Applied Probability Trust (24 August 2022)

ASYMPTOTIC PERSISTENCE TIME FORMULAE FOR MULTI-TYPE BIRTH-DEATH PROCESSES

FRANK G BALL,* *University of Nottingham*

DAMIAN CLANCY,** *Heriot-Watt University*

Abstract

We consider a class of processes describing a population consisting of k types of individuals. The process is almost surely absorbed at the origin within finite time, and we study the expected time taken for such extinction to occur. We derive simple and precise asymptotic estimates for this expected persistence time, starting either from a single individual or from a quasi-equilibrium state, in the limit as a system size parameter N tends to infinity. Our process need not be a Markov process on $\mathbb{Z}_+^k$; we allow the possibility that individuals' lifetimes may follow more general distributions than the exponential distribution.

Keywords: Large deviations; population processes; stochastic epidemic models

2020 Mathematics Subject Classification: Primary 60J27
Secondary 92D30;92D40

1. Introduction

A fundamental issue in modelling biological populations is the risk of a population or species becoming extinct. The corresponding issue in modelling infectious spread is extinction of infection from a population. In either case, a random variable of particular

---

* Postal address: School of Mathematical Sciences, University of Nottingham, Nottingham, NG7 2RD, UK

* Email address: frank.ball@nottingham.ac.uk

** Postal address: Department of Actuarial Mathematics and Statistics, Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, UK

** Email address: d.clancy@hw.ac.uk

interest is the persistence time until extinction occurs.

A substantial body of recent research addresses the issue via the semi-rigorous WKB (Wentzel, Kramers, Brillouin) approach, see review papers [20, 2]. Such work focuses specifically on the supercitical case (defined precisely in section 2 below), so that the expected persistence time grows exponentially with population size. For many naturally one-dimensional models (processes that can be approximated, in a sense made precise in section 2.3, by the solution of a one-dimensional ordinary differential equation), it can be shown that in the supercritical case, as a system size parameter N tends to infinity, the expected persistence time τ starting from quasi-equilibrium satisfies

$$\tau \quad \sim \quad \frac{K}{\sqrt{N}} \exp(AN), \tag{1}$$

where we use $\sim$ to denote that the ratio of the two sides converges to 1 as $N \to \infty$. Here A and K are constants whose values do not depend upon N, with explicit formulae available for A and K in terms of parameters of the process.

For multidimensional models, it is usually only possible to establish results of the cruder form $\ln \tau \sim AN$, and to evaluate the leading-order constant A via numerical solution of a system of ordinary differential equations. An exception is provided by [9], where a relation of the form (1) and explicit formulae for A and K are obtained using the WKB approach for a susceptible-infectious-susceptible (SIS) infection model in a heterogeneous population, a naturally multidimensional process.

In a different vein, a rather more rigorous approach is taken in [7]. For a general class of multitype birth-death processes, the existence of a quasi-stationary distribution is proved and a bound established for the total variation distance between the process conditioned to non-extinction before time t and the quasi-stationary distribution, As a by-product of this analysis, it is shown that there exist constants $d_1 > d_2 > 0$ such that for all sufficiently large N, the mean persistence time τ starting from quasi-stationarity satisfies $\exp(d_2 N) \leq \tau \leq \exp(d_1 N)$.

A limitation of both the WKB approach and the approach of [7] is that the models considered generally assume that individuals' lifetimes or infectious periods are exponentially distributed. This is not usually biologically realistic, the assumption being rather for reasons of mathematical tractability. In [5] the effects of relaxing this assumption were considered for a class of (one-dimensional) birth-death processes, via application of a network insensitivity result of [23]. The subcritical and critical cases were considered in addition to the more widely-studied supercritical case. In the critical case it was found that the expected persistence time τ_0, starting from one initial individual, depends logarithmically upon the system size parameter N as $N \to \infty$, whereas in the subcritical case, τ_0 was shown to converge to a finite limit as $N \to \infty$. In [9], the multitype version of the insensitivity result of [23] was applied to allow for non-exponentially distributed infectious periods in the SIS infection model in a heterogeneous population, in the supercritical case.

In the current paper, we consider a particular class of multitype birth-death processes, defined in Section 2. We build on the approach of [5], so we do not need to assume exponentially distributed lifetimes. For this class of processes we are able to obtain unusually precise results, including a result of the form (1) in the supercritical case, with explicit formulae for the constants A and K. The only multidimensional process for which such a precise result has previously been available is the heterogeneous population SIS model of [9]. Not only is the class of processes considered here considerably more general than in [9], but the methods of proof are quite different, and in contrast to [9] we consider the subcritical and critical cases as well as the supercritical case. The class of processes is more restricted than that of [7] but unlike in [7] our results do not require exponentially distributed lifetimes. In estimating the mean persistence time starting from a single individual, our approach is fully rigorous, unlike the WKB method of [9]. In estimating the mean persistence time starting from quasi-stationarity, we present only a heuristic sketch proof, referring the reader to [4] for full details of the proof in a specific one-dimensional setting.

The remainder of the paper is structured as follows. In section 2, we define our general modelling framework and give two illustrative examples, before describing two standard approximating processes that provide motivation and intuition for what follows. In section 3, we define a re-started version of our general process and show how this re-started process may be used to analyse extinction times for the original process of interest. In sections 4 and 5, we consider the subcritical and critical cases, respectively, obtaining limiting results (theorems 1, 2 and 3) for the expected extinction time starting from a single initial individual. We move on in section 6 to the supercritical case, for which we again obtaining a limiting result (theorem 4) for the expected extinction time starting from a single initial individual, before going on to outline the corresponding result (formula (49)) for the expected extinction time starting from quasi-equilibrium. Finally, in section 7 we present some concluding discussion.

2. The model

We consider a sequence of multitype birth-death processes indexed by N, where N is an overall system size parameter, our interest being in the limiting behaviour as $N \to \infty$. The state of the process at time $t \geq 0$ is denoted $\boldsymbol{X}^{(N)}(t) = \left(X_1^{(N)}(t), X_2^{(N)}(t), \ldots, X_k^{(N)}(t)\right) \in \mathbb{Z}_+^k$, where k is the number of types of individual, and for $i = 1, 2, \ldots, k$, $X_i^{(N)}(t)$ is the number of type i individuals present at time t. We assume that the state space S is either the whole of $\mathbb{Z}_+^k$, or the finite set $\{\boldsymbol{x} = (x_1, x_2, \ldots, x_k) : 0 \leq x_i \leq N_i \text{ for } i = 1, 2, \ldots, k\}$ for some $\boldsymbol{N} = (N_1, N_2, \ldots, N_k)$, and that the extinction state $\boldsymbol{x} = \boldsymbol{0}$ is absorbing, while $C = S \setminus \{\boldsymbol{0}\}$ forms a single communicating class. In the case of finite state space, noting that N_i is the upper bound for the number of individuals of type i, we take $N = N_1 + N_2 + \cdots + N_k$, set $f_i = N_i/N$ for $i = 1, 2, \ldots, k$, and assume that $f_1, f_2, \ldots, f_k$ are all strictly positive. For simplicity we assume that $f_1, f_2, \ldots, f_k$ do not vary with N, so when we write, for instance, $N \to \infty$, it is implicit that N increases through a subsequence of integers such that Nf_i is always integer-valued for all i. To avoid overly cumbersome notation, we suppress the superscript N from now on.

We assume that all birth and death rates scale with N in the manner of the 'density-dependent' processes of chapter 11 of [13], so that when N is large the scaled process $\boldsymbol{X}(t)/N$ may be approximated by a deterministic process (see section 2.3 for details). We allow the birth and death rates for type i individuals to depend upon (i) the (scaled) number of type i individuals present; and (ii) the (scaled) total number of individuals of all types. More precisely, denoting by $\xi_1, \xi_2, \ldots, \xi_k$ a set of independent homogeneous Poisson processes of rate 1, then for $i = 1, 2, \ldots, k$, new type i individuals are born at the points of the time-transformed process

$$\xi_i \left(N \int_{u=0}^{t} b_0 \left(\sum_{j=1}^{k} X_j(u)/N \right) b_i \left(X_i(u)/N \right) du \right) \tag{2}$$

for some functions $b_0, b_1, \ldots, b_k$ from $\mathbb{R}^+$ to $\mathbb{R}^+$. Each newly born type i individual is assigned a 'lifeforce' distributed as a non-negative random variable Q_i, independent of other individuals' lifeforce variables and of the processes $\xi_1, \xi_2, \ldots, \xi_k$, with $\mathbb{E}[Q_i] = 1$ for $i = 1, 2, \ldots, k$. Each type i individual's lifeforce reduces at rate

$$Nd_0\left(\sum_{j=1}^{k} X_j(t)/N\right) d_i\left(X_i(t)/N\right)/X_i(t) \tag{3}$$

for some functions $d_0, d_1, \ldots, d_k$ from $\mathbb{R}^+$ to $\mathbb{R}^+$, until it reaches zero, at which point the individual is removed from the population (dies). Note that if $d_0(y) = 1$ and $d_i(y) = y/\alpha_i$ for $i = 1, \ldots, k$, for some constants $\alpha_1, \alpha_2, \ldots, \alpha_k > 0$, then the time that a type i individual remains alive in the population is distributed as $\alpha_i Q_i$, so the lifeforce random variables are simply scaled lifetimes. More generally, the functions $d_0, d_1, \ldots, d_k$ may be interpreted as competition effects, so that d_0 represents the effect on an individual's lifespan of competition from the entire population, while for $i = 1, 2, \ldots, k$, d_i represents the effect on a type i individual's lifespan of competition from individuals of its own type. Alternatively, these could be 'safety in numbers' effects, whereby the presence of others enhances an individual's survival chances.

The model that we have defined may be regarded as a piecewise deterministic Markov process [11], in which the state of the process at time t is given by $\boldsymbol{X}(t)$ together with the remaining lifeforce values at time t for every individual alive in the population at that time. However, our approach avoids the need for the general machinery of piecewise-deterministic Markov processes, so we will not pursue this. In the case that Q_i is exponentially distributed for $i = 1, 2, \ldots, k$, then $\{\boldsymbol{X}(t) : t \geq 0\}$ is a continuous-time Markov process with transition rates as follows (where $\boldsymbol{e}_i$ denotes the unit vector with ith element equal to 1 and all other elements equal to 0).

$$\begin{aligned}
\boldsymbol{x} \to \boldsymbol{x} + \boldsymbol{e}_i \quad &\text{at rate} \quad Nb_0\left(\sum_{j=1}^{k} x_j \Bigg/ N\right) b_i(x_i/N), \\
\boldsymbol{x} \to \boldsymbol{x} - \boldsymbol{e}_i \quad &\text{at rate} \quad Nd_0\left(\sum_{j=1}^{k} x_j \Bigg/ N\right) d_i(x_i/N).
\end{aligned}$$

To see this, we use an argument based on a construction of [21]. Suppose that $\boldsymbol{X}(t) = \boldsymbol{x}$ for some $\boldsymbol{x} \in \mathbb{Z}_+^k$. Since each Q_i has constant hazard rate equal to 1, it follows from expression (3) that the probability that a particular type i individual is removed in the time interval $[t, t + \delta t)$ is $Nd_0\left(\sum_{j=1}^{k} x_j/N\right)\left(d_i\left(x_i/N\right)/x_i\right)\delta t + o(\delta t)$. Independence of the lifeforce random variables of different individuals then implies that the probability that exactly one individual is removed in the time interval $[t, t+\delta t)$, and they are of type i, is $Nd_0\left(\sum_{j=1}^{k} x_j/N\right) d_i\left(x_i/N\right)\delta t + o(\delta t)$. Similarly, expression (2) together with independence of the processes $\xi_1, \xi_2, \ldots, \xi_k$ implies that the probability that exactly one individual is born in the time interval $[t, t+\delta t)$, and they are of type i, is $Nb_0\left(\sum_{j=1}^{k} x_j \Big/ N\right) b_i(x_i/N)\delta t + o(\delta t)$. Independence between the lifeforce random variables and the processes ξ_i then yields the required result.

2.1. Examples

Two examples of processes that illustrate the biological usefulness of the framework described above are as follows.

Example 1. The susceptible-infectious-susceptible (SIS) infection model of [8, 9] with heterogeneous susceptibilities and infectious periods. This process has finite state space, with N_i now representing the (constant) total number of type i individuals in the population, while $X_i(t)$ is the number of infectious type i individuals at time t. Rate functions are $b_0(y) = \beta y$, $d_0(y) = 1$, and for $i = 1, 2, \ldots, k$, $b_i(y) = \mu_i(f_i - y)$, $d_i(y) = y/\alpha_i$. Here β is an overall infection rate parameter, while for $i = 1, 2, \ldots, k$, μ_i gives the level of susceptibility of type i individuals, α_i is the mean infectious period of type i individuals, and $f_i = N_i/N$ is the proportion of all individuals that are of type i. We assume that $\beta > 0$ and that $\alpha_i, \mu_i, f_i > 0$ for $i = 1, 2, \ldots, k$, and without loss of generality we scale the μ_i values so that $\sum_{i=1}^{k} \mu_i f_i = 1$.

Example 2. The multitype birth-death process with linear birth rates and quadratic death rates described in section 2.2 of [7]. This process has state space $\mathbb{Z}_+^k$, and may be obtained by setting $b_0(y) = \lambda y$, $d_0(y) = \mu + \kappa y$, and for $i = 1, 2, \ldots, k$, $b_i(y) = 1$, $d_i(y) = y$, with parameters $\lambda, \mu, \kappa > 0$. In [7] it is additionally assumed that the lifeforce variables Q_i are exponentially distributed.

2.2. Assumptions

In order to prove our results, we adopt the following assumptions throughout.

In the case that $S = \mathbb{Z}_+^k$ we set $\tilde{C}_0 = \tilde{C}_1 = \cdots = \tilde{C}_k = (0, \infty)$, while in the case of finite state space we set $\tilde{C}_0 = (0, 1)$ and $\tilde{C}_i = (0, f_i)$ for $i = 1, 2, \ldots, k$. To ensure that $\mathbf{0}$ is an absorbing state, that $C = S \setminus \{\mathbf{0}\}$ is a communicating class, and that the process cannot leave S, we require that

$$b_0(0) = 0,\ b_0(y) > 0 \text{ for } y \in \tilde{C}_0,\ d_0(y) > 0 \text{ for } y \in \tilde{C}_0, \tag{4}$$

and for $i = 1, 2, \ldots, k$,

$$b_i(0) > 0,\ b_i(y) > 0 \text{ for } y \in \tilde{C}_i,\ d_i(0) = 0,\ d_i(y) > 0 \text{ for } y \in \tilde{C}_i. \tag{5}$$

In the case of finite state space, we additionally require

$$d_0\,(1) > 0 \text{ and for } i = 1, 2, \ldots, k,\ b_i\,(f_i) = 0,\ d_i\,(f_i) > 0. \tag{6}$$

We further assume that

$$b_0, b_1, \ldots, b_k, d_0, d_1, \ldots, d_k \text{ are differentiable throughout their domains,} \tag{7}$$

$$d_0(0) > 0 \text{ and } b_0'(0), d_1'(0), d_2'(0), \ldots, d_k'(0) > 0. \tag{8}$$

Finally, define

$$a(\boldsymbol{x}) \quad = \quad \frac{1}{N d_0(1/N)} \prod_{r=1}^{\sum_j x_j - 1} \left(\frac{b_0(r/N)}{d_0((r+1)/N)} \right) \prod_{i=1}^{k} \prod_{r=0}^{x_i - 1} \left(\frac{b_i(r/N)}{d_i((r+1)/N)} \right). \tag{9}$$

In section 3 we will define a modified version of our process, re-started in a particular way whenever the original process is absorbed at the state $\boldsymbol{x} = \mathbf{0}$. The stationary distribution of this re-started process is found to be proportional to $\{a(\boldsymbol{x}) : \boldsymbol{x} \in S\}$, see section 3. We assume that for all sufficiently large N,

$$\sum_{\boldsymbol{x} \in C} a(\boldsymbol{x}) \quad < \quad \infty \tag{10}$$

$$\text{and } \sum_{\boldsymbol{x} \in C} a(\boldsymbol{x}) b_0 \left(\frac{\sum_{j=1}^{k} x_j}{N} \right) \sum_{i=1}^{k} b_i(x_i/N) \quad < \quad \infty. \tag{11}$$

The significance of conditions (10) and (11) is discussed in section 3. Note that in the case of a finite state space, these two conditions are trivially satisfied. For example 2 of section 2.1, we verify condition (10) as follows. Recalling the multinomial formula

$k^l = (1+1+\cdots+1)^l = \sum_{\left\{\boldsymbol{x}\in\mathbb{Z}_+^k:\sum_i x_i=l\right\}} \frac{l!}{\prod_{i=1}^k x_i!}$, then we have

$$\begin{aligned}
\sum_{\boldsymbol{x}\in C} a(\boldsymbol{x}) &= \frac{1}{\mu+(\kappa/N)} \sum_{l=1}^{\infty} \left(\frac{\lambda}{\kappa}\right)^{l-1} (l-1)! \prod_{r=1}^{l-1} \left(\frac{1}{(\mu/\kappa)+((r+1)/N)}\right) \\
&\quad \times \sum_{\left\{\boldsymbol{x}\in\mathbb{Z}_+^k:\sum_i x_i=l\right\}} \prod_{i=1}^{k} \frac{1}{x_i!} \\
&= \frac{k}{\mu+(\kappa/N)} \sum_{l=1}^{\infty} \left(\frac{\lambda k}{\kappa}\right)^{l-1} \frac{1}{l} \prod_{r=1}^{l-1} \left(\frac{1}{(\mu/\kappa)+((r+1)/N)}\right) \\
&< \frac{k}{\mu+(\kappa/N)} \sum_{l=1}^{\infty} \left(\frac{\lambda k N}{\kappa}\right)^{l-1} \left(\frac{1}{l}\right)\left(\frac{1}{l!}\right) \\
&< \frac{\kappa}{\lambda N\left(\mu+(\kappa/N)\right)} \left(\exp\left(\lambda k N/\kappa\right)-1\right) \ < \ \infty.
\end{aligned}$$

Condition (11) follows by similar arguments.

We assume conditions (4)-(8), (10), (11) throughout. For the supercritical case, we require further assumptions (35)-(39), set out in section 6.

2.3. Deterministic and branching process approximations

We now describe two approximating processes, valid in different regimes, that are useful in motivating our methods and giving some intuition behind our results.

Firstly, when N is large and for $i = 1, 2, \ldots, k$, $X_i(t)$ is of comparable order to N, then the scaled process $\boldsymbol{X}(t)/N$ may be approximated by the deterministic process $\boldsymbol{y}(t) = (y_1(t), y_2(t), \ldots, y_k(t))$ defined by

$$\begin{aligned}
y_i(t) &= \int_{s=0}^{\infty} b_0\left(\sum_{j=1}^{k} y_j(t-s)\right) b_i(y_i(t-s)) \\
&\quad \times \mathbb{P}\left(Q_i > \int_{u=0}^{s} \frac{d_0\left(\sum_{j=1}^{k} y_j(t-u)\right) d_i(y_i(t-u))}{y_i(t-u)}\, du\right) ds \qquad (12)
\end{aligned}$$

for $i = 1, 2, \ldots, k$. That is, in the deterministic process, the (scaled) number of individuals of type i at time t, $y_i(t)$, consists of all those individuals born into group i at time $t-s$ who remain alive at time t, integrated over $s \geq 0$. Individuals are born into group i at time $t-s$ at rate $b_0\left(\sum_{j=1}^k y_j(t-s)\right) b_i(y_i(t-s))$. The proportion of such individuals remaining alive at time t is equal to $\mathbb{P}\left(Q_i > \int_{u=0}^{s} \eta_i(t-u)\, du\right)$, where $\eta_i(t-u)$ is the rate of reduction of each type i individual's lifeforce at time $t-u$, given by $\eta_i(t-u) = d_0\left(\sum_{j=1}^k y_j(t-u)\right) d_i(y_i(t-u)) \Big/ y_i(t-u)$.

When the Q_i are exponentially distributed, then (12) is equivalent to the system

$$\frac{dy_i}{dt} = b_0\left(\sum_{j=1}^{k} y_j\right) b_i(y_i) - d_0\left(\sum_{j=1}^{k} y_j\right) d_i(y_i) \text{ for } i = 1, 2, \ldots, k. \tag{13}$$

Assumptions (4) and (5) ensure that the system (13) has an equilibrium point at $\boldsymbol{y} = \boldsymbol{0}$, corresponding to extinction. Assuming that $\boldsymbol{X}(0)/N \to \boldsymbol{y}_0$ as $N \to \infty$, where $\boldsymbol{y}_0 \neq \boldsymbol{0}$, it follows immediately from section 11.2 of [13] that, if the Q_i are exponentially distributed, then under certain conditions on the functions $b_0, b_1, \ldots, b_k$ and $d_0, d_1, \ldots, d_k$, in the limit as $N \to \infty$ the scaled process $\boldsymbol{X}(t)/N$ converges in probability over finite time intervals to the process $\boldsymbol{y}(t)$ satisfying the ordinary differential equations (13) with initial condition $\boldsymbol{y}(0) = \boldsymbol{y}_0$. We will not need to make use of this result, rather the approximating deterministic process is used merely to provide intuitive motivation.

Secondly, when N is large and for $i = 1, 2, \ldots, k$, $X_i(t)$ is small compared to N, then the process $\boldsymbol{X}(t)$ may be approximated by a (linear) multitype branching process as follows. Suppose that $\boldsymbol{X}(t) = \boldsymbol{x}$ for some $\boldsymbol{x} \in \mathbb{Z}_+^k$. Then, recalling assumptions (5) and (7), the rate (3) at which each type i individual's lifeforce reduces may be approximated by

$$\lim_{\boldsymbol{x}/N \to \boldsymbol{0}} d_0\left(\sum_{j=1}^{k} x_j/N\right) \frac{d_i(x_i/N)}{x_i/N} = d_0(0)d_i'(0).$$

Hence each type i individual in the approximating branching process lives for a time distributed as $\left(d_0(0)d_i'(0)\right)^{-1} Q_i$. During its lifetime, the rate at which each type i individual gives birth to type j individuals in the approximating branching process is given, from expression (2), recalling assumptions (4) and (7), by

$$\lim_{\boldsymbol{x}/N \to \boldsymbol{0}} \frac{b_0\left(\sum_{l=1}^{k} x_l/N\right)}{\sum_{l=1}^{k} x_l/N} b_j(x_j/N) = b_0'(0)b_j(0)$$

for $j = 1, 2, \ldots, k$.

The mean number of type j offspring produced by a type i individual over its lifetime in this approximating process is

$$m_{ij} = \frac{b_0'(0)b_j(0)}{d_0(0)d_i'(0)} \text{ for } i, j = 1, 2, \ldots, k. \tag{14}$$

Denoting by M the matrix with elements m_{ij}, assumptions (5) and (8) ensure that all elements of M are finite and strictly positive. Further, it follows from the expression (14) that M is of rank 1, with Perron-Frobenius eigenvalue R_0 given by its trace, viz.

$$R_0 \quad = \quad \frac{b_0'(0)}{d_0(0)} \sum_{j=1}^{k} \frac{b_j(0)}{d_j'(0)}. \tag{15}$$

This type of branching process approximation is made rigorous for certain epidemic models in, for example, [3, 6]. We will not make use of such results, as we employ the approximating branching process only for intuitive motivation and in the heuristic argument of section 6.2. In the context of epidemic modelling, the matrix M is referred to as the *next generation matrix* and its Perron-Frobenius eigenvalue, R_0, as the *basic reproduction number.* In general, the next generation matrix M need not be of rank 1; this is a consequence of our model assumptions, in particular the fact that in the approximating branching process, the rate at which type i individuals give birth to type j individuals depends only upon j.

Denoting by ω_i the probability that the multitype branching process initiated by a single individual of type i produces a finite number of progeny, and by $\phi_i(\cdot)$ the moment generating function of the lifeforce random variable Q_i, then from standard theory of multitype branching processes, eg [19], the probabilities ω_i satisfy

$$\omega_i \quad = \quad \phi_i\left(-\frac{b_0'(0)}{d_0(0)d_i'(0)} \sum_{j=1}^{k} (1-\omega_j) b_j(0) \right) \text{ for } i = 1, 2, \ldots, k. \tag{16}$$

For $R_0 \leq 1$, we have $\omega_1 = \omega_2 = \cdots = \omega_k = 1$, while for $R_0 > 1$, $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_k)$ is the unique solution of equations (16) in $[0,1]^k \setminus \{\mathbf{1}\}$, and $\omega_i < 1$ for $i = 1, 2, \ldots, k$ ([19], section 1.7).

This approximating multitype branching process can alternatively be viewed as a single-type branching process, as follows. Each individual has lifetime distributed according to a mixture distribution, distributed as $\left(d_0(0)d_i'(0)\right)^{-1} Q_i$ with probability $b_i(0)/\sum_j b_j(0)$ for $i = 1, 2, \ldots, k$, and during its lifetime gives birth at rate $b_0'(0)\sum_j b_j(0)$. Denote by ω the probability that this single-type branching process produces a finite number of progeny. Then for $R_0 \leq 1$, we have $\omega = 1$, while for $R_0 > 1$, ω is the unique solution in $[0,1)$ of

$$\omega \;=\; \frac{1}{\sum_{j=1}^{k} b_j(0)} \sum_{i=1}^{k} b_i(0)\phi_i \left(-\frac{b_0'(0)}{d_0(0)d_i'(0)} \left(\sum_{j=1}^{k} b_j(0) \right) (1-\omega) \right). \tag{17}$$

This single-type branching process is equivalent to the preceding multitype branching process provided that the multitype process is initiated by a single individual which is of type i with probability $b_i(0)/\sum_j b_j(0)$ for $i = 1, 2, \ldots, k$. Consequently, we have the relationship

$$\omega \;=\; \left(\sum_{i=1}^{k} b_i(0)\omega_i \right) \Bigg/ \sum_{j=1}^{k} b_j(0). \tag{18}$$

Note that combining relationship (18) with equations (16) gives

$$\omega_i \;=\; \phi_i \left(-\frac{b_0'(0)}{d_0(0)d_i'(0)} \left(\sum_{j=1}^{k} b_j(0) \right) (1-\omega) \right) \quad \text{for } i = 1, 2, \ldots, k,$$

so that each probability ω_i $(i = 1, 2, \ldots, k)$ can be expressed in terms of the single probability ω.

A link between the approximating deterministic process and the approximating branching process is as follows. Recalling assumptions (4) and (5), then with δ_{ij} denoting the Kronecker delta, the Jacobian $J(\mathbf{0})$ of the deterministic system (13) at $\boldsymbol{y} = \mathbf{0}$ has elements

$$J_{ij}(\mathbf{0}) \;=\; b_0'(0)b_i(0) - d_0(0)d_i'(0)\delta_{ij}. \tag{19}$$

From theorem A.1 of [12] or theorem 2 of [22], all eigenvalues of $J(\mathbf{0})$ have strictly negative real part precisely when $R_0 < 1$, and conversely $J(\mathbf{0})$ has at least one eigenvalue with strictly positive real part precisely when $R_0 > 1$. That is, subcriticality of the approximating branching process implies local stability of the extinction equilibrium point $\boldsymbol{y} = \mathbf{0}$ of the deterministic system (13), while supercriticality of the branching process implies local instability of the extinction point.

3. The re-started process

In order to analyse the long-term behaviour of our process, we follow [14, 5] in considering first the stationary behaviour of a modified process without extinction.

Specifically, we introduce a regeneration step as follows. Whenever the process reaches the state $\boldsymbol{x} = \mathbf{0}$, it remains there for an exponentially distributed time of mean 1, after which a birth occurs, being of type i with probability ρ_i for some distribution $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_k)$, and the process then continues as before. Taking

$$\rho_i = \frac{b_i(0)}{\sum_{j=1}^k b_j(0)} \text{ for } i = 1, 2, \ldots, k \tag{20}$$

ensures that for the case of exponentially distributed lifeforce variables Q_i, the re-started process is reversible [16]. We can then apply detailed balance conditions to find its stationary distribution $\boldsymbol{\pi} = \{\pi(\boldsymbol{x}) : \boldsymbol{x} \in S\}$. Recalling the definition (9) of $a(\boldsymbol{x})$, we find that

$$\pi(\boldsymbol{x}) = \frac{a(\boldsymbol{x})}{\sum_{j=1}^k b_j(0)} \pi(\mathbf{0}) \text{ for } \boldsymbol{x} \in C, \tag{21}$$

with

$$\pi(\mathbf{0}) = \left(1 + \frac{\sum_{\boldsymbol{x} \in C} a(\boldsymbol{x})}{\sum_{j=1}^k b_j(0)}\right)^{-1}. \tag{22}$$

Assumptions (10) and (11) imply, respectively, that $\boldsymbol{\pi}$ is a proper distribution and that condition (11) of [23] is satisfied. From theorem 2 of [23] it now follows that $\boldsymbol{\pi}$ is the stationary distribution of numbers of individuals of each type in the system, *regardless* of the distributions of the Q_i (recall $\mathbb{E}[Q_i] = 1$ for $i = 1, 2, \ldots, k$).

The stationary distribution $\boldsymbol{\pi}$ of the re-started process, restricted to C, may be used to approximate the quasi-stationary distribution $\boldsymbol{q}$ of the original process (assuming that a unique quasi-stationary distribution exists). For the classic (single type) SIS epidemic model, this gives the approximation of section 3.1 of [17]. It has been shown [15, 10] that for general single-type birth-death processes (with exponentially distributed lifetimes), $\boldsymbol{\pi}$ restricted to C is a lower bound for $\boldsymbol{q}$ in the sense of likelihood ratio ordering of distributions.

The mean time between regenerations is $1/\pi(\mathbf{0})$. Denoting by τ_i the mean time taken for the process to first hit state $\mathbf{0}$ starting from a single type i individual (which is the same for the original process as for the re-started version) and recalling the form (20) of the probabilities ρ_i and formula (22), we thus have

$$\begin{aligned}
\frac{1}{\pi(\mathbf{0})} &= 1+\sum_{i=1}^{k}\rho_i\tau_i \\
\Rightarrow \quad \sum_{i=1}^{k} b_i(0)\tau_i &= \left(\frac{1}{\pi(\mathbf{0})}-1\right)\sum_{j=1}^{k} b_j(0) \\
&= \sum_{\boldsymbol{x}\in C} a(\boldsymbol{x}). \qquad (23)
\end{aligned}$$

To investigate mean persistence time, we consider the asymptotic behaviour of the sum (23) as $N\to\infty$.

4. The subcritical case

For the subcritical case $R_0<1$, we have the following.

Theorem 1. *Suppose that* $R_0<1$*, that assumptions (4)-(8) are satisfied, and that* $b_0', b_1, b_2, \ldots, b_k$ *are non-increasing functions, while* $d_0, d_1', d_2', \ldots, d_k'$ *are non-decreasing functions.*

Then conditions (10), (11) are automatically satisfied, and recalling that τ_i *denotes the mean time to extinction starting from a single type* i *individual, we have*

$$\sum_{i=1}^{k} b_i(0)\tau_i \quad \to \quad -\frac{1}{b_0'(0)}\log\left(1-R_0\right) \;\; \textit{as } N\to\infty.$$

Proof. First, note that the conditions of the theorem imply that

$$\begin{aligned}
\sum_{\boldsymbol{x}\in C} a(\boldsymbol{x}) \quad &= \quad \frac{1}{Nd_0(1/N)}\sum_{l=1}^{\infty}\prod_{r=1}^{l-1}\left(\frac{b_0(r/N)}{d_0((r+1)/N)}\right) \\
&\quad \times \sum_{\{\boldsymbol{x}\in C:\sum_i x_i=l\}}\prod_{i=1}^{k}\prod_{r=0}^{x_i-1}\left(\frac{b_i(r/N)}{d_i((r+1)/N)}\right) \\
&\leq \quad \frac{1}{d_0(0)}\sum_{l=1}^{\infty}(l-1)!\left(\frac{b_0'(0)}{d_0(0)}\right)^{l-1}\sum_{\{\boldsymbol{x}\in C:\sum_i x_i=l\}}\prod_{i=1}^{k}\frac{1}{x_i!}\left(\frac{b_i(0)}{d_i'(0)}\right)^{x_i}.
\end{aligned}$$

Now from the multinomial theorem, we have

$$\begin{aligned}\left(\sum_{i=1}^{k}\frac{b_i(0)}{d_i'(0)}\right)^l &= \sum_{\left\{\boldsymbol{x}\in\mathbb{Z}_+^k:\sum_i x_i=l\right\}}\frac{l!}{\prod_{i=1}^k x_i!}\prod_{i=1}^{k}\left(\frac{b_i(0)}{d_i'(0)}\right)^{x_i}\\ &\geq l!\sum_{\left\{\boldsymbol{x}\in C:\sum_i x_i=l\right\}}\prod_{i=1}^{k}\frac{1}{x_i!}\left(\frac{b_i(0)}{d_i'(0)}\right)^{x_i}\end{aligned}$$

and so

$$\begin{aligned}\sum_{\boldsymbol{x}\in C}a(\boldsymbol{x}) &\leq \frac{1}{b_0'(0)}\sum_{l=1}^{\infty}\frac{1}{l}\left(\frac{b_0'(0)}{d_0(0)}\right)^l\left(\sum_{i=1}^{k}\frac{b_i(0)}{d_i'(0)}\right)^l\\ &= \frac{1}{b_0'(0)}\sum_{l=1}^{\infty}\frac{R_0^l}{l}\\ &= -\frac{1}{b_0'(0)}\log\left(1-R_0\right) \;<\; \infty,\end{aligned}$$

so that condition (10) is satisfied, and from equation (23) it follows that

$$\limsup_{N\to\infty}\sum_{i=1}^{k}b_i(0)\tau_i \;\leq\; -\frac{1}{b_0'(0)}\log\left(1-R_0\right).$$

A similar argument shows that under the conditions of the theorem,

$$\begin{aligned}&\sum_{\boldsymbol{x}\in C}a(\boldsymbol{x})b_0\left(\sum_{j=1}^{k}x_j\Big/N\right)\sum_{i=1}^{k}b_i(x_i/N)\\ &\quad\leq\; (1/N)\sum_{i=1}^{k}b_i(0)R_0/\left(1-R_0\right) \;<\; \infty,\end{aligned}$$

so that condition (11) is satisfied.

Next, for any $m\in\mathbb{N}$, we have

$$\begin{aligned}\sum_{i=1}^{k}b_i(0)\tau_i \;\geq\; &\frac{1}{Nd_0(1/N)}\sum_{l=1}^{m}\prod_{r=1}^{l-1}\left(\frac{b_0(r/N)}{d_0((r+1)/N)}\right)\\ &\times\sum_{\left\{\boldsymbol{x}\in C:\sum_i x_i=l\right\}}\prod_{i=1}^{k}\prod_{r=0}^{x_i-1}\left(\frac{b_i(r/N)}{d_i((r+1)/N)}\right).\end{aligned}$$

Since m is fixed, then as $N\to\infty$, for each $r\leq m$ we have $Nb_0(r/N)\to rb_0'(0)$ and $d_0((r+1)/N)\to d_0(0)$, and for $i=1,2,\ldots,k$ we have $b_i(r/N)\to b_i(0)$ and $Nd_i((r+1)/N)\to(r+1)d_i'(0)$. Note also that for $l\leq m$, for all sufficiently large N,

$\{\boldsymbol{x} \in C : \sum_i x_i = l\} = \left\{\boldsymbol{x} \in \mathbb{Z}_+^k : \sum_i x_i = l\right\}$. Hence

$$\begin{aligned}
\liminf_{N\to\infty} \sum_{i=1}^{k} b_i(0)\tau_i \quad &\geq \quad \frac{1}{d_0(0)} \sum_{l=1}^{m} \left(\frac{b_0'(0)}{d_0(0)}\right)^{l-1} (l-1)! \\
&\quad \times \sum_{\left\{\boldsymbol{x}\in\mathbb{Z}_+^k : \sum_i x_i = l\right\}} \prod_{i=1}^{k} \frac{1}{x_i!} \left(\frac{b_i(0)}{d_i'(0)}\right)^{x_i} \\
&= \quad \frac{1}{b_0'(0)} \sum_{l=1}^{m} \frac{1}{l} \left(\frac{b_0'(0)}{d_0(0)}\right)^{l} \left(\sum_{i=1}^{k} \frac{b_i(0)}{d_i'(0)}\right)^{l} \\
&= \quad \frac{1}{b_0'(0)} \sum_{l=1}^{m} \frac{R_0^l}{l},
\end{aligned}$$

and letting $m \to \infty$ we obtain

$$\liminf_{N\to\infty} \sum_{i=1}^{k} b_i(0)\tau_i \quad \geq \quad -\frac{1}{b_0'(0)} \log\left(1 - R_0\right).$$

The result follows. □

Remark 1. Note that examples 1 and 2 of section 2.1 both satisfy the conditions of theorem 1.

Remark 2. In the symmetric case that $b_1 = b_2 = \cdots = b_k$ and $d_1 = d_2 = \cdots = d_k$, then $\tau_1 = \tau_2 = \cdots = \tau_k$, and so for $i = 1, 2, \ldots, k$ we have

$$\tau_i \quad \to \quad -\frac{1}{k b_0'(0) b_1(0)} \log\left(1 - R_0\right) \text{ as } N \to \infty.$$

In particular, for example 2 of section 2.1, with $R_0 = \lambda k/\mu$, we have that for $i = 1, 2, \ldots, k$,

$$\tau_i \quad \to \quad -\frac{1}{\lambda k} \log\left(1 - \frac{\lambda k}{\mu}\right) \text{ as } N \to \infty.$$

Remark 3. For example 1 of section 2.1, the heterogeneous SIS model, with $R_0 = \beta \sum_{i=1}^{k} \alpha_i \mu_i f_i$, theorem 1 yields that

$$\sum_{i=1}^{k} \mu_i f_i \tau_i \quad \to \quad -\frac{\log\left(1 - R_0\right)}{\beta} \text{ as } N \to \infty. \tag{24}$$

It seems intuitively clear that for this model, τ_i will depend upon i only through the mean infectious period α_i, and that larger α_i values will correspond to larger values of τ_i. Given the form of (24), a natural conjecture then is that $\tau_i \to -\left(\alpha_i/R_0\right) \log\left(1 - R_0\right)$ as $N \to \infty$. For more general processes satisfying the conditions of theorem 1, we correspondingly conjecture that $\tau_i \to -\left(\log\left(1 - R_0\right)\right) / \left(d_0(0) d_i'(0) R_0\right)$ as $N \to \infty$.

5. The critical case

For the critical case $R_0 = 1$, we present results for two models, example 1 of section 2.1 and an extension of example 2. The algebra is rather more intricate and model-specific than in the subcritical case.

Theorem 2. *For the SIS infection model with heterogeneous susceptibilities and infectious periods, example 1 of section 2.1, if* $R_0 = 1$ *then*

$$\sum_{i=1}^{k} \mu_i f_i \tau_i = \sum_{i=1}^{k} b_i(0)\tau_i \quad \sim \quad \frac{\log N}{2\beta} \quad \text{as } N \to \infty.$$

Proof. For $\boldsymbol{x} \in C$ and $\sum_{i=1}^{k} x_i = l$,

$$\begin{aligned} a(\boldsymbol{x}) &= \frac{1}{N} \prod_{r=1}^{l-1} \left(\beta \frac{r}{N}\right) \prod_{i=1}^{k} \prod_{r=0}^{x_i-1} \left(\frac{\alpha_i \mu_i \left(f_i - \frac{r}{N}\right)}{\frac{r+1}{N}} \right) \\ &= \frac{\beta^{l-1}}{l} \frac{l!}{\prod_{i=1}^{k} x_i!} \prod_{i=1}^{k} \left[(\alpha_i \mu_i f_i)^{x_i} \prod_{r=0}^{x_i-1} \left(1 - \frac{r}{f_i N}\right) \right]. \end{aligned} \tag{25}$$

Now,

$$\sum_{\boldsymbol{x} \in C} a(\boldsymbol{x}) = A_N + B_N,$$

where

$$A_N = \sum_{l=1}^{\lfloor \sqrt{N} \rfloor} \sum_{\left\{\boldsymbol{x} \in C : \sum_i x_i = l\right\}} a(\boldsymbol{x}) \qquad \text{and} \qquad B_N = \sum_{l=\lfloor \sqrt{N} \rfloor + 1}^{N} \sum_{\left\{\boldsymbol{x} \in C : \sum_i x_i = l\right\}} a(\boldsymbol{x}).$$

(For $x \in \mathbb{R}$, $\lfloor x \rfloor$ denotes the greatest integer $\leq x$ and, for future reference, $\lceil x \rceil$ denotes the smallest integer $\geq x$.) Using (25) and the multinomial theorem,

$$A_N \leq \sum_{l=1}^{\lfloor \sqrt{N} \rfloor} \frac{\beta^{l-1}}{l} \left(\sum_{i=1}^{k} \alpha_i \mu_i f_i \right)^{l} = \beta^{-1} \sum_{l=1}^{\lfloor \sqrt{N} \rfloor} \frac{1}{l}, \tag{26}$$

since $R_0 = \beta \sum_{i=1}^{k} \alpha_i \mu_i f_i = 1$.

Turning to B_N, note that by the AM/GM inequality, if $x_i > 0$ then

$$\prod_{r=0}^{x_i-1} \left(1 - \frac{r}{f_i N}\right) \leq \left[\frac{1}{x_i} \sum_{r=0}^{x_i-1} \left(1 - \frac{r}{f_i N}\right) \right]^{x_i} = \left(1 - \frac{x_i - 1}{2 f_i N}\right)^{x_i},$$

and that this inequality holds trivially if $x_i = 0$. Therefore,

$$\sum_{\{\boldsymbol{x}\in C:\sum_i x_i = l\}} \frac{l!}{\prod_{i=1}^k x_i!} \prod_{i=1}^k \left[(\alpha_i\mu_i f_i)^{x_i} \prod_{r=0}^{x_i-1} \left(1 - \frac{r}{f_i N}\right)\right]$$

$$\leq \sum_{\{\boldsymbol{x}\in C:\sum_i x_i = l\}} \frac{l!}{\prod_{i=1}^k x_i!} \prod_{i=1}^k (\alpha_i\mu_i f_i)^{x_i} \left(1 - \frac{x_i - 1}{2 f_i N}\right)^{x_i}. \quad (27)$$

Further, if $l > \lfloor\sqrt{N}\rfloor$ and $x_1 + x_2 + \cdots + x_k = l$, then at least one of $x_1, x_2, \ldots, x_k$ is $\geq \sqrt{N}/k$. Hence, for all sufficiently large N, for such l,

$$\prod_{i=1}^k (\alpha_i\mu_i f_i)^{x_i} \left(1 - \frac{x_i - 1}{2 f_i N}\right)^{x_i} \leq \sum_{i=1}^k \left[(\alpha_i\mu_i f_i)\left(1 - \frac{(\sqrt{N}/k) - 1}{2 f_i N}\right)\right]^{x_i} \prod_{\substack{j=1\\ j\neq i}}^k (\alpha_j\mu_j f_j)^{x_j}$$

and, using the multinomial theorem, it follows from (27) that

$$\sum_{\{\boldsymbol{x}\in C:\sum_i x_i = l\}} \frac{l!}{\prod_{i=1}^k x_i!} \prod_{i=1}^k \left[(\alpha_i\mu_i f_i)^{x_i} \prod_{r=0}^{x_i-1} \left(1 - \frac{r}{f_i N}\right)\right]$$

$$\leq \sum_{i=1}^k \left[\alpha_i\mu_i f_i \left(1 - \frac{(\sqrt{N}/k) - 1}{2 f_i N}\right) + \sum_{\substack{j=1\\ j\neq i}}^k \alpha_j\mu_j f_j\right]^l$$

$$= \sum_{i=1}^k \left[\frac{1}{\beta} - \frac{[(\sqrt{N}/k) - 1]\alpha_i\mu_i}{2N}\right]^l,$$

since $R_0 = 1$. Thus,

$$\begin{aligned} B_N &\leq \sum_{l=\lfloor\sqrt{N}\rfloor+1}^{N} \frac{1}{\beta l} \sum_{i=1}^k \left(1 - \frac{\beta\alpha_i\mu_i[(\sqrt{N}/k) - 1]}{2N}\right)^l \\ &\leq \frac{1}{\beta\sqrt{N}} \sum_{i=1}^k \frac{2N}{\beta\alpha_i\mu_i[(\sqrt{N}/k) - 1]} \\ &\leq \frac{4k}{\beta^2} \sum_{i=1}^k \frac{1}{\alpha_i\mu_i} \end{aligned}$$

for all sufficiently large N, which together with (26) implies

$$\limsup_{N\to\infty} \frac{\sum_{\boldsymbol{x}\in C} a(\boldsymbol{x})}{\log N} \leq \frac{1}{2\beta}. \quad (28)$$

Fix $V > 0$. Then it follows from (25) that for any $\boldsymbol{x} \in C$ with $\sum_{i=1}^k x_i = l$, where

$l \leq \lfloor V\sqrt{N} \rfloor$,

$$a(\boldsymbol{x}) \geq \left[\prod_{i=1}^{k} \prod_{r=0}^{\lfloor V\sqrt{N} \rfloor} \left(1 - \frac{r}{f_i N}\right) \right] \frac{\beta^{l-1}}{l} \frac{l!}{\prod_{i=1}^{k} x_i!} \prod_{i=1}^{k} (\alpha_i \mu_i f_i)^{x_i}.$$

Thus, for $l \leq \lfloor V\sqrt{N} \rfloor$ and $R_0 = 1$, application of the multinomial theorem yields

$$\sum_{\left\{\boldsymbol{x} \in C : \sum_i x_i = l\right\}} a(\boldsymbol{x}) \geq \frac{1}{\beta l} \left[\prod_{i=1}^{k} \prod_{r=0}^{\lfloor V\sqrt{N} \rfloor} \left(1 - \frac{r}{f_i N}\right) \right],$$

whence

$$\sum_{\boldsymbol{x} \in C} a(\boldsymbol{x}) \geq \sum_{l=1}^{\lfloor V\sqrt{N} \rfloor} \sum_{\left\{\boldsymbol{x} \in C : \sum_i x_i = l\right\}} a(\boldsymbol{x}) \geq \left[\prod_{i=1}^{k} \prod_{r=0}^{\lfloor V\sqrt{N} \rfloor} \left(1 - \frac{r}{f_i N}\right) \right] \frac{1}{\beta} \sum_{l=1}^{\lfloor V\sqrt{N} \rfloor} \frac{1}{l}.$$

Straightforward analysis (see Appendix A) yields

$$\lim_{N \to \infty} \prod_{i=1}^{k} \prod_{r=0}^{\lfloor V\sqrt{N} \rfloor} \left(1 - \frac{r}{f_i N}\right) = \exp\left(-\frac{V^2}{2} \sum_{i=1}^{k} f_i^{-1}\right). \tag{29}$$

Therefore,

$$\liminf_{N \to \infty} \frac{\sum_{\boldsymbol{x} \in C} a(\boldsymbol{x})}{\log N} \geq \frac{1}{2\beta} \exp\left(-\frac{V^2}{2} \sum_{i=1}^{k} f_i^{-1}\right)$$

and letting $V \downarrow 0$,

$$\liminf_{N \to \infty} \frac{\sum_{\boldsymbol{x} \in C} a(\boldsymbol{x})}{\log N} \geq \frac{1}{2\beta}. \tag{30}$$

Recalling (23), the theorem follows from (28) and (30). □

Remark 4. Theorem 2 for the case $k = 1$ corresponds to relation (3.7) of [5].

We consider now an extension of example 2 of section 2.1, in which $d_0(y) = \mu + \kappa y^\eta$, where $\eta > 0$. Note from (15) that $R_0 = k\lambda/\mu$.

Theorem 3. *For the above extension of example 2, if $R_0 = 1$ then for $i = 1, 2, \ldots, k$,*

$$\tau_i \sim \frac{\eta \log N}{\mu(1+\eta)} \quad \text{as } N \to \infty.$$

Proof. For $\boldsymbol{x} \in C$ and $\sum_{i=1}^{k} x_i = l$,

$$\begin{aligned} a(\boldsymbol{x}) &= \frac{1}{N\left(1 + \kappa(1/N)^\eta\right)} \left(\prod_{r=1}^{l-1} \frac{\lambda(r/N)}{(\mu + \kappa[(r+1)/N]^\eta)} \right) \prod_{i=1}^{k} \prod_{r=0}^{x_i - 1} \left(\frac{1}{(r+1)/N} \right) \\ &= \frac{1}{\mu} \left(\frac{\lambda}{\mu} \right)^{l-1} \frac{(l-1)!}{\prod_{i=1}^{k} x_i!} \frac{1}{\prod_{i=1}^{l} \left(1 + \theta(i/N)^\eta\right)}, \end{aligned}$$

where $\theta = \kappa/\mu$. Thus, using the multinomial theorem,

$$\sum_{\{\boldsymbol{x}\in C:\sum_i x_i = l\}} a(\boldsymbol{x}) \quad = \quad \frac{1}{\mu l}\left(\frac{\lambda}{\mu}\right)^{l-1} \frac{k^l}{\prod_{i=1}^{l}\left(1+\theta(i/N)^{\eta}\right)}$$

and, since $R_0 = 1$,

$$\sum_{\boldsymbol{x}\in C} a(\boldsymbol{x}) \quad = \quad \frac{k}{\mu}\sum_{l=1}^{\infty} \frac{1}{l\prod_{i=1}^{l}\left(1+\theta(i/N)^{\eta}\right)}.$$

Now

$$\sum_{\boldsymbol{x}\in C} a(\boldsymbol{x}) = A_N + B_N,$$

where

$$A_N = \frac{k}{\mu}\sum_{l=1}^{\lceil N^{\frac{\eta}{1+\eta}}\rceil} \frac{1}{l\prod_{i=1}^{l}\left(1+\theta(i/N)^{\eta}\right)} \qquad \text{and} \qquad B_N = \frac{k}{\mu}\sum_{l=\lceil N^{\frac{\eta}{1+\eta}}\rceil+1}^{\infty} \frac{1}{l\prod_{i=1}^{l}\left(1+\theta(i/N)^{\eta}\right)}.$$

Clearly,

$$A_N \le \frac{k}{\mu}\sum_{l=1}^{\lceil N^{\frac{\eta}{1+\eta}}\rceil}\frac{1}{l}. \tag{31}$$

Also,

$$\prod_{i=1}^{l}\left(1+\theta(i/N)^{\eta}\right) \ge 1 + \frac{\theta}{N^{\eta}}\sum_{i=1}^{l} i^{\eta} \ge 1 + \frac{\theta}{N^{\eta}}\int_0^l u^{\eta}\,du \ge \frac{\theta l^{\eta+1}}{(\eta+1)N^{\eta}},$$

so

$$\begin{aligned}\sum_{l=\lceil N^{\frac{\eta}{1+\eta}}\rceil+1}^{\infty} \frac{1}{l\prod_{i=1}^{l}\left(1+\theta(i/N)^{\eta}\right)} &\le \frac{(\eta+1)N^{\eta}}{\theta}\sum_{l=\lceil N^{\frac{\eta}{1+\eta}}\rceil+1}^{\infty}\frac{1}{l^{\eta+2}} \\ &\le \frac{(\eta+1)N^{\eta}}{\theta}\int_{\lceil N^{\frac{\eta}{1+\eta}}\rceil}^{\infty} u^{-(\eta+2)}\,du \le \theta^{-1}.\end{aligned}$$

Note that in conjunction with (31) this shows that condition (10) is satisfied and a similar argument shows that condition (11) is satisfied. Further, it follows that

$$\limsup_{N\to\infty}\frac{\sum_{\boldsymbol{x}\in C} a(\boldsymbol{x})}{\log N} \le \frac{k\eta}{\mu(1+\eta)}. \tag{32}$$

Fix $L > 0$. Then

$$\begin{aligned}\sum_{\boldsymbol{x}\in C} a(\boldsymbol{x}) \quad &\ge \quad \frac{k}{\mu}\sum_{l=1}^{\lfloor LN^{\frac{\eta}{1+\eta}}\rfloor} \frac{1}{l\prod_{i=1}^{l}\left(1+\theta(i/N)^{\eta}\right)} \\ &\ge \quad \frac{k}{\mu}\left(\sum_{l=1}^{\lfloor LN^{\frac{\eta}{1+\eta}}\rfloor}\frac{1}{l}\right)\left(\frac{1}{\prod_{i=1}^{\lfloor LN^{\frac{\eta}{1+\eta}}\rfloor}\left(1+\theta(i/N)^{\eta}\right)}\right).\end{aligned}$$

Straightforward analysis (see Appendix A) yields that, for $\eta, L > 0$,

$$\lim_{N\to\infty} \prod_{i=1}^{\lfloor LN^{\frac{\eta}{1+\eta}}\rfloor} (1+\theta(i/N)^{\eta}) = \exp\left(\frac{\theta L^{\eta+1}}{\eta+1}\right), \tag{33}$$

so

$$\liminf_{N\to\infty} \frac{\sum_{\boldsymbol{x}\in C} a(\boldsymbol{x})}{\log N} \geq \frac{k\eta}{\mu(1+\eta)} \exp\left(-\frac{\theta L^{\eta+1}}{\eta+1}\right). \tag{34}$$

From (32) and letting $L \downarrow 0$ in (34) it follows that

$$\sum_{i=1}^{k} \tau_i = \sum_{i=1}^{k} b_i(0)\tau_i \sim \frac{k\eta \log N}{\mu(1+\eta)} \text{ as } N \to \infty.$$

Recalling remark 2, from the symmetry of the process we have $\tau_1 = \tau_2 = \cdots = \tau_k$, and the result follows. □

6. The supercritical case

Our assumptions of section 2.2 ensure that system (13) has an equilibrium point at $\boldsymbol{y} = \boldsymbol{0}$. We now consider the supercritical case $R_0 > 1$ under the following further assumptions.

1. System (13) has a unique equilibrium point $\boldsymbol{y}^*$ in $C = S \setminus \{\boldsymbol{0}\}$, with $y_i^* > 0$ for $i = 1, 2, \ldots, k$. (35)
2. The eigenvalues of the Jacobian of (13) at $\boldsymbol{y}^*$ all have strictly negative real part. (36)
3. The functions $\log\left(\frac{b_0(u)}{ud_0(u)}\right)$ and $\log\left(\frac{ub_i(u)}{d_i(u)}\right)$ for $i = 1, 2, \ldots, k$ are twice differentiable throughout their domains. (37)
4. For every $y \in \tilde{C}_0$, $\sup_{0\leq u\leq y} \left|\frac{d^2}{du^2} \log\left(\frac{b_0(u)}{ud_0(u)}\right)\right| < \infty$. (38)
5. For $i = 1, 2, \ldots, k$, for every $y \in \tilde{C}_i$, $\sup_{0\leq u\leq y} \left|\frac{d^2}{du^2} \log\left(\frac{ub_i(u)}{d_i(u)}\right)\right| < \infty$. (39)

We have already seen that when $R_0 > 1$ the Jacobian of the system (13) at $\boldsymbol{y} = \boldsymbol{0}$ has at least one eigenvalue with strictly positive real part, so that the extinction equilibrium point is locally unstable; condition (36) ensures that the equilibrium point $\boldsymbol{y}^*$ is locally stable.

6.1. Time to extinction starting from a single individual

Corresponding to theorem 1 for the subcritical case, in the supercritical case we have the following.

Theorem 4. *Suppose that $R_0 > 1$, and that assumptions (4)-(8), (10), (11) together with assumptions (35)-(39) are all satisfied. With δ_{ij} denoting the Kronecker delta, denote by H the matrix with elements h_{ij} given by*

$$h_{ij} \quad = \quad \frac{b_0'\left(\sum_l y_l^*\right)}{b_0\left(\sum_l y_l^*\right)} - \frac{d_0'\left(\sum_l y_l^*\right)}{d_0\left(\sum_l y_l^*\right)} + \left(\frac{b_i'\left(y_i^*\right)}{b_i\left(y_i^*\right)} - \frac{d_i'\left(y_i^*\right)}{d_i\left(y_i^*\right)}\right)\delta_{ij} \tag{40}$$

for $i, j = 1, 2, \ldots, k$. Recalling that τ_i denotes the mean time to extinction starting from a single individual of type i, then as $N \to \infty$,

$$\sum_{i=1}^{k} b_i(0)\tau_i \quad \sim \quad \frac{K_0}{\sqrt{N}}\exp(AN)$$

where

$$A \quad = \quad \int_0^{\sum_j y_j^*} \log\left(\frac{b_0(u)}{d_0(u)}\right) du + \sum_{i=1}^{k}\int_0^{y_i^*} \log\left(\frac{b_i(u)}{d_i(u)}\right) du \tag{41}$$

and

$$K_0 \quad = \quad \sqrt{\frac{2\pi d_0(0)}{b_0'(0) b_0\left(\sum_j y_j^*\right) d_0\left(\sum_j y_j^*\right) det(-H)} \prod_{i=1}^{k} \frac{b_i(0)d_i'(0)}{b_i\left(y_i^*\right) d_i\left(y_i^*\right)}}, \tag{42}$$

with

$$\begin{aligned} det(-H) \quad = \quad & \prod_{i=1}^{k}\left(\frac{d_i'\left(y_i^*\right)}{d_i\left(y_i^*\right)} - \frac{b_i'\left(y_i^*\right)}{b_i\left(y_i^*\right)}\right) \\ & + \left(\frac{d_0'\left(\sum_l y_l^*\right)}{d_0\left(\sum_l y_l^*\right)} - \frac{b_0'\left(\sum_l y_l^*\right)}{b_0\left(\sum_l y_l^*\right)}\right) \sum_{i=1}^{k}\prod_{j\neq i}\left(\frac{d_j'\left(y_j^*\right)}{d_j\left(y_j^*\right)} - \frac{b_j'\left(y_j^*\right)}{b_j\left(y_j^*\right)}\right). \end{aligned} \tag{43}$$

Proof. Letting $\boldsymbol{y} = \boldsymbol{x}/N$ and $\tilde{a}(\boldsymbol{y}) = a(N\boldsymbol{y})$, then formula (9) may be written as

$$\tilde{a}(\boldsymbol{y}) \quad = \quad \frac{1}{N d_0(1/N)} \prod_{r=1}^{N\sum_j y_j - 1}\left(\frac{b_0(r/N)}{d_0((r+1)/N)}\right) \prod_{i=1}^{k}\prod_{r=0}^{Ny_i - 1}\left(\frac{b_i(r/N)}{d_i((r+1)/N)}\right),$$

so that

$$\begin{aligned}
\log\left(\tilde{a}(\boldsymbol{y})\right) &= -\log\left(Nd_0(1/N)\right) + \sum_{r=1}^{N\sum_j y_j - 1} \log\left(\frac{b_0(r/N)}{d_0((r+1)/N)}\right) \\
&\quad + \sum_{i=1}^{k} \sum_{r=0}^{Ny_i - 1} \log\left(\frac{b_i(r/N)}{d_i((r+1)/N)}\right) \\
&= -\log\left(Nd_0(1/N)\right) - \log\left(\frac{b_0\left(\sum_j y_j\right)}{d_0(1/N)}\right) + \sum_{r=1}^{N\sum_j y_j} \log\left(\frac{b_0(r/N)}{d_0(r/N)}\right) \\
&\quad + \sum_{i=1}^{k} \left(\log\left(\frac{b_i(0)}{b_i\left(y_i\right)}\right) + \sum_{r=1}^{Ny_i} \log\left(\frac{b_i(r/N)}{d_i(r/N)}\right)\right) \\
&= -\log\left(Nb_0\left(\sum_j y_j\right)\right) + \sum_{i=1}^{k} \log\left(\frac{b_i(0)}{b_i\left(y_i\right)}\right) \\
&\quad + \sum_{r=1}^{N\sum_j y_j} \log\left(\frac{b_0(r/N)}{d_0(r/N)}\right) + \sum_{i=1}^{k} \left(\sum_{r=1}^{Ny_i} \log\left(\frac{b_i(r/N)}{d_i(r/N)}\right)\right). \qquad (44)
\end{aligned}$$

For some $i = 1, 2, \ldots, k$, consider the sum $\sum_{r=1}^{Ny_i} \log\left(b_i(r/N)/\, d_i(r/N)\right)$ for large N. We can approximate this sum by an integral via the trapezium rule, with the complication that $d_i(0) = 0$, so that the integrand diverges at the lower end. We deal with this by regularising the integrand as follows.

$$\begin{aligned}
\sum_{r=1}^{Ny_i} \log\left(\frac{b_i(r/N)}{d_i(r/N)}\right) &= \sum_{r=1}^{Ny_i} \log\left(\frac{(r/N)b_i(r/N)}{d_i(r/N)}\right) - \sum_{r=1}^{Ny_i} \log(r/N) \\
&= \sum_{r=1}^{Ny_i} \log\left(\frac{(r/N)b_i(r/N)}{d_i(r/N)}\right) - \log\left((Ny_i)!\right) + Ny_i \log N \\
&= N\int_0^{y_i} \log\left(\frac{ub_i(u)}{d_i(u)}\right) du + \frac{1}{2}\log\left(\frac{y_i b_i\left(y_i\right)/d_i\left(y_i\right)}{b_i(0)\lim_{u\to 0}(u/d_i(u))}\right) \\
&\quad - \log((Ny_i)!) + Ny_i \log N + O(1/N),
\end{aligned}$$

where we have made use of assumption (39) to control the order of the approximation error in the trapezium rule. From assumption (8) we have that $d_i'(0) > 0$, so applying l'Hôpital's rule to the term $\lim_{u\to 0}(u/d_i(u))$ and Stirling's formula to the factorial term yields

$$\begin{aligned}
&\sum_{r=1}^{Ny_i} \log\left(\frac{b_i(r/N)}{d_i(r/N)}\right)\\
&= N\int_0^{y_i} \log\left(\frac{ub_i(u)}{d_i(u)}\right) du + \frac{1}{2}\log\left(\frac{y_i b_i\left(y_i\right)/d_i\left(y_i\right)}{b_i(0)/d_i'(0)}\right)\\
&\quad - Ny_i \log\left(Ny_i\right) + Ny_i - (1/2)\log\left(2\pi N y_i\right) + Ny_i\log N + o(1)\\
&= N\left(\int_0^{y_i} \log\left(\frac{b_i(u)}{d_i(u)}\right) du + \int_0^{y_i} \log u\, du\right) + \frac{1}{2}\log\left(\frac{y_i b_i\left(y_i\right) d_i'(0)}{b_i(0) d_i\left(y_i\right)}\right)\\
&\quad - Ny_i \log y_i + Ny_i - (1/2)\log\left(2\pi N y_i\right) + o(1)\\
&= N\int_0^{y_i} \log\left(\frac{b_i(u)}{d_i(u)}\right) du + N\left(y_i\log y_i - y_i\right) + \frac{1}{2}\log\left(\frac{y_i b_i\left(y_i\right) d_i'(0)}{b_i(0) d_i\left(y_i\right)}\right)\\
&\quad - Ny_i \log y_i + Ny_i - (1/2)\log\left(2\pi N y_i\right) + o(1)\\
&= N\int_0^{y_i} \log\left(\frac{b_i(u)}{d_i(u)}\right) du + \frac{1}{2}\log\left(\frac{y_i b_i\left(y_i\right) d_i'(0)}{b_i(0) d_i\left(y_i\right)}\right)\\
&\quad - (1/2)\log\left(2\pi N y_i\right) + o(1)\\
&= N\int_0^{y_i} \log\left(\frac{b_i(u)}{d_i(u)}\right) du + \frac{1}{2}\log\left(\frac{b_i\left(y_i\right) d_i'(0)}{2\pi N b_i(0) d_i\left(y_i\right)}\right) + o(1). \qquad (45)
\end{aligned}$$

The term $\sum_{r=1}^{N\sum_j y_j} \log\left(b_0(r/N)/d_0(r/N)\right)$ in equation (44) may be treated similarly, noting that from assumptions (4), (8) we have $b_0(0) = 0$, $d_0(0) > 0$ and $b_0'(0) > 0$, and making use of assumption (38) in invoking the trapezium rule. Equation (44) now becomes

$$\begin{aligned}
\log\left(\tilde{a}(\boldsymbol{y})\right) &= N\int_0^{\sum_j y_j} \log\left(\frac{b_0(u)}{d_0(u)}\right) du + \frac{1}{2}\log\left(\frac{2\pi N b_0\left(\sum_j y_j\right) d_0(0)}{b_0'(0) d_0\left(\sum_j y_j\right)}\right) \\
&\quad + \sum_{i=1}^{k}\left(N\int_0^{y_i}\log\left(\frac{b_i(u)}{d_i(u)}\right) du + \frac{1}{2}\log\left(\frac{b_i\left(y_i\right) d_i'(0)}{2\pi N b_i(0) d_i\left(y_i\right)}\right)\right) \\
&\quad - \log\left(N b_0\left(\sum_j y_j\right)\right) + \sum_{i=1}^{k}\log\left(\frac{b_i(0)}{b_i\left(y_i\right)}\right) + o(1) \\
&= N\left(\int_0^{\sum_j y_j}\log\left(\frac{b_0(u)}{d_0(u)}\right) du + \sum_{i=1}^{k}\int_0^{y_i}\log\left(\frac{b_i(u)}{d_i(u)}\right) du\right) \\
&\quad - \left(\frac{k+1}{2}\right)\log N - \left(\frac{k-1}{2}\right)\log(2\pi) \\
&\quad + \frac{1}{2}\log\left(\frac{d_0(0)}{b_0'(0) b_0\left(\sum_j y_j\right) d_0\left(\sum_j y_j\right)}\right) \\
&\quad + \frac{1}{2}\sum_{i=1}^{k}\log\left(\frac{b_i(0) d_i'(0)}{b_i\left(y_i\right) d_i\left(y_i\right)}\right) + o(1).
\end{aligned}$$

That is, as $N \to \infty$,

$$\begin{aligned}
&\log\left(\tilde{a}(\boldsymbol{y})\right) \\
&\quad = N f(\boldsymbol{y}) - \left(\frac{k+1}{2}\right)\log N - \left(\frac{k-1}{2}\right)\log(2\pi) + g(\boldsymbol{y}) + o(1),
\end{aligned} \tag{46}$$

where

$$f(\boldsymbol{y}) = \int_0^{\sum_j y_j}\log\left(\frac{b_0(u)}{d_0(u)}\right) du + \sum_{i=1}^{k}\int_0^{y_i}\log\left(\frac{b_i(u)}{d_i(u)}\right) du$$

and

$$g(\boldsymbol{y}) = \frac{1}{2}\log\left(\frac{d_0(0)}{b_0'(0) b_0\left(\sum_j y_j\right) d_0\left(\sum_j y_j\right)}\right) + \frac{1}{2}\sum_{i=1}^{k}\log\left(\frac{b_i(0) d_i'(0)}{b_i\left(y_i\right) d_i\left(y_i\right)}\right).$$

It is clear from the form of equation (46) that for large N, contributions to the sum (23) will be dominated by terms $a(N\boldsymbol{y}) = \tilde{a}(\boldsymbol{y})$ where $f(\boldsymbol{y})$ achieves its maximal value. Recalling from equation (21) that the stationary distribution of the scaled re-started process is proportional to $\{\tilde{a}(\boldsymbol{y}) : N\boldsymbol{y} \in S\}$, one would intuitively expect this stationary distribution to be peaked around a unique mode that converges to the deterministic stable equilibrium point $\boldsymbol{y}^*$ as $N \to \infty$, and to become increasingly strongly peaked as N increases. We will now show that $f(\boldsymbol{y})$ does indeed have a unique maximum point at $\boldsymbol{y}^*$.

For $i = 1, 2, \ldots, k$,

$$\frac{\partial f}{\partial y_i} = \log \left(\frac{b_0 \left(\sum_j y_j \right) b_i \left(y_i \right)}{d_0 \left(\sum_j y_j \right) d_i \left(y_i \right)} \right),$$

so the condition for a stationary point of $f(\boldsymbol{y})$ is precisely the condition for an equilibrium point of the deterministic system (13). We have assumed that system (13) has exactly two equilibrium points, at $\mathbf{0}$ and $\boldsymbol{y}^*$. We now show that $\boldsymbol{y}^*$ is a local maximum of $f(\boldsymbol{y})$, and hence the unique global maximum.

The elements $h_{ij} = \frac{\partial^2 f}{\partial y_i \partial y_j}$ of the Hessian matrix H of $f(\boldsymbol{y})$ at $\boldsymbol{y}^*$ are given by equation (40). Since all off-diagonal elements of H are equal, application of the matrix determinant lemma shows that the determinant $\det(-H)$ is given by formula (43). Noting that $\boldsymbol{y}^*$ is an equilibrium point of (13), then for $i = 1, 2, \ldots, k$ we have $b_0 \left(\sum_l y_l^* \right) b_i(y_i^*) = d_0 \left(\sum_l y_l^* \right) d_i(y_i^*)$, and so

$$\begin{aligned} h_{ij} = {} & \frac{1}{b_0 \left(\sum_l y_l^* \right) b_i \left(y_i^* \right)} \left(b_0' \left(\sum_l y_l^* \right) b_i \left(y_i^* \right) - d_0' \left(\sum_l y_l^* \right) d_i \left(y_i^* \right) \right. \\ & \left. + \left(b_0 \left(\sum_l y_l^* \right) b_i' \left(y_i^* \right) - d_0 \left(\sum_l y_l^* \right) d_i' \left(y_i^* \right) \right) \delta_{ij} \right) \end{aligned}$$

for $i, j = 1, 2, \ldots, k$. That is, denoting by B the diagonal matrix with entries $b_{ii} = b_0 \left(\sum_l y_l^* \right) b_i \left(y_i^* \right)$ for $i = 1, 2, \ldots, k$, then $H = B^{-1} J(\boldsymbol{y}^*)$, where $J(\boldsymbol{y}^*)$ is the Jacobian of system (13) evaluated at $\boldsymbol{y}^*$. Setting $\Sigma = -H^{-1}$ and noting that $\Sigma = \Sigma^T$, it follows that Σ satisfies the Lyapunov equation $J(\boldsymbol{y}^*)\Sigma + \Sigma J(\boldsymbol{y}^*)^T = -2B$. From theorem 13.21 of [18], this Lyapunov equation has a unique solution Σ. Since all diagonal elements of B are strictly positive and, by assumption (36), the eigenvalues of $J(\boldsymbol{y}^*)$ all have negative real part, it follows from theorem 13.24 of [18] that Σ is positive definite. Hence H is negative definite, and $\boldsymbol{y}^*$ is a local maximum point of $f(\boldsymbol{y})$.

Writing $\tilde{C} = \tilde{C}_1 \times \tilde{C}_2 \times \cdots \times \tilde{C}_k$, then as $N \to \infty$, from equation (46) we obtain

$$\begin{aligned}\sum_{\boldsymbol{x}\in C} a(\boldsymbol{x}) &\sim N^k \int_{\tilde{C}} \tilde{a}(\boldsymbol{y})\, d\boldsymbol{y} \\ &= N^k \int_{\tilde{C}} \exp\left(Nf(\boldsymbol{y}) - \left(\frac{k+1}{2}\right)\log N - \left(\frac{k-1}{2}\right)\log(2\pi) + g(\boldsymbol{y}) + o(1)\right) d\boldsymbol{y} \\ &= \left(\frac{N}{2\pi}\right)^{(k-1)/2} \int_{\tilde{C}} \exp\left(Nf(\boldsymbol{y}) + g(\boldsymbol{y}) + o(1)\right) d\boldsymbol{y}. \qquad (47)\end{aligned}$$

Now condition (10) implies that for sufficiently large N the integral on the right hand side of equation (47) is convergent, and conditions (38), (39) ensure that the $o(1)$ term convergences uniformly to zero within a neighbourhood of $\boldsymbol{y}^*$. We can therefore apply the multivariate Laplace approximation to this integral, yielding

$$\sum_{\boldsymbol{x}\in C} a(\boldsymbol{x}) \sim \sqrt{\frac{2\pi}{N\det(-H)}} \exp\left(Nf(\boldsymbol{y}^*) + g(\boldsymbol{y}^*)\right).$$

So finally,

$$\sum_{i=1}^{k} b_i(0)\tau_i = \sum_{\boldsymbol{x}\in C} a(\boldsymbol{x}) \sim \frac{K_0}{\sqrt{N}} \exp(AN)$$

where

$$\begin{aligned} A &= f(\boldsymbol{y}^*), \\ K_0 &= \sqrt{\frac{2\pi}{\det(-H)}} \exp(g(\boldsymbol{y}^*)), \end{aligned}$$

and the result follows. □

6.2. Time to extinction starting from quasi-equilibrium

In the supercritical case, under assumptions (35)-(36), as well as the mean time to extinction starting from a single individual, another quantity of interest is the mean time to extinction starting from close to the quasi-equilibrium state $N\boldsymbol{y}^*$. Denoting by τ this mean extinction time starting close to quasi-equilibrium, we follow [5] in outlining a heuristic derivation of the asymptotic form of τ.

Recall from section 2.3 that in its initial phase, the process initiated from a single individual may be approximated by a multitype branching process. In the supercritical case $R_0 > 1$, this branching process either goes extinct quickly, or takes off to produce an infinite number of progeny. When the approximating branching process takes off, the original multitype birth-death process will move quickly towards the quasi-equilibrium state $N\boldsymbol{y}^*$, and spend a long time fluctuating close to this state before eventually going extinct. This can be seen from the form of the equilibrium distribution $\pi(\boldsymbol{x})$, since we have seen in section 6.1 that for large N, the probability mass is concentrated close to $N\boldsymbol{y}^*$.

For a process initiated by a single individual of type i ($i = 1, 2, \ldots, k$), denote by ζ_i the expected time to extinction of the process conditional upon rapid extinction, and by σ_i the expected time taken to first attain a small neighbourhood of $N\boldsymbol{y}^*$ conditional upon the process taking off, and recall that ω_i denotes the extinction probability of the approximating branching process. Then in the limit as $N \to \infty$, we have

$$\tau_i \quad \sim \quad \omega_i \zeta_i + (1 - \omega_i)\,(\sigma_i + \tau)\,,$$

where both ζ_i and σ_i are negligible in comparison to τ, so

$$\tau_i \quad \sim \quad (1 - \omega_i)\,\tau. \tag{48}$$

Recalling the relationship (18), then from theorem 4 we now obtain

$$\tau \quad \sim \quad \frac{K}{\sqrt{N}}\,\exp(AN) \tag{49}$$

where A is given by equation (41), and

$$K \quad = \quad \frac{K_0}{(1-\omega)\sum_i b_i(0)}, \tag{50}$$

with K_0 given by equation (42) and ω being the solution to equation (17).

The above heuristic argument is made rigorous for the case of the SIS infection model in $k = 1$ dimension, with the additional condition that lifeforce random variables are of finite variance, in Appendix B of [4]. The proof is rather lengthy, even in this particular case. It is conjectured in [4] that the finite variance condition may not be necessary.

Example 1. For the heterogeneous-population SIS infection model (example 1 of section 2.1), formula (49) reduces, after some algebraic simplification, to formula (6) of [9].

Example 2. For example 2 of section 2.1, the multitype birth-death process with linear birth rates and quadratic death rates, we obtain

$$A = \frac{k\lambda - \mu}{\kappa} + \frac{\mu}{\kappa} \log\left(\frac{\mu}{k\lambda}\right), \tag{51}$$

$$K = \frac{\sqrt{2\pi\mu\kappa}}{(1-\omega)\, k\lambda\, (k\lambda - \mu)}. \tag{52}$$

Note that for this process, if lifeforce variables are exponentially distributed then the total number of individuals $\sum_{i=1}^{k} X_i(t)$ is a 1-dimensional Markov process, and formulae (51), (52) can alternatively be obtained from equation (57) of [1].

6.3. Exponentially distributed lifeforces

In the case that the lifeforce random variables Q_i are exponentially distributed, so that $\boldsymbol{X}(t)$ is a Markov process, equations (16) become

$$\omega_i = \frac{d_0(0)d_i'(0)}{d_0(0)d_i'(0) + b_0'(0)\sum_{j=1}^{k}(1-\omega_j)b_j(0)} \text{ for } i = 1, 2, \ldots, k. \tag{53}$$

Setting $D = (b_0'(0)/\, d_0(0)) \sum_{j=1}^{k}(1-\omega_j)b_j(0)$, then equations (53) may be written as

$$\omega_i = \frac{d_i'(0)}{d_i'(0) + D} \text{ for } i = 1, 2, \ldots, k. \tag{54}$$

Substituting from (54) back into both sides of equations (53) and rearranging we find that for $R_0 > 1$, D is the unique positive solution of

$$\frac{b_0'(0)}{d_0(0)} \sum_{j=1}^{k} \left(\frac{b_j(0)}{d_j'(0) + D} \right) = 1.$$

Recalling the relationship (18) it is then straightforward to show that

$$\omega = 1 - \frac{Dd_0(0)}{b_0'(0)\sum_j b_j(0)},$$

so that equation (50) becomes

$$K = \frac{K_0 b_0'(0)}{Dd_0(0)}. \tag{55}$$

Formula (49) with K given by equation (55) can alternatively be obtained via the WKB approximation method, following the same steps as in [9]. Although neither the argument of section 6.2 nor the WKB method is entirely rigorous, the fact that the two methods are in agreement provides some reassurance.

7. Discussion

In this paper, we have greatly extended the class of multidimensional processes for which precise asymptotic formulae such as (1) are available. We have found that in the subcritical case, expected persistence time remains bounded as the system size parameter N grows to infinity; in the examples considered in the critical case, expected persistence time grows logarithmically with N; and in the supercritical case, expected persistence time grows exponentially with N. This is as one might expect. What is more surprising is that we were able to obtain such precise and explicit formulae as those provided by theorem 4 and formula (49). The key to our approach is that the restarted process of section 3 is time-reversible, although the details of our proofs are also dependent upon the particular form of the transition rate functions. The extension of our results to more general restart-reversible processes is the subject of ongoing work.

It is worth noting that while formulae such as (41) and (42) may appear formidable, their evaluation is essentially trivial. This contrasts with the standard characterization of mean persistence time from quasi-stationarity in terms of an eigenvalue of the transition rate matrix, in which the eigenvalue equation appears simple, but evaluation of the relevant eigenvalue can present a formidable computational challenge.

For supercritical processes, arguably of greatest interest is the mean extinction time from quasi-equilibrium, τ. For the mean extinction time τ_i starting from a single individual of type i, theorems 1, 2 and 4 provide limiting and asymptotic formulae only for the linear combination $\sum_{i=1}^{k} b_i(0)\tau_i$. In the subcritical case, we conjectured in section 4 that $\tau_i \to -(\log(1-R_0)) / (d_0(0)d_i'(0)R_0)$ as $N \to \infty$ for $i = 1, 2, \ldots, k$. In the supercritical case, the heuristic arguments of section 6.2 imply that $\tau_i \sim (1-\omega_i)\tau$ as $N \to \infty$ (relationship (48)), which together with formula (49) gives the asymptotic form of τ_i for $i = 1, 2, \ldots, k$.

We have focused here upon processes with a single absorbing state at the origin. Another interesting class of population processes is those for which absorption occurs at the boundary $A_0 = \{\boldsymbol{x} \in \mathbb{Z}_+^k : x_i = 0 \text{ for some } i\}$. The study of such processes is a topic for future work.

Appendix A. Proofs of equations (29) and (33)

Consider first equation (29). Fix $V, \gamma > 0$ and for $N > (V\gamma)^2$ let

$$S_N = \prod_{r=0}^{\lfloor V\sqrt{N}\rfloor} \left(1 - \frac{r\gamma}{N}\right),$$

so

$$\log S_N = \sum_{r=0}^{\lfloor V\sqrt{N}\rfloor} \log\left(1 - \frac{r\gamma}{N}\right).$$

By Taylor's theorem with the Lagrange form of the remainder, for $x \in (0,1)$,

$$-x - \frac{x^2}{2(1-x)^2} < \log(1-x) < -x.$$

Hence,

$$\limsup_{N\to\infty} \log S_N \leq \limsup_{N\to\infty} \left(-\frac{\gamma}{N}\sum_{r=0}^{\lfloor V\sqrt{N}\rfloor} r\right) = -\frac{\gamma V^2}{2}. \tag{56}$$

Also, $1 - \frac{r\gamma}{N} \geq 1 - \frac{V\gamma}{N^{\frac{1}{2}}}$, for $r = 0, 1, \dots, \lfloor V\sqrt{N}\rfloor$, so

$$\lim_{N\to\infty} \sum_{r=0}^{\lfloor V\sqrt{N}\rfloor} \left(\frac{\frac{r\gamma}{N}}{1 - \frac{r\gamma}{N}}\right)^2 \leq \lim_{N\to\infty} \left(\frac{1}{1 - \frac{V\gamma}{N^{\frac{1}{2}}}}\right)^2 \frac{\gamma^2}{N^2} \frac{\lfloor V\sqrt{N}\rfloor(\lfloor V\sqrt{N}\rfloor + 1)(2\lfloor V\sqrt{N}\rfloor + 1)}{6} = 0.$$

Thus,

$$\liminf_{N\to\infty} \log S_N \geq -\frac{\gamma V^2}{2},$$

which together with (56) implies

$$\lim_{N\to\infty} S_N = \exp\left(-\frac{\gamma V^2}{2}\right),$$

and (29) follows.

Turning to equation (33), recall that $\eta, L > 0$, let $c_N = \lfloor LN^{\frac{\eta}{1+\eta}}\rfloor$ and

$$R_N = \prod_{i=1}^{\lfloor LN^{\frac{\eta}{1+\eta}}\rfloor} \left(1 + \theta(i/N)^\eta\right).$$

For $x > 0$,

$$x - \frac{x^2}{2} < \log(1+x) < x,$$

so

$$d_N - \frac{e_N}{2} < \log R_N < d_N,$$

where

$$d_N = \frac{\theta}{N^\eta} \sum_{i=1}^{c_N} i^\eta \qquad \text{and} \qquad e_N = \frac{\theta^2}{N^{2\eta}} \sum_{i=1}^{c_N} i^{2\eta}.$$

Now,

$$\frac{c_N^{\eta+1}}{\eta+1} = \int_0^{c_N} x^\eta dx < \sum_{i=1}^{c_N} i^\eta < \int_0^{c_N+1} x^\eta dx = \frac{(c_N+1)^{\eta+1}}{\eta+1}.$$

Also, $\lim_{N\to\infty} N^{-\eta} c_N^{\eta+1} = L^{\eta+1}$, so $\lim_{N\to\infty} d_N = \frac{\theta L^{\eta+1}}{\eta+1}$. Further,

$$0 < N^{-2\eta} c_N^{2\eta+1} \leq L^{2\eta+1} N^{\frac{\eta(2\eta+1)}{\eta+1} - 2\eta} = L^{2\eta+1} N^{-\frac{\eta}{\eta+1}},$$

so $\lim_{N\to\infty} e_N = 0$. Thus,

$$\lim_{N\to\infty} \log R_N = \frac{\theta L^{\eta+1}}{\eta+1}$$

and (33) follows.

Acknowledgement

It is a pleasure to thank two anonymous reviewers for detailed remarks that improved the presentation of this paper.

Funding information

There are no funding bodies to thank relating to the creation of this article.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] Assaf, M. and Meerson, B. (2010). Extinction of metastable stochastic populations. *Phys. Rev. E* **81,** 021116.

[2] Assaf, M. and Meerson, B. (2017). WKB theory of large deviations in stochastic populations. *J. Phys. A Math. Theor.* **50,** 263001.

[3] Ball, F. G. (1983). The threshold behaviour of epidemic models. *J. Appl. Probab.* **20,** 227–241.

[4] Ball, F. G., Britton, T. and Neal, P. (2014). On expected durations of birth-death processes, with applications to branching processes and SIS epidemics. *arXiv* 1408.0641.

[5] Ball, F. G., Britton, T. and Neal, P. (2016). On expected durations of birth-death processes, with applications to branching processes and SIS epidemics. *J. Appl. Probab.* **53,** 203–215.

[6] Ball, F. G. and O'Neill, P. D. (1994). Strong convergence of stochastic epidemics. *Adv. Appl. Probab.* **26,** 629–655.

[7] Chazottes, J. B., Collet, P. and Méléard, S. (2019). On time scales and quasi-stationary distributions for multitype birth-and-death processes. *Ann. Inst. H. Poincaré Probab. Statist.* **55,** 2249–2294.

[8] Clancy, D. (2018). Persistence time of SIS infections in heterogeneous populations and networks. *J. Math. Biol.* **77,** 545–570.

[9] Clancy, D. (2018). Precise estimates of persistence time for SIS infections in heterogeneous populations. *Bull. Math. Biol.* **80,** 2871–2896.

[10] Clancy, D. and Pollett, P. K. (2003). A note on quasi-stationary distributions of birth-death processes and the SIS logistic epidemic. *J. Appl. Probab.* **40,** 821–825.

[11] Davis, M. H. A. (1993). *Markov Models and Optimization.* Chapman and Hall, London.

[12] Diekmann, O., Heesterbeek, J. A. P. and Roberts, M. G. (2010). The construction of next-generation matrices for compartmental epidemic models. *J. R. Soc. Interface* **7,** 873–885.

[13] Ethier, S. N. and Kurtz, T. G. (2005). *Markov Processes: Characterization and Convergence.* Wiley, New York.

[14] Hernàndez-Suàrez, C. M. and Castillo-Chavez, C. (1999). A basic result on the integral for birth-death markov processes. *Math. Biosci.* **161,** 95–104.

[15] Keilson, J. and Ramaswamy, R. (1984). Convergence of quasistationary distributions in birth-death processes. *Stoch. Proc. Applic.* **18,** 301–312.

[16] Kelly, F. P. (2011). *Reversibility and Stochastic Networks.* Cambridge University Press, Cambridge.

[17] Kryscio, R. J. and Lefèvre, C. (1989). On the extinction of the S-I-S stochastic logistic epidemic. *J. Appl. Probab.* **26,** 685–694.

[18] Laub, A. J. (2005). *Matrix Analysis for Scientists and Engineers.* SIAM publications, Philadelphia.

[19] Mode, C. J. (1971). *Multitype Branching Processes: Theory and Applications.* Elsevier, New York.

[20] Ovaskainen, O. and Meerson, B. (2010). Stochastic models of population extinction. *Trends in Ecology & Evolution* **25,** 643–652.

[21] Sellke, T. (1983). On the asymptotic distribution of the size of a stochastic epidemic. *J. Appl. Probab.* **20,** 390–394.

[22] van den Driessche, P. and Watmough, J. (2002). Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosci.* **180,** 29–48.

[23] Zachary, S. (2007). A note on insensitivity in stochastic networks. *J. Appl. Probab.* **44,** 238–248.