Electronic Journal of Statistics
Vol. 11 (2017) 5406–5451
ISSN: 1935-7524
https://doi.org/10.1214/17-EJS1381

Exchangeable Markov survival processes and weak continuity of predictive distributions

Walter Dempsey

Department of Statistics
Harvard University
Cambridge, MA 02138
e-mail: wdempsey@fas.harvard.edu

and

Peter McCullagh

Department of Statistics
University of Chicago
Chicago, IL 60637
e-mail: pmcc@galton.uchicago.edu

Abstract: We study exchangeable, Markov survival processes – stochastic processes giving rise to infinitely exchangeable non-negative sequences $(T_1, T_2, \ldots)$. We show how these are determined by their characteristic index $\{\zeta_n\}_{n=1}^{\infty}$. We identify the harmonic process as the family of exchangeable, Markov survival processes that compose the natural set of statistical models for time-to-event data. In particular, this two-dimensional family comprises the set of exchangeable, Markov survival processes with weakly continuous predictive distributions. The harmonic process is easy to generate sequentially, and a simple expression exists for both the joint probability distribution and multivariate survivor function. We show a close connection with the Kaplan-Meier estimator of the survival distribution. Embedded within the process is an infinitely exchangeable ordered partition. Aspects of the process, such as the distribution of the number of blocks, are investigated.

Keywords and phrases: infinite exchangeability, survival analysis, characteristic index, Bayesian nonparametrics, Gibbs-type splitting rules.

Received April 2017.

1. Introduction

A fundamental principle underlying statistical modeling is that arbitrary choices such as sample size or observational unit labels[1] should not affect the sense of a model and meaning of parameters. For example, a priori the labels attached to observational units (unit $1, 2, \ldots$) carry no meaning other than to distinguish

[1]Labels are not always arbitrary (e.g., a time series indexed by the integers).

among them. This guiding precept gives rise to the study of exchangeable distributions. Infinite exchangeability captures the idea that the data generating process should not be affected by choice of sample size or arbitrary unit labeling.

In this paper we are interested in models for infinitely exchangeable non-negative sequences $(T_1, T_2, \ldots)$ which arise in survival analysis where the observation values are survival times. We start by introducing exchangeable Markov survival processes. We show how these are determined by their characteristic index $\{\zeta_n\}_{n=1}^{\infty}$. We then identify the family of exchangeable Markov survival processes whose predictive distributions are weakly continuous. We provide a sequential description for this family which we call the *harmonic process*, whose asymptotic behavior is also derived. The harmonic process is part of a larger family of *beta-splitting processes*, which we identify as the set of exchangeable, Markov survival processes of Gibbs-type. We show that the process exhibits markedly distinct asymptotic behavior as a function of a single parameter $\beta > -1$.

Research on nonparametric Bayesian survival analysis has been based around the de Finetti approach to constructing exchangeable survival processes by generating survival times conditionally independent and identically distributed given a completely random measure, i.e., the cumulative conditional hazard is a Lévy process (Cornfield and Detre, 1977; Kalbfleisch, 1978; Hjort, 1990; Clayton, 1991). Such processes are sometimes called *neutral to the right* (Doksum, 1974; James, 2006). The harmonic process corresponds to a particular choice of Lévy process and thus fits naturally within the existing literature. It has the added advantage that the probability distribution function can be computed analytically, and the process can be simulated directly from the predictive distributions, allowing us to bypass the Lévy process entirely.

2. Exchangeable non-negative sequences

We start by formally defining exchangeable, Markov survival processes. To do this we introduce the following notation. Define $\mathbf{t} = (t_1, t_2, \ldots)$ to be a fixed non-negative sequence. That is, $t_i \in \mathbb{R}_+$ for each $i \in \mathbb{N}$. The infinite sequence $\mathbf{t}$ takes values in $\mathbb{R}_+^{\mathbb{N}}$. We define several operations on $\mathbf{t}$. First, for any finite permutation $\sigma : \mathbb{N} \to \mathbb{N}$, the *relabeling* of $\mathbf{t}$ by σ is

$$\mathbf{t}^\sigma := (t_{\sigma(1)}, t_{\sigma(2)}, \ldots).$$

Second, for any finite $n = 1, 2, \ldots$ the *restriction* of $\mathbf{t}$ to $[n] := \{1, \ldots, n\}$ is

$$\mathbf{t}[n] := (t_1, \ldots, t_n).$$

Third, for any $s \in \mathbb{R}_+$ we write $s + \mathbf{t}$ to mean the addition of s entrywise; that is,

$$s + \mathbf{t} = (s + t_1, s + t_2, \ldots).$$

We write $s + \mathbf{t}[n]$ to mean the restriction of $(s + \mathbf{t})$ to $[n]$, (i.e., $(s + \mathbf{t})[n]$). Finally, for $m > n$, define the restriction operator $\Xi_{m,n}(\mathbf{t}[m])$ to be the further restriction of $\mathbf{t}[m]$ to the first n components.

A *survival process* is a sequence of probability distributions $(P_1, P_2, \ldots)$, where P_n is a distribution on $\mathbb{R}_+^n$. We write $\mathbf{T}[n] := (T_1, \ldots, T_n)$ to be a random variable with probability distribution P_n. The process is called

- *exchangeable* if, for each integer n, $\mathbf{T}^\sigma[n]$ is equal in distribution to $\mathbf{T}[n]$.
- *(Kolmogorov) consistent* if, for any $m > n$, $\Xi_{m,n}(\mathbf{T}[m])$ is equal in distribution to $\mathbf{T}[n]$.
- *Markovian* if, for each integer n and $s \in \mathbb{R}_+$, the distribution P_n satisfies the following memoryless property

$$P(\mathbf{T}[n] \geq s + \mathbf{t}[n] \mid \mathbf{T}[n] \geq s) = P_n(\mathbf{T}[n] \geq \mathbf{t}[n]).$$

We define a survival process satisfying all three of the above properties to be an *exchangeable, Markov survival process*. In Section 3, we provide an equivalent representation of exchangeable, Markov survival processes via their risk set trajectories.

Under the consistency assumption, $T[n]$ satisfies *lack of interference*; mathematically, for integers $m > n$ and sequence of Borel sets $A_1, \ldots, A_n$

$$P(T[n] \in A \mid H_{[m]}(s)) = P(T[n] \in A \mid H_{[n]}(s)).$$

where $A = (A_1, \ldots, A_n)$, and $H_{[l]}(s)$ is the σ-field generated by the Boolean variables $\mathbf{1}[T_i < u]$ for $i \in [l]$ and $u \leq s$. Lack of interference is essential, ensuring the conditional distribution of $T[n]$ given the joint history of all m particles up to time s is unaffected by the history for subsequent particles (i.e., by particles $n+1, \ldots, m$).

2.1. Censoring

Consider a sample of size n. Let $\mathbf{T}[n]$ denote the random survival times and $\mathbf{t}[n]$ their corresponding realizations for the sample $[n]$. To each unit $i \in [n]$, there often corresponds an observational interval $[0, c_i]$ where c_i is some arbitrary positive censoring time. The event time recorded for unit i is $Y_i = \min(T_i, c_i)$. So if $Y_i = c_i$ then the event is a censoring time; otherwise, if $Y_i < c_i$ then the event is known to be death or failure. For simplicity the description of exchangeable Markov survival processes below presumes $c_i = \infty$ for all units i. We revisit censoring in Section 9 as part of a discussion on parameter estimation. In particular we show that censoring has a relatively trivial effect on probability calculations and thus on associated statistical procedures.

3. Risk set trajectories

For each unit i, an equivalent representation of the random variable T_i is the random Boolean function $R_i : \mathbb{R}_+ \mapsto \{0, 1\}$ defined by $R_i(t) = \mathbf{1}[t < T_i]$. We write

$$R_{[n]}(t) = \{i \in [n] : R_i(t) = 1\}$$

to define the random subset of units of $[n]$ known to be alive at time t, i.e., the *risk set* at time t. The risk set is a stochastic process that is in one-to-one correspondence with $\mathbf{T}[n]$ almost surely.

The stochastic process $R_{[n]} = \{R_{[n]}(t)\}_{t\geq 0}$ satisfies $R_{[n]}(t+s) \subseteq R_{[n]}(t)$ almost surely for all $t, s \geq 0$. We write $r_{[n]} = \{r_{[n]}(t)\}_{t\geq 0}$ to denote a realization of the stochastic process $R_{[n]}$. At each time $t \geq 0$, we write $B_{[n]}(t) = R_{[n]}(t-)\backslash R_{[n]}(t)$ to denote the random subset of particles that fail at time t. That is, for each $i \in B_{[n]}(t)$ we have $T_i = t$ almost surely. We write $b_{[n]} = \{b_{[n]}(t)\}_{t\geq 0}$ to denote a realization of the stochastic process $B_{[n]} = \{B_{[n]}(t)\}_{t\geq 0}$. It is worth emphasizing that the subsets $B_{[n]}(t)$ and $R_{[n]}(t)$ are disjoint for every t almost surely. We write $R^\sharp_{[n]}(t)$, $r^\sharp_{[n]}(t)$, $B^\sharp_{[n]}(t)$, $b^\sharp_{[n]}(t)$ to denote the cardinality of the risk set (both the random variable and realization) and failure set (both the random variable and realization) respectively.

We define two operations on $r_{[n]}(t)$. First, for each finite permutation $\sigma : \mathbb{N} \to \mathbb{N}$, the *relabeling* of $r_{[n]}(t)$ by σ is

$$r^\sigma_{[n]}(t) = \{i \in [n] : r_{\sigma(i)}(t) = 1\}$$

Second, for $m > n$, we overload notation and define the restriction operator $\Xi_{m,n}$ to be the further restriction of $r_{[m]}(t)$ to the first n components.

A survival process induces a distribution on risk set trajectories. That is for each $n \in \mathbb{N}$ the random variable $T[n]$ with probability distribution P_n can be mapped to a random risk set trajectory $R_{[n]} = \{R_{[n]}(t)\}_{t=1}^\infty$. Each realization $t[n]$ is also in one-to-one correspondence with a realization of the risk set trajectory $r_{[n]}$. We call this process the *risk set process*. This process is called

- *exchangeable* if, for each integer n, $R^\sigma_{[n]} = \{R^\sigma_{[n]}(t)\}_{t\geq 0}$ is a version of $R_{[n]}$.
- *(Kolmogorov) consistent* if, for integers $m > n$, $\Xi_{m,n}(R_{[m]})$ is a version of $R_{[n]}$.
- *Markovian* if, for each integer n, the process $R_{[n]}$ evolves as a continuous-time Markov chain.

A risk set process built from an exchangeable, Markov survival process is exchangeable, consistent and Markovian. Since $R_{[n]}$ evolves as a continuous-time Markov chain, we can characterize exchangeable, Markovian survival processes by their corresponding exchangeable, Markov risk process. In the next section, we use the connection to continuous-time Markov chains to describe the *embedded process*. Using the embedded process, we fully characterize each exchangeable, Markov survival process by its characteristic index.

4. Embedded process

In this section, we will show how the continuous-time Markov chain $R_{[n]}$ started at $R_{[n]}(0) = [n]$ (i.e., all particles are at risk initially) can be built as follows: first, generate a holding time H. In Section 4.3 we will show this can be done by generating an exponential random variable with the parameter only depending on the number of particles at risk. The Markov chain $R_{[n]}$ will remain in state $[n]$ until time $T = H$. Second, choose a state $R_{[n]}(T)$ at random according to the distribution $P^{[n]}(R_{[n]}(T) = \cdot)$, where P^r denotes the probability measure under which the Markov chain starts in state $r \subseteq [n]$. Third, given that $R_{[n]}(T) = r' \subset [n]$, build an independent Markov chain started at r', and attach it to the initial segment. Iterating on this procedure yields a sequence of holding times $(H_1, H_2, \dots, H_k)$ the length of which is random but finite almost surely as $r = \emptyset$ is an absorbing state. Define $\bar{H}_j = \sum_{i=1}^{j} H_i$ to be the the partial sums of the holding times. Then the procedure also yields a sequence of successive states visited

$$[n] = R_{[n]}(0), R_{[n]}(\bar{H}_1), R_{[n]}(\bar{H}_2), \dots, R_{[n]}(\bar{H}_k) = \emptyset.$$

This sequence is a discrete-time Markov chain with $\emptyset$ as the absorbing state. Note that the risk process being exchangeable and consistent implies the discrete-time Markov process of successive states will be exchangeable and consistent.

By the theory of continuous-time Markov chains, each successive state is chosen independently of the previous holding time conditional on the previous state. That is, $R_{[n]}(\bar{H}_j) \perp\!\!\!\perp H_j \mid R_{[n]}(\bar{H}_{j-1})$. Therefore, we construct the exchangeable, Markov risk trajectories in two parts: (1) we first show how to generate the sequence of successive states, and (2) conditional on the current state we show how to generate the holding time in that state.

We start by investigating the embedded discrete-time Markov chain of successive states. By design, each successive state will be the remaining set of particles at risk. This sequence of successive states is in one-to-one correspondence with the random sequence

$$B = (B_{[n]}(\bar{H}_1), B_{[n]}(\bar{H}_2) \dots, B_{[n]}(\bar{H}_k))$$

where $B_{[n]}(\bar{H}_i)$ is the random subset of particles failing at the ith transition time. The correspondence is due to the fact that the subsets $B_{[n]}(t)$ and $R_{[n]}(t)$ are disjoint for every t almost surely and satisfy $R_{[n]}(t-) = B_{[n]}(t) \cup R_{[n]}(t)$. We use this fact below to define the distribution of the discrete-time Markov chain of successive states via exchangeable, Markov partial rankings.

4.1. Exchangeable, Markov partial rankings

A partial ranking of $[n]$ is an *ordered list* $b_{[n]} = (b_1, \dots, b_k)$ satisfying

$$|b_i| \geq 1, \text{ and } b_i \cap b_j = \emptyset \text{ for } i \neq j, \text{ and } \cup_{i=1}^{k} b_i = [n].$$

That is $b_{[n]}$ consists of disjoint non-empty subsets of $[n]$ whose union is $[n]$. The elements of $[n]$ are unordered within blocks, but b_1 is the subset ranked first, b_2 is the subset ranked second, and so on. We let $b_i^\sharp$ to denote the cardinality of the ith block.

Let $\mathcal{OP}_n$ denote the finite set of partial rankings of $[n]$. A random partial ranking of $[n]$, denoted $B_{[n]}$, is a random variable taking values in $\mathcal{OP}_n$ whose probability distribution will be denoted by p_n. The sequence of distributions $(p_1, p_2, \ldots)$ is called

- *exchangeable* if, for each integer n, the probability distribution p_n only depends on block sizes and block order. In general $p_n(B_{[n]} = b_{[n]}) \neq p_n(B^\sigma_{[n]} = b^\sigma_{[n]})$ for a permutation $\sigma : [n] \to [n]$.
- *Markovian* if, for each integer n, non-negative integer $r \geq 0$, and positive integer $d \geq 1$ such that $r + d = n$ there exists a positive number $q(r, d) \in [0, 1]$ (i.e., a *splitting rule*) such that

$$p_n(B_{[n]} = b_{[n]}) = q(n - b^\sharp_1, b^\sharp_1) \times p_{n-b^\sharp}((B_2, \ldots, B_k) = (b_2, \ldots, b_k))$$
$$= \prod_{j=1}^{\#b} q\left(n - \sum_{i=1}^{j} b^\sharp_i, b^\sharp_i\right). \quad (1)$$

  Every splitting rule q satisfies $q(0, 1) = 1$.
- *(Kolmogorov) consistent* if, for each integer n, p_n is the marginal distribution of p_{n+1} under the restriction map $\mathcal{OP}_{n+1} \to \mathcal{OP}_n$ in which the element $n + 1$ is ignored or deleted.

As we will see below, the Markovian assumption helps lead to mathematically tractable conclusions. The following proposition is a direct consequence of the Markovian property.

Proposition 4.1. *Every non-negative function q subject to normalization conditions*

$$\sum_{d=1}^{n} \binom{n}{d} q(n - d, d) = 1 \quad (2)$$

for all $n \geq 0$ defines an exchangeable and Markovian sequence of distribution functions $(p_1, p_2, \ldots)$.

Not all splitting rules q lead to a sequence of consistent distributions $(p_1, p_2, \ldots)$. The consistency condition is necessary in order for the sequence of distributions to determine a process. Without it, there is no partial ranking of the population and there is no possibility of inference using conditional distributions. Proposition 4.2 details a condition on the splitting rule that guarantees the associated sequence of distributions are consistent. A sequence of probability distributions on partial rankings $(p_1, p_2, \ldots)$ that are consistent, exchangeable, and Markovian defines an *exchangeable, Markov partial ranking process.*

Proposition 4.2. *A splitting rule q gives rise to an exchangeable, Markov partial ranking process if and only if*

$$(1 - q(n, 1))q(n - d, d) = q(n - d, d + 1) + q(n - d + 1, d) \quad (3)$$

for all integers $n \geq d \geq 1$. In particular, equation (3) implies that the splitting rules are determined by the sequence of singleton splits $(q(0, 1), q(1, 1), \ldots, q(n, 1), \ldots)$.

Proof. To derive the conditions for consistency proceed by induction supposing that $p_1, \dots, p_n$ are mutually consistent in the Kolmogorov sense. Then, in order for p_{n+1} to be consistent with p_n, it must be for each ordered partition $b_{[n]}$ of $[n]$,

$$\begin{aligned} p_n(B_{[n]} = b_{[n]}) =& q(n,1)p_n(b_{[n]}) + q(n - b_1^\sharp, b_1^\sharp + 1)p_{n-b_1^\sharp}((b_2,\dots,b_k)) \\ &+ q(n - b_1^\sharp + 1, b_1^\sharp) \sum_{\tilde{b} \in A} p_{n-b_1^\sharp+1}(\tilde{b}) \end{aligned}$$

where A is the set of ordered partitions where $n+1$ is appended to a block of $(b_2, \dots, b_k)$ or inserted as a singleton. That is, removal of $n+1$ from $\tilde{b}$ yields $(b_2, \dots, b_k)$ for all $\tilde{b} \in A$.

The terms on the right are partial rankings in $\mathcal{OP}_{n+1}$ such that deletion of $n+1$ gives rise to the ordered partition $b_{[n]} \in \mathcal{OP}_n$. Either $n+1$ occurs in the first block as a singleton, which has probability $q(n,1)p_n(B_{[n]} = b_{[n]})$; or it occurs appended to b_1 as a non-singleton, which has probability $q(n - b_1^\sharp + 1, b_1^\sharp + 1)p_{n-b_1^\sharp}(B_{[n-b_1^\sharp]} = (b_2, \dots, b_k))$; or it occurs elsewhere either as a singleton or appended to one of the other blocks of $b_{[n]}$ other than b_1, which occurs with probability $q(n - b_1^\sharp + 1, b_1^\sharp) \sum_{\tilde{b} \in A} p_{n-b_1^\sharp+1}(\tilde{b})$.

By the induction hypothesis that $p_1, \dots, p_n$ are mutually consistent, the sum is equal to the probability $p_{n-b_1^\sharp}((b_2, \dots, b_k))$. Hence, a splitting rule gives rise to a consistent, exchangeable Markov partial ranking process if and only if

$$\begin{aligned} q(n - b_1^\sharp, b_1^\sharp)p_{n-b_1^\sharp}(b_2, \dots) =& q(n,1)q(n - b_1^\sharp, b_1^\sharp)p_{n-b_1^\sharp}(b_2, \dots) \\ &+ q(n - b_1^\sharp, b_1^\sharp + 1)p_{n-b_1^\sharp}(b_2, \dots) \\ &+ q(n - b_1^\sharp + 1, b_1^\sharp)p_{n-b_1^\sharp}(b_2, \dots). \end{aligned}$$

Cancelling $p_{n-b^\sharp,1}(b_2, \dots)$ from both sides yields equation (3). □

A splitting rule corresponding to a consistent Markov partial ranking process is said to be a *consistent splitting rule*. A nice property of consistent splitting rules is that they admit an integral representation.

Proposition 4.3. *Every consistent splitting rule admits an integral representation*

$$q(n-d, d) = \frac{1}{Z_n} \left(\int_0^1 x^{n-d}(1-x)^d \varpi(dx) + c\, \mathbf{1}_{d=1} \right)$$

where $Z_n = \int_0^1 (1-x)^n \varpi(dx) + n \cdot c$. *The measure* $\varpi(\cdot)$ *is defined on* $[0,1)$ *and satisfies*

$$\int_0^1 (1-x)\varpi(dx) < \infty. \tag{4}$$

The constant, c, is called the erosion coefficient and ϖ *is called the dislocation measure.*

In Proposition 4.3, d represents the number of particles observed to fail and therefore must be positive while $n-d$ is the number of particles still at risk and therefore must only be non-negative. That is, $d \in \{1, \ldots, n\} = [n]$ and $n - d \in \{0, 1, \ldots, n-1\}$. Both the condition and integral representation are similar to equation (4) in McCullagh, Pitman, and Winkel (2008) and to proposition 41 of Ford (2005). However, the splitting probabilities in these papers are symmetric and subject to a normalization condition different from (2), so the probabilities are different.

4.2. The characteristic index

To each exchangeable, Markov partial ranking process we associate a *characteristic index*. A relation between the defined characteristic index and the splitting rule via kth order differences is shown.

Definition 4.4. A *characteristic index*, ζ, is a sequence $\zeta_0, \zeta_1, \zeta_2, \ldots$ beginning with $\zeta_0 = 0$, $\zeta_1 > 0$, and subsequently

$$\zeta_{n+1} = \frac{\zeta_n}{1 - q(n,1)} = \zeta_1 \prod_{j=1}^{n} (1 - q(j,1))^{-1} \tag{5}$$

for $n \geq 1$. Alternatively we write $\zeta(n)$ to denote the characteristic index when more convenient.

The sequence is said to be in *standardized form* if $\zeta_1 = 1$. Since the standardized sequence is in one-to-one correspondence with the singleton splits, each standardized sequence determines an exchangeable, Markov partial ranking process provided that the splitting rule in (3) is non-negative.

This multiplicative construction implies that ζ is non-negative and strictly increasing. A natural question is whether the splitting rules can be reconstructed from the characteristic index. Proposition 4.5 shows the splitting rule can be recovered from the characteristic index using kth order forward differences, $\Delta^k \zeta$, defined as

$$(\Delta^k \zeta)_n = \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} \zeta_{n+j},$$

for $n \geq 0$, so that $\Delta \zeta_n = \zeta_{n+1} - \zeta_n$ is the first difference, $(\Delta^2 \zeta)_n = \zeta_{n+2} - 2\zeta_{n+1} + \zeta_n$ is the second, and so on.

Proposition 4.5. *Given a characteristic index, ζ, the corresponding splitting rule is given by*

$$q(r,d) = \frac{(-1)^{d-1} (\Delta^d \zeta)_r}{\zeta_{r+d}} \tag{6}$$

for $r \geq 0$ and $d \geq 1$. The splitting probabilities are therefore determined by the standardized sequence ζ/ζ_1.

The proof of Proposition 4.5 can be found in Appendix A. The main advantage of working with the characteristic index is that the Kolmogorov consistency condition (3) may be written in a more convenient form as a non-negativity condition on forward differences.

Corollary 4.6. *A sequence $\zeta = (\zeta_0, \zeta_1, \zeta_2, \dots)$ beginning with $\zeta_0 = 0$ and $\zeta_1 > 0$ defines a consistent splitting rule via equation (6) if for $r \geq 0$ and $d \geq 1$ it satisfies the following non-negativity condition on forward differences:*

$$(-1)^{d-1}(\Delta^d \zeta)_r \geq 0.$$

Under this condition, the sequence ζ is the characteristic index for some exchangeable, Markov partial ranking process.

4.2.1. The harmonic process

Since each exchangeable, Markov partial ranking process is associated with a characteristic sequence (modulo scalar multiplication), the space of Markov partial rankings may be identified with the space of non-negative sequences satisfying (6). Evidently this space is a convex cone, closed under positive linear combinations.

Based on the integral representation in Proposition 4.3 a natural family of dislocation measures are conjugate measures to the binomial distribution, $\varpi(dx) = x^{\rho-1}(1-x)^{\beta-1}dx$. The integrability condition on the measure given by equation (4) implies $\rho > 0$ and $\beta > -1$. The characteristic index depends on the second parameter, β. Here we focus on the setting where $\beta = 0$ as we will show this family composes the natural set of statistical models for time-to-event data. For $\beta = 0$, the characteristic index is

$$\begin{aligned} \zeta_n &= \int_0^1 x^{\rho-1}(1 + x + \cdots + x^{n-1})\, dx \\ &= \sum_{j=0}^{n-1} \frac{1}{\rho + j} = \psi(n+\rho) - \psi(\rho), \end{aligned} \tag{7}$$

where ψ is the derivative of the log gamma function. This process is called the *harmonic process*.

A very similar process to the harmonic process with dislocation measure $\varpi(dx) = x^{\rho-1}dx/(-\log(x))$ has characteristic index

$$\begin{aligned} \zeta_n &= \int_0^1 (1-x^n)x^{\rho-1}\, dx/(-\log x) = \int_0^\infty (1-e^{-nz})z^{-1}e^{-\rho z}\, dz \\ &= \log(1 + n/\rho). \end{aligned} \tag{8}$$

This is the characteristic index of the gamma process, which is explored in more detail in Section 9.2.

4.3. Holding times and continuous-time embedding

We now consider how to generate the holding times conditional on the current state $r \subset [n]$ in order to generate exchangeable, Markov risk set trajectories. In the prior section, we used the fact that the random sequence of successive states visited is in one-to-one correspondence with the random partial ranking B of the set of particles $[n]$. In this section, we use the fact that the random sequence of successive states is a strictly decreasing sequence of subsets almost surely:

$$[n] \equiv R_{[n]}(0) \supset R_{[n]}(\bar{H}_1) \supset R_{[n]}(\bar{H}_2) \cdots \supset R_{[n]}(\bar{H}_k) = \emptyset.$$

As discussed in Section 4, by assuming each H_i is an independent exponentially distributed holding time, a continous-time Markov process is constructed which represents the random risk set trajectory $R_{[n]}$.

The question is how to choose the rate functions for the holding times to ensure $\Xi_{m,n}\left(R_{[m]}\right)$ is a version of $R_{[n]}$ for all integers $m \geq n$. Recall the rate function is a mapping from the current state to the positive reals $\tau : \mathcal{P}([n]) \to \mathbb{R}_+$ where $\mathcal{P}([n])$ is the power set of $[n]$. Exchangeability implies the rate function must only be a function of the cardinality of the current state $r \subseteq [n]$. Therefore the rate function is the mapping from the natural numbers to the positive reals $\tau : \mathbb{N} \to \mathbb{R}_+$. We write this equivalently as the sequence $\{\tau_n\}_{n=0}^{\infty}$. An argument essentially equivalent to that used in Section 4 of McCullagh, Pitman, and Winkel (2008) leads to the following consistency condition on the rate function.

Proposition 4.7. *A sequence of rate functions, $\{\tau_n\}_{n=0}^{\infty}$, gives rise to an exchangeable, Markov risk set process if*

$$\tau_{n+1}(1 - q(n,1)) = \tau_n.$$

In other words, the characteristic index is proportional to the exponential failure rate needed to ensure consistency of the continuous-time Markov process $R_{[n]}$.

Therefore the characteristic index fully characterizes the distribution of the continuous-time Markov process $R_{[n]}$. Given the risk set trajectory $R_{[n]}$ is in one-to-one correspondence with $T[n]$ almost surely the characteristic index also fully characterizes the probability distributions P_n for each $n \geq 1$.

5. Probability distribution function

In Section 4 we described the embedded discrete-time Markov chain within the continuous-time Markov chain $R_{[n]}$. Section 4.1 showed how this process is in one-to-one correspondence with an exchangeable, Markov partial ranking process. Section 4.3 proves the characteristic index to be the rate function necessary to ensure the continuous-time Markov process $R_{[n]}$ is both consistent and exchangeable. Here we use this construction of $R_{[n]}$ to give an expression for the probability distribution function, P_n, of $T[n]$.

Let $B \in \mathcal{OP}_n$ be a random partial ranking with associated consistent splitting rule $q(\cdot,\cdot)$. Recall the probability distribution function is given by equation (1). We re-write here in a more convenient form:

$$p_n(B = b) = \prod_{i=1}^{\#b} q(r_i^\sharp, b_i^\sharp) = \prod_{i=1}^{\#b} \frac{\lambda(r_i^\sharp, b_i^\sharp)}{\zeta(r_i^\sharp + b_i^\sharp)} = \prod_{i=1}^{\#b} \frac{\lambda(r_i^\sharp, b_i^\sharp)}{\zeta(r_{i-1}^\sharp)}.$$

where $r_i^\sharp = n - \sum_{j=1}^{i} b_j^\sharp$. The numerator term $\lambda(\cdot,\cdot)$ is the un-normalized splitting rule, equal to $(-1)^{d-1}(\Delta^d\zeta)_r$. The second equality is due to $r_{i-1}^\sharp = r_i^\sharp + b_i^\sharp$ for $i = 1, \ldots, \#b$ and $r_0^\sharp = n$.

The partial ranking b represents the sequence of subsets of particles that fail at each consecutive failure time. Therefore $r_i = [n] \backslash \cup_{j=1}^{i} b_i$ is the set of particles still at risk after the ith failure time. Conditional on the partial ranking $B = b$, proposition 4.7 yields the exponential failure rate for each holding time. Suppose the current state R equals $r \subseteq [n]$. Then for any Borel set A the holding time H has probability distribution function

$$P(H \in A \mid R = r) = \int_{u \in A} \zeta(r^\sharp) \exp\left(-\zeta(r^\sharp)\, u\right) du$$

where $r^\sharp$ is the cardinality of the set r. The above distribution is absolutely continuous with respect to Lebesgue measure and therefore has the density $g(u \mid r) = \zeta(r^\sharp) \exp\left(-\zeta(r^\sharp) u\right)$. Based on this, we can define

$$\begin{aligned} f_n(b, s_1, \ldots, s_{\#b}) :=& P_n(B = b) \prod_{i=1}^{\#b} g(s_i - s_{i-1} \mid r_{i-1}) \\ =& \prod_{i=1}^{\#b} q(r_i^\sharp, b_i^\sharp) \zeta(r_{i-1}^\sharp) \exp\left(-\zeta(r_{i-1}^\sharp)(s_i - s_{i-1})\right) \\ =& \exp\left(-\int_0^\infty \zeta(r^\sharp(t)) dt\right) \prod_{i=1}^{\#b} \lambda(r_i^\sharp, b_i^\sharp) \end{aligned} \tag{9}$$

where $r_0 = [n]$, $s_0 = 0 < s_1 < \cdots < s_{\#b}$ are realizations of the partial sums of the holding times, and the function $r^\sharp(t)$ is the cardinality of the risk set at time t.

We now define the probability distribution, P_n, of $T[n] = (T_1, \ldots, T_n)$, which can be easily described in terms of the function f_n introduced in (9). For every sequence of Borel sets $A_1, \ldots, A_n$,

$$P_n(T_1 \in A_1, \ldots, T_n \in A_n) = \sum_{b \in OP_n} \int_{A^\star(b)} f_n(b, s_1, \ldots, s_{\#b}) ds_1 \ldots ds_{\#b} \tag{10}$$

where

$$\begin{aligned} A^\star(b) = & \{s_1, \ldots, s_{\#b} \in \mathbb{R}_+^{\#b} : s_1 < \ldots < s_{\#b}\} \\ & \cap \{\cap_{j \in b_1} A_j \times \cap_{j \in b_2} A_j \cdots \times \cap_{j \in b_{\#b}} A_j\}. \end{aligned}$$

The n-dimensional joint probability distribution given by equation (10) is continuous in the sense that it has no atoms. For $n \geq 2$, it is not absolutely continuous with respect to Lebesgue measure in $\mathbb{R}_n$ because the distribution has condensations on all diagonals implying that $P(T_1 = T_2) = q(0, 2) > 0$, and likewise for arbitrary subsets. The one-dimensional marginal distributions are exponential with rate ζ_1. However, under a monotone continuous temporal transformation that sends Lebesgue measure to the measure $\nu(\cdot)$ the risk set $R_{[n]}$ evolves as an exchangeable, semi-Markov process. If $g(t) = \nu((0, t))$ is the associated monotone continuous function then $g^{-1}(T_i)$ is exponential with rate ζ_1.

5.1. Sequential description

We now construct the conditional distribution function for the random variable T_{n+1} given $T[n] = t[n]$. Let $\bar{b} \in \mathcal{OP}_n$ be the partial ranking induced by $t[n]$ and $\bar{s}_1 < \cdots < \bar{s}_{\#\bar{b}}$ be the unique elements of $t[n]$. We write $\bar{\mathbf{s}} = (\bar{s}_1, \ldots, \bar{s}_{\#\bar{b}})$ to denote the vector of unique elements of $t[n]$.

Let $t[n+1] \in \mathbb{R}_+^{n+1}$ satisfy $\Xi_{n+1,n}(t[n+1]) = t[n]$. That is, the restriction to the first n particles is equal to $t[n]$. Let $b \in \mathcal{OP}_{n+1}$ be the partial ranking induced by $t[n+1]$. Then the restriction of b to the first n particles is equal to $\bar{b}$ (i.e., $\Xi_{n+1,n}(b) = \bar{b}$). For each $b \in \Xi_{n+1,n}^{-1}(\bar{b})$ we write $i^\star(b)$ to denote the block that contains the $(n+1)$st particle. Each ordered partition b in the set $\Xi_{n+1,n}^{-1}(\bar{b})$ either (1) has the same number of blocks as $\bar{b}$ or (2) has one additional block with the single element $\{n+1\}$. We write $\Xi_{n+1,n}^{-1}(\bar{b}) = \Phi_1 \cup \Phi_2$ to denote these two disjoint sets. Let $s_1 < \cdots < s_{\#b}$ be the unique elements of $t[n+1]$. We write $\mathbf{s} = (s_1, \ldots, s_{\#b})$ to denote the vector of unique elements of $t[n+1]$. If the induced partial ranking b is an element of Φ_1 then $\mathbf{s} = \bar{\mathbf{s}}$. If instead $b \in \Phi_2$ then $\mathbf{s}_{-i^\star(b)} = (s_1, \ldots, s_{i^\star(b)-1}, s_{i^\star(b)+1}, \ldots, s_{\#b}) = \bar{\mathbf{s}}$. We now define

$$g(b, s_1, \ldots, s_{\#b}) = f_{n+1}(b, s_1, \ldots, s_{\#b}) / f_n(\bar{b}, \bar{s}_1, \ldots, \bar{s}_{\#\bar{b}}).$$

Then for any Borel set A the conditional probability distribution function is given by

$$\begin{aligned} P(T_{n+1} \in A \mid T[n] = t[n]) = \sum_{b \in \Phi_1} & g(b, \bar{s}_1, \ldots, \bar{s}_{\#b}) \mathbf{1}[\bar{s}_{i^\star(b)} \in A] \\ & + \sum_{b \in \Phi_2} \int_{A^\star(b, \bar{\mathbf{s}})} g(b, s_1, \ldots, s_{\#b}) ds_{i^\star(b)} \end{aligned} \tag{11}$$

where $\mathbf{1}[\cdot]$ is the indicator function and

$$\begin{aligned} A^\star(b, \bar{\mathbf{s}}) = \{ \mathbf{s} = (s_1, \ldots, s_{\#b}) \in \mathbb{R}_+^{\#b} : s_1 < \ldots < s_{\#b} \\ \text{and } \ \mathbf{s}_{-i^\star(b)} = \bar{\mathbf{s}} \ \text{ and } \ s_{i^\star(b)} \in A \} \end{aligned}$$

Based on equation (11), we can provide a simple description of the predictive distribution $P(T_{n+1} > t \mid T[n] = t[n])$. See Appendix B for a detailed derivation. First let $X_1^{(n+1)}$ and $X_2^{(n+1)}$ be independent random variables given $T[n] = t[n]$. Let the probability distribution for $X_1^{(n+1)}$ be absolutely continuous with respect to Lebesgue measure. Therefore the hazard function is well-defined and set equal to

$$h_1(s \mid t[n]) = \zeta(r^\sharp(s) + 1) - \zeta(r^\sharp(s)) = (\Delta\zeta)(r^\sharp(s))$$

where $r^\sharp(s) = \#\{i \in [n] : t_i > s\}$. Then the survival distribution is

$$P(X_1^{(n+1)} > t \mid T[n] = t[n]) = \exp\left(- \int_0^t (\Delta\zeta)(r^\sharp(s)) ds \right)$$

Recall $\bar{\mathbf{s}}$ denotes the unique elements of $t[n]$ and $\bar{b}$ the associated ordered partition. The probability distribution for $X_2^{(n+1)}$ has discrete support $\bar{\mathbf{s}} \cup \{\infty\}$. At each $\bar{s}_i$ define $r_i = r^{\sharp}_{[n]}(\bar{s}_i)$ and $b_i = b^{\sharp}_{[n]}(\bar{s}_i)$. We define the probability distribution via its associated discrete hazard function

$$P(X_2^{(n+1)} = \bar{s}_i \mid T[n] = t[n], X_2^{(n+1)} \geq \bar{s}_i) = 1 - \frac{(\Delta^{b_i}\zeta)(r_i + 1)}{(\Delta^{b_i}\zeta)(r_i)}. \tag{12}$$

The discrete hazard is set equal to one at $\bar{s}_{\#\bar{b}+1} = \infty$.

Equation (11) implies the distribution of T_{n+1} is equal in law to $X_1^{(n+1)} \wedge X_2^{(n+1)}$ conditional on $T[n] = t[n]$. This implies the predictive distribution can be written as:

$$\begin{aligned} P(T_{n+1} > t \,|T[n] = t[n]) &= P(X_1^{(n+1)} \wedge X_2^{(n+1)} > t \mid T[n] = t[n]) \\ &= P(X_1^{(n+1)} > t \mid T[n] = t[n]) \cdot P(X_2^{(n+1)} > t \mid T[n] = t[n]) \\ &= \exp\left(-\int_0^t (\Delta\zeta)(r^{\sharp}(s))ds\right) \cdot \prod_{j:\bar{s}_j \leq t} \frac{(\Delta^{b_j}\zeta)(r_j + 1)}{(\Delta^{b_j}\zeta)(r_j)} \end{aligned} \tag{13}$$

5.2. Weak continuity of predictions

As seen above, tied failures are an intrinsic aspect of every exchangeable Markov survival process; in practice, ties usually occur as a result of rounding, and are not intrinsic. If exchangeable Markov survival processes are to have any role in applied work, it is essential that the model should not be sensitive to rounding. Sensitivity to rounding is addressed in this section by asking for exchangeable Markov survival processes whose predictive distributions are weakly continuous.

Section 5.1 shows the predictive distribution of T_{n+1} given the sequence $T[n] = t[n]$ of previous failures can be described by multiplying the predictive distributions of two independent random variables: one a discrete probability distribution with support on the unique failure times of $t[n]$ and ∞, and the other absolutely continuous with respect to Lebesgue measure. The predictive distribution is weakly continuous if a small perturbation of the failure times gives rise to a small perturbation of the predictive distribution. Formally, for each $n \geq 1$, and for each non-negative vector $t[n]$,

$$\lim_{\epsilon \to 0} P(T_{n+1} > t \mid T[n] = t[n] + \epsilon) = P(T_{n+1} > t \mid T[n] = t[n])$$

at each continuity point, i.e., $t > 0$ not equal to one of the prior times $t[n]$.

Weak continuity is an additivity condition on the discrete hazard function given in equation (12). In particular, the conditional survival distribution is weakly continuous if and only if for every integer $r \geq 0$ and $d \geq 1$

$$\frac{(\Delta^d\zeta)_{r+1}}{(\Delta^d\zeta)_r} = \frac{(\Delta\zeta)_{r+d}}{(\Delta\zeta)_r}.$$

The above condition implies equality between d failures occurring as singletons ϵ-apart and d failures occurring simultaneously with r particles remaining at risk in both cases.

Theorem 5.1 (Continuity of predictions)**.** *An exchangeable, Markov survival process has weakly continuous predictive distributions if and only if it is a harmonic process: $\zeta_n \propto \psi(n+\rho) - \psi(\rho)$ for some $\rho > 0$. The iid exponential model is included as a limit point.*

To see that it is not satisfied by any other exchangeable, Markov survival process, it is sufficient to consider a sequence in standard form beginning with $(\Delta\zeta)_0 = 1$, $(\Delta\zeta)_1 = \rho/(\rho+1) < 1$. Then the key continuity condition determines the subsequent sequence $(\Delta\zeta)_r = \rho/(\rho+r)$ in conformity with the harmonic series. The only exception is the iid exponential process, which arises in the limit $\rho \to \infty$ in which tied failures occur with probability zero. All other exchangeable Markov survival processes have conditional survival distributions that are discontinuous as a function of the configuration $t[n]$. A detailed proof of Theorem 5.1 can be found in Appendix C.

6. Key behavior

In this section, we describe key behavior of the harmonic process. Questions of interest are mostly related to the behavior of the sequences $T_1, \ldots, T_n$ both for finite n and the limit as $n \to \infty$. Some very natural questions are related to the number and the sizes of the blocks. Proofs of these results can be found in the Appendix.

Remark 6.1. By Sections 5 and 4.2.1, the function $f_n(b, s_1, \ldots, s_{\#b})$ for the harmonic process is equal to

$$\nu^{\#b} \rho^{-\uparrow n} \exp\left(-\nu \int_0^\infty \big(\psi(r^\sharp(s)+\rho) - \psi(\rho)\big)\, ds\right) \prod_{i=1}^{\#b} \Gamma(b_i)$$

where $\rho^{\uparrow n} = \rho(\rho+1)\ldots(\rho+n-1)$ is the rising factorial, $\psi(\cdot)$ is the derivative of the log gamma function, and $\Gamma(\cdot)$ is the gamma function.

Following Section 5.1, the hazard function for $X_1^{(n+1)}$ for the harmonic process is $h(t \mid t[n]) = \nu/(\rho + r^\sharp(t))$, implying that the cumulative hazard

$$H(t \mid t[n]) = \nu \frac{t - \bar{s}_j}{\rho + r^\sharp(\bar{s}_j)} + \sum_{i:\bar{s}_i < t} \nu \frac{\bar{s}_i - \bar{s}_{i-1}}{\rho + r^\sharp(\bar{s}_{i-1})}$$

where $\bar{s}_0 = 0$ and $\bar{s}_j = \max\{\bar{s}_i : \bar{s}_i < t\}$.

The cumulative discrete hazard function for $X_2^{(n+1)}$ for the harmonic process is

$$\prod_{i:\bar{s}_i \le t} \frac{(\Delta^{b_i}\zeta)(r_i+1)}{(\Delta^{b_i}\zeta)(r_i)} = \prod_{i:\bar{s}_i \le t} \frac{\rho + r_i}{\rho + r_i + b_i}.$$

For small ρ, the discrete component is essentially the same as the right-continuous version of the Kaplan-Meier product limit estimator. Note that exchangeability implies the joint conditional survival probability, $P(T_{n+1} > t, T_{n+2} > t' \mid T[n] = t[n])$ is distinct from the Kaplan-Meier product estimate. Figure 1 shows several simulated harmonic survival processes for various choices of $\rho > 0$ along with the predictive distribution function.

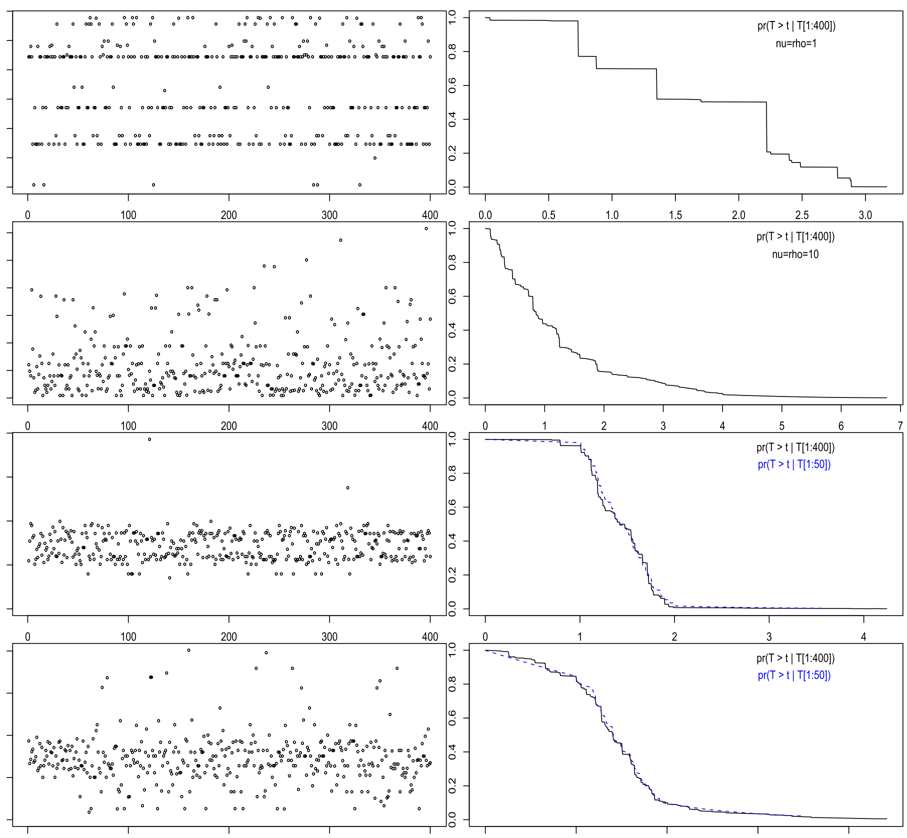


FIG 1. *Simulated harmonic process, together with the conditional survival function (right panel). The lower two series are seeded with 50 initial values, independent, uniform on* $(1, 2)$*. Parameter values:* $\nu = \rho = 1$ *in rows 1, 3;* $\nu = \rho = 10$ *in rows 2, 4. Among the first 400 failure times, the number of distinct values was 15, 93, 58, 110, respectively, including the initial seeds.*

6.1. Number of blocks & block sizes

Here we investigate the expected number of blocks $\mu_n = E[\#B]$ for the random ordered partition $B \in \mathcal{OP}_n$ with probability distribution p_n. This depends critically on the associated splitting rule $q(\cdot, \cdot)$. In particular, the mean number of blocks satisfies the recurrence relation

$$\mu_n = 1 + \sum_{d=1}^{n} \binom{n}{d} q(n-d, d)\mu_{n-d}$$

with $\mu_0 = 0$ and $\mu_1 = 1$. The interest here is in the behavior of μ_n for large n.

Lemma 6.2. *For the harmonic process, the mean number of blocks μ_n is a non-decreasing function of n that satisfies the recursion*

$$\mu_n = 1 + \zeta_n^{-1} \sum_{d=1}^{n} \frac{\Gamma(n+1)}{\Gamma(n+\rho)} \frac{\Gamma(n-d+\rho)}{\Gamma(n-d+1)} \frac{\mu_{n-d}}{d}.$$

This implies $\mu_n \sim \log^2(n)$ where $x_n \sim y_n$ if $\lim_{n\to\infty} x_n/y_n = C \in (0,\infty)$ for x_n and y_n two non-negative sequences.

See Appendix D for the proof. Remark 6.3 gives an approximate description of how the number of blocks of a given size grows as a function of n.

Remark 6.3. The number of blocks of size j is approximately Poisson with asymptotic rate proportional to $\log(n)/j$ with independent components for $j \neq j'$.

See Appendix E for a sketch of the proof for Remark 6.3. Lemma 6.4 answers the question of interest of how the block size grows as a function n.

Lemma 6.4. *The expected number of particles in the first block satisfies $\mathbb{E}[\#B_1] \sim n/(\rho\log(n))$. Asymptotically,*

$$\frac{\log(\#B_1)}{\log(n)} \xrightarrow{D} U$$

where U is the uniform distribution on $(0,1)$.

See Appendix D for the proof.

7. Beta-splitting processes

As stated in Section 4.2.1, a natural family of dislocation measures are those conjugate to the binomial splitting probabilities, $\varpi(dx) = x^{\rho-1}(1-x)^{\beta-1}dx$ for $\rho > 0$ and $\beta > -1$. We refer to exchangeable Markov survival processes arising from this conjugate choice of dislocation measure with zero erosion measure ($c = 0$) as *beta-splitting processes*. While the harmonic process (i.e., $\beta = 0$) composes the natural subset of beta-splitting processes for time-to-event data, one may ask whether alternative choices of β may well approximate the behavior of the harmonic process. In this section, we answer in the negative. Lemma 7.1 shows that the asymptotic behavior is a discontinous function of β. In particular, we see marked difference in asymptotic behavior for $\beta > 0$ and $\beta < 0$.

Lemma 7.1. *Let $\mu_n(\rho,\beta)$ denote the expected number of blocks given n particles for the beta-splitting process with $\rho \in (0,\infty)$ and $\beta \in (-1,\infty)$. Then*

1. *For $\beta > 0$, the characteristic index is*

$$\zeta_n = B(\rho, \beta) - B(\rho + n, \beta) = B(\rho, \beta)\left(1 - \frac{\rho^{\uparrow n}}{(\rho + \beta)^{\uparrow n}}\right)$$

where $\rho^{\uparrow k} = \rho \cdot (\rho + 1) \ldots (\rho + k - 1)$ is the ascending factorial and $B(\rho, \beta)$ is the beta function. The expected number of blocks satisfies

$$\mu_n(\rho, \beta) \sim \log(n)$$

where $x_n \sim y_n$ if $\lim_{n\to\infty} x_n/y_n = C \in (0, \infty)$ for x_n and y_n two non-negative sequences. For $\beta > 0$ we have $C = \frac{B(\rho,\beta)}{\psi(\rho+\beta)-\psi(\rho)}$.
The fraction of particles in the first block, $n^{-1}\#B_{1,n}$, is asymptotically distributed Beta(β, ρ). Therefore, the relative frequencies within each block are given by

$$(P_1, P_2, \ldots) = (W_1, \bar{W}_1 W_2, \bar{W}_1 \bar{W}_2 W_3, \ldots)$$

where W_i are independent beta variables with parameters (β, ρ), and $\bar{W}_i = 1 - W_i$. The number of blocks of size j is approximately Poisson with asymptotic rate proportional to $1/j$ with independent components for $j \neq j'$.

2. *For $\beta \in (-1, 0)$, the characteristic index is*

$$\zeta_n = \sum_{j=0}^{n-1} \frac{\Gamma(j + \rho)\Gamma(\beta + 1)}{\Gamma(j + 1 + \beta + \rho)}.$$

The expected number of blocks satisfies $\mu_n \sim n^{-\beta}$ where $C = -\Gamma(\rho + \beta + 1)/(\Gamma(\rho)\beta)$. For all $\rho > 0$,

$$\lim_{n\to\infty} P(\#B_{1,n} = d) = g_\beta(d)$$

where $g_\beta(d)$ is defined by

$$g_\beta(d) = \frac{-\beta \cdot \Gamma(d + \beta)}{\Gamma(d + 1) \cdot \Gamma(1 + \beta)} = \frac{-\beta}{\Gamma(1 + \beta)} d^{\beta - 1}$$

for large d. So the number of particles in the first block has a power law distribution of degree $1 - \beta$.

Proof can be found in Appendix F.

7.1. Splitting rules of Gibbs-type

A question of interest is whether the family of beta-splitting processes can be characterized in a similar manner to Aldous' beta-splitting models for Gibbs fragmentation trees (McCullagh, Pitman, and Winkel, 2008). Proposition 7.2 below states that like in the fragmentation tree setting, the family of beta-splitting processes are the only exchangeable, Markov survival processes with Gibbs-type splitting rules.

Proposition 7.2. *Define splitting rules of Gibbs-type to be consistent splitting rules of the form*

$$q(r,d) = \frac{w_1(r) w_2(d)}{Z(r+d)} \qquad \text{for all } r \geq 0,\ d \geq 1 \tag{14}$$

for two non-negative sequences of weights $w_1(r) \geq 0$ with $r \geq 0$, $w_2(d)$ with $d \geq 1$, and normalization constants $Z(n)$, $n \geq 1$. Then the beta-splitting rules are the only consistent splitting rules of Gibbs-type.

By the integral representation, the beta-splitting process has splitting rule:

$$\begin{aligned} q(r,d) &= \frac{1}{Z(r+d)} \int_0^1 x^{r+\rho-1}(1-x)^{d+\beta-1} dx \\ &= \frac{\Gamma(r+\rho)\Gamma(d+\beta)}{\Gamma(r+d+\rho+\beta)\cdot\zeta_{r+d}} = \frac{\Gamma(r+\rho)\Gamma(d+\beta)}{Z(r+d)}. \end{aligned}$$

for all $r \geq 0$, $d \geq 1$. The coefficients satisfy $\rho > 0$ and $\beta > -1$. Therefore, we can see the beta-splitting rules are a two-parameter family of Gibbs-type. It rests to show that these are the only splitting rules of Gibbs-type.

Proof. Start from equation (14) and assume $w_1(0) = w_2(1) = 1$ (i.e., $Z(1) = 1$ and $q(0,1) = 1$ which is true for all splitting rules). For any consistent splitting rule, this implies $w_l(j) > 0$ for all $j \geq 1$. The consistency criterion (3) in terms of $W_l(j) = w_l(j+1)/w_l(j)$ now gives

$$W_1(r) + W_2(d) = \frac{Z(r+d+1) - w_1(r+d)}{Z(r+d)} = f(r+d)$$

for all $r \geq 0$ and $d \geq 1$. The righthand side is a function of $r+d$, $f(r+d)$. This implies that for $j' \geq 0$ and $j \geq 1$

$$\begin{aligned} &W_1(j') + W_2(j+1) = f(j'+j+1) = W_1(j'+1) + W_2(j) \\ \Rightarrow\ &W_2(j+1) - W_2(j) = W_1(j'+1) - W_1(j). \end{aligned}$$

Therefore $W_2(j+1) - W_2(j)$ for $j \geq 1$ and $W_1(j'+1) - W_1(j')$ for $j' \geq 0$ are all equal to the same constant $b \in \mathbb{R}$. Hence, $W_1(j') = a_1 + b \cdot j'$ for $j' \geq 0$ and $a_1 > 0$, and $W_2(j) = a_2 + b \cdot j$ for $j \geq 1$ and $a_2 > -b$. Now, either $b = 0$ or we have

$$\begin{aligned} w_1(j) &= W_1(0) \cdots W_1(j-1) = \prod_{q=0}^{j-1} (a_1 + b \cdot q) \\ &= b^j \prod_{q=0}^{j-1} \left(\frac{a_1}{b} + q \right) = b^j \frac{\Gamma(a_1/b + j)}{\Gamma(a_1/b)} \end{aligned}$$

and

$$w_2(j) = W_2(1) \cdots W_2(j-1) = \prod_{q=1}^{j-1} (a_2 + b \cdot q)$$

$$= b^{j-1}\prod_{q=1}^{j-1}\left(\frac{a_2}{b}+q\right) = b^{j-1}\frac{\Gamma(a_2/b+j)}{\Gamma(a_2/b+1)}$$

Reparameterizing by $\beta = a_2/b$ and $\rho = a_1/b$ yields

$$q(r,d) = \frac{w_1(r)w_2(d)}{Z(r+d)} = \frac{1}{Z(r+d)}\frac{b^r\Gamma(r+\rho)}{\Gamma(\rho)}\frac{b^{d-1}\Gamma(d+\beta)}{\Gamma(1+\beta)}.$$

Setting $\tilde{Z}(r+d) = Z(r+d)b^{1-r-d}\Gamma(\rho)\Gamma(1+\beta)$ to be the normalization constant then $q(r,d) = \tilde{Z}^{-1}(r+d)\Gamma(r+\rho)\cdot\Gamma(d+\beta)$. The case $b=0$ is the limiting case of the beta-splitting rules where $\beta \to \infty$ and $\rho \to \infty$ such that $\beta/\rho \to \theta$. Then $q(r,d) \propto \theta^{d-1}$. □

8. Connections to the literature

8.1. Bayesian survival analysis

In this paper we have characterized the class of exchangeable, Markov survival processes – stochastic processes giving rise to infinitely exchangeable non-negative sequences $(T_1, T_2, \ldots)$. By de Finetti's theorem (see, e.g., Aldous (1985)), the distribution of every such countably infinite sequence can be expressed as a mixture of independent, identically distributed (i.i.d.) sequences. Based on the de Finetti approach, Doksum (1974) introduced a class of random distribution functions $F(t)$ which are said to be *neutral to the right* (NTR) if the normalized increments

$$F(t_1), \frac{F(t_2)-F(t_1)}{1-F(t_1)}, \ldots, \frac{F(t_k)-F(t_{k-1})}{1-F(t_{k-1})}$$

are independent for all $t_1 < t_2 < \cdots < t_k$. A key property of random NTR distribution functions, is that the posterior distribution is also NTR (Doksum, 1974, Theorem 4.2). Doksum (1974, Theorem 3.1) showed that an NTR distribution function F on $\mathbb{R}_+$ can be written in terms of a positive Lévy process Z on $\mathbb{R}_+$:

$$F(t) = 1 - \exp(-Z(t))$$

The Lévy process Z is assumed to have non-negative independent increments, be non-decreasing a.s., be right continuous a.s., $Z(0) = 0$ a.s., and $\lim_{t\to\infty} Z(t) = \infty$ a.s.. We limit focus to *stationary* Lévy processes which are invariant under translation in $\mathbb{R}_+$. The non-negative random sequence $(T_1, T_2, \ldots)$ given Z is then a sequence of independent random variables each with cumulative distribution function $F(\cdot)$.

As these processes are positive, it is natural to work with the cumulant function

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

for $t \geq 0$ where $X = Z(1)$ is an infinitely divisible distribution. The Lévy-Khintchine characterization for positive, stationary, Lévy processes implies

$$K(t) = -\left[\gamma \cdot t + \int_0^\infty (1 - e^{-ty}) w(dy)\right] \tag{15}$$

for some $\gamma \geq 0$ and some measure w on $\mathbb{R}_+$, called the Lévy measure, such that the integral is finite for $t > 0$. The unconditional multivariate survival function is

$$\begin{aligned} P(T_1 > t_1, \ldots, T_n > t_n) =& E\left[P(T_1 > t_1, \ldots, T_n > t_n \mid Z)\right] \\ =& E\left[\prod_{i=1}^n \exp(-Z(t_i))\right] \\ =& \exp\left(\int_0^\infty K(r_n^\sharp(s)) ds\right) \end{aligned}$$

where $r_n^\sharp(t) = \#\{i \in [n] : t_i > t\}$. For example, the cumulant function and the Lévy measure for the gamma process considered by Kalbfleisch (1978) and Clayton (1991) are

$$K^\star(t) = -\nu \log(1 + t/\rho) \quad \text{and} \quad w^\star(dy) = \nu y^{-1} e^{-\rho y}\, dy.$$

The cumulant function and the Lévy measure for the harmonic process are

$$K(t) = -\nu\left(\psi(\rho + t) - \psi(\rho)\right) \quad \text{and} \quad w(dy) = \nu e^{-\rho y}\, dy/(1 - e^{-y}),$$

where ψ is the derivative of the log gamma function.

Processes of this type have been investigated by Doksum (1974), Ferguson and Phadia (1979), Kalbfleisch (1978), Clayton (1991), Walker and Muliere (1997), and James (2006). In particular, Hjort (1990) built the associated Lévy process Z indirectly by constructing a prior distribution, termed the beta process, on the space of cumulative hazard functions. Walker and Muliere (1997) instead constructed priors directly on the cumulative distribution function (cdf) F. Termed beta-Stacy processes, Walker and Muliere construct the random cumulative distribution function via the associated Lévy process Z. In both instances, the process is defined via the Lévy process, its associated Lévy measure, and the representation of NTR processes introduced by Doksum (1974).

By standard Bayesian calculations it can be shown that the sequence of unconditional distributions for the random variables $T[n]$ generated via the above Lévy process construction are exchangeable, Markov survival processes with characteristic index $\zeta_n = -K(n)$ evaluated at $n = 1, 2, \ldots$. See, for instance, the proof of Proposition 1 in Lijoi, Prunster, and Walker (2008). The converse is also true. Namely every exchangeable, Markov survival process can be generated via the above Lévy process construction. See Appendix G for details. Therefore every exchangeable Markov survival process corresponds to a particular choice of Lévy process and thus fits naturally within the existing literature.

8.2. Regenerative composition structures

As stated previously, an exchangeable Markov survival process is an exchangeable Markov partial ranking process embedded in continuous time. The unlabelled exchangeable Markov partial ranking counting only block sizes embedded within the survival process is equivalent to a particular regenerative composition structure (Gnedin and Pitman, 2005). A *composition* of n is an order-dependent integer partition $\lambda = (n_1, \ldots, n_k)$ where $\sum_j n_j = n$. A random composition of n is a random variable $\mathcal{C}_n$ taking values in the space of compositions of n. A composition structure is a sequence of random compositions $(\mathcal{C}_n)_{n=1}^{\infty}$. The structure is *regenerative* if given the first part of $\mathcal{C}_n$ is m, then the remaining composition of $n-m$ is distributed like $\mathcal{C}_{n-m}$.

9. Parameter estimation

Consider the problem of parameter estimation for a two parameter Markov survival process with characteristic index of the form $\zeta_n = \nu\Psi(n;\rho)$ where $\Psi(n;\rho) = \Phi(n+\rho)-\Phi(\rho)$ for $\nu > 0$, $\rho > 0$ with $\Phi(\cdot)$ given. Such a family is generated from a family of Lévy measures proportional to $w(dy)e^{-\rho y}$, so the Lévy process Z is expected to have larger atoms if ρ is small. The gamma and harmonic processes are of this form with $w((0,1))$ and $w((1,\infty))$ both infinite.

The goal in this section is estimation of the parameters ρ and ν from observations $t[n] = (t_1, \ldots, t_n)$. In this paper, we consider maximum likelihood estimation. Let b and $\mathbf{s} = (s_1, \ldots, s_{\#b})$ be the ordered partition and the ordered unique times (i.e., $s_1 < \cdots < s_{\#b}$) induced by $t[n]$. Then the log-likelihood is given by:

$$\begin{aligned}&\log(f_n(b, s_1, \ldots, s_{\#b}; \nu, \rho)) \\ &\quad = (\#b)\log(\nu) - \nu\int_0^\infty \left(\Psi(r^\sharp(u);\rho)\right) du + \sum_{i=1}^{\#b} \log\left(\lambda(r_i, b_i;\rho)\right)\end{aligned}$$

where $r_i = r^\sharp(s_i)$ and $b_i = b^\sharp(s_i)$. Note that the second term is only a function of ρ. Therefore, for fixed ρ, the log-likelihood as a function of ν has a two-dimensional sufficient statistic:

$$\#b, \int_0^\infty \Psi(r^\sharp(u);\rho)du.$$

Taking the first derivative of the log-likelihood the maximum-likelihood estimate is the ratio

$$\hat{\nu} = \left[\frac{1}{\#b}\int_0^\infty \Psi(r^\sharp(u);\rho)du\right]^{-1}.$$

The Fisher information for $\log(\nu)$ is $E[\#B]$, suggesting the asymptotic variance of $\log(\hat{\nu})$ is $1/E[\#B]$. As shown in section 6.1, for the gamma and harmonic process $E[\#B] \propto \log^2(n)$. Therefore, in the absence of censoring, the estimator is consistent but the rate of convergence is very slow.

One natural alternative to maximum likelihood estimation for the parameter ρ is to consider the product of the per-particle death rate $\nu\Psi(1;\rho)$ and the total particle time at risk $\int R^\sharp(t)\,dt$, and to estimate ρ by setting the product to the observed number of deaths, i.e.,

$$\#b = \hat{\nu}\hat{\Psi}(1;\rho)\int_0^\infty r^\sharp(t)\,dt. \tag{16}$$

In the absence of censoring, this is equivalent to setting the mean survival time $\bar{T}_n$ to its expected value $1/\hat{\zeta}_1$. But $\mathrm{cov}(T_i, T_j) = 1/\zeta_2^2$ for each pair $i \neq j$ implies that

$$\mathrm{var}(\bar{T}_n) = 1/(n\zeta_1^2) + (n-1)/(n\zeta_2^2) \to 1/\zeta_2^2$$

does not tend to zero as $n \to \infty$. Nonetheless, this second equation is less sensitive than the first to rounding of survival times, which is a desirable property for applied work. The parameter pair can then be estimated by iteration. We present both maximum likelihood estimates for the parameter pair and estimates using the natural relation in the numerical example below.

9.1. Impact of censoring on parameter estimation

To each unit $i \in [n]$, there often corresponds an observational interval $[0, c_i]$ where c_i is some arbitrary positive censoring time. The event time recorded for unit i is $Y_i = \min(T_i, c_i)$. So if $Y_i = c_i$ the event is a censoring time; otherwise, if $Y_i < c_i$, the event is known to be death or failure. To each unit, let $\Delta_i = \mathbf{1}[Y_i = c]$ be the indicator of whether the event is a censoring time.

Let $(Y[n], \Delta[n])$ be the random pair of event times and censoring indicators. Then we re-define the random Boolean function $R_i : \mathbb{R}_+ \to \{0, 1\}$ to be a function of (Y_i, Δ_i)

$$R_i(t) = \mathbf{1}[Y_i > t]\mathbf{1}[\Delta_i = 0] + \mathbf{1}[Y_i \geq t]\mathbf{1}[\Delta_i = 1].$$

That is, the particle i is "at risk" at time t if it is known to be alive at time t, which includes being censored at time t.

Given a realization $(y[n], \delta[n])$, we now consider how to construct the inputs to $f_n(b, s_1, \ldots, s_{\#b})$. First, the order partition b is induced by the restriction of $y[n]$ to the set of failure times $\{i \in [n] : \delta_i = 0\}$. Similarly the ordered sequence $\mathbf{s} = (s_1, \ldots, s_{\#b})$ is also induced by this restricted set. The cardinality of the risk set is given by $r^\sharp(t) = \sum_{i=1}^n r_i(t)$ where r_i the re-defined risk trajectory. Then equation (9) is correct with the above adjustments to the definition of the risk set trajectory $r(t)$, ordered partition b, and ordered sequence $\mathbf{s}$.

Thus censoring has a relatively trivial effect on probability calculations and thus on associated statistical procedures. A more general condition on the censoring mechanism is introduced in Appendix H. The condition includes independent censoring (Anderson et al., 1993) and allows generalization from deterministic censoring.

9.2. Numerical example

Consider parameter estimation for a set of failure and censoring times (in weeks) of the 6-MP subset of leukemia patients taken from Gehan (1965):

$$6, 6, 6, 6^\star, 7, 9^\star, 10, 10^\star, 11^\star, 13, 16, 17^\star, 19^\star, 20^\star, 22, 23, 25^\star, 32^\star, 32^\star, 34^\star, 35^\star$$

There are 9 uncensored observations, and a total risk time of 359 weeks. Assuming the survival times are iid exponential with rate parameter, θ, then the maximum likelihood estimate of θ is given by $9/359$, or an expected survival time of 39.89 weeks.

Consider the two-parameter Markov survival process defined in Section 9, specifically the harmonic and gamma processes. Table 1 provides maximum likelihood estimates for ρ and ν. For the gamma process, the empirical Bayes estimate of the rate is then $\hat{\nu} \cdot \log(1+\hat{\rho}^{-1}) \approx 2.47 \times 10^{-2}$, implying the expected survival time is 40.52 weeks. The expected time is the same for the harmonic process.

Table 1
Maximum likelihood estimates for two processes

	Harmonic process		Gamma process	
Parameter	Est.	Std. Error	Est.	Std. Error
ρ	21.45	19.63	20.95	19.61
ν	0.53	0.44	0.53	0.44

Estimation using the maximum likelihood estimate of ν given ρ and the natural relation between the marginal survival rate associated with the gamma process and $\hat{\theta}$

$$\hat{\nu}(\rho)\,(\Phi(1+\rho) - \Phi(\rho)) = \hat{\theta} = 9/359$$

yields $(\hat{\rho}, \hat{\nu}) = (19.24, 0.49)$ for the gamma process and $(19.73, 0.49)$ for the harmonic process. Supplementary figure 1 shows that the profile likelihood for ρ is relatively flat for values sufficiently removed from the origin. For the harmonic process, the figure suggests a 95% confidence interval of approximately $[1.3, 5.1]$ for $\log(\rho)$, while under the gamma process, there is an approximate confidence interval of $[1.2, 5.1]$. Twice the difference in the log-likelihoods at their respective maxima is 8.12×10^{-5}.

Figure 2 shows the conditional survival distribution given the observed risk set trajectory. The empirical Bayes estimate of the conditional distribution for the harmonic process is approximately equal to that of the gamma process. Both are approximately an average of the Kaplan-Meier product limit estimator and the maximum likelihood exponential estimator of the conditional survival distribution.

Appendix I discusses the Markov survival process and parameter estimation when covariate effects are included.

9.3. Ties as a result of numerical rounding

In Section 9, individuals having the same recorded survival time are regarded as failing simultaneously. This viewpoint assumes that the data are exact survival times generated by the underlying exchangeable Markov survival process. In practice, grouped data are observed due to rounding of intrinsically continuous-time processes. Rounding reflects the granularity at which the data is collected; for example, failure may only be known up to the day. In our numerical exam-

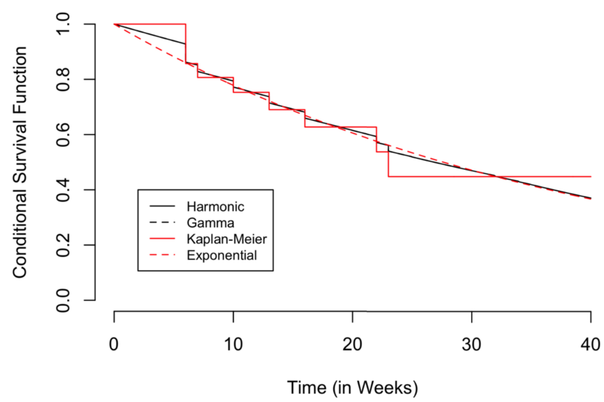


FIG 2. *Conditional survival distribution for leukemia patients*

ple, remission was known up to the week. Ideally exact survival times would be observed, but this is often not the case unless patients are under intense supervision.

Here we consider the effect on the likelihood when actual failure times are in fact distinct, so that tied values arise solely as a result of numerical rounding (e.g., grouping). The integral in the exponent of the joint density is a continuous function of the risk set trajectory $R(t)$, so an ϵ-perturbation of failure times has an $O(\epsilon)$ effect on the integral, which is ignored. However, the remaining term is not a continuous function of the observations, so an ϵ-perturbation by rounding may have an appreciable effect on the likelihood. Most obviously, the statistic $\#b$, the number of distinct failure times, is not continuous as a function of $t[n]$; if ties are an artifact of rounding, then $\#b$ is the total number of failures.

While the likelihood and parameter estimation are affected by ties as a result of numerical rounding, the conditional survival distribution for the harmonic process given ρ and ν is unaffected due to the weak continuity of predictive distributions. This suggests it may be best to regard ρ as a fixed "tuning parameter". As all other processes have discontinuous predictive distributions, the use of the harmonic process in applications where ties are the result of numerical rounding (e.g., grouping) seems most natural.

Appendix A: Proof of Proposition 4.5

Proof. By Proposition 4.2, we know consistent splitting rules are completely determined by the set of singleton splitting rules, $\{q(n,1) \mid n \geq 0\}$ where $q(0,1) = 1$ by definition. By construction in Definition 4.4, we have $q(r,1) = \Delta\zeta_r/\zeta_{r+1} = 1 - \zeta_r/\zeta_{r+1}$. Considering the standardized sequence and using this relation yields a one-to-one correspondence between the set of singleton splitting rules and the characteristic index. Therefore the characteristic index completely determines a splitting rule.

Showing equation (6) yields a splitting rule that satisfies the consistency condition (3) completes the proof. The left hand side is given by

$$\begin{aligned}(1-q(n,1))q(n-d,d) &= \frac{\zeta_n}{\zeta_{n+1}} \frac{(-1)^{d-1}(\Delta^d \zeta)_{n-d}}{\zeta_n} \\ &= \frac{(-1)^{d-1}(\Delta^d \zeta)_{n-d}}{\zeta_{n+1}}\end{aligned}$$

The right hand side is given by

$$\begin{aligned}q(n-d+1,d)+q(n-d,d+1) &= \frac{(-1)^{d-1}(\Delta^d \zeta)_{n-d+1}}{\zeta_{n+1}} + \frac{(-1)^{d}(\Delta^{d+1} \zeta)_{n-d}}{\zeta_{n+1}} \\ &= \frac{(-1)^{d-1}}{\zeta_{n+1}} \cdot \left[(\Delta^d \zeta)_{n-d+1} - (\Delta^{d+1}\zeta)_{n-d}\right] \\ &= \frac{(-1)^{d-1}}{\zeta_{n+1}} \cdot (\Delta^d \zeta)_{n-d}\end{aligned}$$

So the splitting rule satisfies the consistency condition (3). By definition, the splitting rule must be non-negative, and therefore the sequence defines a consistent splitting rule if the forward differences are non-negative – i.e., $(-1)^{d-1}(\Delta^d\zeta)_r \geq 0$ for all $r \geq 0$, $d \geq 1$. □

Appendix B: Sequential description – technical details

Here we provide a detailed derivation of the conditional survival distribution $P(T_{n+1} > t \mid T[n] = t[n])$. First, for $A = (t, \infty)$ each term within the first sum in equation (11) satisfies

$$\begin{aligned}&g(b, \bar{s}_1, \ldots, \bar{s}_{\#b}) \mathbf{1}[\bar{s}_{i^\star(b)} > t] \\ &\quad = \exp\left(-\int_0^t (\Delta\zeta)(r^\sharp(s))ds\right) \prod_{j:\bar{s}_j \leq t} \left[\frac{\lambda(r_j+1, b_j)}{\lambda(r_j, b_j)}\right] \\ &\quad \times \exp\left(-\int_t^{\bar{s}_{i^\star(b)}} (\Delta\zeta)(r^\sharp(u))du\right) \frac{\lambda(r_{i^\star(b)}, b_{i^\star(b)}+1)}{\lambda(r_{i^\star(b)}, b_{i^\star(b)})} \\ &\quad \times \prod_{j<i^\star(b):t>\bar{s}_j} \frac{\lambda(r_j+1, b_j)}{\lambda(r_j, b_j)} \mathbf{1}[\bar{s}_{i^\star(b)} > t]\end{aligned}$$

The second term is equal to

$$\begin{aligned}&\int_{A^\star(b)} g(b, s_1, \ldots, s_{\#b}) ds_{i^\star(b)} \\ &\quad = \exp\left(-\int_0^t (\Delta\zeta)(r^\sharp(s))ds\right) \prod_{j:\bar{s}_j \leq t} \left[\frac{\lambda(r_j+1, b_j)}{\lambda(r_j, b_j)}\right]\end{aligned}$$

$$\times \int_{A^\star(b)} \exp\left(-\int_t^{s_{i^\star(b)}} (\Delta\zeta)(r^\sharp(u)du)\right) \lambda(r_{i^\star(b)}, 1)$$
$$\times \prod_{j<i^\star(b):t>\bar{s}_j} \frac{\lambda(r_j+1, b_j)}{\lambda(r_j, b_j)} ds_{i^\star(b)}.$$

So each term contains

$$\exp\left(-\int_0^t (\Delta\zeta)(r^\sharp(s))ds\right) \prod_{j:\bar{s}_j \le t} \left[\frac{\lambda(r_j+1, b_j)}{\lambda(r_j, b_j)}\right] \tag{17}$$

We start by factoring out this common term. We then note that the denominator for *every* term can be written in terms of the function f from equation (9). First, let $r(t) = \{i \in [n] : t[n] > t\}$. Then define $\bar{b}_t = \bar{b}\,\big|_{r(t)}$ to be the restriction of the ordered partition $\bar{b}$ to the particles in $r(t)$. If $r(t) = \emptyset$ then the denominator is equal to one (i.e., $f_0(\emptyset) = 1$). Otherwise, define $j^\star = \arg\min_{i\in[n]}\{t_i : t_i > t\}$. Then the denominator is equal to

$$f_{r^\sharp(t)}(\bar{b}_t, \bar{s}_{j^\star} - t, \ldots, \bar{s}_{j^\star+\#\bar{b}_t} - t) = \exp\left(-\int_t^\infty (\Delta\zeta)(r^\sharp(s))ds\right) \prod_{j:\bar{s}_j > t} \lambda(r_j, b_j). \tag{18}$$

In other words, the denominator is the function f evaluated at the ordered partition and unique times induced by $t[r(t)]-t$. We factor out this denominator as well. The remaining terms in the sum are given by

$$\sum_{b\in\Phi_1} \exp\left(-\int_t^{\bar{s}_{i^\star(b)}} \zeta(r^\sharp(u)+1)du\right) \lambda(r_{i^\star(b)}, b_{i^\star(b)}+1)$$
$$\times \prod_{j<i^\star(b):t>\bar{s}_j} \lambda(r_j+1, b_j)\mathbf{1}[\bar{s}_{i^\star(b)} > t]$$
$$\times \prod_{j>i^\star(b):t>\bar{s}_j} \lambda(r_j, b_j) \cdot \exp\left(-\int_{\bar{s}_{i^\star(b)}}^\infty \zeta(r^\sharp(u))du\right)$$
$$+ \sum_{b\in\Phi_2} \int_{A^\star(b)} \exp\left(-\int_t^{s_{i^\star(b)}} \zeta(r^\sharp(u)+1)du\right) \lambda(r_{i^\star(b)}, 1)$$
$$\times \prod_{j<i^\star(b):t>\bar{s}_j} \lambda(r_j+1, b_j)$$
$$\times \prod_{j>i^\star(b):t>\bar{s}_j} \lambda(r_j, b_j) \cdot \exp\left(-\int_{s_{i^\star(b)}}^\infty \zeta(r^\sharp(u))du\right) ds_{i^\star(b)}.$$

The above is equivalent to marginalizing over the $n+1$st particle in the joint distribution of T_{n+1} and $T[r(t)]$ with origin t instead of 0. The linear time-change preserves the consistency and Markovian properties and so the above sum is equal to equation (18). Therefore, these terms cancel and we are left with equation (17) equal to $P(T_{n+1} > t \mid T[n] = t[n])$.

Appendix C: Proof of Theorem 5.1

Here we present the conditions for the predictive distribution, $P(T_{n+1} > t | T[n] = t[n])$, to be a weakly continuous function of $t[n]$. For the predictive distribution to be right continuous, the atomic component (second term in equation (13)) of the predictive distribution must satisfy

$$\frac{(\Delta^{d-1}\zeta)(r+2)}{(\Delta^{d-1}\zeta)(r+1)} \cdot \frac{(\Delta\zeta)(r+1)}{(\Delta\zeta)(r)} = \frac{(\Delta^{d}\zeta)(r+1)}{(\Delta^{d}\zeta)(r)}$$

for $r \geq 0$ and $d > 1$. On the other hand, for the function to be left continuous, the atomic component must satisfy

$$\frac{(\Delta\zeta)(r+d)}{(\Delta\zeta)(r+d-1)} \cdot \frac{(\Delta^{d-1}\zeta)(r+1)}{(\Delta^{d-1}\zeta)(r)} = \frac{(\Delta^{d}\zeta)(r+1)}{(\Delta^{d}\zeta)(r)}$$

Recursive substitution shows the two conditions to be equal and weak continuity holds if

$$\frac{(\Delta\zeta)(r+d)}{(\Delta\zeta)(r)} = \frac{(\Delta^{d}\zeta)(r+1)}{(\Delta^{d}\zeta)(r)} \tag{19}$$

It is now shown that such a condition is uniquely satisfied by the harmonic process.

Proof. Recall the definition of the kth order forward differences

$$\lambda(r,d) = (\Delta^{d}\zeta)(r) = \sum_{j=0}^{d} (-1)^{d-j} \binom{d}{j} \zeta_{r+j}$$

which implies that this forward difference is a linear function of the set $\{\zeta_r, \ldots, \zeta_{r+d}\}$. Start by considering the standardized characteristic index ($\zeta_1 = 1$). First, let $\zeta_2 > 0$ be fixed.

Let $n = r + d > 2$ and $d \geq 1$. Then equation (19) becomes

$$\frac{(\Delta\zeta)(n)}{(\Delta\zeta)(r)} = \frac{(\Delta^{d}\zeta)(r+1)}{(\Delta^{d}\zeta)(r)}$$
$$\Rightarrow (\Delta\zeta)(n) \cdot (\Delta^{d}\zeta)(n-d) = (\Delta^{d}\zeta)(n-d+1) \cdot (\Delta\zeta)(n-d)$$

By definition, $(\Delta\zeta)(n)$ is a linear function of $\{\zeta_n, \zeta_{n+1}\}$ while $(\Delta^{d}\zeta)(r)$ is a function of $\{\zeta_r, \ldots, \zeta_n\}$. Therefore, the left hand side is a linear function of ζ_{n+1} given $\{\zeta_i\}_{i \leq n}$. On the right hand side, $(\Delta^{d}\zeta)(n-d+1)$, is a function of $\{\zeta_{n-d+1}, \ldots, \zeta_{n+1}\}$ while $(\Delta\zeta)(n-d)$ is a function of $\{\zeta_{n-d}, \zeta_{n-d+1}\}$. Since $d \geq 1$, both sides are linear in ζ_{n+1}. Therefore solving for ζ_{n+1} shows the characteristic index is a deterministic function of the previous characteristic index values $\{\zeta_1, \ldots, \zeta_n\}$.

The above argument holds if the coefficient of ζ_{n+1} is nonzero. The coefficient is equivalently zero if $\lambda(n-d, d) = \lambda(n-d, 1)$. Now, suppose equality holds for all $d \in [n]$. By definition, the splitting rule satisfies

$$\begin{aligned}
1 &= \sum_{d=1}^{n} \binom{n}{d} q(n-d, d) \\
\zeta_n &= \sum_{d=1}^{n} \binom{n}{d} \lambda(n-d, d) \quad \text{by definition} \\
\zeta_n &= \sum_{d=1}^{n} \binom{n}{d} \lambda(n-d, 1) \quad \text{by assumption} \\
\zeta_n &= \sum_{d=1}^{n} \binom{n}{d} (\zeta_{n-d+1} - \zeta_{n-d}) \quad \text{by definition} \\
\zeta_n &= \frac{1}{n-1} \left[n \cdot \zeta_{n-1} - \sum_{d=2}^{n} \binom{n}{d} (\zeta_{n-d+1} - \zeta_{n-d}) \right]
\end{aligned}$$

which implies that ζ_n is again a function of the previous characteristic indices.

The above argument shows ζ_k is a deterministic function of $\{\zeta_1, \zeta_2\}$ for $k \geq 3$. However, ζ_2 is constrained by choice of ζ_1. In particular, the holding times satisfy $\zeta_{n+1} = \zeta_n/(1 - q(n, 1))$. Moreover, the splitting rule must satisfy $0 \leq q(n, 1) \leq 1/(n + 1)$. Therefore

$$\zeta_n \leq \zeta_{n+1} \leq \left(1 + \frac{1}{n}\right) \zeta_n$$

This translates to $\zeta_2 = \zeta_1(1 + c)$ for some $c \in [0, 1]$. It rests to show a correspondence between c and the parameter ρ controlling the harmonic process.

For the harmonic process,

$$\nu\Gamma(1)/\rho = \zeta_1$$

so that $\zeta_1 \rho = \nu$. Then due to the normalization of the splitting rule

$$\begin{aligned}
\zeta_2 &= \binom{2}{1} \frac{\nu\Gamma(1)}{(1+\rho)} + \binom{2}{2} \frac{\nu\Gamma(2)}{\rho \cdot (1+\rho)} \\
&= \zeta_1 \left[1 + \frac{\rho}{1+\rho} \right]
\end{aligned}$$

This establishes the correspondence between c and ρ and therefore between all $\{\zeta_1, \zeta_2\}$ that define continuous survival functions and the set of all harmonic processes with parameters ρ and ν. □

Proposition C.1. *If $\lambda(m - d, d) = 0$ for fixed $m \geq 1$ and for all $2 \leq d \leq m$, then $\zeta_n \propto n$ for all n. Thus the only process with trivial forward differences is when T_i are iid exponential.*

Proof. Consider the standardized sequence, $\zeta_1 = 1$, and start by showing that $\lambda(m - d, d) = 0$ implies that $\zeta_k = k$ for $k \leq m$. Indeed, the condition implies that $\zeta_m = m$ and $q(m - 1, 1) = 1/m$. Consistency implies

$$(1 - q(m-1, 1)) \cdot q(m-1-d, d) = q(m-1-d, d+1) + q(m-1-d+1, d)$$

$$
\begin{aligned}
&= q(m-(d+1),(d+1)) + q(m-d,d) \\
&= q(m-d,d) \\
&= \begin{cases} 0 & \text{if } d \geq 2 \\ 1/m & \text{if } d = 1 \end{cases}
\end{aligned}
$$

This implies $q(m-1-d,d) = 0$ for $d \geq 2$ which implies $q((m-1)-1,1) = 1/(m-1)$ which is equivalent to $\zeta_{m-1} = m-1$. Recursively applying this argument yields $\zeta_k = k$ for all $k \leq m$.

It rests to show that $\zeta_k = k$ for $k < n$ implies $\zeta_n = n$. The normalization condition of the splitting rule implies

$$\sum_{d=1}^{n} \binom{n}{d} (\Delta^d \zeta)(n-d) = \zeta_n$$

Now study the dth forward differences:

$$
\begin{aligned}
(\Delta^d \zeta)(n-d) &= \sum_{k=0}^{d} (-1)^{d-k} \binom{d}{k} \zeta_{n-d+k} \\
&= \zeta_n + \sum_{k=0}^{d-1} (-1)^{d-k} \binom{d}{k} (n-d+k) \\
&= \zeta_n - n + \mathbf{1}[d=1] \quad \text{by Lemma C.2}
\end{aligned}
$$

Plugging into the normalization condition yields

$$
\begin{aligned}
\zeta_n &= \sum_{d=1}^{n} \binom{n}{d} \left[\zeta_n - n + \mathbf{1}[d=1]\right] \\
&= 2^n \zeta_n - n \cdot (2^n - 1) \\
\Rightarrow \zeta_n &= n
\end{aligned}
$$

which completes the proof. □

Lemma C.2.

$$\sum_{k=0}^{d-1} (-1)^{d-k} \binom{d}{k} (n-d+k) = -n + \mathbf{1}[d=1]$$

Proof. First note $\sum_{k=0}^{d} (-1)^{d-k} \binom{d}{k} = 0$ and therefore

$$\sum_{k=0}^{d-1} (-1)^{d-k} \binom{d}{k} (n-d) = -(n-d)$$

It was previously shown that $\sum_{k=0}^{d} (-1)^{d-k} \binom{d}{k} k = \mathbf{1}[d=1]$ which implies

$$\sum_{k=0}^{d-1} (-1)^{d-k} \binom{d}{k} k = \begin{cases} 0 & \text{if } d = 1 \\ -d & \text{if } d \neq 1 \end{cases}$$

Adding these together proves the lemma. □

Appendix D: Proofs of Lemma 6.2 and 6.4

We begin with the following lemma.

Lemma D.1. *Let $a_{n,d}$ and $b_{n,d}$ be two non-negative, double-indexed sequences. That is, for each $n \geq 1$ we have two sequences $a_{n,1}, \ldots, a_{n,n}$ and $b_{n,1}, \ldots, b_{n,n}$. Suppose for each n that there exists a non-decreasing sequence $m_n \geq 1$ such that*

$$\lim_{n\to\infty} \sum_{d=m_n}^{n} a_{n,d} b_{n,d} = 1.$$

Then suppose there exists another non-negative doubly-indexed sequence $\tilde{a}_{n,d}$ such that $\frac{\tilde{a}_{n,d}}{a_{n,d}}$ converges uniformly to one for $d \geq m_n$. That is, for all $\epsilon > 0$ there exists $N_\epsilon > 0$ such that for all $n > N_\epsilon$

$$\left| \frac{\tilde{a}_{n,d}}{a_{n,d}} - 1 \right| < \epsilon. \tag{20}$$

for each $d \geq m_n$. Then

$$\lim_{n\to\infty} \sum_{d=m_n}^{n} \tilde{a}_{n,d} b_{n,d} = 1.$$

Proof. Let $\epsilon > 0$ and $n > N_\epsilon$ then equation (20) implies

$$|\tilde{a}_{n,d} - a_{n,d}| < a_{n,d}\epsilon.$$

for all $d \geq m_n$. Then

$$\begin{aligned}
\sum_{d=m_n}^{n} \tilde{a}_{n,d} b_{n,d} &= \sum_{d=m_n}^{n} (\tilde{a}_{n,d} - a_{n,d} + a_{n,d}) b_{n,d} \\
&\leq \sum_{d=m_n}^{n} |\tilde{a}_{n,d} - a_{n,d}| b_{n,d} + \sum_{d=m_n}^{n} a_{n,d} b_{n,d} \\
&< \sum_{d=m_n}^{n} \epsilon a_{n,d} b_{n,d} + \sum_{d=m_n}^{n} a_{n,d} b_{n,d} \\
&= (1+\epsilon) \sum_{d=m_n}^{n} a_{n,d} b_{n,d}.
\end{aligned}$$

where the second inequality is due to equation (20). On the other hand,

$$\begin{aligned}
\sum_{d=m_n}^{n} \tilde{a}_{n,d} b_{n,d} &= \sum_{d=m_n}^{n} (\tilde{a}_{n,d} - a_{n,d} + a_{n,d}) b_{n,d} \\
&\geq - \sum_{d=m_n}^{n} |\tilde{a}_{n,d} - a_{n,d}| b_{n,d} + \sum_{d=m_n}^{n} a_{n,d} b_{n,d}
\end{aligned}$$

$$> - \sum_{d=m_n}^{n} \epsilon a_{n,d} b_{n,d} + \sum_{d=m_n}^{n} a_{n,d} b_{n,d}$$
$$= (1-\epsilon) \sum_{d=m_n}^{n} a_{n,d} b_{n,d}.$$

Therefore, for every $\epsilon > 0$ there exists N_ϵ such that for $n > N_\epsilon$

$$(1-\epsilon) \sum_{d=m_n}^{n} a_{n,d} b_{n,d} < \sum_{d=m_n}^{n} \tilde{a}_{n,d} b_{n,d} < (1+\epsilon) \sum_{d=m_n}^{n} a_{n,d} b_{n,d}$$

and therefore

$$\lim_{n\to\infty} \sum_{d=m_n}^{n} \tilde{a}_{n,d} b_{n,d} = \lim_{n\to\infty} \sum_{d=m_n}^{n} a_{n,d} b_{n,d} = 1. \qquad \square$$

The following theorem will also be employed in the below.

Theorem D.2 (The Stolz-Cesaro Theorem)**.** *If* $(b_n)_{n=1}^{\infty}$ *is a strictly increasing sequence with* $\lim_{n\to\infty} b_n = \infty$*, then for any sequence* $(\tilde{b}_n)_{n=1}^{\infty}$ *the following inequalities hold:*

$$\limsup_{x\to 0} \frac{\tilde{b}_n}{b_n} \leq \limsup_{x\to 0} \frac{\tilde{b}_n - \tilde{b}_{n-1}}{b_n - b_{n-1}}$$
$$\liminf_{x\to 0} \frac{\tilde{b}_n}{b_n} \geq \liminf_{x\to 0} \frac{\tilde{b}_n - \tilde{b}_{n-1}}{b_n - b_{n-1}}$$

In particular, if the sequence $\left(\frac{\tilde{b}_n - \tilde{b}_{n-1}}{b_n - b_{n-1}}\right)_{n=1}^{\infty}$ *has a limit, then*

$$\lim_{n\to\infty} \frac{\tilde{b}_n}{b_n} = \lim_{n\to\infty} \frac{\tilde{b}_n - \tilde{b}_{n-1}}{b_n - b_{n-1}}$$

We now prove Lemma 6.2, which we re-write below as Proposition D.3.

Proposition D.3. *Let* $\mu_n(\rho)$ *denote the expected number of blocks given* n *individuals and* $\rho \in (0,\infty)$ *for the harmonic process. As* $n \to \infty$*,*

$$\mu_n(\rho) \sim \log^2(n)$$

where the ratio tends to a constant equal to $\frac{1}{2\cdot\Psi_1(\rho)}$ *where* Ψ_1 *is the trigamma function. The same holds for the gamma process.*

Proof. In this proof we write $x_n \sim y_n$ if $\lim_{n\to\infty} x_n/y_n = C \in (0,\infty)$ for x_n and y_n two non-negative sequences.

For the gamma process, the recurrence relation of the expected number of blocks is given by:

$$\mu_n = 1 + \sum_{d=1}^{n} \binom{n}{d} q(n-d,d) \cdot \mu_{n-d}$$

$$= 1 + \sum_{d=1}^{n} \binom{n}{d} \frac{(-1)^{d-1}(\Delta^d \zeta)_{n-d}}{\log(1+\frac{n}{\rho})} \cdot \mu_{n-d}$$

which gives the following approximation:

$$\begin{aligned}
(-1)^{d-1}(\Delta^d \zeta)_{n-d} &= \sum_{i=0}^{d-1} \binom{d-1}{i} (-1)^i \log\left(1 + \frac{1}{n-d+i+\rho}\right) \\
&\sim \sum_{i=0}^{d-1} \binom{d-1}{i} (-1)^i \frac{1}{n-d+i+\rho} \\
&= \frac{\Gamma(d)\Gamma(n-d+\rho)}{\Gamma(n+\rho)} = \frac{\Gamma(d)}{(n-d+\rho)^{\uparrow d}}.
\end{aligned}$$

This is the exact expression for the harmonic process. The second line is due to Taylor's remainder theorem. In particular the approximation of $\log(1+x)$ by x has error bounded by $|x|^2$. Therefore, the error for $d \le n$ can be bounded by

$$\sum_{i=1}^{d-1} (n-d+i+\rho)^{-2} < \sum_{x=1}^{n-1} (x+\rho)^{-2} < \sum_{x=1}^{\infty} (x+\rho)^{-2} < \infty.$$

Therefore, the approximation error is finite and so the ratio of the left hand side and right hand side will tend one as n goes to infinity (i.e., $(-1)^{d-1}(\Delta^d \zeta)_{n-d} \sim \frac{\Gamma(d)}{(n-d+\rho)^{\uparrow d}}$ with $C = 1$ for the characteristic index of the gamma process).

So for both processes, the above implies that the probability of d individuals in block 1 is proportional to:

$$\begin{aligned}
\binom{n}{d} (-1)^{d-1}(\Delta^d \zeta)_{n-d} &= \frac{\Gamma(n+1)}{\Gamma(n+\rho)} \cdot \frac{\Gamma(n-d+\rho)}{\Gamma(n-d+1)} \cdot \frac{1}{d} \\
&\sim \left(\frac{n-d}{n}\right)^{\rho-1} \cdot \frac{1}{d}
\end{aligned}$$

where the approximation

$$\frac{\Gamma(x+a)}{\Gamma(x+b)} \sim x^{1-\rho}$$

is used. Again the constant in this case is $C = 1$.

We now investigate the recurrence relation:

$$\mu_n = 1 + \frac{1}{\psi(n+\rho) - \psi(\rho)} \sum_{d=1}^{n} \left(\frac{n-d}{n}\right)^{\rho-1} \cdot \frac{\mu_{n-d}}{d}$$

Substitute $c \log^2(n)$ in for μ_n. and approximating the sum by an integral yields:

$$1 + \frac{c}{\log\left(1+\frac{n}{\rho}\right)} \int_0^1 \frac{(1-x)^{\rho-1}\log(1-x)}{x} \left(2\log(n) + \log(1-x)\right) dx$$

$$\begin{aligned}&\to 1 + 2 \cdot c \cdot \int_0^1 \frac{(1-x)^{\rho-1}\log(1-x)}{x} dx \\ &= 1 - 2 \cdot c \cdot \Psi_1(\rho)\end{aligned}$$

Giving us $c = \frac{1}{2 \cdot \Psi_1(\rho)}$. Then $b_{n,d} = c\left(\log^2(n-d) - \log^2(n)\right)$ satisfies

$$\lim_{n\to\infty} \sum_{d=1}^{n} a_{n,d} b_{n,d} = 1$$

where

$$a_{n,d} = \frac{1}{\log\left(1 + \frac{n}{\rho}\right)} \cdot \left(\frac{n-d}{n}\right)^{\rho-1} \cdot \frac{1}{d}.$$

In order to invoke lemma D.1, we must show that for

$$\tilde{a}_{n,d} = \frac{1}{\psi(n+\rho) - \psi(\rho)} \frac{\Gamma(n+1)}{\Gamma(n+\rho)} \cdot \frac{\Gamma(n-d+\rho)}{\Gamma(n-d+1)} \cdot \frac{1}{d}$$

the uniform convergence statement holds for $d \geq m_n$. Here we set $m_n = \lfloor n^\gamma \rfloor$ for $0 < \gamma < 1$. Then for $\epsilon > 0$ there exists N' such that for $n \geq N'$

$$(1-\epsilon) \leq \frac{(n-d)^{\rho-1}}{\left(\frac{\Gamma(n-d+\rho)}{\Gamma(n-d+1)}\right)} \leq (1+\epsilon)$$

for $m_n \leq d \leq n$. Then

$$\begin{aligned}&\frac{\log\left(1+\frac{n}{\rho}\right)}{\psi(n+\rho)-\psi(\rho)} \frac{1}{n^{\rho-1}} \cdot \frac{\Gamma(n+1)}{\Gamma(n+\rho)}(1-\epsilon) \\ &\quad \leq \frac{a_{n,d}}{\tilde{a}_{n,d}} \leq \frac{\log\left(1+\frac{n}{\rho}\right)}{\psi(n+\rho)-\psi(\rho)} \frac{1}{n^{\rho-1}} \cdot \frac{\Gamma(n+1)}{\Gamma(n+\rho)}(1+\epsilon)\end{aligned}$$

for all $n > N'$ and $m_n \leq d \leq n$. For all $\epsilon > 0$ there exists N such that for $n \geq N$

$$(1-\epsilon) \leq \frac{\log\left(1+\frac{n}{\rho}\right)}{\psi(n+\rho)-\psi(\rho)} \leq (1+\epsilon)$$

and N'' such that for $n \geq N''$,

$$(1-\epsilon) \leq \frac{1}{n^{\rho-1}} \cdot \frac{\Gamma(n+1)}{\Gamma(n+\rho)} \leq (1+\epsilon)$$

Let $n \geq \max(N, N', N'')$ then

$$(1-\epsilon)^3 \leq \frac{a_{n,d}}{\tilde{a}_{n,d}} \leq (1+\epsilon)^3.$$

Therefore, for all $\epsilon > 0$ there exists N such that for $n > N$ and $m_n < d < n$

$$\left| \frac{a_{n,d}}{\tilde{a}_{n,d}} - 1 \right| < \epsilon$$

as desired.

For $m_n = n^\gamma$ for $0 < \gamma < 1$, we have already shown that the sum can be well-approximated by the integral. In particular, this implies

$$\lim_{n\to\infty} \sum_{d=1}^{m_n} a_{n,d} b_{n,d} = 0 \text{ and } \lim_{n\to\infty} \sum_{d=m_n}^{n} a_{n,d} b_{n,d} = 1.$$

That is, the left-tail of the sum converges to zero and so can be ignored. By Lemma D.1 we therefore have

$$\lim_{n\to\infty} \frac{1}{\psi(n+\rho) - \psi(\rho)} \sum_{d=m_n}^{n} \frac{\Gamma(n+1)}{\Gamma(n+\rho)} \cdot \frac{\Gamma(n-d+\rho)}{\Gamma(n-d+1)} \cdot \frac{1}{d} b_{n,d} = 1$$

where $b_{n,d} = c \log^2(n-d) - c \log^2(n)$. Now we address the component of the sum for $1 \le d \le m_n$. First we have

$$\frac{\Gamma(n+1)}{\Gamma(n+\rho)} \cdot \frac{\Gamma(n-d+\rho)}{\Gamma(n-d+1)} \le \frac{\Gamma(n+1)}{\Gamma(n+\rho)} \left(\frac{\Gamma(n-n^\gamma+\rho)}{\Gamma(n-n^\gamma+1)} \vee \frac{\Gamma(n-1+\rho)}{\Gamma(n-1+1)} \right) \to 1$$

for $1 \le d \le n^\gamma$ and

$$\begin{aligned} 0 \le \log^2(n) - \log^2(n-d) &\le \log^2(n) - \log^2(n-n^\gamma) \\ &= -2\log(n)\log(1-n^{\gamma-1}) - \log^2(1-n^{\gamma-1}) \to 0. \end{aligned}$$

Moreover,

$$\frac{1}{\psi(n+\rho) - \psi(\rho)} \sum_{d=1}^{m_n} d^{-1} = \frac{\psi(n^\gamma+\rho) - \psi(\rho)}{\psi(n+\rho) - \psi(\rho)}.$$

For $0 < \gamma < 1$ by above we have each term tends to either a constant or zero. Therefore we have

$$\lim_{n\to\infty} \frac{1}{\psi(n+\rho) - \psi(\rho)} \sum_{d=1}^{m_n} \frac{\Gamma(n+1)}{\Gamma(n+\rho)} \cdot \frac{\Gamma(n-d+\rho)}{\Gamma(n-d+1)} \cdot \frac{1}{d} b_{n,d} = 0.$$

This implies

$$\lim_{n\to\infty} \frac{1}{\psi(n+\rho) - \psi(\rho)} \sum_{d=1}^{n} \frac{\Gamma(n+1)}{\Gamma(n+\rho)} \cdot \frac{\Gamma(n-d+\rho)}{\Gamma(n-d+1)} \cdot \frac{1}{d} b_{n,d} = 1.$$

as desired. Note the approximation error of the ratio between $a_{n,d}$ and $\tilde{a}_{n,d}$ for $d \le m_n$ does not go to one. Instead it is simply the case that both left-tail sums converge to zero and therefore the approximation error is unimportant.

Recall that the recursion relation for all $n \geq 0$ implies

$$\sum_{d=1}^{n} \frac{\Gamma(n+1)}{\Gamma(n+\rho)} \cdot \frac{\Gamma(n-d+\rho)}{\Gamma(n-d+1)} \cdot \frac{1}{d}(\mu_{n-d} - \mu_n) = 1.$$

Therefore both $\{b_{n,d}\}$ and $\{\tilde{b}_{n,d} = \mu_{n-d} - \mu_n\}$ satisfy the relation for large n. This implies

$$\lim_{n\to\infty} \frac{(\mu_n - \mu_{n-d})}{c \log^2(n) - c \log^2(n-d)} = 1$$

for any $d \geq 1$. By the Stolz-Cesaro Theorem, this implies

$$\lim_{n\to\infty} \frac{\mu_n}{c \cdot \log^2(n)} = 1$$

as desired. □

A similar argument can be used to obtain the asymptotic expected fraction of particles in the first block. We omit the technical details (invoking Lemma D.1 and the Stolz-Cesaro theorem) as it follows in the same manner as in the proof above. We now prove Lemma 6.4, which we re-write below as Proposition D.4.

Proposition D.4. *The expected number of particles in the first block satisfies* $E[\#B_1] \sim n/(\rho \log n)$ *for the harmonic and gamma process. Asymptotically, for integer values of* ρ,

$$\frac{\log B_1}{\log n} \xrightarrow{D} U$$

where U *has the uniform distribution on* $(0, 1)$.

Proof. By the same rationale as above, the expected number of particles in the first block is given by

$$\begin{aligned} \sum_{d=1}^{n} d \binom{n}{d} q(n-d, d) &\sim \frac{1}{\log(1+\frac{n}{\rho})} \sum_{d=1}^{n} d \left(\frac{n-d}{n}\right)^{\rho-1} \frac{1}{d} \\ &\to \frac{n}{\log(1+\frac{n}{\rho})} \int (1-x)^{\rho-1}\, dx \\ &\sim \frac{n}{\log(n)} \frac{1}{\rho} \end{aligned}$$

So the fraction of particles in the first block is roughly $(\rho \log(n))^{-1}$. For $z \in (0, 1)$, the asymptotic distribution is given by

$$\begin{aligned} P\left[\frac{\log B_1}{\log n} \leq z\right] &= \frac{1}{\log(n)} \int_{1/n}^{n^{z-1}} x^{-1}(1-x)^{\rho-1} dx \\ &= \frac{1}{\log n} \left[\sum_{j=0}^{\rho-1} c_j x^j + \log(x)\right]_{1/n}^{n^{z-1}} \end{aligned}$$

$$\to \frac{1}{\log n}\left[(z-1)\log(n) - \log(1/n))\right] = z$$

where the second line is true for $\rho \in \mathbb{Z}$. The result holds for general $\rho > 0$ by the squeezing theorem. □

Appendix E: Sketch proof of Remark 6.3

The splitting rule for the pilgrim process is

$$q(r,d) = \frac{1}{\psi(r+d+\rho) - \psi(\rho)} \cdot \left(1 - \frac{d}{n}\right)^{\rho-1} \frac{1}{d}$$

When $n = r+d$ is large, $\psi(n+\rho) - \psi(\rho) \approx \log(n)$. Moreover, the number of blocks grows at a rate of $\log^2(n)$. Let $m = \log^2(n)$. Then if we thin the process by an additional $1/\log(n) = m^{-1/2}$, for large n the sequence of splits is approximately binomial with success probability $m^{-1/2}q(r,d)$. Since $m \cdot m^{-1/2} q(r,d) \to c/d$, we have the number of blocks of size d is approximately Poisson with rate parameter c/d. This implies that the rate parameter for the un-thinned process is

$$\lambda_n \approx \frac{c \cdot m^{1/2}}{d} = \frac{c \cdot \log(n)}{d}.$$

The constant can be derived via the expected number of blocks. In this case, $c = \frac{1}{2\Psi_1(\rho)}$ is a constant dependent on ρ, where Ψ_1 is the trigamma function.

Appendix F: Proof of Lemma 7.1

We now consider the number of blocks and block sizes for the general beta-splitting rules. We now prove Lemma 7.1, which we write as a series of propositions below.

Proposition F.1. *Define $\mu_n(\rho,\beta)$ denote the number of blocks given n individuals for the beta process. For $\beta > 0$ as $n \to \infty$,*

$$\mu_n(\rho,\beta) \sim \log(n)$$

where the ratio tends to a constant $c = \frac{1}{\psi(\rho+\beta)-\psi(\rho))}$. The fraction of edges in the first block, $\#B_1$, is distributed Beta(ρ,β). Therefore, the relative frequencies within each block is given by

$$(P_1, P_2, \ldots) = (B_1, \bar{B}_1 B_2, \bar{B}_1 \bar{B}_2 B_3, \ldots)$$

where B_i are independent beta variables with parameters (ρ,β). The number of blocks of size j is approximately Poisson with asymptotic rate proportional to $1/j$ with independent components for $j \neq j'$.

Proof. The probability of d individuals in the first block is given by

$$\binom{n}{d} q(n-d,d) \sim \frac{1}{Z_n} \frac{\Gamma(n+1)}{\Gamma(d+1)\Gamma(n-d+1)} \frac{\Gamma(n-d+\rho)\Gamma(d+\beta)}{\Gamma(n+\rho+\beta)}$$

where $Z_n = \int_0^1 (1-s^n) s^{\rho-1}(1-s)^{\beta-1} ds = B(\rho,\beta) - B(\rho+n,\beta)$. As $n \to \infty$, for $\rho, \beta > 0$ the normalization constant converges to $B(\rho,\beta)$. Therefore, the expected number of blocks is given by

$$0 = 1 + \frac{1}{Z_n} \sum_{d=1}^{n} \left(\frac{n-d}{n}\right)^{\rho-1} \left(\frac{d}{n}\right)^{\beta-1} n^{-1} \left(\mu_{n-d} - \mu_n\right)$$

Writing $\mu_n \sim c\log(n)$ then as $n \to \infty$ we have the above equation becomes

$$\begin{aligned} 0 &= 1 + \frac{c}{B(\rho,\beta)} \int_0^1 (1-x)^{\rho-1} x^{\beta-1} \log(1-x) dx \\ &= 1 + c \cdot E[\log(1-X)] \\ &= 1 + c \cdot (\psi(\rho) - \psi(\rho+\beta)) \end{aligned}$$

as desired. That $\mu_n \sim c\log(n)$ for the beta-splitting rule with $\beta > 0$ follows from the same argument provided in the prior section for $\beta = 0$.

The probability of d out of n particles in block one is given by

$$\frac{n}{Z_n} \left(\frac{n-d}{n}\right)^{\rho-1} \left(\frac{d}{n}\right)^{\beta-1} n^{-1} \to \frac{n}{B(\rho,\beta)} (1-x)^{\rho-1} x^{\beta-1} dx$$

So the fraction of particles in block one is distributed beta with parameters (ρ, β).

We end by proving the number of blocks of size j is approximately Poisson with asymptotic rate proportional to $1/j$ with independent components for $j \neq j'$. The splitting rule for $\beta > 0$ is given by

$$\begin{aligned} q(r,d) &= \frac{1}{B(\rho,\beta)(1-\rho^{\uparrow n}/(\rho+\beta)^{\uparrow n})} \cdot \left(1 - \frac{d}{n}\right)^{\rho-1} \left(\frac{d}{n}\right)^{\beta-1} \frac{1}{n} \\ &\sim d^{-1} \frac{1}{B(\rho,\beta)} \cdot \left(1-\frac{d}{n}\right)^{\rho-1} \left(\frac{d}{n}\right)^{\beta} \end{aligned}$$

Let $n_i = r_i + d_i$ and $d > 0$ a fixed constant. Then define

$$a_i = \frac{1}{B(\rho,\beta)} \cdot \left(1 - \frac{d}{n_i}\right)^{\rho-1} \left(\frac{d}{n_i}\right)^{\beta}$$

Then

$$\frac{1}{d} \sum a_i$$

denotes the expected number of blocks of size d where the random sequence $(n_1, n_2, \ldots)$ satisfies $\sum n_i = n$. As $n \to \infty$, we want to show $\sum_i a_i < \infty$. By the ratio test

$$(a_i)^{1/i} \to \left(\frac{1}{n_i}\right)^{\beta/i}$$

$$\approx n_i^{-\beta/(c\cdot\log(i))}$$
$$\to e^{-\beta/c} < 1 \text{ for } \beta, c > 0$$

where we have used $n_i \sim c\log(i)$ for $c > 0$. Therefore, the sequence converges almost surely and we have the expected number of blocks of size d is proportional to d^{-1}.

Since the expected number of blocks is $1/(\psi(\beta+\rho)-\psi(\rho))\log(n)$ we see that the expected number of blocks of size d is

$$\frac{1}{\psi(\beta+\rho)-\psi(\rho)}\, d^{-1}$$

In particular, for $\beta = 1$ then we have the expected number of blocks is ρ/d which is equivalent to the one parameter Chinese restaurant process. ☐

The final case is when $\beta \in (-1, 0)$. Then the number of blocks grows polynomially in n.

Proposition F.2. *Define $\mu_n(\rho, \beta)$ denote the number of blocks given n individuals for the beta process. For $\beta \in (-1, 0)$, as $n \to \infty$,*

$$\mu_n(\rho, \beta) \sim n^{-\beta}$$

where the ratio tends to a constant $c = -\Gamma(\rho+\beta+1)/(\Gamma(\rho)\beta^2)$. The fraction of edges in the first block is asymptotically proportional to n^{β}. Asympotically,

$$\text{pr}[B_1 = d] = \frac{-\beta\cdot\Gamma(d+\beta)}{\Gamma(d+1)\cdot\Gamma(1+\beta)} = \frac{-\beta}{\Gamma(1+\beta)} d^{\beta-1}$$

for large d.

Proof. For $\beta \in (-1, 0)$, the characteristic index is given by

$$\begin{aligned}
\zeta_n &= \int_0^1 (1-s^n)s^{\rho-1}(1-s)^{\beta-1}ds \\
&= \int_0^1 \sum_{j=0}^{n-1} s^{j+\rho-1}(1-s)^{\beta}ds \\
&= \sum_{j=0}^{n-1} \frac{\Gamma(j+\rho)\Gamma(\beta+1)}{\Gamma(j+1+\rho+\beta)} \\
&\sim \frac{\Gamma(1+\beta)}{-\beta} n^{-\beta}
\end{aligned}$$

Plugging this into the recursive formula, while assuming $\mu_n \sim cn^{-\beta}$ yields

$$0 = 1 + c\frac{-\beta\, n^{\beta}}{\Gamma(1+\beta)} \sum_{d=1}^{n} \left(\frac{n-d}{n}\right)^{\rho-1} \left(\frac{d}{n}\right)^{\beta-1} n^{-1}\left((n-d)^{-\beta} - n^{-\beta}\right)$$

$$
\begin{aligned}
&= 1 - \frac{c\,\beta}{\Gamma(1+\beta)} \sum_{d=1}^{n} \left(\frac{n-d}{n}\right)^{\rho-1} \left(\frac{d}{n}\right)^{\beta-1} n^{-1} \left((1-\frac{d}{n})^{-\beta} - 1 \right) \\
&= 1 - \frac{c\,\beta^2}{\Gamma(1+\beta)} \sum_{d=1}^{n} \left(\frac{n-d}{n}\right)^{\rho-1} \left(\frac{d}{n}\right)^{\beta} n^{-1} \\
&\to 1 - \frac{c\,\beta^2}{\Gamma(1+\beta)} \int (1-x)^{\rho-1} (x)^{\beta}\, dx \\
&= 1 + c \frac{\beta^2\, \Gamma(\rho)}{\Gamma(\rho+\beta+1)}
\end{aligned}
$$

where for large n the approximation $(1-\frac{d}{n})^{-\beta} = 1 + \beta \frac{d}{n}$ is used. The result is immediate. The fraction of edges in the first block is given by

$$
\begin{aligned}
\frac{-\beta\, n^{\beta}}{\Gamma(1+\beta)} \sum_{i=1}^{n} \left(\frac{d}{n}\right)^{\beta} \left(\frac{n-d}{n}\right)^{\rho-1} n^{-1} &\sim \frac{-\beta\, n^{\beta}}{\Gamma(1+\beta)} \int (1-x)^{\rho-1} (x)^{\beta}\, dx \\
&= \frac{-\beta\, \Gamma(\rho)}{\Gamma(\rho+\beta+1)} n^{\beta}
\end{aligned}
$$

Finally, the asymptotic distribution comes from

$$
\begin{aligned}
\text{pr}(B_1 = d) &= \zeta_n^{-1} \frac{\Gamma(d+\beta)\Gamma(n-d+\rho)}{\Gamma(n+\rho+\beta)} \cdot \binom{n}{d} \\
&\sim \frac{-\beta \cdot n^{\beta}}{\Gamma(1+\beta)} \cdot \frac{\Gamma(n+1)}{\Gamma(n+\rho+\beta)} \frac{\Gamma(d+\beta)}{\Gamma(d+1)} \frac{\Gamma(n-d+\rho)}{\Gamma(n-d+1)} \\
&\sim \frac{-\beta}{\Gamma(1+\beta)} \frac{\Gamma(d+\beta)}{\Gamma(d+1)} \left(1 - \frac{d}{n}\right)^{\rho-1} \\
&\to \frac{-\beta}{\Gamma(1+\beta)} \frac{\Gamma(d+\beta)}{\Gamma(d+1)}
\end{aligned}
$$

as $n \to \infty$. □

Appendix G: Bayesian survival analysis

Proposition G.1. *Every exchangeable, Markov survival process can be generated via the Lévy process construction.*

Proof. Every exchangeable, Markov survival process is determined by the splitting rule $q(\cdot,\cdot)$ and parameter ν. Without loss of generality we consider the case where $\nu = 1$. For a particular splitting rule, we show how to construct a Lévy process Z such that the unconditional risk set evolves as an exchangeable, Markov survival process with splitting rule q.

In Section 8.2, we stated the Lévy-Khintchine representation for non-negative Lévy processes. The converse is also true. Namely any pair (γ, w) such that $\gamma \geq 0$ and w such that

$$
\int_0^{\infty} (1 - e^{-y}) w(dy) < \infty
$$

determine a non-negative Lévy process Z such that $X = Z(1)$ has cumulant function determined by (15). Therefore for a particular exchangeable, Markov survival process, we must construct a pair (γ, w) that satisfies the above requirements. This will ensure the pair (γ, w) determine a non-negative Lévy process.

We already know that the sequence of unconditional distributions for the random variables $T[n]$ generated via the Lévy process construction are exchangeable, Markov survival processes with characteristic index $\zeta_n = -K(n)$. So then for any Lévy process Z with cumulant generating function $K(\cdot)$ by the Lévy-Khintchine characterization we must have

$$\cdot q(n,1) = \frac{(\Delta\zeta)_n}{\zeta_{n+1}} = \frac{\zeta_{n+1} - \zeta_n}{\zeta_{n+1}} = \frac{K(n+1) - K(n)}{K(n+1)}.$$

The first equality is due to Proposition 4.5. Then the Lévy-Khintchine representation implies

$$K(n) = -\left(\gamma n + \int_0^\infty (1 - e^{-ny}) w(dy)\right)$$
$$\Rightarrow -(K(n+1) - K(n)) = \left(\gamma + \int_0^\infty (e^{-(n+1)y} - e^{-ny}) w(dy)\right)$$

The second equality holds as the integrals are both non-negative and so

$$\int_0^\infty (1 - e^{-(n+1)y}) w(dy) - \int_0^\infty (1 - e^{-ny}) w(dy) = \int_0^\infty (e^{-ny} - e^{-(n+1)y}) w(dy).$$

Recall that consistent splitting rules admit an integral representation (i.e., Proposition 4.3). The integral representation of the singleton split is given by

$$q(n,1) = \frac{1}{Z_{n+1}} \left(\int_0^1 x^n (1-x) \varpi(dx) + c\right).$$

Given the erosion measure $c \geq 0$ and dislocation measure $\varpi(\cdot)$, we set $\gamma = c$, $Z_n = -K(n)$, and define $w(dy) = e^{-z} \varpi(-\log(x) \in dy)$. Then re-writing the integral representation in these terms we have

$$q(n,1) = \frac{1}{Z_{n+1}} \left(\gamma + \int_0^\infty e^{-ny} (1 - e^{-y}) w(dy)\right).$$

We thus have selected a pair (γ, w) such that $\gamma \geq 0$. It rests to check the integrability condition for the measure w to ensure the associated process Z is a Lévy measure. First,

$$\int_0^\infty (1 - e^{-y}) w(dy) = \int_0^1 (1-x) \varpi(dx) < \infty$$

Equality is due to the change of variables $y \to \log(x)$. The integral is finite due to the integrability condition for the dislocation measure in Proposition 4.3. Finally,

$$\int_0^\infty (1 - e^{-y}) w(dy) < \infty \Rightarrow \int_0^\infty (1 - e^{-yt}) w(dy) < \infty$$

for all $t > 0$. So we have found the pair (γ, w) such that the associated process Z is (1) a Lévy process and (2) the unconditional risk sets evolve with the singleton splitting rules $q(n, 1)$. But by Proposition 4.2 we know that the splitting rule $q(\cdot, \cdot)$ is completely determined by the singleton splits. Therefore, the unconditional risk sets evolve with the correctly specified splitting rule q and therefore we have generated an exchangeable, Markov survival process with associated splitting rule q as desired. □

Appendix H: Stochastic censoring patterns

In section 2.1, we considered the setting where censoring times are arbitrary positive numbers and are known for each particle. Here we define a general notion of stochastic censoring under which the likelihood construction is preserved.

Definition H.1 (Relative exchangeability and censoring). Let $R[n] = \{R_{[n]}(t)\}_{t \geq 0}$ be the risk set trajectory restricted to the set of particles, $[n]$. Then the censoring mechanism is *exchangeability preserving* if for each t, the particles still at risk, $R_n(t)$, are exchangeable.

In other words: the additional knowledge of the right-censoring times up until time t should not affect the exchangeability of particles still at risk at time t. We require the particles that survive are homogeneous and thus only distinguished by their labelling. Simple type I censoring obviously preserves exchangeability, as does independent censoring.

Equation (9) in Section 5 is still correct if we assume that the parameters underlying the exchangeable, Markov survival process remain unchanged after each censoring time (e.g., the process is still a harmonic process with the same parameter values). This *statistical* assumption is valid under noninformative right-censoring.

Appendix I: Covariate effects

We briefly introduce a limiting case of the harmonic process. Specifically, the inverse linear characteristic index, ξ, is defined as

$$\begin{aligned} \xi_n &= \lim_{\rho \to 0} \rho \cdot \zeta(\rho \cdot n) = \lim_{\rho \to 0} \rho \cdot [\psi(\rho(n+1)) - \psi(\rho)] \\ &= -\frac{1}{n+1} + 1 = \frac{n}{n+1} \end{aligned}$$

This corresponds to the beta-splitting rule with $\rho = \beta = 1$. The inverse linear process arises in connection with the proportional conditional hazards described below.

I.1. Proportional conditional hazards

Now consider the proportional conditional hazards model as described by Kalbfleisch (1978), Hjort (1990) and Clayton (1991). It is sufficient to consider only the stationary version because the non-stationary version involves a relatively straightforward monotone temporal transformation.

Let Λ be a stationary, completely independent random measure on $\mathbb{R}$ with characteristic exponent $\zeta(t)$. In the proportional conditional hazards model, the cumulative hazard for individual i is $w_i\Lambda((0,t])$ for some $w_i > 0$, typically $w_i = e^{x_i\beta}$ depending on covariate x_i. Thus, the ratio $w_i/w_j = e^{(x_i-x_j)\beta}$ of conditional hazards for particles i and j is non-random and constant over time; the marginal distributions are exponential with rates $\zeta(w_i), \zeta(w_j)$, so the hazard ratio $\zeta(w_i)/\zeta(w_j)$ is also constant over time. If Λ is a nonstationary measure then the marginal hazard rate for particle i is $\zeta(w_i)\nu(ds)$. Therefore, the marginal distributions satisfy the proportional hazards assumption independent of assuming the measure, Λ, is stationary. However, the marginal and conditional hazard ratios need not be equal. For ρ sufficiently large, the gamma and harmonic processes satisfy approximate equality, $w_i/w_j \approx \zeta(w_i)/\zeta(w_j)$.

Survival times are conditionally independent given Λ, and the conditional survival density for particle i is

$$e^{-w_i H(t)}\big(1 - \exp(-w_i\Lambda(dt))\big).$$

Consequently, the conditional joint density is

$$\exp\bigg(-\int_0^\infty R^\sharp(t)\,d\Lambda\bigg)\prod_{r=1}^{k}\prod_{i\in D_r}\big(1 - e^{-w_i\Lambda(dt_r)}\big)$$

where $R(t)$ is the risk set as previously defined, $D_r = R(t_r^-)\setminus R(t_r)$ is the set of individuals failing at time t_r, and

$$R^\sharp(t) = w(R(t)) = \sum_{i\in R(t)} w_i$$

is the sum of the risk-set weights. The argument used to obtain the joint marginal density (2.9) is essentialy unchanged. The only difference occurs in the definition of the intensities associated with a failure time at which $R \equiv R(t)$ and D are disjoint subsets

$$\begin{aligned}\lambda(R,D) &= E\bigg(e^{-R^\sharp\Lambda(dt)}\prod_{i\in D}\big(1 - e^{-w_i\Lambda(dt)}\big)\bigg)\\ &= dt\sum_{d\subset D}(-1)^{\#d-1}\zeta^\sharp(R\cup d),\end{aligned}$$

where $\zeta^\sharp(R) = \zeta(R^\sharp)$. Note that λ is a function of two disjoint subsets, whose value depends only on the weights assigned by w to R and the various subsets of D. With this modification, the joint marginal density (2.9) applies also to the inhomogeneous case:

$$f_n(R) = \exp\Big(-\int_0^\infty \zeta^\sharp(R(s))\,ds\Big) \times \prod_{j=1}^k \lambda(R_j, D_j).$$

The Bayes estimate of the survival distribution depends on the value w_{n+1} attached to the new particle:

$$\mathrm{pr}(T_{n+1} > t \mid R[n]) = \exp\Big(-\int_0^t (\Delta\zeta^\sharp)(R(s))\,ds\Big) \times \prod_{j:t_j \le t} \frac{\lambda(R(t_j) \cup \{n+1\}, D_j)}{\lambda(R(t_j), D_j)}$$

where $(\Delta\zeta^\sharp)(R) = \zeta^\sharp(R \cup \{n+1\}) - \zeta^\sharp(R)$ is the increment associated with the new particle.

I.2. Numerical example

The leukemia data of Gehan (1965) reproduced in Section 3.5 in the paper and Table 2 in this supplementary material is used to illustrate parameter estimation in the proportional conditional hazards model. Each patient is assigned to either the control or 6-MP treatment group specificed below by an indicator, Z. Table 2 is the survival and censoring times (in weeks) associated with the control group.

TABLE 2
Times of remission in weeks of leukemia patients

Control ($Z = 0$)	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

Consider the three-parameter Markov survival processes with characteristic index

$$\zeta(t) = \kappa\rho\,(\Phi(\rho(\gamma t + 1)) - \Phi(\rho))$$

where Φ is given. The gamma, harmonic, and inverse linear processes are all of this form. As ρ tends to zero, the harmonic process tends to the inverse linear process with $\rho = 1$. At the other extreme, as ρ tends to infinity, the harmonic process tends to the gamma process with $\rho = 1$. The parameter ρ is interpreted as a measure of proximity of the harmonic process to these limiting cases.

The weight for individual i is taken to be $w_i = \exp(\beta_0 + \beta_1 Z_i)$. The parameter γ is equal to $\exp(\beta_0)$ implying β_0 and ρ are not separately identifiable under the gamma process. Therefore set β_0 to zero and find maximum likelihood estimates of the remaining parameters.

For the harmonic process, the log-likelihood is maximized at the inverse linear boundary. Table 3 summarizes the parameter estimates for (β, κ, γ). The estimate of the treatment parameter, β_1, for the inverse linear process is close to the estimate under the gamma process. Supplementary figure 3 shows the profile

Table 3
Maximum likelihood estimates for two processes

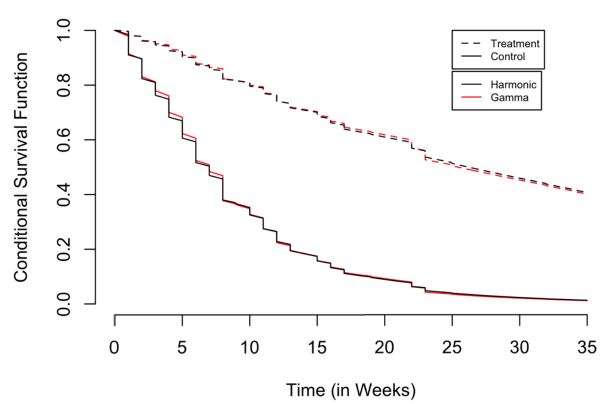
	Inverse linear process		Gamma process	
Parameter	Est.	Std. Error	Est.	Std. Error
κ	1.76	0.57	0.20	0.11
γ	0.07	0.03	0.13	0.05
β_0	0.00	-	0.00	-
β_1	-1.61	0.41	-1.63	0.41

likelihood for β_1 under the inverse linear process and gamma process. It suggests a 95% confidence interval of approximately $[-2.4, -0.9]$ in both instances.

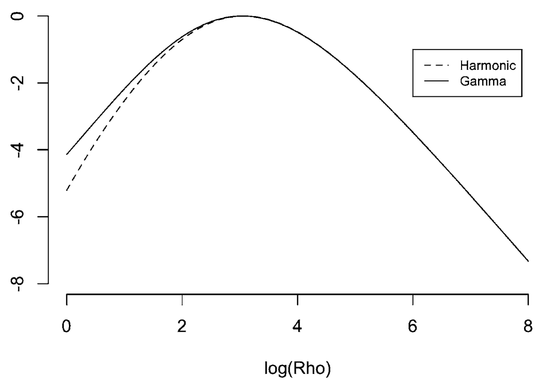
The maximum partial likelihood coefficient is -1.63 using the exact method for breaking ties with a standard error of 0.43. This is comparable to estimates under both processes, with the estimate under the gamma process closer to the partial likelihood estimate.

The empirical Bayes estimate of the conditional survival function for each group is shown in Figure 1. Each curve has discontinuities at observed survival times. At the maximum observed survival time, the estimated conditional survival function is 4.8% (4.3%) and 53.7% (52.7%) for the control and treatment groups respectively under the inverse linear (gamma) process. After the maximum observed time, the conditional distribution for the treatment group is exponential with rate 2.55×10^{-2} and 2.59×10^{-2} for the inverse linear and gamma process respectively, corresponding to an additional expected survival times of approximately 39.14 and 38.63. The expected survival time for the control group is 8.28 and 8.17 under the inverse linear and gamma processes respectively.

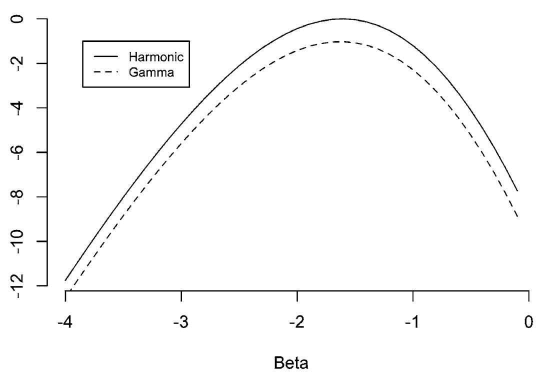
Appendix J: Supplementary figures



Supplementary Figure 1. *Proportional hazards conditional survival distributions*



Supplementary Figure 2. *Profile Log-Likelihood of* $\log(\rho)$



Supplementary Figure 3. *Profile log-likelihood of* β_1

References

Aldous, D. (1985) *Exchangeability and related topics.* In *Ecole D'été de Probáilités de Saint-Flour, 1983. Lecture Notes in Math* **1117**, 1–198. MR0883646

Aldous, D. (1996) *Probability distributions on cladograms.* In *Random Discrete Structures* **76**, 1–18. MR1395604

Anderson, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical models based on counting processes.* New York: Springer. MR1198884

Clayton, D. G. (1991) A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* **47**, 467–485.

Cornfield, J., and Detre, K. (1977) *Bayesian life table analysis. J. Roy. Statist. Soc. B* **39**, 264–296. MR0501462

Doksum, K. A. (1974) *Tailfree and neutral random probabilities and their posterior distributions. Annals of Probability* **2**, 183–201. MR0373081

Cox, D. R. and Oakes, D. (1978) *Analysis of Survival Data.* London, Chapman & Hall. MR0751780

Ford, D. J. (2005) Probabilities on cladograms: Introduction to the alpha model. Preprint. arXiv:math.PR/0511246 MR2708802

Ferguson, T. S. (1973) A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**, 209–230. MR0350949

Gehan, E. A. (1975) A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* **52**, 203–233. MR0207130

Gnedin, A. and Pitman, J. (2005) *Regenerative composition structures. Annals of Probability* **33**, 445–479. MR2122798

Ghosal, L. and van der Vaart, A. (2017) Fundamentals of Bayesian nonparametric inference. Cambridge University Press. MR3587782

Hjort, N. L. (1990) Nonparametric Bayes estimators based on beta processes in models for life history data. *Annals of Statistics* **18**, 1259–1294. MR1062708

James, Lancelot F. (2006) Poisson calculus for spatial neutral to the right processes. Preprint. arXiv:math.PR/0604266

James, Lancelot F. (2006) Spatial Neutral to the Right Species Sampling Mixture Model *The Annals of Statistics* **34**, 416–440 MR2275248

James, Lancelot F. (2007) Neutral-to-the-right species sampling mixture models. In: *Advances in Statistical Modeling and Inference*, 425–439. World Sci. Publ., Hackensack, N.J. MR2416127

Kalbfleisch, J. D. (1978) *Nonparametric Bayesian analysis of survival time data. J. Roy. Statist. Soc. B* **40**, 214–221. MR0517442

Kaplan, E. L. and Meier, P. (1958) *Nonparametric estimation from incomplete observations. J. Amer. Statist. Assn.* **53**, 457—481. MR0093867

Kingman, J. F. C (1993) *Poisson Processes.* Oxford Scientific Publications. MR1207584

Lijoi, A., Prunster, I., and Walker, S. (2008) *Posterior analysis for some classes of nonparametric models. J. Nonparametr. Statist.* **20**, 447–457. MR2424252

McCullagh, P., Pitman, J., and Winkel, M. (2008) *Gibbs fragmentation trees. Bernoulli* **14**, 988–1002. MR2543583

Müller, P., and Quintana, F. (2004) *Nonparametric Bayesian data analysis. Statistical Science* **19**, 95–110. MR2082149

Pitman, J. (2006) *Combinatorial Stochastic Processes.* Lecture Notes Math. Springer, New York. MR2245368

Walker, S., Damien, P., Laud, P., and Smith, A. (1999) *Bayesian nonparametric inference for random functions and related functions. J. Roy. Statist. Soc. B* **61**, 485–527. MR1707858

Walker, S. and Muliere, P. (1997) *Beta-Stacy processes and a generalization of the Polya-urn scheme. Ann. Statist.* **25**, 1762–1780. MR1463574