Finite delayed branching processes

Andrew Hart* Servet Martínez†

Center for Mathematical Modeling, IRL 2807 CNRS-UCHILE, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile.

February 24, 2021

Abstract

We describe and study the finite-time delayed multi-type branching process, a delayed multi-type branching process in which individuals are active (can reproduce offspring) during a finite time interval of random length bounded by D. We show that the criterion for extinction is similar to that for the non-delayed case but is based on the sum of the mean matrices rather than a single mean matrix. We shall impose the condition that the mean matrices at each delay offset share the right and left Perron-Frobenius eigenvectors. In this case we are able to give explicit analytic expressions for various quantities derived from the limit of the geometrically weighted mean evolution of the process. Finally, we discuss the relationship of this process to Fibonacci numbers. In particular, when the mean of the offspring distribution is 1 at every delay offset in the period of time an individual is actively reproducing, the mean evolution of the process is described by a D-Fibonacci sequence.

Keywords: delayed branching process, Multi-type branching process, Perron-Frobenius theory, Fibonacci sequence, Renewal theory.
2010 MSC: 60J80; 92D30.

1 Introduction

The aim of this work is to present a class of multi-type discrete-time branching processes (we shall write $\mathfrak{bp}$ for branching process in future) in which individuals are able to produce offspring during a bounded period of time immediately following their births. Individuals that are able to produce offspring are considered to be active while the interval of time during which an individual can reproduce will be called the individual's active period. Thus, an individual born at time s has the chance to reproduce a set of offspring at every time point in the active period $\{s+1, s+2, \ldots, s+D\}$, where D is the maximum number of time points at which an individual is able to reproduce and the number of individuals produced at time $s+d$ follows a law that depends on the time offset $d \leq D$. Also a latency period is allowed. In addition, each individual has a random lifetime that is independent of the active period, and the type and number of offspring reproduced. We shall call processes with these characteristics delayed $\mathfrak{bp}$'s. They appear naturally in the context of epidemics and the modelling of disease spread, for instance, see [4]. A prominent example of this is the 2019 SARS-CoV2 pandemic. Agent-based simulation studies of the spread of this virus (such as that reported in [9]) inspired the model presented here.

Common to all infectious diseases, individuals who contract the disease only exhibit contagiousness within a certain window of time of some length, say, D. Firstly, this window corresponds to the active period (or delay) of the model we shall study. The kind of impact this active period phenomenon can have on the evolution of an epidemic was brought into sharp relief by the SARS-CoV2 pandemic. The number of days

*E-mail: ahart@dim.uchile.cl
†E-mail: smartine@dim.uchile.cl

that an infected person is contagious while not exhibiting symptoms is important for understanding the spread of a disease and for tailoring strategies of containment and/or mitigation to bring it under control. For many diseases, infection starts with a non-infectious incubation period during which the bacteria or virus multiplies until it masses enough for the infected individual to be contagious to others. The delayed $\mathfrak{bp}$ model we develop here allows the offspring law to vary throughout the active period so that the offspring law depends on the time offset into the active period. This allows the model to explicitly vary the level of contagiousness to distinguish between symptomatic and asymptomatic states, take into account effects such as an initial incubation/latent phase or incorporate a final phase in which the degree of infectiousness falls as an individual recovers from the disease.

Secondly, individuals may continue suffering the effects of the disease long after they have ceased to be infectious and the delayed $\mathfrak{bp}$ captures this persistence by assigning each individual in the population a random lifetime.

Within a country or city, there are inhomogeneities in the geographic and socioeconomic distribution of the population as well as in the daily patterns of movement that govern the frequency and location of contacts between individuals. Commonly employed measures of mitigation, such as quarantines and sanitary cordons, are based on the principle of reducing the frequency of contact between members of the population by interrupting such patterns. The primary means of doing this is by restricting mobility between geographic regions. In order to capture such spatial disparity, we consider multi-type processes where the types represent regions and the numbers of offspring of all types produced by individuals of the distinct types can have different distributions, that is, offspring distributions are allowed to depend on both the type of the parent and the type of the offspring. In terms of modelling an epidemic, this means that the spread of a contagion between regions is allowed to be heterogeneous.

This paper is organized as follows. Section 2 gives a precis of the theory of discrete-time multi-type $\mathfrak{bp}$'s and establishes basic notation that will be needed later on. The delayed variant of the multi-type $\mathfrak{bp}$ is introduced in Section 3. Delayed multi-type $\mathfrak{bp}$'s incorporate two new elements: the individuals can produce new offspring during a bounded period of time according to a possibly time-dependent law and each individual is alive for a random period of time. We consider two closely related processes evolving in time: the offspring process which gives the number of offspring produced at each time (which can be identified with a process in which individuals are alive for exactly one unit of time) and the population size or alive process which tracks the number of individuals that are alive.

There is an extensive literature on processes with delay. Delayed multi-type $\mathfrak{bp}$'s are described an studied in full generality in [7], [8], [15], [5] and [6]. This body of work encompasses more general frameworks than that used here and mostly deal with the continuous-time setting. In these studies, individuals have random continuous lifetimes (as opposed to discrete lifetimes) and there is a period of time during which individuals are active and produce offspring (but this period can be unbounded whereas we only consider a bounded period). One of the more general cases is studied in [5], where the authors also considered the multi-type case with countably infinite types. One of the main problems in this setting is the need to handle products of mean matrices and the authors imposed a series of conditions in order to describe the limiting behavior. The results in [8] pertain to the super-critical 1-type delayed $\mathfrak{bp}$. There the multi-type delayed $\mathfrak{bp}$ is also studied by using the process for each individual type.

In the multi-type setting, we study the asymptotic behavior of the mean population size and there, the main problem is to control a matrix product. This was done in [5] under very general conditions on supercritical processes. Our framework is more restrictive (we assume that all the matrices share the same Perron-Frobenius (P-F) eigenvectors), but we are able to describe the limit distribution similarly to the non-delayed multi-type , in particular, giving the limit of the mean evolution of types. We also describe the combinatorics of the mean evolution in some detail and interpret the quantities that are involved in the limit behavior.

Proposition 1 of Section 4 shows that the criterion for extinction is the same as in the non-delayed multi-type $\mathfrak{bp}$. Afterwards, Section 5 is concerned with deriving an evolution equation for the mean number of individuals that are alive as a function of time which is governed by the set of mean matrices at the delay times. We are interested in the case where the family of mean matrices share the right and left P-F eigenvectors, which generalizes the case where the mean matrices only differ by scalar factors. Our main result is presented in Proposition 4, where we show that when the mean matrices at different delays share P-F eigenvectors, then the mean population size (taking into account the lifetimes of individuals) weighted by an exponential of time and the Malthusian parameter (which is the geometric rate of growth) converges to a multiple of the left P-F eigenvector, a result that is similar to what happens in a non-delayed multi-type $\mathfrak{bp}$. We notice that an extra condition is needed when the process is not supercritical: the critical case requires the expected lifetime to be finite while in the subcritical case, the exponential moment of the lifetime, with parameter equal to the negative of the Malthusian parameter, must be finite.

In Section 6, we examine families of matrices that share P-F eigenvectors. Lemma 6 shows that when the family of mean matrices is commutative, then the matrices necessarily have the same P-F eigenvectors. This section also examines the limiting combinatorics of sequences of symbols in which the length of runs is restricted. Such sequences are required in the proof of Proposition 4 which is presented in Section 7.

Finally, Section 8 compares our results to the Fibonacci-type behavior previously studied in branching processes. In [12], the author studied a class of $\mathfrak{bp}$'s where individuals live forever and produce offspring ad infinitum starting two time units after birth. When the mean number of offspring is 1, the rate of growth is given by the golden ratio, the 2-Fibonacci constant. This model was also studied in [2] for estimating the geometric rate of growth. We compare this model with ours, finding that they coincide when the mean number of offspring produced by each individual is 1.

2 Multi-type branching processes

To set the scene and establish some basic notation that will be used throughout the following, we begin by giving a brief overview of multi-type $\mathfrak{bp}$'s. In a multi-type $\mathfrak{bp}$, each individual takes on a type i from a finite set of types I and the individuals produce offspring independently of all other individuals. Type i individuals produce $\xi^{i,j}$ offspring of type j according to the law $p^{i,j}$ on the non-negative integers. Each individual is born at some time s and produces all of its offspring at time $s+1$. Individuals of type i born at the same time s are enumerated in a way that is independent of individuals born to other types or at other times, so that any individual may be uniquely identified by a triple of the form (s,l,i). The random variable $\xi^{i,j}_{s,l}$ with distribution $p^{i,j}$ indicates the number of type j offspring produced by the l-th individual of type i at time s. All of these random variables are independent. To exclude uninteresting behaviour, we assume that the process is non-singular. Singularity means that each individual has exactly one offspring during its lifetime.

A multi-type $\mathfrak{bp}$ is a process $Z(s) = (Z_j(s) : j \in I)$ for $s \in \mathbb{N}_0\{0,1,2,\ldots\}$, where $Z_j(s)$ is the number of individuals of type j born at time s, which satisfies

$$Z_j(s+1) = \sum_{i\in I} \sum_{l=1}^{Z_i(s)} \xi^{i,j}_{s,l}. \tag{1}$$

The initial configuration $Z(0)$ of individuals present at time 0 is assumed to be non-trivial, that is, $\mathbb{P}(Z(0) = 0) = 0$. Since we will impose irreducibility conditions on the dynamics governing the propagation of types we will assume the simplest initial configuration which comprises a single individual of a fixed type i_0. Thus, $\mathbb{P}(Z(0) = e_{i_0}) = 1$ where e_{i_0} is the vector having 1 as its i_0-th coordinate and vanishing everywhere else in I.

The mean number of offspring of type j generated by an individual of type i is $M(i,j) = \mathbb{E}\left(\xi^{i,j}\right) = \sum_{n\geq 0} n\, p^{i,j}(n)$, $i,j \in I$. Then, the matrix of means is defined to be $M = (M(i,j) : i,j \in I)$. We shall assume that M is irreducible, which suffices to guarantee that the process cannot be decomposed into multiple separate processes by type. Equivalently, irreducibility means that an individual of any type may have a descendant of any type after sufficiently many generations. The Perron-Frobenius theorem then asserts that M has a simple eigenvalue $\rho > 0$ which is the spectral radius of M and has (up to a constant multiple) unique, strictly positive left and right eigenvectors $\nu = (\nu(i) : i \in I)$ and $h = (h(i) : i \in I)$, that is,

$$\nu' M = \rho \nu' \text{ and } Mh = \rho h.$$

See [16, Chapter 1]. Here the vectors are column vectors and $'$ signifies the transpose so that ν' is a row vector. The vectors ν and h are normalized such that $\nu' h = \sum_{i\in I} \nu(i)h(i) = 1$ and $\nu\mathbf{1} = \sum_{i\in I} \nu(i) = 1$. Since ρ is the spectral radius of M, it satisfies $\rho = \lim\limits_{t\to\infty} \|M^t\|^{\frac{1}{t}}$, for any norm $\|\cdot\|$. The matrix $\big(\rho^{-1} M(i,j)\, h(j)/h(i) : i,j \in I\big)$ is stochastic with stationary distribution $(\nu(j)h(j) : j \in I)$. Hence, $\nu(j)h(j) = \lim\limits_{t\to\infty} \rho^{-t} M^t(i,j)h(j)/h(i)$ and convergence to the stationary distribution is geometrically fast. Thus for all norms $\|\cdot\|$,

$$\exists C < \infty, \delta \in (0,1) \text{ such that } \forall t \geq 0 : \quad \|\rho^{-t} M^t - h\nu'\| \leq C\delta^t. \tag{2}$$

This yields the componentwise limit

$$\lim_{t\to\infty} \rho^{-t} M^t = h\nu'. \tag{3}$$

In [4], the author proved the following result which provides a method for estimating ρ in a supercritical multi-type $\mathfrak{bp}$: $\rho = \lim\limits_{t\to\infty} \big(\sum_{s=1}^{t+1} \sum_{j\in I} Z_j(s)\big)/\big(\sum_{s=1}^{t} \sum_{j\in I} Z_j(s)\big)$ almost surely on the non-extinction set. This extends a result proved in [11] for the 1-type $\mathfrak{bp}$.

Various key properties of multi-type $\mathfrak{bp}$'s can be found in Chapter V in [3], in particular, the characterization of extinction. Let $q_j = \mathbb{P}\left(\lim\limits_{s\to\infty} Z_j(s) = 0\right)$ be the probability that the population of type j becomes extinct. In [10] (also see Theorem 2 in Section V.3 of [3]), it was shown that

$$\begin{aligned} \rho \leq 1 \qquad &\Rightarrow\ q_j = 1 \text{ for all } j \in I, \\ \rho > 1 \qquad &\Rightarrow\ q_j < 1 \text{ and } \mathbb{P}\left(\lim_{s\to\infty} Z_j(s) = \infty\right) = 1 - q_j > 0 \text{ for all } j \in I. \end{aligned} \tag{4}$$

In other words, if $\rho \leq 1$ the process almost surely becomes extinct while if $\rho > 1$, then each of the type sub-populations has a positive probability of exploding.

Assume $\sum_{i,j\in I} \sum_{n\geq 1} n \log n\, p^{i,j}(n) < \infty$. When $\rho < 1$ the processes $(Z_j(s) : j \in I)$ are absorbed geometrically fast (see [13] and [3, Theorems 1 and 2 in Section V.4]). Theorem 1 in Section V.6 of [3] deals with the supercritical case ($\rho > 1$) where $\lim\limits_{s\to\infty} \rho^{-s} Z(s) = \nu W$ holds $\mathbb{P}$-a.s. for some non-negative random variable W satisfying $\mathbb{P}(W > 0) > 0$.

We shall use $\mathbb{E}_{i_0}$ to denote the expected value of a multi-type $\mathfrak{bp}$ which starts with a single individual of type i_0, but if there is no confusion we simply denote it by $\mathbb{E}$. The expected value of the number of individuals at time s is $\mathbb{E}_{i_0}(Z(s)) = (\mathbb{E}_{i_0}(Z_j(s)) : j \in I)$. Let $F(s)$ be the σ-field of events defining individuals born up to time s, that is

$$F(s) = \sigma(\xi^{i,j}_{t,l_i} : i,j \in I, 1 \leq l_i \leq Z_i(t), 0 \leq t \leq s).$$

The random variables $\xi^{i,j}_{s,l}$ are independent of $F(s-1)$. Then, by conditioning relation (1) with respect to $F(s-1)$ and taking expectations, we obtain $\mathbb{E}(Z_j(s)) = \sum_{i\in I} \mathbb{E}(Z_i(s-1))M(i,j)$ for $s \geq 1$. Written in matrix form, this is $\mathbb{E}(Z(s))' = \mathbb{E}(Z(s-1))'M$, for $s \geq 1$ and iterating yields the evolution equation,

$$\mathbb{E}(Z(s))' = \mathbb{E}(Z(0))' M^s, \text{ for } s \geq 0, \tag{5}$$

with initial condition $\mathbb{E}(Z(0)) = e_{i_0}$. From (5) and (2), we have $\lim\limits_{s\to\infty} \mathbb{E}(Z(s)) = 0$ if and only if $\rho < 1$. In contrast, the mean size of the population explodes geometrically when $\rho > 1$ while if $\rho = 1$, it converges to a positive constant.

The probability-normalized left eigenvector ν describes the stationary mean behavior of the types. More precisely, if $\mathbb{E}(Z(0)) = \kappa\,\nu$, where κ is the mean number of individuals at time 0, then $\mathbb{E}(Z(s))' = \kappa\rho^s\nu'$ so that ν gives the mean limit behavior of the types. From (2) and (5), one deduces that $\rho^{-t}\mathbb{E}(Z_i(t)) = \mathbb{E}(Z(0))'\rho^{-t}M^t e_{i_0} = \mathbb{E}(Z(0))'\, h\,\nu(j) + o(1)$ where $o(1)$ denotes a function that vanishes at infinity. Hence, $\lim\limits_{t\to\infty} \mathbb{E}(Z_j(t))/(\sum_{i\in I} \mathbb{E}(Z_i(t)) = \nu(j)$. Later, we will recover the same result of ν giving the limiting mean type distribution for the class of finite time-delayed multi-type$\mathfrak{bp}$'s we study as relation (18) in Proposition 4.

Finally, we mention that a pointwise limit result on the type distribution of the population was presented in [14] in the form of the Theorem on the convergence of types which states that, almost surely on non-extinction for a supercritical multi-type $\mathfrak{bp}$, one has $\lim\limits_{t\to\infty} Z_j(t)/(\sum_{i\in I} Z_i(t)) = \nu(j)$ for $j \in I$.

3 Delayed multi-type branching processes

3.1 Definitions

Now we can present the model of interest which we call a finite delayed multi-type $\mathfrak{bp}$. In this model, each individual is born at some time $s \geq 0$ and generates offspring independently of all other individuals in the process. The offspring can be of any type $i \in I$ and each is born during a finite period of time $\mathcal{D} \subset \mathbb{N} = \{1, 2, \ldots\}$ called the active period. The number of offspring of each type in I produced at each time point in the active period is independent. So, an individual born at some time s generates offspring at times in $s + \mathcal{D}$, and the offspring generated at $s + d$ are said to have been generated when the individual had age d.

Let $D = \max \mathcal{D} < \infty$ be the maximum age at which an individual is able to produce offspring. This quantity is assumed to be the same for all individuals, independent of type. The length of the active period could also be modelled as a bounded random variable whose distribution is the same for all individuals. We shall see later that, in terms of analyzing the evolution of the mean number of individuals, this does not add any complexity to the model and this generalization can be managed within the present framework. We shall therefore continue to treat the active period as having a fixed finite extent.

Furthermore, we shall suppose that $|\mathcal{D}| > 1$ to guarantee that the process is distinct from a multi-type $\mathfrak{bp}$ in which each individual is only active for one unit of time. In fact, when $\mathcal{D} = \{D\}$ for some $D \geq 1$, then the process observed at times in the lattice $D\,\mathbb{N}_0$ is a multi-type $\mathfrak{bp}$. The canonical form of the multi-type $\mathfrak{bp}$ is recovered by rescaling the time index set by $1/D$. This is also the reason why one can assume g.c.d. $\mathcal{D} = 1$. The set $\mathcal{D}$ is of the form $\mathcal{D} = \{d_1, d_2, \ldots, d_\ell\}$, where $1 \leq d^1 < d^2 < \cdots < d^\ell = D$ and $\ell > 1$. In particular, we may permit a latency period of duration $d_1 - 1$ to exist when $d_1 > 1$. In terms of epidemic modelling, this characteristic allows a delayed $\mathfrak{bp}$ to incorporate an incubation time, that is, a period of time following infection during which and individual is not yet contagious.

Next, let individuals of type i born at time s be enumerated by indices $l \in \mathbb{N}$. The l-th individual of type i born at time s generates $\xi^{i,j}_{s,l;d}$ offspring of type j at time $s + d$ for $d \in \mathcal{D}$. The random variables $(\xi^{i,j}_{s,l;d} : i, j \in I, l \geq 1, s \geq 0, d \in \mathcal{D})$ are independent and for fixed $i, j \in I$ and $d \in \mathcal{D}$, $(\xi^{i,j}_{s,l;d} : l \geq 1, s \geq 0)$ is a sequence of identically distributed random variables. We shall use $p^{i,j}_d$ to denote their common distribution while $\xi^{i,j}_d$ will denote a random variable with this distribution. Thus, $p^{i,j}_d(n)$ is the probability a type i individual gives birth to n type j offspring d time units after its own birth. Naturally, $\sum_{n \geq 0} p^{i,j}_d(n) = 1$. As was the case for multi-type $\mathfrak{bp}$'s, here we require non-singularity. Recall singular means each individual has exactly one offspring.

Delayed multi-type $\mathfrak{bp}$'s generalize multi-type $\mathfrak{bp}$'s to have multiple mean matrices, one for each offset in the active period $\mathcal{D}$. The mean number of offspring of type j produced by an individual of type i after d units of time is given by

$$M_d(i,j) = \mathbb{E}\left(\xi^d(i,j)\right) = \sum_{n \geq 0} n\, p^{i,j}_d(n), \text{ for } i, j \in I, d \in \mathcal{D}.$$

Then, the matrix of the mean number of offspring of each type born after an offset of d time units is $M_d = (M_d(i,j) : i, j \in I)$. We shall assume M_d is irreducible for all $d \in \mathcal{D}$. By convention, we set $M_d = 0$ if $d \notin \mathcal{D}$. Moreover, the left and right eigenvectors ν_d and h_d of each mean matrix M_d are assumed to be normalized such that $\nu'_d h_d = 1$ and $\nu'_d \mathbf{1} = 1$ for all $d \in \mathcal{D}$. Let ρ_d denote the P-F eigenvalue (spectral radius) of M_d, so $M_d h_d = \rho_d h_d$ and $\nu'_d M_d = \rho_d \nu'_d$.

Now let us consider the sequence $\mathcal{X}(s) = (\mathcal{X}_j(s) : j \in I)$ for $s \in \mathbb{N}_0$, defined by $\mathcal{X}(s) = \mathbf{0}$ for $s < 0$ and

$$\mathcal{X}_j(s) = \mathbf{1}(j = i_0, s = 0) + \sum_{i \in I} \sum_{d \in \mathcal{D}} \sum_{l=1}^{\mathcal{X}_i(s-d)} \xi^{i,j}_{s-d,l;d} \text{ for } s \geq 0. \tag{6}$$

Evidently the initial condition is $\mathcal{X}(0) = e_{i_0}$ and for $s > 0$, $\mathcal{X}_j(s)$ represents the number of type $j$offspring born at time s. In the process $\mathcal{X} = (\mathcal{X}(s) : s \geq 0)$ there is the implicit assumption that individuals are only *alive* for a single unit of time, in other words, at the time they are born. Therefore, this process reduces to a multi-type $\mathfrak{bp}$ when $\mathcal{D} = \{1\}$.

Next, we shall endow each individual with a lifetime during which it is considered to be alive. Let $\mathcal{L}$ be a non-negative random variable such that $\mathbb{P}(\mathcal{L} < \infty) = 1$. The law of $\mathcal{L}$ will be used to determine the number of time units an individual lives following its birth. Then, the lifetimes of individuals are given by $\mathcal{L}(t, l, i)+1$ where the random variables $\mathcal{L}(t, l, i)$'s are independent copies of $\mathcal{L}$. An individual (t, l, i) is thus alive during the timespan $\{t, \ldots, t+\mathcal{L}(t, l, i)\}$. Define $\mathcal{Z}(s) = (\mathcal{Z}_j(s) : j \in I)$ to be the number of individuals of each type alive at time s. To derive an expression for $\mathcal{Z}(s)$, recall that when $s \geq 1$, the set $\{(s, l, j) : l = 1, \ldots, \mathcal{X}_j(s)\}$ enumerates the offspring of type j born at time s when $s \geq 1$. For $s = 0$ and $j = i_0$, $(0, 1, i_0)$ denotes the initial individual. Then,

$$\mathcal{Z}_j(s) = \sum_{b \geq 0} \sum_{l=1}^{\mathcal{X}_j(s-b)} \mathbf{1}(\mathcal{L}(s-b, l, j) \geq b) \text{ for } s \geq 0, \tag{7}$$

and $\mathcal{Z}(s) = 0$ for $s < 0$. We call $\mathcal{Z} = (\mathcal{Z}(s) : s \in \mathbb{N}_0)$ a (non-singular) delayed multi-type $\mathfrak{bp}$. The active period is $\mathcal{D}$, individuals' lifetimes are independent copies of $\mathcal{L} + 1$ and the mean offspring matrices are M_d for $d \in \mathcal{D}$.

Observe that (7) gives rise to the following componentwise inequalities between $\mathcal{X}(s)$ and $\mathcal{Z}(s)$:

$$\mathcal{X}_j(s) \leq \mathcal{Z}_j(s) \leq \sum_{b=0}^{s} \mathcal{X}_j(s-b) \text{ for } s \geq 0. \tag{8}$$

Finally, as promised earlier, we briefly touch on generalizing the model from a deterministic active period to a bounded random period. Assume individuals are active for a random bounded amount of time $\mathcal{A}$ independently of the generation process with the number of offspring $\xi_d^{i,j}$ distributed according to law $p_d^{i,j}$. An appropriate reformulation of the offspring laws allows us to recover a constant active-period model. Set D to be the least upper bound for $\mathcal{A}$ and redefine the random number of offspring to be $\widehat{\xi}_d^{i,j} = \xi_d^{i,j}\mathbf{1}(\mathcal{A} \geq d)$. Then, the new offspring law $\widehat{p}_d^{i,j}(n) = \mathbb{P}(\widehat{\xi}_d^{i,j} = n)$ would be

$$\widehat{p}_d^{i,j}(n) = p_d^{i,j}(n)\mathbb{P}(\mathcal{A} \geq d) \text{ for } n > 0 \text{ and } \widehat{p}_d^{i,j}(0) = p_d^{i,j}(0)\mathbb{P}(\mathcal{A} \geq d) + \mathbb{P}(\mathcal{A} < d). \tag{9}$$

Now, if $(M_d\colon\ d \in \mathcal{D})$ is the family of mean matrices in the random case, then the mean matrices in the deterministic case would be $\widehat{M_d} = a_d M_d$, where $a_d = \mathbb{P}(\mathcal{A} \geq d)$ for $d \in \mathcal{D}$. So, when the quantities to be analyzed depend on the mean matrices, a simple change of offspring law as in (9) allows us to assume that the active period is constant and this change only introduces rescalings of the mean matrices.

4 Extinction

Our first result concerns criteria for determining extinction of the process. Define the probability of extinction of the population of type j by $\mathfrak{q}_j = \mathbb{P}(\lim_{t\to\infty} \mathcal{Z}_j(t) = 0)$.

Proposition 1. *Let $\overline{\rho}$ be the spectral radius of $\sum_{d\in\mathcal{D}} M_d$. On one hand, if $\overline{\rho} \leq 1$, then $\mathcal{Z}$ almost surely becomes extinct, that is, $\mathfrak{q}_j = 1$ for all $j \in I$. On the other hand, if $\overline{\rho} > 1$, then $\mathfrak{q}_j < 1$ for all $j \in I$ and $\mathbb{P}(\lim_{t\to\infty} \mathcal{Z}_j(t) = \infty) = 1 - \mathfrak{q}_j > 0$ for all $j \in I$.*

Proof: Based on (7) and since lifetimes are almost surely finite, we first make the claim that $\mathcal{Z}$ almost surely becomes extinct if and only if $\mathcal{X}$ does, that is $\mathbb{P}(\lim_{t\to\infty} \mathcal{X}_j(t) = 0) = 1$. From (8), it is clear that extinction of $\mathcal{Z}$ implies extinction of $\mathcal{X}$. Next consider a realization of $\mathcal{X}$ and assume it becomes extinct almost surely. Recall that each individual produced by the process can be enumerated by the time s of its birth, together with its type j and some index $l \in \mathbb{N}$ that distinguishes individuals of the same type born at the same time, so an individual is identified by the triplet (s, l, j). Since $\mathcal{X}$ almost surely becomes extinct, the extinction time $T = \min\{t \geq 0 : \sum_{j\in I} \mathcal{X}_j(t+s) = 0$ for all $s \geq 0\}$ is almost surely finite and conditioning on this obtains

$$\mathbb{P}(\mathcal{Z}(T+s) \neq 0 \,|\, T < \infty) = \mathbb{P}\left(\sum_{j\in I}\sum_{t=0}^{T-1}\sum_{l=1}^{\mathcal{X}_j(t)} \mathbf{1}(t + \mathcal{L}(t,l,j) \geq T+s) > 0 \;\middle|\; T < \infty\right) \to 0 \text{ as } s \to \infty$$

and so $\mathcal{Z}$ becomes extinct. The claim that $\mathcal{Z}$ becomes extinct if and only if $\mathcal{X}$ becomes extinct holds and henceforth we need only deal with the extinction of $\mathcal{X}$.

Consider a realization of $\mathcal{X}$ and select an individual, say, (s,l,i). Let $\mathfrak{a}(s,l,i)$ be the parent of (closest ancestor to) individual (s,l,i). We fix $\widehat{d}(s,l,i)$ to be the age of $\mathfrak{a}(s,l,i)$ at the instant (s,l,i) was born, and so the parent $\mathfrak{a}(s,l,i)$ was born at time $s-\widehat{d}(s,l,i)$. Set $\iota(s,l,i)$ to the type of (s,l,i)'s parent and let $\widehat{l}(s,l)$ be the index that distinguishes $\mathfrak{a}(s,l,i)$ among all those individuals of type $\iota(s,l,i)$ born at time $s-\widehat{d}((s,l,i)$. Then $\left(s-\widehat{d}(s,l,i), \widehat{l}(s,l,i), \iota(s,l,i)\right)$ is the label that uniquely identifies the parent of individual (s,l,i).

Next define the function $\varphi : \mathbb{N}_0 \times \mathbb{N} \times I \to \mathbb{N}_0$ inductively as follows:

- For individuals labelled by $(0,l,i)$, where $i \in I$ and $l \in \mathbb{N}$, that is, initial individuals without a parent, we set $\varphi(0,l,i) = 0$.
- For an individual (s,l,i) born at time $s \geq 1$, we set

$$\varphi(s,l,i) = \varphi\left(s-\widehat{d}(s,l,i), \widehat{l}(s,l,i), \iota(s,l,i)\right) + 1.$$

The function $\varphi(s,l,i)$ counts the number of ancestors that individual (s,l,i) has. By definition, one has $\varphi(s,l,i) \leq s$ and $\widehat{d}(s,l,i) \leq D$ so that $\varphi(s,l,i) \geq \lceil s/D \rceil$, the smallest integer greater than or equal to s/D.

Now, define a new process $Z = (Z_j(s) : j \in I, s \geq 0)$ by

$$\begin{aligned} Z_j(s+1) &= \sum_{i\in I}\ \sum_{(t,l)\in\mathbb{N}_0\times\mathbb{N}:\varphi(t,l,i)=s} \left(\sum_{d\in\mathcal{D}} \xi^{i,j}_{t,l;d}\right) \text{ for } s \geq 0 \text{ and} \\ Z_j(0) &= |\{(t,l,i) \in \mathbb{N}_0 \times \mathbb{N} \times I : \varphi(t,l,i) = 0\}|. \end{aligned} \tag{10}$$

Based on the characteristics of the delayed multi-type $\mathfrak{bp}$ set out in Section 3.1, the number of offspring of type j produced by individuals of type i are independent and identically distributed as $\sum_{d\in\mathcal{D}} \xi^{i,j}_d$. Therefore, in the new process Z, the law of the number of offspring of type j attributable to a type i individual is the convolution $p^{i,j}_{d^1} * \cdots * p^{i,j}_{d^\ell}$, where $\mathcal{D} = \{d^1, \ldots, d^\ell\}$. Furthermore, the number of offspring of different types are independent, whether or not they share the same parent. By relabelling individuals so that the offspring of (s,l,i) have labels of the form $(s+1,u,j)$ where u is chosen appropriately to index all of its type j offspring for $j \in I$ irrespective of the time they were born in the delayed multi-type $\mathfrak{bp}$, (10) can be written in the same form as (1). Hence, Z is a multi-type $\mathfrak{bp}$ with mean matrix $\sum_{d\in\mathcal{D}} M_d$.

From the definition, we have that if $\mathcal{X}(t) = 0$ for $t \geq s$, then $Z(t) = 0$ for $t \geq s$, and conversely, if $Z(t) = 0$ for $t \geq s$, then $\mathcal{X}(t) = 0$ for $t \geq \lceil s/D \rceil$. So, $\mathfrak{q}_j$ is equal to the probability of extinction q_j of the multi-type $\mathfrak{bp}$ Z with mean matrix $\sum_{d\in\mathcal{D}} M_d$. The result then follows from the analysis of q_j in (4). □

The spectral radius $\overline{\rho}$ of $\sum_{d\in\mathcal{D}} M_d$ specifies the value that delineates between extinction and non-extinction. Observe that

$$\overline{\rho} \leq \sum_{d\in\mathcal{D}} \rho_d.$$

To see this, let $\|\cdot\|$ be a matrix norm. Then, $\overline{\rho} \leq \|(\sum_{d\in\mathcal{D}} M_d)^t\|^{\frac{1}{t}}$ and the inequality follows by making use of $\|\sum_{d\in\mathcal{D}} M_d\| \leq \sum_{d\in\mathcal{D}} \|M_d\|$.

Remark 2. *For general positive matrices $(M_d : d \in \mathcal{D})$, the spectral radius $\overline{\rho}$ of $\sum_{d\in\mathcal{D}} M_d$ may satisfy $\overline{\rho} < \sum_{d\in\mathcal{D}} \rho_d$, in which case , it is possible that $\overline{\rho} < 1 < \sum_{d\in\mathcal{D}} \rho_d$. Hence there are situations where a positive parameter solves (16) while the process $\mathcal{Z}$ is subcritical, becoming extinct geometrically fast at rate $\overline{\rho}$.* □

5 Mean evolution and limit behavior

Now we seek recursion formulae for the mean evolution $\mathbb{E}_{i_0}(\mathcal{X}(s))$ and $\mathbb{E}_{i_0}(\mathcal{Z}(s))$, after which we will obtain the long-term limit of the latter when the mean matrices (M_d) share left and right P-F eigenvectors.

5.1 Mean evolution

To study the evolution of the mean number of individuals $\mathbb{E}_{i_0}(\mathcal{Z}(s)) = (\mathbb{E}(\mathcal{Z}_j(s)) : j \in I)$ in a delayed multi-type $\mathfrak{bp}$, we begin by computing the mean $\mathbb{E}_{i_0}(\mathcal{X}(s))$ of the process $(\mathcal{X}(s) : s \geq 0)$. Let $\mathcal{F}(s)$ be the σ-field of events up to time s generated by the random variables defining individuals born up to time s, that is

$$\mathcal{F}(s) = \sigma(\xi^{i,j}_{t,l_i;d} : i, j \in I, 1 \leq l_i \leq \mathcal{X}_i(t), 0 \leq t < s, d \in \mathcal{D}, d \leq s-t).$$

By conditioning relation (6) with respect to $\mathcal{F}(s-1)$ and taking expectations, we obtain

$$\begin{aligned}
\mathbb{E}_{i_0}(\mathcal{X}_j(s)) &= \mathbf{1}(j = i_0, s = 0) + \sum_{i\in I}\sum_{d\in\mathcal{D}} \mathbb{E}_{i_0}(\mathcal{X}_i(s-d))\mathbb{E}(\xi^{i,j}_d) \\
&= \mathbf{1}(j = i_0, s = 0) + \sum_{i\in I}\sum_{d\in\mathcal{D}} \mathbb{E}_{i_0}(\mathcal{X}_i(s-d))M_d(i,j).
\end{aligned}$$

The sequence of means satisfy $\mathbb{E}(\mathcal{X}(s)) = 0$ for $s < 0$ and we obtain the following evolution equation,

$$\mathbb{E}_{i_0}(\mathcal{X}(s))' = e'_{i_0}\mathbf{1}(s = 0) + \sum_{d\in\mathcal{D}} \mathbb{E}_{i_0}(\mathcal{X}(s-d))'M_d, \text{ for } s \geq 0. \tag{11}$$

To evaluate the expected number of individuals of each type alive at time s, $\mathbb{E}_{i_0}(\mathcal{Z}_j(s))$, we can use relation (7) together with $M_d = 0$ for $d \notin \mathcal{D}$ to get

$$\mathbb{E}_{i_0}(\mathcal{Z}_j(s)) = \sum_{b=0}^{s} \mathbb{E}_{i_0}\left(\sum_{l=1}^{\mathcal{X}_j(s-b)} \mathbf{1}\big(\mathcal{L}(s-b,l,j) \geq b\big)\right) = \sum_{b=0}^{s} \mathbb{E}_{i_0}(\mathcal{X}_j(s-b))\mathbb{P}(\mathcal{L} \geq b).$$

Then, substituting $\mathbb{E}(\mathcal{X}_j(s))$ into this gives

$$\begin{aligned}
\mathbb{E}_{i_0}(\mathcal{Z}_j(s)) &= \sum_{b=0}^{s}\left[\mathbf{1}(j = i_0, s-b = 0) + \sum_{i\in I}\sum_{d\in\mathcal{D}} \mathbb{E}_{i_0}(\mathcal{X}_i(s-d-b))M_d(i,j)\right]\mathbb{P}(\mathcal{L} \geq b) \\
&= \sum_{b=0}^{s} \mathbf{1}(j = i_0, s = b)\mathbb{P}(\mathcal{L} \geq b) + \sum_{b=0}^{s}\sum_{i\in I}\sum_{d\in\mathcal{D}} \mathbb{E}_{i_0}(\mathcal{X}_i(s-d-b))M_d(i,j)\mathbb{P}(\mathcal{L} \geq b) \\
&= \mathbf{1}(j = i_0)\mathbb{P}(\mathcal{L} \geq s) + \sum_{i\in I}\sum_{d\in\mathcal{D}}\left(\sum_{b=0}^{s} \mathbb{E}_{i_0}(\mathcal{X}_i(s-d-b))\mathbb{P}(\mathcal{L} \geq b)\right)M_d(i,j) \\
&= \mathbf{1}(j = i_0)\mathbb{P}(\mathcal{L} \geq s) + \sum_{i\in I}\sum_{d\in\mathcal{D}} \mathbb{E}_{i_0}(\mathcal{Z}_i(s-d))M_d(i,j).
\end{aligned}$$

In vectorial form, we therefore have

$$\mathbb{E}_{i_0}(\mathcal{Z}(s))' = e'_{i_0}\mathbb{P}(\mathcal{L} \geq s) + \sum_{d\in\mathcal{D}} \mathbb{E}_{i_0}(\mathcal{Z}_i(s-d))'M_d, \quad s \geq 0. \tag{12}$$

Let $s \geq 1$. We can iterate equations (11) and (12) backwards through time from s to 0. This is done by taking a path with elements in $\mathcal{D}$. For $r \leq s$, the set of paths of length r from s to 0 with elements in $\mathcal{D}$ is defined by

$$\Gamma(s,r) = \{(d_1,\ldots,d_r) \in \mathcal{D}^r : \sum_{l=1}^{r} d_l = s\}. \tag{13}$$

The elements in a path $(d_1, \dots, d_r)$ represent step sizes in $\mathcal{D}$ so that a path is represented as a sequence of r steps that span the range from 0 to s, so d_r connects $s - d_r$ to s and so on. Note that if $\Gamma(s,r) \neq \emptyset$ then $r \geq \lceil s/D \rceil$. So, for a family of functions (f_r) we have

$$\sum_{1\leq r\leq s} \sum_{(d_1,\dots,d_r)\in\Gamma(s,r)} f_r(d_1,\dots,d_r) = \sum_{\lceil s/D\rceil \leq r\leq s} \sum_{(d_1,\dots,d_r)\in\Gamma(s,r)} f_r(d_1,\dots,d_r).$$

Therefore, from (11) and (12) we find

$$\mathbb{E}(\mathcal{X}(s))' = \mathbf{1}(s=0)\mathbb{E}(\mathcal{X}(0))' + \mathbb{E}(\mathcal{X}(0))' \left(\sum_{1\leq r\leq s} \sum_{(d_1,\dots,d_r)\in\Gamma(s,r)} \prod_{l=1}^{r} M_{d_l} \right) \tag{14}$$

and

$$\mathbb{E}(\mathcal{Z}(s))' = \mathbb{E}(\mathcal{Z}(0))'\mathbb{P}(\mathcal{L} \geq s) + \mathbb{E}(\mathcal{Z}(0))' \sum_{b=0}^{s-1} \mathbb{P}(\mathcal{L} \geq b) \left(\sum_{1\leq r\leq s-b} \sum_{(d_1,\dots,d_r)\in\Gamma(s-b,r)} \prod_{l=1}^{r} M_{d_l} \right). \tag{15}$$

Remark 3. *If all the matrices $(M_d : d \in \mathcal{D})$ all have the same left eigenvector ν, then $\nu' \prod_{l=1}^{r} M_{d_l} = \prod_{l=1}^{r} \rho_{d_l} \nu'$. Further, since we assume ν is normalized to sum to unity, it is stationary for the mean evolution of types, that is,*

$$\mathbb{E}(\mathcal{Z}(0))' = \nu' \;\Rightarrow\; \forall s \geq 0,\; \frac{\mathbb{E}(\mathcal{Z}(s))'}{\mathbb{E}(\mathcal{Z}(s))'\mathbf{1}} = \nu'. \quad \square$$

We shall see later in Proposition 4 and its proof that this coincides with the limit distribution for the mean evolution of types.

5.2 Limit results for matrices sharing P-F eigenvectors

Now, we first introduce the class of matrices where our result will apply. a family of matrices $(M_d : d \in \mathcal{D})$ is said to share P-F eigenvectors when $h_d = h$ and $\nu_d = \nu$ for $d \in \mathcal{D}$. We assume that h and ν are scaled so that $\nu' h = 1$. Obviously if all the M_d's only differ by scalar factors, then they share P-F eigenvectors. When $(M_d : d \in \mathcal{D})$ belongs to this class of matrices, we will be able to handle the product of matrices $\prod_{l=1}^{r} M_{d_l}$ in expression (15).

Since $M_d h = \rho_d\, h$ for $d \in \mathcal{D}$, then $\sum_{d\in\mathcal{D}} \rho_d$ is the spectral radius of $\sum_{d\in\mathcal{D}} M_d$. More precisely,

$$(\sum_{d\in\mathcal{D}} M_d)h = (\sum_{d\in\mathcal{D}} \rho_d)h, \quad \nu'(\sum_{d\in\mathcal{D}} M_d) = (\sum_{d\in\mathcal{D}} \rho_d)\nu'.$$

From Proposition 1, we know that if $\sum_{d\in\mathcal{D}} \rho_d \leq 1$, then the process $\mathcal{Z}$ almost surely becomes extinct and if $\sum_{d\in\mathcal{D}} \rho_d > 1$, there is a positive probability that it survives forever.

We want to examine the behaviour of the evolution equation (12) when the matrices share P-F eigenvectors. The key notion is the so-called Malthusian parameter $\theta \in \mathbb{R}$ which uniquely solves the following equation:

$$\sum_{d\in\mathcal{D}} \rho_d e^{-\theta d} = 1. \tag{16}$$

The parameter θ is positive, zero or negative according as $\sum_{d\in\mathcal{D}} \rho_d$ is greater than, equal to or less than unity, respectively. Hence, the subcritical, critical or supercritical nature of the process is determined by the Malthusian parameter being negative, zero or positive, respectively. Now, one can define a probability vector $\vec{\beta} = (\beta_d : d \in \mathcal{D})$ by

$$\beta_d = \rho_d e^{-\theta d}, d \in \mathcal{D}.$$

The mean of $\vec{\beta}$ is $\mu(\vec{\beta}) = \sum_{d\in\mathcal{D}} d\, \beta_d$.

Proposition 4. *Assume that the matrices* $(M_d : d \in \mathcal{D})$ *share P-F eigenvectors and let* $\theta \in \mathbb{R}$ *be the unique solution to (16). If the process* $\mathcal{Z}$ *is subcritical (that is,* $\theta < 0$*), assume that* $\mathbb{E}\left(e^{-\theta \mathcal{L}}\right) < \infty$*; and if the process is critical (that is,* $\theta = 0$*), assume that* $\mathbb{E}(\mathcal{L}) < \infty$*. Define the probability vector* $\vec{\beta} = \left(\beta_d = \rho_d e^{-\theta d} : d \in \mathcal{D}\right)$*. Then*

$$\lim_{s\to\infty} e^{-\theta s}\mathbb{E}(\mathcal{Z}(s))' = (\mu(\vec{\beta}))^{-1}\mathbb{E}((\mathcal{Z}(0))'h\left(\sum_{t\geq 0}\mathbb{P}(\mathcal{L}\geq t)e^{-\theta t}\right)\nu'. \tag{17}$$

Moreover, ν' *is the limit of the mean evolution of types,*

$$\lim_{s\to\infty}\frac{\mathbb{E}(\mathcal{Z}(s))'}{\mathbb{E}(\mathcal{Z}(s))'\mathbf{1}} = \nu' \tag{18}$$

and the asymptotic behaviour of the expected number of offspring is described by

$$\lim_{s\to\infty} e^{-\theta s}\mathbb{E}(\mathcal{X}(s))' = (\mu(\vec{\beta}))^{-1}\mathbb{E}(\mathcal{X}(0))'h\nu'. \tag{19}$$

Finally the proportion of the expected active population of type j *that is of age* d *at time* s*, which is given by*

$$\mathcal{E}_d^j(s) = \frac{\mathbb{E}(\mathcal{X}_j(s-d))\mathbb{P}(\mathcal{L}\geq d)}{\sum_{d'\in\mathcal{D}}\mathbb{E}(\mathcal{X}_j(s-d'))\mathbb{P}(\mathcal{L}\geq d')}$$

satisfies

$$\lim_{s\to\infty}\mathcal{E}_d^j(s) = \frac{\mathbb{P}(\mathcal{L}\geq d)e^{-\theta d}}{\sum_{d'\in\mathcal{D}}\mathbb{P}(\mathcal{L}\geq d')e^{-\theta d'}} \text{ for } d\in\mathcal{D},\ j\in I. \tag{20}$$

Remark 5. *The conditions* $\mathbb{E}\left(e^{-\theta\mathcal{L}}\right) < \infty$ *in the subcritical case and* $\mathbb{E}(\mathcal{L}) < \infty$ *in the critical case pose no significant limitation in real-world applications because they are satisfied for any bounded lifetime* $\mathcal{L} \leq L < \infty$*. That is, the conditions hold for any population of organisms having bounded lifetimes.*

In the following sections, we discuss the main elements involved in this proposition and furnish its proof.

6 Matrices sharing P-F eigenvectors and runs

First we will briefly discuss the property of a family of matrices sharing the same P-F eigenvectors and then we will study runs in sequences of symbols in $\mathcal{D}$ that will provide control over the product of matrices that appears in (15).

6.1 Matrices sharing P-F eigenvectors

Let M be an irreducible non-negative matrix whose right and left P-F eigenvectors are ν and h respectively. The associated stochastic matrix P_M defined by $P_M(i,j) = \rho^{-1}M(i,j)h(j)/h(i)$ has $\nu h = (\nu(i)h(i) : i \in I)$ as its unique stationary distribution. Conversely, let P be an irreducible stochastic matrix with stationary distribution π. Fix $\rho > 0$, a vector $h > 0$ and define

$$M(i,j) = \rho\, P(i,j)h(i)/h(j). \tag{21}$$

Then, M is an irreducible non-negative matrix. It is straightforward to see that $Mh = \rho\, h$ and so ρ is the P-F eigenvalue corresponding to the right eigenvector h. By definition, its associated stochastic matrix is P. It is easily checked that $\nu = \pi/h = (\pi(i)/h(i) : i \in I)$ is the left P-F eigenvector for M. Therefore, the class of irreducible non-negative matrices with fixed P-F eigenvectors ν and h can be described by an arbitrary $\rho > 0$ and the class of stochastic matrices with stationary distribution $\pi = (\pi(i) = h(i)\nu(i) : i \in I)$. We note that a similar construction to (21) can be made by taking $M(i,j) = \rho\,(\nu(j)/\nu(i))P(j,i)$. This is a time-reversed construction because $P(i,j) = (\pi(j)/\pi(i))P(j,i)$ is the transition matrix of the time reverse of the Markov chain defined by P.

Obviously, when all the M_d's only differ by scalar multiples, then they share P-F eigenvectors. Below, we will show that families of irreducible non-negative matrices which commute share P-F eigenvectors. The matrices $(M_d : d \in \mathcal{D})$ commute if $M_d M_{d'} = M_{d'} M_d$ for $d, d' \in \mathcal{D}$.

Lemma 6. *If the family of matrices $(M_d : d \in \mathcal{D})$ is commutative, then the matrices share the same right and left P-F eigenvectors.*

Proof: Let $d, d' \in \mathcal{D}$. The commutativity property and relation (3) imply

$$h_d \nu'_d h_{d'} \nu'_{d'} = \lim_{t\to\infty} \rho_d^{-t} M_d^t \rho_{d'}^{-t} M_{d'}^t = \lim_{t\to\infty} \rho_{d'}^{-t} M_{d'}^t \rho_d^{-t} M_d^t = h_{d'} \nu'_{d'} h_d \nu'_d.$$

So, $\delta h_d \nu'_{d'} = \epsilon h_{d'} \nu'_d$ where $\delta = \nu'_d h_{d'}$ and $\epsilon = \nu'_{d'} h_d$. Then, $\delta h_d(i) \nu_{d'}(j) = \epsilon h_{d'}(i) \nu_d(j)$ for all $i, j \in I$. Summing over j leads to $h_{d'} = (\delta/\epsilon) h_d$ and hence $\delta = \nu'_d h_{d'} = \nu'_d h_d (\delta/\epsilon) = \delta/\epsilon$. Similarly, one can show that $\epsilon = \epsilon/\delta$. From this we obtain $\delta = 1 = \epsilon$ and hence we can conclude that $h_d = h_{d'}$. It also follows that $\nu_d(j) = \nu_{d'}(j)$ for all $j \in I$, that is, $\nu_d = \nu_{d'}$. $\Box$

Suppose the matrices $(M_d : d \in \mathcal{D})$ share P-F eigenvectors. Since $h_d = h$, $\nu_d = \nu$ and $\nu' h = 1$, we have $\nu'_d h_{d'} = \nu' h = 1$ for all $d, d' \in \mathcal{D}$. So,

$$\prod_{d \in \mathcal{D}} h_d \nu'_d = h \nu'.$$

Let $\|\cdot\|$ be the ∞-norm on vectors as well as the associated spectral norm on matrices, that is, $\|A\| = \max_{i\in I} \sum_{j \in I} |A(i,j)|$. So all stochastic matrices P have $\|P\| = 1$. Assume the matrices $(M_d : d \in \mathcal{D})$ share the right P-F eigenvector h. Then, we claim that

$$\forall r \geq 1, (d_1, \ldots, d_r) \in \mathcal{D}^r : \quad \|\prod_{l=1}^{r} \rho_{d_l}^{-1} M_{d_l}\| \leq \mathfrak{w}(h) \text{ where } \mathfrak{w}(h) = \max_{i,j\in I}(h_i/h_j). \tag{22}$$

In fact, h is a right eigenvector of $A = \prod_{l=1}^r \rho_{d_l}^{-1} M_{d_l}$ corresponding to eigenvalue 1. Then, $P = (P(i,j : i,j \in I)$, where $P(i,j) = A(i,j)h(j)/h(i)$, is a stochastic matrix so $\|P\| = 1$ and $\|A\| = \max_{i \in I} \sum_{j \in I} |P(i,j)h(i)/h(j)| \leq \max_{i,j \in I}(h(i)/h(j))$. Then, (22) holds. Similarly, one shows that if $(M_d : d \in \mathcal{D})$ share the left P-F eigenvector ν, then, $\forall r \geq 1$, $(d_1, \ldots, d_r) \in \mathcal{D}^r$, $\|\prod_{l=1}^r \rho_{d_l}^{-1} M_{d_l}\|_1 \leq \mathfrak{w}(\nu)$, where $\|A\|_1 = \|A'\|$ is the L^1-norm.

6.2 Frequency of runs

Let $\kappa > 1$. We require the concept of a κ-run in a sequence. We say that a sequence $(d_l : l = 1, \ldots, s)$ has a κ-run if there exists some $l_0 \in \{1, \ldots, s - \kappa + 1\}$ such that $d_l = d_{l_0}$ for $l = l_0, \ldots, l_0 + \kappa - 1$. The importance of this notion is to handle products of matrices for families of matrices sharing P-F eigenvectors, as will become apparent in the next result.

Lemma 7. *Assume $(M_d : d \in \mathcal{D})$ share the right and left P-F eigenvectors h and ν respectively. Then, for every $\epsilon \in (0,1)$ there exists $\kappa(\epsilon)$ such that for all $r \geq \kappa(\epsilon)$, the following property holds: if the sequence $(d_l : l = 1, \ldots, r)$ has a $\kappa(\epsilon)$-run, then*

$$\|\prod_{l=1}^{r} \rho_{d_l}^{-1} M_{d_l} - h\nu'\| \leq \epsilon. \tag{23}$$

Proof: Since $\lim_{k\to\infty} \rho_d^{-k} M_d^k = h\nu'$ componentwise and the matrix $h\nu'$ is strictly positive, for every $\epsilon' \in (0,1)$ there exists $k(\epsilon')$ such that

$$\forall d \in \mathcal{D}, \forall k \geq k(\epsilon') : \quad (1 - \epsilon')h\nu' \leq \rho_d^{-k} M_d^k - h\nu' \leq (1 + \epsilon)h\nu' \tag{24}$$

componentwise. Hence, if for $(d_l : l = 1, \dots, r)$ there exists some $l_0 \in \{1, \dots, r - k(\epsilon') - 1\}$ such that $d_l = d_{l_0}$ for $l = l_0, \dots, l_0 + k(\epsilon') - 1$, then

$$\rho_{d_{l_0-1}}^{-1} M_{d_{l_0-1}} (1-\epsilon')h\nu' = (1-\epsilon')h\nu' \text{ for } l_0 > 1,$$
$$(1-\epsilon)h\nu' \rho_{d_{l_0+k(\epsilon')}}^{-1} M_{d_{l_0+k(\epsilon')}} = (1-\epsilon')h\nu' \text{ for } l_0 + k(\epsilon') \le r.$$

Therefore, an inductive argument shows that

$$(1-\epsilon')h\nu' \le \left(\prod_{l=1}^{r} \rho_d^{-l} M_d^l - h\nu' \right) \le (1+\epsilon')h\nu',$$

componentwise. Hence, $\| \prod_{l=1}^{r} \rho_d^{-l} M_d^l - h\nu' \| \le \epsilon' \|h\nu'\|$. We take $\kappa(\epsilon) = k(\epsilon')$ with $\epsilon' = \epsilon / \|h\nu'\|$. □

Next, we will study the frequency of runs on finite sets of sequences contained in $\widetilde{\Gamma} = \bigcup_{r \in \mathbb{N}} \mathcal{D}^r$, the set of finite sequences of symbols in $\mathcal{D}$. Let $(x_1, \dots, x_{r(x)})$ be an element in $\widetilde{\Gamma}$. Let $\kappa > 1$. For a class of sequences $\Gamma \subseteq \widetilde{\Gamma}$, one defines

$$\Gamma^{-\kappa} = \{(x_1, \dots, x_{r(x)}) \in \Gamma : \nexists m \le r(x) - (\kappa - 1) \text{ such that } x_m = \dots = x_{m+\kappa-1}\},$$

the class of sequences in Γ that avoid κ-runs. Notice that $(\bigcup_{r<\kappa} \mathcal{D}^r) \cap \Gamma \subseteq \Gamma^{-\kappa}$. It is also useful to introduce

$$\Gamma^{-\kappa;d} = \{(x_1, \dots, x_{r(x)}) \in \Gamma : \nexists m \le r(x) - (\kappa - 1) \text{ such that } x_m = \dots = x_{m+\kappa-1} = d\} \text{ for } d \in \mathcal{D}.$$

This is the class of sequences that avoid κ-runs of the symbol d. One has $\Gamma^{-\kappa} = \bigcap_{d \in \mathcal{D}} \Gamma^{-\kappa;d}$. Let Γ_1 and Γ_2 be two classes of sequences contained in $\widetilde{\Gamma}$. First, since $(\Gamma_1 \times \Gamma_2)^{-\kappa} \subseteq \Gamma_1^{-\kappa} \times \Gamma_2^{-\kappa}$, one obtains

$$\frac{|(\Gamma_1 \times \Gamma_2)^{-\kappa}|}{|\Gamma_1 \times \Gamma_2|} \le \frac{|\Gamma_1^{-\kappa}|}{|\Gamma_1|} \cdot \frac{|\Gamma_2^{-\kappa}|}{|\Gamma_2|} \text{ and } \forall d \in \mathcal{D}, \; \frac{|(\Gamma_1 \times \Gamma_2)^{-\kappa;d}|}{|\Gamma_1 \times \Gamma_2|} \le \frac{|\Gamma_1^{-\kappa;d}|}{|\Gamma_1|} \cdot \frac{|\Gamma_2^{-\kappa;d}|}{|\Gamma_2|}. \tag{25}$$

On the other hand, when the classes Γ_1 and Γ_2 are disjoint one has $(\Gamma_1 \cup \Gamma_2)^{-\kappa} = \Gamma_1^{-\kappa} \cup \Gamma_2^{-\kappa}$ and $(\Gamma_1 \cup \Gamma_2)^{-\kappa;d} = \Gamma_1^{-\kappa} \cup \Gamma_2^{-\kappa;d}$, where the unions are disjoint. Then,

$$\frac{|(\Gamma_1 \cup \Gamma_2)^{-\kappa}|}{|\Gamma_1 \cup \Gamma_2|} \le \max\left(\frac{|\Gamma_1^{-\kappa}|}{|\Gamma_1|}, \frac{|\Gamma_2^{-\kappa}|}{|\Gamma_2|} \right) \text{ and } \forall d \in \mathcal{D}, \; \frac{|(\Gamma_1 \cup \Gamma_2)^{-\kappa;d}|}{|\Gamma_1 \cup \Gamma_2|} \le \max\left(\frac{|\Gamma_1^{-\kappa;d}|}{|\Gamma_1|}, \frac{|\Gamma_2^{-\kappa;d}|}{|\Gamma_2|} \right). \tag{26}$$

Let us consider $\Lambda = \mathbb{N}_0^{\mathcal{D}}$, which is the class of $\mathcal{D}$-tuples taking non-negative values. Its elements will be denoted by $\vec{k}$, $\vec{\lambda}$ or $\vec{h}$. Let $\vec{k} = (k_d : d \in \mathcal{D})$. Set $|\vec{k}| = \sum_{d \in \mathcal{D}} k_d$ and $\mathfrak{m}(\vec{k}) = \sum_{d \in \mathcal{D}} d\, k_d$. We have $\lceil \mathfrak{m}(\vec{k})/D \rceil \le |\vec{k}| \le \mathfrak{m}(\vec{k})$. We define

$$\Lambda(s) = \{\vec{k} \in \mathbb{N}_0^{\mathcal{D}} : \mathfrak{m}(\vec{k}) = s\}, \; \Lambda(s,r) = \{\vec{k} \in \Lambda(s) : |\vec{k}| = r\}.$$

Note that $\Lambda(s,r) \neq \emptyset$ only when $\lceil s/D \rceil \le r \le s$. To each $\vec{k} \in \Lambda(s,r)$ one associates the set of sequences

$$S(\vec{k}) = \{(x_1, \dots, x_r) : \sum_{i=1}^{r} \mathbf{1}(x_i = d) = k_d, \forall\, d \in \mathcal{D}\}.$$

This is the set of all the different permutations of one fixed sequence $(y_1, \dots, y_r) \in S(\vec{k})$. One has

$$|S(\vec{k})| = \frac{r!}{\prod_{d \in \mathcal{D}} k_d!} \text{ where } r = |\vec{k}|. \tag{27}$$

Next, one associates with the class of sequences $S(\vec{k})$ the classes $S(\vec{k})^{-\kappa}$ and $S(\vec{k})^{-\kappa;d} = S(\vec{k}) \cap \Gamma^{-\kappa;d}$ for $d \in \mathcal{D}$. Then, $|S(\vec{k})^{-\kappa}| \le |S(\vec{k})^{-\kappa;d}|$ for $d \in \mathcal{D}$.

Lemma 8. *Let $\kappa > 1$ be fixed. Then, for all $\epsilon > 0$ there exists $s(\epsilon)$ such that for all $s \geq s(\epsilon)$, the inequality $|S(\vec{k})^{-\kappa}|/|S(\vec{k})| \leq \epsilon$ holds for every tuple $\vec{k} \in \Lambda(s)$.*

Proof: Note that the structure of the classes $S(\vec{k})^{-\kappa}$ and $S(\vec{k})^{-\kappa;d}$ only depend on the unordered multiset $\{k_d : d \in \mathcal{D}\}$. In fact when $\{k_d : d \in \mathcal{D}\} = \{\lambda_d : d \in \mathcal{D}\}$, one can define a permutation π in $\mathcal{D}$, such that $k_d = \lambda_{\pi(d)}$. This permutation induces a bijection between $S(\vec{k})$ and $S(\vec{\lambda})$ while its restriction maps $S(\vec{k})^{-\kappa}$ into $S(\vec{\lambda})^{-\kappa}$ and $S(\vec{k})^{-\kappa;d}$ into $S(\vec{\lambda}^s)^{-\kappa,\pi(d)}$.

A $\vec{k} = (k_d : d \in \mathcal{D}) \in \Lambda(s, r)$ is called uniform if the (k_d) are as equal as they can be. More precisely, the following property is satisfied. If r is a multiple of $|\mathcal{D}$, then $k_d = k_D = r/|\mathcal{D}|$ for all $d \in \mathcal{D}$. If r is not a multiple of $|\mathcal{D}|$, then for a set $\mathcal{D}_0 \subset \mathcal{D}$ one has: $k_d = \lceil r/|\mathcal{D}| \rceil$ for $d \in \mathcal{D}_0$ and $k_d = \lfloor r/|\mathcal{D}| \rfloor$, the largest integer smaller than or equal to $r/|\mathcal{D}|$, for $d \in \mathcal{D} \setminus \mathcal{D}_0$. Naturally, $\mathcal{D}_0$ must satisfy $|\mathcal{D}_0| = r - \lfloor r/|\mathcal{D}| \rfloor \cdot |\mathcal{D}|$.

Let Λ^U denote the set of uniform tuples. For a tuple $\vec{k} \in \Lambda$, there always exists a uniform tuple $\vec{\lambda} \in \Lambda^U$ that satisfies $|\vec{\lambda}| = |\vec{k}|$. In this case, we set $\vec{\lambda} = U(\vec{k})$.

Take $\vec{k} \in \Lambda$ and $d_0 \in \mathcal{D}$ to be such that $k_d \leq k_{d_0}$. We claim that,

$$\forall \vec{\lambda} = U(\vec{k}) : \quad |S(\vec{k})^{-\kappa;d_0}| \leq |S(\vec{\lambda})^{-\kappa;d_0}|. \tag{28}$$

To see this, first observe that $\lambda_{d_0} \leq k_{d_0}$. We set $r = |\vec{k}| = |\vec{\lambda}|$. We enumerate the r symbols in $\vec{k}$ and $\vec{\lambda}$ by assigning a different pair (d, l) to each d. Then we create a bijection π between these pairs in such a way that $\pi((d, l)) = (d, l')$ for $\min(k_d, \lambda_d)$ pairs (d, l) (or (d, l')). The bijection π allows us to map $S(\vec{k})$ into $S(\vec{\lambda})$ by first mapping the different pairs into different pairs and afterwards deleting the second coordinate. This construction together with $\lambda_{d_0} \leq k_{d_0}$ allows us to conclude that if κ-runs of symbol d_0 are avoided in $S(\vec{k})$, then κ-runs of symbol d_0 are also avoided in $S(\vec{\lambda})$. Thus, (28) is proved.

Let

$$\delta_0 = \max\{|S(\vec{k})^{-\kappa;d}|/|S(\vec{k})| : \vec{k} \in \Lambda^U, |\vec{k}| \in \{\kappa D, \kappa D + 1\}, d \in \mathcal{D}\} \tag{29}$$

be the maximal frequency of uniform sequences of length κD and $\kappa D+1$ that avoids κ-runs of symbol $d \in \mathcal{D}$. Note that in the case $\vec{k} \in \kappa|\mathcal{D}|$, then the ratios $|S(\vec{k})^{-\kappa;d}|/|S(\vec{k})|$ are all the same for all $d \in \mathcal{D}$. One has $\delta_0 \in (0, 1)$ and one takes t_0 to be such that $(1 - \delta_0)^{t_0} \leq \epsilon$.

Let $\vec{k} \in \Lambda(s, r)$. Let $1 \leq m < m + v \leq r$ and $u > 0$. We define

$$S(\vec{k}; u; m, m + v) = \Big\{(y_1, \ldots, y_v) : \exists (x_1, \ldots, x_r) \in S(\vec{k}) \text{ such that } (x_m, \ldots, x_{m+v-1}) = (y_1, \ldots, y_v), \sum_{i=m}^{m+v-1} x_i = u\Big\}.$$

Assume $(y_1, \ldots, y_v) \in S(\vec{k}; u; m, m + v)$. Then, for any permutation $(y_{\pi(1)}, \ldots, y_{\pi(v)})$ of $(y_1, \ldots, y_v)$, we also have $(y_{\pi(1)}, \ldots, y_{\pi(v)}) \in S(\vec{k}; u; m, m + v)$. So, $S(\vec{\lambda}) \subseteq S(\vec{k}; u; m, m + v)$ where $\vec{\lambda} \in \Lambda$ is such that $|\vec{\lambda}| = v$, $\mathfrak{m}(\lambda) = u$ and $\lambda_k = \sum_{i=1}^{v} \mathbf{1}(y_i = d)$ for $d \in \mathcal{D}$. In this case we write $\vec{\lambda} \in \Lambda(\vec{k}; u; m, m + v))$ and we have $S(\vec{k}; u; m, m + v) = \bigcup_{\vec{\lambda} \in \Lambda(\vec{k}; u; m, m+v)} S(\vec{\lambda})$, where this last union is disjoint.

Now take a tuple $\vec{k} \in \Lambda(s)$ where $s \geq t_0 \kappa D^2$. Then $r = |\vec{k}| \geq t_0 \kappa D$. Assume $r = t\kappa d + m$ with $t \geq t_0$ and $0 \leq m < \kappa d$. Firstly, we shall assume $m = 0$. We make a partition of the class of sequences $S(\vec{k})$ as follows:

$$S(\vec{k}) = \bigcup_{(s_1, \ldots, s_t) : \sum_{i=1}^{t} s_i = s} \prod_{j=1}^{t} S(\vec{k}; s_j; (j - 1)\kappa D + 1, j\kappa D),$$

where the union is disjoint and we assume the classes in the products are non-empty. From relations (25) and (26) one gets,

$$|S(\vec{k})^{-\kappa}|/|S(\vec{k})| \leq \max_{(s_1, \ldots, s_t) : \sum_{i=1}^{t} s_i = s} \prod_{j=1}^{t} |S(\vec{k}; s_j; (j - 1)\kappa D + 1, j\kappa D)^{-\kappa}|/|S(\vec{k}; s_j; (j - 1)\kappa D + 1, j\kappa D)|.$$

Now, we know that $S(\vec{k}; s_j; (j-1)\kappa D+1, j\kappa D) = \bigcup_{\vec{\lambda}\in\Lambda(\vec{k};s_j;(j-1)\kappa D+1,j\kappa D)} S(\vec{\lambda})$ is a disjoint union. So,

$$\frac{|S(\vec{k}; s_j; (j-1)\kappa D+1, j\kappa D)^{-\kappa}|}{|S(\vec{k}; s_j; (j-1)\kappa D+1, j\kappa D)|} \le \max_{\vec{\lambda}\in\Lambda(\vec{k};s_j;(j-1)\kappa D+1,j\kappa D)} \left\{ \frac{|S(\vec{\lambda})^{-\kappa}|}{|S(\vec{\lambda})|} \right\}.$$

Now, from (28) one has that $|S(\vec{\lambda})^{-\kappa}|/|S(\vec{\lambda})| \le |S(\vec{h})^{-\kappa}|/|S(\vec{h})|$ where $\vec{h}$ is a uniform tuple with the same length as $\vec{\lambda}$. Since $|\vec{\lambda}| = \kappa D$ from (29), we get $|S(\vec{h})^{-\kappa}|/|S(\vec{h})| \le \delta_0$. Since $t \ge t_0$ one finds that

$$\prod_{j=1}^{t} |S(\vec{k}; s_j; (j-1)\kappa D+1, j\kappa D)^{-\kappa}|/|S(\vec{k}; s_j; (j-1)\kappa D+1, j\kappa D)| \le \delta_0^t \le \epsilon.$$

When $m > 0$, instead of the t blocks $[(j-1)\kappa D+1, j\kappa D]$, $j = 1, \ldots, t$, of length κD, one takes $t-m$ blocks of length κD and m of length $\kappa D+1$ and makes a similar construction to that made in the case $m = 0$. This gives the result. □

7 Proof of Proposition 4

First, since we are considering the mean evolution, $\mathcal{Z} = (\mathcal{Z}(s) : s \ge 0)$ we will associate it to a 1-type delayed $\mathfrak{bp}$ $\zeta = (\zeta(s) : s \ge 0)$ with offspring means $(\rho_d : d \in \mathcal{D})$ starting from $\zeta(0) = 1$. We assume each individual is alive during $\mathcal{L}+1$ units of time. The mean $a(s) = \mathbb{E}(\zeta(s))$ of the total number of individuals born at time s satisfies a similar evolution equation to (12), which is

$$a(s) = \mathbb{P}(\mathcal{L} \ge s) + \sum_{d\in\mathcal{D}} a(s-d)\rho_d, \quad s \ge 0.$$

The 1-type process shares the same Malthusian parameter θ as $\mathcal{Z}$ and thus the two processes are either subcritical, critical or supercritical together. Recall the probability vector $\vec{\beta} = (\beta_d = \rho_d e^{-\theta d} : d \in \mathcal{D})$ and Consider the sequence $(c(s) = a(s)e^{-\theta s} : s \ge 0)$. This satisfies the renewal equation

$$c(s) = e^{-\theta s}\mathbb{P}(\mathcal{L} \ge s) + \sum_{k\ge 0} c(s-k)\beta_k,\ s \ge 0.$$

The solution to this equation has the form

$$c(s) = \sum_{b=0}^{s} e^{-\theta(s-b)}\mathbb{P}(\mathcal{L} \ge s-b)u(b),\ s \ge 0,$$

where $\sum_{t=1}^{s} u(t)$ is the renewal function associated with $\vec{\beta}$ and $u(t)$ is a discrete analogue to the renewal density. On one hand, if $\theta = 0$, then $\sum_{s\ge 0}\mathbb{P}(\mathcal{L} \ge s) = \mathbb{E}(\mathcal{L}) < \infty$ by assumption. On the other hand when $\theta \ne 0$, one has

$$\sum_{s\ge 0}\mathbb{P}(\mathcal{L} \ge s)e^{-\theta s} = \frac{e^{-\theta}\mathbb{E}(e^{-\theta\mathcal{L}}) - 1}{e^{-\theta} - 1}.$$

In the supercritical case this is finite because $\theta > 0$ and in the subcritical case this is also finite since $\mathbb{E}(e^{-\theta\mathcal{L}})$ is assumed to be finite. Thus, $\sum_{s\ge 0}\mathbb{P}(\mathcal{L} \ge s)e^{-\theta s} < \infty$ and this makes $\mathbb{P}(\mathcal{L} \ge s)e^{-\theta s}$ directly Riemann integrable. From the renewal theorem, see Proposition 4.7 in Chapter V of [1], we then get

$$\lim_{s\to\infty} a(s)e^{-\theta s} = \frac{1}{\mu(\vec{\beta})}\sum_{t\ge 0}\mathbb{P}(\mathcal{L} \ge t)e^{-\theta t}. \tag{30}$$

Recall that $\mu(\vec{\beta}) = \sum_{d\in\mathcal{D}} d\,\beta_d$ is the mean value of the distribution $\vec{\beta}$. Lemma 1 of [8] showed this for the non-arithmetic case in a general framework. Also see Proposition 1.1 in [15].

Now we can finish the proof of Proposition 4. Take (15) and write it in the form

$$\begin{aligned} \mathbb{E}(\mathcal{Z}(s))' = \mathbb{E}(\mathcal{Z}(0))'\mathbb{P}(\mathcal{L} \geq s) + \mathbb{E}(\mathcal{Z}(0))' \sum_{b=0}^{s-1} \mathbb{P}(\mathcal{L} \geq b) & \left(\sum_{\lceil (s-b)/D \rceil \leq r \leq s-b} \Xi(s-b, r) \right) \\ \text{where } \Xi(s,r) = & \sum_{(d_1,\ldots,d_r)\in\Gamma(s,r)} \left(\prod_{l=1}^{r} M_{d_l} \right). \end{aligned} \tag{31}$$

Then, we will group the paths in $\Gamma(s,r)$ according to the tuples $\vec{k} = (k_d : d \in \mathcal{D})$ where $k_d = |\{l : d_l = d\}|$ is the number of steps of size d in the path. Observe that $\vec{k}$ must satisfy the constraints

$$\vec{k} \in \Lambda(s,r) = \left\{ \vec{k} \in \mathbb{N}_0^{\mathcal{D}} : |\vec{k}| = r, \mathfrak{m}(\vec{k}) = s \right\}. \tag{32}$$

This is the set of all possible tuples of paths of length r from s to 0. Expression (31) also shows that the mean of the associated 1-type delayed $\mathfrak{bp}$ starting from $\zeta(0) = 1$ is

$$\mathbb{E}(\zeta(s)) = \mathbb{P}(\mathcal{L} \geq s) + \sum_{b=0}^{s-1} \mathbb{P}(\mathcal{L} \geq b) \sum_{1\leq r\leq s-b} \sum_{\vec{k}\in\Lambda(s-b,r)} \frac{r!}{\prod_{d\in\mathcal{D}} k_d!} \prod_{d\in\mathcal{D}} \rho_d^{k_d}.$$

Then, by using (27) one obtains

$$\begin{aligned} e^{-\theta s}\mathbb{E}(\zeta(s)) &= \mathbb{P}(\mathcal{L} \geq s)e^{-\theta s} + e^{-\theta s} \sum_{b=0}^{s-1} \mathbb{P}(\mathcal{L} \geq b) \left(\sum_{1\leq r\leq s-b} \sum_{\vec{k}\in\Lambda(s-b,r)} |S(\vec{k})| \prod_{d\in\mathcal{D}} \rho_d^{k_d} \right) \\ &= \sum_{b=0}^{s} \mathbb{P}(\mathcal{L} \geq b) e^{-\theta b} u(s-b) \end{aligned}$$

where $u(0) = 1$ and for $t > 0$,

$$u(t) = \left(\sum_{1\leq r\leq t} \sum_{\vec{k}\in\Lambda(t,r)} |S(\vec{k})| \prod_{d\in\mathcal{D}} \rho_d^{k_d} \right) e^{-\theta t} = \sum_{1\leq r\leq t} \sum_{\vec{k}\in\Lambda(t,r)} |S(\vec{k})| \prod_{d\in\mathcal{D}} \beta_d^{k_d},$$

which is the renewal density for $(\beta_d : d \in \mathcal{D})$. From (30) we have

$$\begin{aligned} \lim_{s\to\infty} e^{-\theta s}\mathbb{E}(\zeta(s)) &= \lim_{s\to\infty} e^{-\theta s} \sum_{b=0}^{s} \mathbb{P}(\mathcal{L} \geq b) \left(\sum_{1\leq r\leq s-b} \sum_{\vec{k}\in\Lambda(s-b,r)} |S(\vec{k})| \prod_{d\in\mathcal{D}} \rho_d^{k_d} \right) \\ &= \mu(\vec{\beta})^{-1} \sum_{t\geq 0} \mathbb{P}(\mathcal{L} \geq t) e^{-\theta t}. \end{aligned} \tag{33}$$

On the other hand, notice that from (31) one has

$$\begin{aligned} \Xi(s-b,r) &= \sum_{(d_1,\ldots,d_r)\in\Gamma(s-b,r)} \left(\prod_{l=1}^{r} M_{d_l} \right) \\ &= \sum_{\vec{k}\in\Lambda(s-b,r)} \prod_{d\in\mathcal{D}} \rho_d^{k_d} \sum_{(d_1,\ldots,d_r)\in S(\vec{k})} \prod_{l=1}^{r} \rho_{d_l}^{-1} M_{d_l}. \end{aligned}$$

We shall assume for the moment that $\mathcal{L}$ is bounded by L, so $\mathbb{P}(\mathcal{L} \geq t) = 0$ for $t > L$. From (24) it holds that for all $k \geq \kappa(\epsilon)$ one has $\|\rho_d^{-k} M_d^k - h\nu'\| \leq \epsilon$ for all $d \in \mathcal{D}$. On the other hand, from Lemma 8 there exists $s(\epsilon)$ such that for all $s \geq s(\epsilon) + L$ one has that every tuple $\vec{k} \in \Lambda(s-b,r)$ with $b \leq L$, satisfies

$|S(\vec{k})^{-\kappa(\epsilon)}|/|S(\vec{k})| \leq \epsilon$. Since for every sequence $(d_1,\ldots,d_r) \in S(\vec{k}) \setminus S(\vec{k})^{-\kappa(\epsilon)}$ there exists a $\kappa(\epsilon)$-run, relation (23) in Lemma 7 guarantees that

$$\|\prod_{l=1}^{r} \rho_{d_l}^{-1} M_{d_l} - h\nu'\| \leq \epsilon.$$

Hence, for all $s \geq s(\epsilon) + L$,

$$\|\Xi(s-b,r) - \left(\sum_{\vec{k}\in\Lambda(s-b,r)} |S(\vec{k})| \prod_{d\in\mathcal{D}} \rho_d^{k_d}\right) h\nu'\| \leq \left(\sum_{\vec{k}\in\Lambda(s-b,r)} |S(\vec{k}) \setminus S(\vec{k})^{-\kappa(\epsilon)}| \prod_{d\in\mathcal{D}} \rho_d^{k_d}\right) \epsilon\|h\nu'\|$$
$$+ \sum_{\vec{k}\in\Lambda(s-b,r)} \prod_{d\in\mathcal{D}} \rho_d^{k_d} \sum_{(d_1,\ldots,d_r)\in S(\vec{k})^{-\kappa(\epsilon)}} \|\prod_{l=1}^{r} \rho_{d_l}^{-1} M_{d_l}\|.$$

From (22) one has $\|\prod_{l=1}^{r} \rho_{d_l}^{-1} M_{d_l}\| \leq \mathfrak{w}(h)$. On the other hand $|S(\vec{k})^{-\kappa(\epsilon)}| \leq \epsilon |S(\vec{k})|$. Hence, for $C = \|h\nu'\| + \mathfrak{w}(h)$ we have

$$\|\Xi(s-b,r) - \left(\sum_{\vec{k}\in\Lambda(s-b,r)} |S(\vec{k})| \prod_{d\in\mathcal{D}} \rho_d^{k_d}\right) h\nu'\| \leq C\epsilon \left(\sum_{\vec{k}\in\Lambda(s-b,r)} |S(\vec{k})| \prod_{d\in\mathcal{D}} \rho_d^{k_d}\right).$$

Therefore,

$$\sum_{b=0}^{L} \mathbb{P}(\mathcal{L} \geq b) \sum_{\lceil (s-b)/D\rceil \leq r \leq s-b} ||\Xi(s-b,r) - \sum_{\vec{k}\in\Lambda(s-b,r)} |S(\vec{k})| \prod_{d\in\mathcal{D}} \rho_d^{k_d} h\nu'||$$
$$\leq C\epsilon \left(\sum_{b=0}^{L} \mathbb{P}(\mathcal{L} \geq b) \sum_{\lceil (s-b)/D\rceil \leq r \leq s-b} \sum_{\vec{k}\in\Lambda(s-b,r)} |S(\vec{k})| \prod_{d\in\mathcal{D}} \rho_d^{k_d}\right).$$

So, from (33) we get

$$\lim_{s\to\infty} e^{-\theta s} \sum_{b=0}^{L} \mathbb{P}(\mathcal{L} \geq b) \sum_{\lceil (s-b)/D\rceil \leq r \leq s-b} \|\Xi(s-b,r) - \left(\sum_{\vec{k}\in\Lambda(s-b,r)} |S(\vec{k})| \prod_{d\in\mathcal{D}} \rho_d^{k_d}\right) h\nu'\|$$
$$\leq C\epsilon\mu(\vec{\beta})^{-1} \sum_{b=0}^{L} \mathbb{P}(\mathcal{L} \geq b) e^{-\theta b}.$$

Letting $\epsilon \to 0$ and using (33) again shows that

$$\lim_{s\to\infty} e^{-\theta s} \mathbb{E}(\mathcal{Z}_j(s)) = \mathbb{E}(\mathcal{Z}(0))' h\nu' e_j \mu(\vec{\beta})^{-1} \left(\sum_{b=0}^{L} \mathbb{P}(\mathcal{L} \geq b) e^{-\theta b}\right). \tag{34}$$

So, we have proved (17) in the case of bounded lifetimes ($\mathcal{L} \leq L$).

Next, we shall extend this to the general case. Take $\mathcal{L}^n$ to be an increasing sequence of bounded lifetimes converging to $\mathcal{L}$. Then, for all $b \geq 0$, $\mathbb{P}(\mathcal{L}^n \geq b)$ increases to $\mathbb{P}(\mathcal{L} \geq b)$. Let us put a superscript n on all quantities in which the lifetime is $\mathcal{L}^n$. We have

$$\begin{aligned}\mathbb{E}(\zeta(s)) - \mathbb{E}(\zeta^n(s)) &= (\mathbb{P}(\mathcal{L} \geq s) - \mathbb{P}(\mathcal{L}^n \geq s)) \\ &\quad + \sum_{b=0}^{s-1} (\mathbb{P}(\mathcal{L} \geq b) - \mathbb{P}(\mathcal{L}^n \geq b)) \left(\sum_{\lceil (s-b)/D\rceil \leq r \leq s-b} \sum_{\vec{k}\in\Lambda(s-b,r)} |S(\vec{k})| \prod_{d\in\mathcal{D}} \rho_d^{k_d}\right).\end{aligned} \tag{35}$$

Observe that the Malthusian parameter θ does not depend on $\mathcal{L}$. By using (33)once again and by taking advantage of the fact that $\mu(\vec{\beta})^{-1}\sum_{t\geq 0}\mathbb{P}(\mathcal{L}^n \geq t)e^{-\theta t} \nearrow \mu(\vec{\beta})^{-1}\sum_{t\geq 0}\mathbb{P}(\mathcal{L} \geq t)e^{-\theta t}$ as $n\to\infty$, we obtain

$$\lim_{n\to\infty}\lim_{s\to\infty} e^{-\theta s}\left(\mathbb{E}(\zeta(s)) - \mathbb{E}(\zeta^n(s))\right) = 0. \tag{36}$$

On the other hand

$$\begin{aligned}&\mathbb{E}(\mathcal{Z}(s))' - \mathbb{E}(\mathcal{Z}^n(s))'\\ &= \left(\mathbb{P}(\mathcal{L}\geq s) - \mathbb{P}(\mathcal{L}^n \geq s)\right) + \sum_{b=0}^{s-1}\left(\mathbb{P}(\mathcal{L}\geq b) - \mathbb{P}(\mathcal{L}^n\geq b)\right)\left(\sum_{\lceil (s-b)/D\rceil \leq r\leq s-b}\Xi(s-b,r)\right),\end{aligned}$$

which is non-negative componentwise. Consequently, we have

$$\begin{aligned}&e^{-\theta s}\left(\mathbb{E}(\mathcal{Z}(s))' - \mathbb{E}(\mathcal{Z}^n(s))'\right)\\ &\quad= \quad e^{-\theta s}\left(\mathbb{P}(\mathcal{L}\geq s) - \mathbb{P}(\mathcal{L}^n\geq s)\right)\mathbb{E}(\mathcal{Z}(0))'\\ &\qquad+\sum_{b=0}^{s-1}\left(\mathbb{P}(\mathcal{L}\geq b) - \mathbb{P}(\mathcal{L}^n\geq b)\right)\sum_{\lceil (s-b)/D\rceil\leq r\leq s-b}\ \sum_{\vec{k}\in\Lambda(s-b,r)}\prod_{d\in\mathcal{D}}\rho_d^{k_d}\left(\sum_{(d_1,\ldots,d_r)\in S(\vec{k})}\prod_{l=1}^{r}\rho_{d_l}^{-1}M_{d_l}\right).\end{aligned}$$

Then,

$$\begin{aligned}&e^{-\theta s}||\mathbb{E}(\mathcal{Z}(s))' - \mathbb{E}(\mathcal{Z}^n(s))'||\\ &\quad\leq\quad e^{-\theta s}\left(\mathbb{P}(\mathcal{L}\geq s) - \mathbb{P}(\mathcal{L}^n\geq s)\right)||\mathbb{E}(\mathcal{Z}(0))'||\\ &\qquad+\mathfrak{w}(h)e^{-\theta s}\left(\sum_{b=0}^{s-1}\left(\mathbb{P}(\mathcal{L}\geq s) - \mathbb{P}(\mathcal{L}^n\geq s)\right)\sum_{\lceil (s-b)/D\rceil\leq r\leq s-b}\ \sum_{\vec{k}\in\Lambda(s-b,r)}|S(\vec{k})|\prod_{d\in\mathcal{D}}\rho_d^{k_d}\right).\end{aligned}$$

From (35) and (36), it follows that

$$\lim_{n\to\infty}\lim_{s\to\infty} e^{-\theta s}\|\mathbb{E}(\mathcal{Z}(s))' - \mathbb{E}(\mathcal{Z}^n(s))'\| = 0$$

and hence

$$\lim_{s\to\infty} e^{-\theta s}\mathbb{E}(\mathcal{Z}(s))' = \lim_{n\to\infty}\lim_{s\to\infty} e^{-\theta s}\mathbb{E}(\mathcal{Z}^n(s))'.$$

At this point, an application of (34) completes the proof of (17).

Next, since $\nu'\mathbf{1} = 1$, the relation (18) follows straightforwardly from (17). We have

$$\lim_{s\to\infty}\frac{\mathbb{E}(\mathcal{Z}(s))'}{\mathbb{E}(\mathcal{Z}(s))'\mathbf{1}} = \frac{\mu(\vec{\beta})^{-1}(\mathbb{E}(\mathcal{Z}(0))'h)\left(\sum_{t\geq 0}\mathbb{P}(\mathcal{L}\geq t)e^{-\theta t}\right)\nu'}{\mu(\vec{\beta})^{-1}(\mathbb{E}(\mathcal{Z}(0))'h)\left(\sum_{t\geq 0}\mathbb{P}(\mathcal{L}\geq t)e^{-\theta t}\right)\nu'\mathbf{1}} = \nu'.$$

Now, the difference between $\mathcal{X}$ and $\mathcal{Z}$ is that $\mathcal{X}$ uses a degenerate lifetime $\mathcal{L}$ satisfying $\mathbb{P}(\mathcal{L}\geq s) = \mathbf{1}(s=0)$. Thus (19) easily follows from (17). Finally, one can write $\mathcal{E}_d^j(s)$ as

$$\mathcal{E}_d^j(s) = \frac{e^{-\theta(s-d)}\mathbb{E}(\mathcal{X}_j(s-d))\mathbb{P}(\mathcal{L}\geq d)e^{-\theta d}}{\sum_{b'\in\mathcal{D}'} e^{-\theta(s-d')}\mathbb{E}(\mathcal{X}_j(s-b'))\mathbb{P}(\mathcal{L}\geq b')e^{-\theta d'}}. \tag{37}$$

As $e^{-\theta(s-d)}\mathbb{E}(\mathcal{X}_j(s-d))$ tends to the same limit for all $d\in\mathcal{D}$ and $j\in I$, taking limits as $s\to\infty$ in (37) yields (20). $\square$

8 Fibonacci sequences and the Malthusian parameter

Finally, we will examine the relationship of the delayed process with Fibonacci sequences as well as a Fibonacci-type process. The final subsection is devoted to providing some insight into the Malthusian parameter in the finite delay setting ($|\mathcal{D}| < \infty$) we have treated here. We will assume that we are in the supercritical case.

8.1 *Relationship to Fibonacci sequences*

Consider the special 1-type supercritical case in which $\rho_d = 1$ for $d \in \mathcal{D}$. From now on assume $\mathcal{D} = \{1, \ldots, D\}$ for a $D > 1$. If we were to take $\mathcal{D}$ to be any subset of $\{1, \ldots, D\}$ as has been done up to this point, we would fail to make the connection with Fibonacci sequences.

Recall that the number of paths of length r from s to 0 with elements in $\mathcal{D}$ is $|\Gamma(s, r)|$, see (13). From (14) $\mathbb{E}(\mathcal{X}(s))' = \mathbb{E}(\mathcal{X}(0))'\Gamma(s)$, where

$$\Gamma(s) = \sum_{\lceil s/D \rceil \le r \le s} |\Gamma(s, r)|$$

is the total number of paths from s to 0. It is straightforward to check that $\Gamma(s)$ satisfies the following properties:

$$\Gamma(s) = \sum_{d=1}^{D} \Gamma(s - d) \text{ with initial conditions } \Gamma(0) = 1 \text{ and } \Gamma(s) = 0 \text{ for } s < 0.$$

Therefore, $(\Gamma(s) : s \ge 1)$ is the D-Fibonacci sequence. By definition it is increasing in s and it may be expressed analytically in the form

$$\Gamma(s) = \sum_{d=1}^{D} c_d \phi_d^s \text{ with } (c_d : d \in \mathcal{D}) \text{ real constants and}$$

$$(\phi_d : d \in \mathcal{D}) \text{ the roots of } g(y) = 0, \text{ where } g(y) = y^D - \sum_{d=0}^{D-1} y^d. \tag{38}$$

Note that $y = 1$ is not a root of $g(y) = 0$ and since $g(y) = (-y^{D+1} + 2y^D - 1)/(1 - y)$ for $y \neq 1$, it follows that

$$g(y) = 0 \Leftrightarrow \left[y^{D+1} - 2y^D + 1 = 0,\ y \neq 1\right]. \tag{39}$$

By Lemma 3.6 of [17], $g(s)$ has a unique real root in the interval $(1, 2)$, say Φ_D, which tends to 2 as $D \to \infty$. All the other roots have moduli in $(3^{-D}, 1)$ and when D is even there is a second real root in $(-1, 0)$. The root Φ_D corresponds to the D-Fibonacci constant. The 2-Fibonacci constant is the golden ratio $\Phi_2 = (1 + \sqrt{5})/2$ and Theorem 3.9 of [17] provides a series representation of Φ_D. The D-Fibonacci constant dominates $\Gamma(s)$ as s becomes large, which gives rise to the following limit behavior:

$$\Phi_D = \lim_{s \to \infty} \frac{\Gamma(s + 1)}{\Gamma(s)} \text{ and } \lim_{s \to \infty} \frac{\Gamma(s)}{\Phi_D^s} = C' \text{ a constant.}$$

See Corollary 3.7 in [17].

In the current set-up, the relation (16) defining the Malthusian parameter becomes $\sum_{d=1}^{D} e^{-d\theta} = 1$, so $\theta > 0$. Let $\eta = e^{-\theta}$. From (38) it is evident that $\eta^{-1} = \Phi_D$, because the D-Fibonacci constant is the unique real root larger than 1. We have

$$\mu(\vec{\beta})^{-1} = \sum_{d=1}^{D} d\, \eta^d = \eta \left(\frac{d}{d\eta} \sum_{d=1}^{D} \eta^d \right) = \frac{\eta}{(1 - \eta)^2} \left((D + 1)(1 - \eta^D) - D + D\eta^{D+1} \right) = \frac{D + 1 - 2D\eta}{1 - \eta},$$

where relation (39) was used to get $1 - 2\eta + \eta^{D+1} = 0$, which is equivalent to $\eta(1 - \eta^D) = 1 - \eta$ for $\eta \in (0, 1)$. Finally, we obtain $\mu(\vec{\beta})^{-1} = \big(\Phi_D - 1\big) / \big((D + 1)\Phi_D - 2D\big)$.

Now, we compare the 1-type delayed $\mathfrak{bp}$ with an active period of 2 time units to the process introduced by Heyde in [12] called a Fibonacci or lagged Bienaymé-Galton-Watson) $\mathfrak{bp}$ in the singular case where the offspring distribution is degenerate and produces exactly one individual. The lagged Fibonacci process, denoted by $(W(s) : s \geq 0)$ starts with initial conditions $W(0)$ and $W(1)$. For $s \geq 2$, it is defined by

$$W(s) - W(s-1) = \sum_{l=1}^{W(s-2)} \xi_{s,l},$$

where the random variables $(\xi_{s,l} : l \geq 1, s \geq 2)$ are i.i.d. with common distribution $p = (p(n) : n \geq \mathbb{N}_0)$ having mean $m = \sum_{n\geq 1} n\, p(n)$. We take the initial conditions to be $W(0) = 1$ and $W(1) \sim p$.

On the other hand, a delayed 1-type $\mathfrak{bp}$ $\mathcal{X}$ with $\mathcal{D} = \{1, 2\}$ and offspring laws $p_1 = p_2 = p$ where each individual is only alive for one unit of time satisfies the relation (6) which becomes

$$\mathcal{X}(s) = \sum_{l=1}^{\mathcal{X}(s-1)} \xi_{s-1,l;1} + \sum_{l=1}^{\mathcal{X}(s-2)} \xi_{s-2,l;2}.$$

Then, when $p(1) = 1$ or equivalently the variables are deterministic with $\xi \equiv 1$, then both processes W and $\mathcal{X}$ coincide and satisfy the 2-Fibonacci evolution $\Psi(s) = \Psi(s-1) + \Psi(s-2)$ with $\Psi(0) = 1 = \Psi(1)$. A similar comparison can be made between the means of the two processes when the mean of the offspring distribution is $m = 1$. In this case, $\mathbb{E}(W(s)) = \mathbb{E}(\mathcal{X}(s)) = \Psi(s)$ for $s \geq 0$.

However, for other cases the processes W and $\mathcal{X}$ behave differently from each other. In fact for $s > 0$ one has the following comparison between mean values: if $m > 1$ then $\mathbb{E}(\mathcal{X}(s)) > \mathbb{E}(W(s))$ and if $m < 1$ then $\mathbb{E}(\mathcal{X}(s)) < \mathbb{E}(W(s))$.

8.2 The Malthusian parameter for finite delay processes

In our case, where the delay (active period) is a finite deterministic set $\mathcal{D}$, we can gain some further insight into the Malthusian parameter θ which solves (16) when we are in the supercritical case $\theta > 0$. Toward this end, it is useful to introduce some quantities related to a probability vector $\vec{b} = (b(d) : d \in \mathcal{D})$, namely,

$$H(\vec{b}) = -\sum_{d\in\mathcal{D}} b_d \log b_d \text{ (the entropy)}\ ,\ \ \mu(\vec{b}) = \sum_{d\in\mathcal{D}} d\, b_d \text{ (the mean)}\ ,\ \ \overline{\log\rho}(\vec{b}) = \sum_{d\in\mathcal{D}} b_d \log \rho_d.$$

For the probability vector $\vec{\beta} = (\beta_d = \rho_d\, e^{-\theta d} : d \in \mathcal{D})$ defined by (16), we have
$H(\vec{\beta}) = \sum_{d\in\mathcal{D}} \rho_d e^{-\theta d}\, \theta\, d - \sum_{d\in\mathcal{D}} \rho_d e^{-\theta d} \log \rho_d = \theta\mu(\vec{\beta}) - \overline{\log\rho}(\vec{\beta})$ and hence

$$\theta = (H(\vec{\beta}) + \overline{\log\rho}(\vec{\beta}))/\mu(\vec{\beta}). \tag{40}$$

A simple argument that serves to explain relation (40) for the Malthusian parameter is as follows. Define $\alpha = r/s$ and $b_d = k_d/r = k_d/(\alpha s)$ for $d \in \mathcal{D}$. From (32), α and $\vec{b} = (b_d : d \in \mathcal{D})$ must satisfy $\sum_{d\in\mathcal{D}} b_d = 1$ and $\sum_{d\in\mathcal{D}} d\, b_d = \alpha^{-1}$. In the analysis of

$$\mathbb{E}(\zeta(s)) = \sum_{1\leq r\leq s} \sum_{\vec{k}\in\Lambda(s,r)} \frac{(\alpha\, s)!}{\prod_{d\in\mathcal{D}} (\alpha b_d\, s)!} \prod_{d\in\mathcal{D}} \rho_d^{\alpha b_d\, s}$$

one can apply Stirling's formula $l! = \sqrt{2\pi} l^{l+1/2} e^{-\kappa}$ for large l and after some further computation, one finds that the exponential term in the resulting expression is $\exp\left(\alpha\,(H(\vec{b}) + \overline{\log\rho}(\vec{b}))\, s\right)$. Thu, the overall growth rate of $\mathbb{E}_{i_0}(\zeta(s))$ as s increases is determined by the term satisfying

$$\alpha^*(H(\vec{b}^*) + \overline{\log\rho}(b^*)) = \max\{\alpha(H(\vec{b}) + \overline{\log\rho}(\vec{b})) : \sum_{d\in\mathcal{D}} b_d = 1,\ \alpha \sum_{d\in\mathcal{D}} d\, b_d = 1\}\,.$$

The Lagrangian is $\mathfrak{L} = \alpha(H(\vec{b}) + \overline{\log\rho}(\vec{b})) - \lambda(\sum_{d\in\mathcal{D}} b_d - 1) - a(\alpha\mu(\vec{b}) - 1)$. Setting $\partial\mathfrak{L}/\partial\alpha = 0$ and letting $\theta = a^*$ denote the value of a given by the optimization of $\mathfrak{L}$, yields

$$(H(\vec{b}^*) + \overline{\log\rho}(\vec{b}^*)) - \theta\mu(\vec{b}^*) = 0, \text{ that is, } \theta = \left(H(\vec{b}^*) + \overline{\log\rho}(\vec{b}^*)\right)/\mu(\vec{b}^*). \tag{41}$$

The solution to $\partial\mathfrak{L}/\partial b_d = 0$ is $-\alpha^*(\log b_d^* + 1 - \log\rho_d) - \lambda - \theta\alpha^* d = 0$. Hence,

$$b_d^* = c^{-1}\rho_d\,\eta^d \text{ where } c = e^{\frac{\lambda}{\alpha^*}+1} \text{ and } \log\eta = -\theta. \tag{42}$$

From the constraint $\sum_{d\in\mathcal{D}} b_d^* = 1$ we have $c = \sum_{d\in\mathcal{D}} \rho_d\,\eta^d\, d$. Computing the entropy of $\vec{b}^*$, we have

$$H(\vec{b}^*) = -\sum_{d\in\mathcal{D}} c^{-1}\rho_d\,\eta^d \log(c^{-1}\rho_d\,\eta^d) = -\log c - \sum_{d\in\mathcal{D}} c^{-1}\rho_d\,\eta^d\log\rho_d - \left(\sum_{d\in\mathcal{D}} c^{-1} d\,\rho_d\,\eta^d\right)\log\eta.$$

Then, with the aid of (41) and (42) one can show that

$$H(\vec{b}^*) = -\log c - \overline{\log\rho}(\vec{b}^*) - \mu(\vec{b}^*)\log\eta = -\log c - \overline{\log\rho}(\vec{b}^*) + H(\vec{b}^*) + \overline{\log\rho}(\vec{b}^*) = -\log c + H(\vec{b}^*),$$

which can hold if and only if $c = 1$. Therefore, $\sum_{d\in\mathcal{D}} \rho_d\,\eta^d = 1$, or equivalently $\sum_{d\in\mathcal{D}} \rho_d\, e^{-\theta d} = 1$. So θ is the Malthusian parameter and $\vec{b}^* = \vec{\beta}$.

Now, $\vec{\beta} = (\beta_d = \eta^d : d \in \mathcal{D})$ is such that η^d is the asymptotic proportion of steps of length d that compose a path from a large time s to 0. Therefore, even if each individual were to generate offspring with the same probability law at each possible offset $d \in \mathcal{D}$, when one counts all the individuals in a single trajectory and one samples the offsets into the active period at which they were generated, then the offsets are not distributed uniformly but rather according to the truncated geometric distribution $\vec{\beta}$. In other words, if we consider the lineage of an individual born at time s, $(\eta^d : d \in \mathcal{D})$ is the asymptotic parent age distribution of the individuals in the lineage, which is truncated geometric irrespective of the offspring distributions.

Acknowledgments

This work was supported by the Center for Mathematical Modeling ANID Basal PIA program AFB 170001.

References

[1] Asmussen, S. *Applied Probability and Queues*; Stochastic Modelling and Applied Probability, Second Edition; Springer: New York, 2010.

[2] Asmussen, S. On the role of certain eigenvalue in estimating the growth rate of a branching process. *Austral. J. Statist.* 1982, 24(2), 151–159.

[3] Athreya, Krishna B.; Ney, Peter N. *Branching Processes*; Springer-Verlag: New York, 1972.

[4] Becker, N. Estimation for Discrete Time Branching Processes with Application to Epidemics. *Biometrics* 1977, 33(3), 515–522.

[5] Biggins, J.D.; Cohn, H.; Nerman, O. Multi-type Branching in varying environment. *Stoch. Process. Applic.* 1999, 83(2), 357–400.

[6] Biggins, J.D.; Kyprianou, A.E. Measure change in multi-type branching. *Adv. Appl. Probab.* 2004, 36(2), 544–581.

[7] Doney, R.A. A limit theorem for a class of supercritical processes. *J. Appl. Probab.* 1972, 9(4), 707–724.

[8] Doney, R.A. On single- and multi-type general age-dependent branching processes. *J. Appl. Probab.* 1976, 13(2), 239–246.

[9] Ferguson, N.M.; Laydon, D.; Nedjati-Gilani, G.; Imai, N.; Ainslie, K.; Baguelin, M.; Bhatia, S.; Boonyasiri, A.; Cucunubá, Z.; Cuomo-Dannenburg, G.; Dighe, A. *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand*. Imperial College COVID-19 Response Team, London, 16 March 2020. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-Colleg

[10] Harris, T. *The theory of branching processes*; Springer-Verlag: Berlin, 1963.

[11] Heyde, C.C. Extensions of a result of Seneta for the super-critical Galton-Watson process. *Ann. Math.Statist.* 1970, 19, 474–494.

[12] Heyde, C.C. On Fibonacci (or lagged Bienaymé-Galton-Watson) branching processes. *J. Appl. Probab.* 1981, 18, 583–591.

[13] Joffe, A.; Spitzer, F. On multitype Markov branching processes with $\rho \leq 1$. *J. Math. Anal. Applic.* 1967, 19, 409–430.

[14] Kurtz, T.G.; Lyons, R.; Pemantle, R.; Peres, Y. A conceptual proof of the Kesten-Stigum theorem for multi-type branching processes. *Classical and Modern Branching Processes*; Springer: New York, 1997; 181–185.

[15] Nerman, O. On the convergence of supercritical general (C-M-J) branching processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 1981, 57, 365–395.

[16] Seneta, E. *Nonnegative Matrices and Markov Chains*; Springer Series in Statistics; Springer-Verlag: New York, 1981.

[17] Wolfram, D.A. Solving Generalized Fibonacci Recurrences. *Fibonacci Quart.* 1998, 36(2), 129–145.