_{\cdot,i} \in [s]^{\mathbb{N}}$ denote the ith column of M (i.e., $M = [M_{\cdot,1} \,|\, \ldots \,|\, M_{\cdot,k}]$. Exchangeability of $\mathbf{Y}$ implies column exchangeability of $\mathbf{M}$. 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

$$\mathbf{M} = [\mathbf{M}_{\cdot,1} \,|\, \ldots \,|\, \mathbf{M}_{\cdot,k}] \overset{D}{=} [\mathbf{M}^{\sigma_1}_{\cdot,1} \,|\, \ldots \,|\, \mathbf{M}^{\sigma_k}_{\cdot,k}] = \mathbf{M}^{\sigma}$$

where $\stackrel{D}{=}$ 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

$$\mathbf{M}^{[n]} = [\mathbf{M}^{[n]}_{\cdot,1} \mid \ldots \mid \mathbf{M}^{[n]}_{\cdot,k}]$$

with such matrices in one-to-one correspondence with elements in $[s]^{[n \cdot k]}$.

We define an action of the matrix representation A on $y \in [s]^{\mathbb{N}}$ by $A(y) = (A_{1,y_1}, A_{2,y_2}, \ldots)$. In other words, the ith row of A, $A_{i,\cdot}$, acts on y_i by sending it to A_{i,y_i}. The identity map I is defined by each row $I_{i,\cdot}$ being equal to $[12 \ldots k]$; then $I(y) = y$ for all $y \in [s]^{\mathbb{N}}$. The equivalent vector representation of I is defined as $\mathrm{id} \in [s]^{\mathbb{N}}$.

We express the *asymptotic frequency* of A by k-vector $|A|_k = (|A_1|, \ldots, |A_k|)$ 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 $\mathcal{P}_G$, the proofs below yield the desired results.

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

$$\mathrm{pr}_n(\mathbf{Y}_{[n]}(t) = y' \mid \mathbf{Y}_{[n]}(t-1) = y) = \mathrm{pr}(\mathbf{Y}(t) = R_n^{-1}(y') \mid \mathbf{Y}(t-1) = y^\star)$$

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

We define a measure η by

$$\eta(\cdot) := \mathrm{pr}(\cdot \mid \mathrm{id})$$

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

$$A^{[n]}(x) = \eta^{[n]}(x) = \mathrm{pr}_n(\cdot \mid I_n(x)) = \mathrm{pr}_n(\cdot \mid x)$$

as we require.

The above argument shows that there exists a measure η such that $\mathbf{Y}^\star$ defined by

$$\mathbf{Y}^\star(t) = (A_t \circ A_{t-1} \circ \dots A_1)(Y_0^\star)$$

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

$\square$

Proof of Corollary S1. Consider the recurrent event process $\mathbf{Y}$ up time $\tau < \infty$. Then $\mathbf{Y}^\star = \tau \wedge \mathbf{Y}$ is a version of $\mathbf{Y}$ on $t \in 0, 1, \dots, \tau$. Let η_τ to denote the measure associated with $\mathbf{Y}^\star$. For $P \in \mathcal{P}_\tau$, let $R_{\tau',\tau}(A)$ be the restriction of this $\mathcal{P}_{\tau'}$. Then for $\tau' < \tau$

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

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

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

is the unique measure such that $\mathbf{Y}^\star$ is a version of $\mathbf{Y}$. □

S2 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 2. Like in the discrete-case, we construct the measure η from the transition rule which governs $\mathbf{Y}$. This will connect $\mathbf{Y}^\star$ to $\mathbf{Y}$ such that they are equal in law.

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

$$Q_n(y, y') = \lim_{t \downarrow 0} \frac{1}{t} \text{pr}(\mathbf{Y}_{[n]}(t) = y' \mid \mathbf{Y}_{[n]}(0) = y).$$

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 $m \leq n$,

$$Q_m(y, y') = Q_n(y^\star, R^{-1}_{m,n}(y'))$$

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

We then define the measure for $A \in [s]^{[n] \times s} \backslash \{\mathrm{id}_{k,n}\}$ as

$$\eta_n(A) = Q_n(\mathrm{id}_{k,n}, A)$$

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

$$\eta_n(A) = \eta\left(\{A^\star : [s]^{[n] \times s} \mid (A^\star)^{[n]} = A\}\right) \tag{S2.1}$$

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

The measure $\eta(A) = Q_n(\mathrm{id}_k, A)$ is also column-wise exchangeable and satisfies

- $\eta(\{\mathrm{id}_k\}) = 0$ (i.e., a transition must occur) and

- $\eta(\{A \mid A^{[n]} \neq \mathrm{id}_{k,n}) < \infty$ (i.e., finite, restricted activity)

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

- if $A_t^{[n]} \neq \mathrm{id}_{k,n}$, then set $\mathbf{Y}^\star_{[n]}(t) = A_t^{[n]}(\mathbf{Y}^\star_{[n]}(t-))$

- otherwise $\mathbf{Y}^\star_{[n]}(t) = \mathbf{Y}^\star_{[n]}(t-)$

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 $Y^\star_{[m]}$ to $[n]$ is consistent with $Y^\star_{[n]}$ so we have $\mathbf{Y}^\star$ is a unique $[s]^{\mathbb{N}}$-valued process.

Lemma S1. *The process $\mathbf{Y}^\star$ 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 $\mathbf{Y}$. Exchangeability is due to η being a column-wise exchangeable measure since $\eta_{[n]}$ are finite, column-wise exchangeable measures. $\square$

The final concern before showing that $\mathbf{Y}^\star$ is stochastically equivalent to $\mathbf{Y}$ is the uniqueness of the measure η related to the restricted measures η_n.

Lemma S2. *There exists unique measure η on $[s]^{\mathbb{N}\otimes s}$ which satisfies (1) $\eta(\{id_k\}) = 0$, (2) $\eta(\{A \in [s]^{\mathbb{N}\otimes s} \mid A^{[n]} \neq id_{k,n}\}) < \infty$, and*

$$\eta(\{A^\star \in [s]^{\mathbb{N}\otimes s} \mid (A^\star)^{[n]} = A\}) = \eta_n(A) \tag{S2.2}$$

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

Proof. The sets

$$\{A^\star \in [s]^{\mathbb{N}\otimes s} \mid (A^\star)^{[n]} = A\}$$

are a π-system generating the σ-field on $[s]^{\mathbb{N}\otimes s}$. The above discussion proves equation S2.2; and the measure is additive. Therefore, uniqueness is a consequence of any measure extended to a σ-algebra being unique if the measure is σ-finite. $\square$

Lemma S3. $\mathbf{Y}^\star$ *is a version of* $\mathbf{Y}$.

Proof. First, the finite restrictions η_n satisfies

$$\begin{aligned}\eta_n(\{A \in [s]^{[n]\otimes s} \mid A(y) = y'\}) &= \sum_{A: A(y)=y'} Q_{nk}(I_{k,n}, A) \\ &= Q_n(I_{k,n}(y), A(y)) = Q_n(y, y').\end{aligned}$$

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 $(\mathbf{Y}^\star)^{[n]}$ is an Markov, exchangeable process with jump rates $Q_n(\cdot, \cdot)$. By Kolmogorov's extension theorem, the unique process $\mathbf{Y}^\star$ is a version $\mathbf{Y}$. □

We still need to show η can be decomposed into the respective components:

- **Dislocation measure**: measure on $s \times s$ transition matrices$\sim \Sigma$ which satisfies

$$\Sigma(\{I_k\}) = 0$$
$$\int_{\mathcal{P}_k} (1 - P_{\min})\Sigma(dP) < \infty$$

  where $P_{\min} = \min_i P_{i,i}$.

- **Erosion measures**: Let $A \in [s]^{\mathbb{N}\times s}$ then we call this $A = \text{id}$ and we flip a single unite $u \in \mathbb{N}$ (i.e., $A_{u,i} = i'$). Let $\mu^u_{i,i'}$ be this point mass measure. Define

$$\mu_{i,i'} = \sum_{u \in \mathbb{N}} \mu^u_{i,i'}$$

- The combined measure is given by:

$$\eta_{\Sigma,c}(\cdot) = \mu_\Sigma(\cdot) + \sum_{i \neq i' \in [s]} c_{ii'}\mu_{i,i'}(\cdot)$$

Lemma S4. *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, $\mu_\Sigma(\{\mathrm{id}_k\}) = 0$ by construction. Moreover, for $P \in \mathcal{P}_k$

$$\begin{aligned}\mu_P(\{A \mid A^{[n]} \neq \mathrm{id}_{s,n}\}) &\leq \mu_P(\{A \mid A^{[n]} \neq \mathrm{id}_{s,n}) \\ &\leq \sum_{j=1}^{s} \mu_P(\{A \mid A^{[n]}_{u,j} \neq j \text{for all } u \in [n]\}) \\ &\leq k(1 - p^n_{\min}) \leq n \cdot k(1 - p^n_{\min})\end{aligned}$$

which implies

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

by the above assumptions.

Second, $\mu_{i,i'}(\{\mathrm{id}_k\}) = 0$ by construction. Moreover,

$$\begin{aligned}\sum_{i \neq i' \in [s]} c_{ii'} \mu_{i,i'}(\{A \mid A^{[n]} \neq \mathrm{id}_{s,n}\}) &\leq c_{\max} \sum_{i \neq i' \in [s]} \sum_{u \in [n]} \mu^u_{i,i'}(\{A \mid A^{[n]} \neq \mathrm{id}_{s,n}\}) \\ &\leq c_{\max} \binom{s}{2} n < \infty.\end{aligned}$$

□

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

Lemma S5. *For η constructed from Q, η-almost every $A \in [s]^{\mathbb{N}\otimes s}$ possesses asymptotic frequencies $|A|_s \in \mathcal{P}_s$.*

Proof. The η by construction satisfies the necessary conditions. We set $\eta_n = \eta$ on $\{A \mid A^{[n]} \neq \mathrm{id}_{s,n}\}$. Then η_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 η_n' be measure obtained from η after applying the *n-shift function* $\phi_n : A \to A'$. Then η_n^{ϕ} 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., $|A|_s = |\phi_n(A)|_s$); therefore, η_n-almost every A has asymptotic frequencies.

To prove η-almost every A has asymptotic frequencies we simply note that $\eta_n \uparrow \eta$ and therefore the monotone convergence theorem completes proof. $\square$

Lemma S6. *There exists a measure Σ such that the restriction of η to $\{|A|_s \neq I_s\}$ is equivalent to μ_{Σ}.*

Proof. Let $\phi_n(A)^{[m]}$ denote the restriction of $\phi_n(A)$ to $[m]$. Then

$$\begin{aligned}
\eta_n(\{\phi_n(A)^{[2]} \neq \mathrm{id}_{s,2}\} \,|\, |A|_s = P) &= \eta_n^{\phi}(\{\phi_n(A)^{[2]} \neq \mathrm{id}_{s,2}\} \,|\, |A|_s = P) \\
&= \eta_n^{\phi}(\{A^{[2]} \neq \mathrm{id}_{s,2}\} \,|\, |A|_s = P) \\
&= \mu_P(\{A^{[2]} \neq \mathrm{id}_{s,2}\}) \\
&\geq 1 - p_{\min}^2 \geq 1 - p \\
\Rightarrow \eta_n(\{A^{[2]} \neq \mathrm{id}_{s,2}\}) &\geq \int_{\mathcal{P}_k} (1 - p_{\min})\Sigma_n(dP)
\end{aligned}$$

with $\Sigma_n = \eta_n \mathbf{1}[|A|_k \neq I_k]$. As $n \to \infty$ this yields,

$$\infty > \eta(\{\phi(A)^{[2]} \neq \mathrm{id}_{s,2}\}) \geq \int_{\mathcal{P}_s} (1 - p_{\min})\Sigma(dP)$$

and $\Sigma(\{\mathrm{id}_s\}) = 0$ by construction.

It rests to show that $\mu_\Sigma = \mathbf{1}[|A|_s \neq \mathrm{id}_s]\eta$. We have

$$\eta(\{A^{[n]} = A^\star, |A|_s \neq \mathrm{id}_s\}) = \lim_{m \uparrow \infty} \eta_m(\{A^{[n]} = A^\star, A^{[m]} \neq \mathrm{id}_{s,m}, |A|_s \neq \mathrm{id}_s\})$$

The right hand side is equivalent to

$$\begin{aligned}
\eta_m^{\phi}(\{A^{[n]} = A^\star, |A|_s \neq \mathrm{id}_s\}) &= \int_{\mathcal{P}_s} \mu_P(\{A^{[n]} = A^\star\})|\eta_m^{\phi}|(|A|_s \in dP) \\
&= \int_{\mathcal{P}_k} \mu_P(\{A^{[n]} = A^\star\})\Sigma(dP) \\
&= \mu_\Sigma(\{A^{[n]} = A^\star\})
\end{aligned}$$

$\square$

Lemma S7. *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 $\mu_c(\cdot) = \sum_{i\neq i'\in[s]} c_{ii'}\mu_{i,i'}(\cdot)$.*

Proof. We restrict our attention to the set of A where $A^{[2]} \neq \mathrm{id}_{s,2}$ but $\phi_3(A) = \mathrm{id}_s$. This set B contains all single unit transition. As the measure η_3^ϕ is proportional to the point mass at id_s, then η restricted to the event $\{A^{[2]} \neq \mathrm{id}_{s,2}, \phi_3(A) = \mathrm{id}_s, |A|_s = \mathrm{id}_s\}$ 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 $\eta(\{A \mid A^{[2]} \neq \mathrm{id}_{s,2}\}) < \infty$. 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. □

This concludes the proof.

□

S3 Examples

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

Example S1 (Survival process)**.** A survival process has state space { Alive, Dead } with transitions governed by the simple graph shown in Figure S1. 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 Σ 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 $\beta/(\alpha + \beta)$.

Figure S1: Graph representation of survival process

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

$$\int_0^1 p^{n-d}(1-p)^d \Sigma(dp) + \delta(d = 1)c_{1,2}$$

where Σ 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 S2 (Illness-death process)**.** The illness-death process has state space { Healthy, Unhealthy, Dead} with transitions governed by the simple graph shown in Figure S2. The state "Dead" (i.e., $s = 3$) is absorbing, the space $\mathcal{P}_G$ is equivalent to a three-dimensional space. The *bi-directional* illness-death process includes the additional edge (Unhealthy,Health), allowing the patient to recover. Both processes can be viewed as refinements of the survival process.

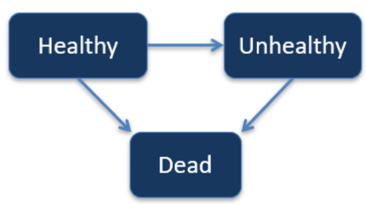


Figure S2: Graph representation of the illness-death process

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 S3 (Comorbidities)**.** Comorbidities are multiple stochastic processes experienced simultaneously by the same patient. Figure S3, 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.

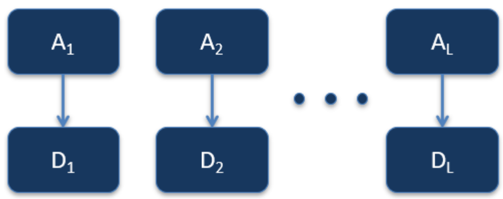


Figure S3: Graph representation of co-morbidities process

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 S4 (Competing risks)**.** A patient may experience failure for a multitude of reasons. Figure S4 shows a setting where failure can be caused by L risks. Unlike comorbidities, a patient may only experience one of the competing risks.

Example S5 (Recurrent events)**.** Recurrent events are events that occur more than once per patient. Examples include recurring hospital admis-

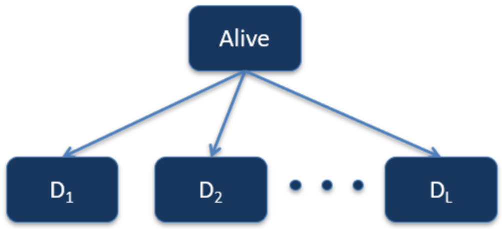


Figure S4: Graph representation of competing risks process

sions, 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 S1 (Discrete-time characterization)**.** *Let* $\mathbf{Y} = (\mathbf{Y}(t), t \in \mathbb{N})$ *be a discrete-time Markov, exchangeable recurrent event process. Then there exists a probability measure* Σ *on* $[0,1]^{\mathbb{N}}$ *such that* $\mathbf{Y}^\star_\Sigma$ *is a version of* $\mathbf{Y}$*.*

Proof of Corollary S1. Consider the recurrent event process $\mathbf{Y}$ up time $\tau < \infty$. Then $\mathbf{Y}^\star = \tau \wedge \mathbf{Y}$ is a version of $\mathbf{Y}$ on $t \in 0, 1, \ldots, \tau$. Let η_τ to denote the measure associated with $\mathbf{Y}^\star$. For $P \in \mathcal{P}_\tau$, let $R_{\tau',\tau}(A)$ be the restriction of this $\mathcal{P}_{\tau'}$. Then for $\tau' < \tau$

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

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

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

is the unique measure such that $\mathbf{Y}^\star$ is a version of $\mathbf{Y}$. $\square$

Corollary S2 (Continuous-time characterization). *Let* $\mathbf{Y} = (\mathbf{Y}(t), t \in \mathbb{R}^+)$ *be a continuous-time Markov, exchangeable recurrent event process such that* $\mathbf{Y}_u(0) = 0$ *for all* $u \in \mathbb{N}$. *Let* I *denote the infinite identity matrix. Then there exists a probability measure* Σ *on* $[0,1]^{\mathbb{N}}$ *satisfying*

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

and a set of constants $\mathbf{c} = \{c_{i,i+1} \,|\, i \in \mathbb{N}\}$ *such that* $\mathbf{Y}^\star_{\Sigma,c}$ *is a version of* $\mathbf{Y}$.

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

S4 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 $t > 0$.

Dempsey and McCullagh [2017] showed that every exchangeable, Markov survival process can be generated via a Lévy process construction. The proof stems from connecting γ 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 nonnormalized transition rules are

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

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 $p_i \to p^{w_{x_i}}$. 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, $\log\left(p_i/(1-p_i)\right) = w_{x_i}\log\left(p/(1-p)\right)$. We do not pursue such alternatives in this paper.

S5 MCMC simulation example: details

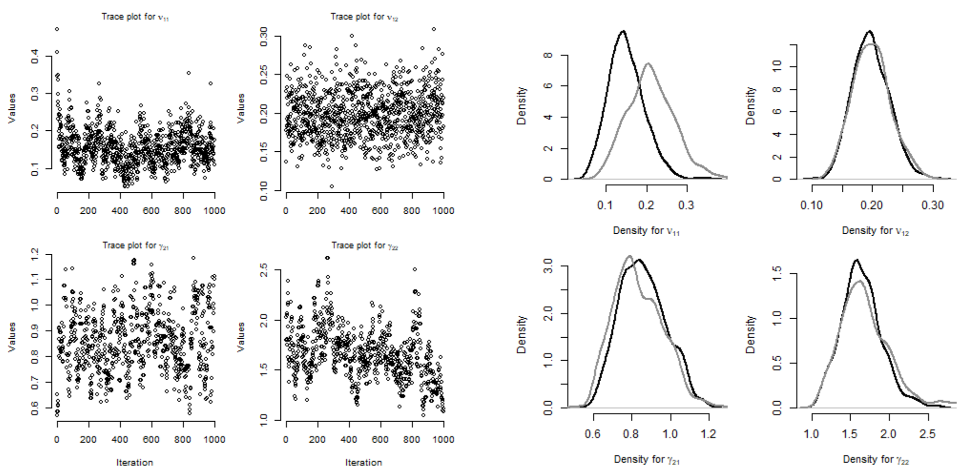
Below we describe the simulation example in detail. We set parameters as follows: first, we assume that marginally a healthy participant transitions to ill and dead after 2 and 5 years respectively (on average); ill participants to both healthy and dead on average every 3 years respectively. Both healthy 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 unhealthy, were generated.

		Maximum Likelihood			Posterior distribution		
Parameter	True Value	Estimate	Lower CI	Upper CI	Mean	5% Quantile	95% Quantile
ν_{11}	0.50	0.53	0.43	0.64	0.15	0.09	0.23
ν_{12}	0.20	0.19	0.13	0.25	0.20	0.15	0.25
γ_{21}	0.70	0.75	0.64	0.86	0.85	0.67	1.07
γ_{22}	1.71	1.71	1.28	2.13	1.23	1.62	2.09

Table S1: Parameter estimation

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 S5a suggest convergence of the MCMC procedure after the first 100 iterations. The MCMC sampler gives posteriors for the parameters. Table S1 contains these estimates. We see good performance for ν_{12}, γ_{21}, and γ_{22}. The posterior for ν_{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.

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.



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

Figure S5: MCMC traceplots and densities for simulation example

Removing the first 100 iterations as burn-in, posterior distributions are presented in Figure S5b. Black curves are the MCMC sampler procedure; grey 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 ν_{11}. This supports the aforementioned difficulty in estimating ν_{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 ν_{11}.

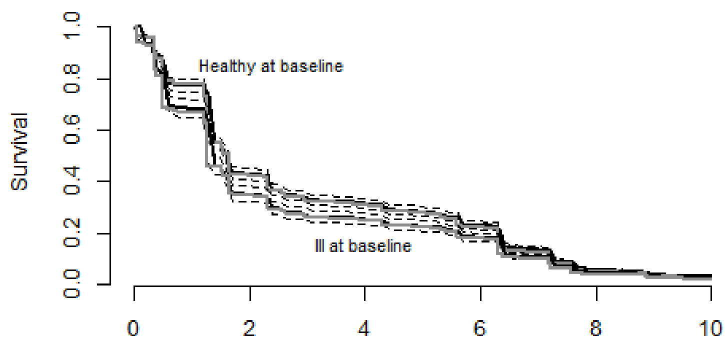


Figure S6: Survival functions given baseline state; median (black), 5% and 95% quantiles (dotted black), and true survival function (grey)

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 S6 presents the point-wise median, 5%, and 95% survival at every time since recruitment when the individual is healthy and ill at baseline respectively. The grey 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.

S6 Code repository

All code related to the CAV analysis and simulation study can be found at `https://github.com/wdempsey/multi-state`.

Bibliography

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

David Aldous. Probability Distributions on Cladograms. In D. Aldous and R. Pemantle, editors, *Random Discrete Structures*, pages 1–18. Springer, 1996.

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

W. Dempsey and P. McCullagh. Weak continuity of predictive distributions for markov survival processes. *Electronic Journal of Statistics*, 11(2): 5406–5451, 2017.

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

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

Jim Pitman. Poisson-kingman partitions. In Darlene R. Goldstein, editor, *Statistics and science: a Festschrift for Terry Speed*, volume Volume 40 of *Lecture Notes–Monograph Series*, pages 1–34. Institute of Mathematical Statistics, 2003. doi: 10.1214/lnms/1215091133. URL `https://doi.org/10.1214/lnms/1215091133`.