

MULTI-TYPE BRANCHING PROCESSES WITH TIME-DEPENDENT BRANCHING RATES

D. DOLGOPYAT,* *Univeristy of Maryland*

P. HEBBAR,* *Univeristy of Maryland*

L. KORALOV,* *Univeristy of Maryland*

M. PERLMAN,** *Stanford Univeristy*

Abstract

Under mild non-degeneracy assumptions on branching rates in each generation, we provide a criterion for almost-sure extinction of a multi-type branching process with time-dependent branching rates. We also provide a criterion for the total number of particles (conditioned on survival and divided by the expectation of the resulting random variable) to approach an exponential random variable as time goes to infinity.

Keywords: Multi-type branching; extinction probability; exponential limit law; nonnegative matrix product

2010 Mathematics Subject Classification: Primary 60J80
Secondary 60F05; 60F10

1. Introduction

Mathematical study of branching processes goes back to the work of Galton and Watson [23] who were interested in the probabilities of long-term survival of family names. Later it was realized that similar mathematical models could be used to describe the evolution of a variety of biological populations, in genetics [9, 10, 11, 13], and in the study of certain chemical and nuclear reactions [21, 14]. Branching processes are central in the study of evolution of various populations such as bacteria, cancer cells, carriers of a particular form of a gene, etc., where each member of the population may die or produce offspring independently of the rest.

The individuals involved in the process are referred to as particles. In many models, the particles may be of different types, representing individuals with different characteristics. For example, in epidemiology, multi-type continuous time Markov branching process may be used to describe the dynamics of the spread of parasites of two types that can mutate into each other in a common host population [6]; when modeling cancer, particles of different types may represent cells that have accumulated different numbers of mutations [8]; in physics, cosmic ray cascades, which involve electrons producing photons and photons producing electrons, can be modeled by a 2-type branching process [17]. In addition, a vast number of applications of multi-type branching processes in biology can be found in [12, 19].

The current paper concerns the long-time behavior of multi-type branching processes with time-dependent branching rates. Let us stress that the temporal inhomogeneity is due to the dependence of the branching rates not on the ages of the particles (which is a well-studied model), but on time (this dependence may model a varying environment for the entire process). We believe that the methods of our paper could be used to handle more general models such as those where, in addition to time dependence, the branching rate may depend on the age of the particles and/or on their spatial location if the spatial motion in a bounded domain is allowed. This may be a subject of future work.

* Postal address: Dept. of Mathematics, University of Maryland, College Park, MD 20742

** Postal address: Dept of Mathematics, Stanford University, Stanford, CA 94305

For multi-type processes with constant branching rates, according to classical results (see Chapter 5 of [2] and references therein), three different cases can be distinguished. In the super-critical case, the expectation of the total population size grows exponentially, and the total population grows exponentially with positive probability as time goes to infinity. In the sub-critical case, the expectation of the total population size decays exponentially, and the population goes extinct with overwhelming probability, i.e., the probability that the population at time n is non-zero decays exponentially in n . In the critical case, the population also goes extinct, but the expectation of the total population size remains bounded away from zero and infinity, and the probability of survival decays as c/n for some $c > 0$. Moreover, after conditioning on survival, the size of the population divided by its expectation tends to an exponential random variable. Whether the process is super-critical, sub-critical, or critical, can be easily determined by examining the (constant) branching rates.

The question we address in the case of time-dependent branching rates is how to distinguish between different kinds of asymptotic behavior of the process based on the behavior of the branching rates. Our first result gives a criterion for almost sure extinction of the process in terms of the asymptotic behavior of the branching rates, under mild non-degeneracy assumptions on the branching rates at each time step. In the case of single-type branching processes, a similar result was obtained by Agresti [1]. An earlier partial result in this direction (for single-type branching processes) was obtained by Jagers [15], who also provided a sufficient condition for the exponential limit (in distribution) of the size of the population (after conditioning on survival and dividing by the expectation of the resulting random variable). For single type branching processes, a necessary and sufficient condition for exponential distribution of particle number conditioned on survival in terms of the branching rates was obtained independently in [4, 18]. Our second result gives a necessary and sufficient condition for the existence of such an exponential limit in the case of multi-type branching processes.

Based on our results, it is natural to classify all the branching processes with time-dependent branching rates (under the non-degeneracy assumptions) into three categories, based on their asymptotic behavior. Processes in the first category (which includes super-critical processes with time-independent rates), are distinguished by a positive probability of survival for infinite time. Processes in the second category (which includes critical processes with time-independent rates) go extinct with probability one, and the size of the population, after conditioning on survival and normalization, tends to the exponential limit. Processes in the third category (which includes sub-critical processes with time-independent rates) go extinct with probability one, but do not have the exponential limit.

It should be stressed that, in contrast to the case of time-independent rates (when the expected population size either grows exponentially, decays exponentially, or is asymptotically constant), now the expected population size may fluctuate greatly in each of the cases, which makes the analysis more complicated.

Let us also remark that some of the classical results on the asymptotic behavior of branching processes in the time-independent case carry over to the case at hand, while others do not. For example, in the time-independent case, super-critical processes have the property that the process normalized by expected population size tends to a random limit. An analogue of this statement still holds in the case of time-dependent branching rates, as follows from the results of [16]. Further results on L^p and almost sure convergence, including those in the case of countably many particle types, can be found in [3]. Sufficient conditions for the continuity of the limiting distribution function were given in [7].

On the other hand, in the time-independent case, a sub-critical process conditioned on survival tends to a random limit. Now, our processes in the third category do not necessarily have this property (e.g., the population, conditioned on survival, may grow along a subsequence). A more detailed analysis of the near-critical behavior of processes with time-dependent rates will be the subject of a subsequent paper.

In the next section, we introduce the relevant notation and formulate the main results. The proofs are presented in Sections 3 and 4. In Section 5, we briefly discuss an application of our results to the case of continuous time branching.

2. Notation and statement of main results

Let $S = \{1, \dots, d\}$ be the set of possible particles types. Suppose that for each $i \in S$ and $n \geq 0$ there is a distribution $P_n(i, \cdot)$ on \mathbb{Z}_+^d . For $a = (a_1, \dots, a_d) \in \mathbb{Z}_+^d$, $P_n(i, a)$ represents the probability that a particle of type i that is alive at time n is replaced in the next generation by $a_1 + \dots + a_d$ particles: a_1 particles of type one, a_2 particles of type two, etc. A d -type branching process $\{Z_n\}$ is obtained by starting with a positive finite number of particles at time zero, and then replacing each particle of each type i , $i \in S$, that is alive at time n , $n \geq 0$, by particles of various types according to the distribution $P_n(i, \cdot)$ independently of the other particles alive at time n and of the past, thus obtaining the population at time $n + 1$.

We write $Z_n = (Z_n(1), \dots, Z_n(d))$, where $Z_n(i)$ is the number of particles of type i at time n . When the initial population consists of one particle of type j , we may write ${}_jZ_n(i)$ to represent the number of particles of type i at time n . Thus $\mathbb{E}({}_jZ_n(i))$ means the same as $\mathbb{E}(Z_n(i) | Z_0 = e_j)$, where e_j is the unit vector in the j -th direction. Let ${}_jX_n$ denote a generic random vector with distribution $P_n(j, \cdot)$.

For $s = (s_1, \dots, s_d) \in [0, 1]^d$, let

$$f_n^j(s) = \mathbb{E}\left(\prod_{i=1}^d s_i^{Z_n(i)} | Z_0 = e_j\right),$$

$$g_n^j(s) = \mathbb{E}\left(\prod_{i=1}^d s_i^{Z_{n+1}(i)} | Z_n = e_j\right).$$

At times, we may drop the superscript from either of those expressions, and then $f_n(s)$ and $g_n(s)$ become vectors. Note that

$$f_n(s) = f_{n-1}(g_{n-1}(s)) = (g_0 \circ g_1 \circ \dots \circ g_{n-1})(s), \text{ and } f_n(\mathbf{1}) = \mathbf{1}$$

where $\mathbf{1} = (1, \dots, 1)$. We also define

$$f_{k,n}(s) = (g_k \circ \dots \circ g_{n-1})(s).$$

Thus $f_{0,n} = f_n$. We denote

$$M_n(j, i) = \frac{\partial f_n^j}{\partial s_i}(\mathbf{1}) = \mathbb{E}(Z_n(i) | Z_0 = e_j),$$

$$A_n(j, i) = \frac{\partial g_n^j}{\partial s_i}(\mathbf{1}) = \mathbb{E}(Z_{n+1}(i) | Z_n = e_j),$$

Then,

$$M_n = A_0 \dots A_{n-1},$$

where A_n and M_n are viewed as matrices. Also define

$$M_{k,n} = A_k \dots A_{n-1}.$$

Let $\|\cdot\|$ denote the following norm of a d -dimensional vector: $\|v\| = |v_1| + \dots + |v_d|$. We will use certain non-degeneracy assumptions on the distribution of descendants at each step. We assume that there are $\varepsilon_0, K_0 > 0$ such that for all $i, j \in S$ the following bounds hold.

1. $\mathbb{P}(Z_{n+1}(i) \geq 2 | Z_n = e_j) \geq \varepsilon_0$.
2. $\mathbb{P}(Z_{n+1} = \mathbf{0} | Z_n = e_j) \geq \varepsilon_0$.
3. $\mathbb{E}(\|Z_{n+1}\|^2 | Z_n = e_j) \leq K_0$.

The following proposition is a generalization of the Perron-Frobenius theorem to the case when the positive matrices forming a product are allowed to be distinct.

Proposition 2.1. *Under Assumptions 1 and 3, there are two sequences of vectors $v_n, u_n \in \mathbb{R}^d$, $n \geq 0$, such that*

- (a) $\|u_n\| = \|v_n\| = 1$.
- (b) $v_n(i), u_n(i) \geq \bar{\varepsilon}$, for some $\bar{\varepsilon} > 0$ and all $n \geq 0, i \in S$,
- (c) *There are sequences of positive numbers λ_n and $\tilde{\lambda}_n$ and a positive constant a such that $\lambda_n, \tilde{\lambda}_n \in (a^{-1}, a)$ for $n \geq 0$ and*

$$A_{n-1}v_n = \lambda_{n-1}v_{n-1}, \quad A_{n-1}^T u_{n-1} = \tilde{\lambda}_{n-1}u_n.$$

- (d) *For each $\delta > 0$ there is $k' \in \mathbb{N}$ such that*

$$(1 - \delta)v_n \leq \frac{M_{n,n+k}v}{\|M_{n,n+k}v\|} \leq (1 + \delta)v_n, \quad (1 - \delta)u_{n+k} \leq \frac{M_{n,n+k}^T u}{\|M_{n,n+k}^T u\|} \leq (1 + \delta)u_{n+k}$$

whenever $k \geq k'$, v and u are non-zero vectors with non-negative components, and inequality between vectors is understood as the inequality between their components.

- (e) *There is $K > 0$ such that if we define $\Lambda_n = \prod_{i=0}^{n-1} \lambda_i$ and $\tilde{\Lambda}_n = \prod_{i=0}^{n-1} \tilde{\lambda}_i$, then*

$$\frac{1}{K} \leq \frac{\Lambda_n}{\tilde{\Lambda}_n} \leq K, \quad \frac{1}{K} \leq \frac{M_{k,n}(j,i)}{(\Lambda_n/\Lambda_k)} \leq K, \quad j, i \in S.$$

This proposition can be derived from the results of Chapter 3 of [22], for example. Indeed, from our Assumptions 1-3, it follows that the matrices A_n have the Birkhoff's contraction coefficient (in the terminology of [22]) uniformly bounded away from one. This implies that the conditions of Lemma 3.4. of [22] are met (which, in particular, implies that the family $M_{k,n}$ is weakly ergodic (see [22])). This lemma and Exercise 3.5 of [22] easily imply the existence of vectors u_n and v_n . Their required properties are also not difficult to establish. For the sake of completeness we provide an independent proof in Appendix A.

Remark 2.1. The vectors v_n and the numbers λ_n are uniquely defined by the above conditions, as seen from the proof of the Proposition. The vectors u_n and the numbers $\tilde{\lambda}_n$ will be defined uniquely by specifying u_0 , which we assume to be fixed as an arbitrary vector satisfying conditions (a) and (b).

The probabilistic meaning of vectors u_n and v_n is the following. The vector u_n gives the asymptotic proportions of different particles in the population provided that Z_n is large (see (4.11), (4.12) for the precise statement). To see the meaning of v_n , consider the total number of particles at time n , $z_n^* = \langle Z_n, \mathbf{1} \rangle$. It will be apparent from the proof of Proposition 2.1 that

$$\lim_{N \rightarrow \infty} \frac{\mathbb{E}(z_N^* | Z_n = u')}{\mathbb{E}(z_N^* | Z_n = u'')} = \frac{\langle v_n, u' \rangle}{\langle v_n, u'' \rangle}$$

for each $u', u'' \in \mathbb{Z}^d$ and each $n \in \mathbb{Z}_+$. Thus v_n controls the expected future size of the population.

Our first result gives a necessary and sufficient condition for the almost sure extinction of $\{Z_n\}$.

Theorem 2.1. *Under Assumptions 1-3, if extinction of the process $\{Z_n\}$ occurs with probability one for some initial population, then $\sum_{k=1}^{\infty} (1/\Lambda_k) = \infty$. If $\sum_{k=1}^{\infty} (1/\Lambda_k) = \infty$, then extinction with probability one occurs for every initial population.*

Remark 2.2. Here and below, when we talk about initial population, we mean that $Z_0 = u$ for some deterministic vector u .

Remark 2.3. The first statement of the theorem can be deduced from the results of [16]. In fact, the assumptions needed for the first part are weaker than our assumptions above. For example, weak ergodicity (see [16]) is sufficient. However, the assumption that the matrices A_k are uniformly bounded from below plays an important role in our proof of the second statement, as well as in the proof of Theorem 2.2 below. We note that finding the least restrictive conditions for the validity of Theorems 2.1 and 2.2 remains an interesting open problem. We refer the reader to the paper [18] for recent results in the case of single-type branching processes.

The following lemma easily follows from Theorem 2.1.

Lemma 2.1. *Suppose Assumptions 1-3 hold.*

(a) *Given $l \geq 0$, consider the process $\{Z'_n\}$ that starts with one particle of type j alive at time l followed by branching with the distributions P_l, P_{l+1}, \dots (where the distributions P_i are used in the definition of the branching process in the beginning of the section). Extinction for this process occurs with probability one if and only if $\sum_{k=1}^{\infty} (1/\Lambda_k) = \infty$.*

(b) *Given $l > 0$, the extinction of $\{Z_n\}$ (or, equivalently, $\{Z'_n\}$) occurs with probability one if and only if $\sum_{k=1}^{\infty} \frac{1}{\Lambda_{lk}} = \infty$.*

Remark 2.4. The divergence of $\sum_{k=1}^{\infty} \frac{1}{\Lambda_{lk}}$ is the extinction condition for the process $\{Z_{ln}\}$ obtained by observing our process only at the moments of time that are multiples of l .

Proof of Lemma 2.1. For part (a) it suffices to note that $\sum_{k=1}^{\infty} \frac{1}{\Lambda_k} = \infty$ if and only if $\sum_{k=l+1}^{\infty} \frac{\Lambda_l}{\Lambda_k} = \infty$, while the latter is equivalent to the almost sure extinction of the process $\{Z'_n\}$ by Theorem 2.1.

To prove part (b), we observe that, under Assumption 3, there exists a constant C such that for each k and each $lk \leq n < l(k+1)$ we have $\frac{\Lambda_{lk}}{C} \leq \Lambda_n \leq C\Lambda_{lk}$. \square

The following lemma will be derived in the end of the next section using the formulas encountered in the proof of Theorem 2.1.

Lemma 2.2. *Under Assumptions 1-3, for each initial population of the branching process, there is a constant $C > 0$ such that*

$$\frac{\Lambda_n}{C} \leq \mathbb{E}\|Z_n\| \leq C\Lambda_n, \quad n \geq 1, \quad (2.1)$$

$$\frac{1}{C} \left(\sum_{k=1}^n \frac{1}{\Lambda_k} \right)^{-1} \leq \mathbb{P}(Z_n \neq \mathbf{0}) \leq C \left(\sum_{k=1}^n \frac{1}{\Lambda_k} \right)^{-1}, \quad n \geq 1. \quad (2.2)$$

To formulate the next theorem, we will make use of the following assumptions:

4. The random variables $\|jX_n\|^2$, $j \in S$, $n \geq 0$, are uniformly integrable.
5. $\mathbb{P}(Z_n \neq \mathbf{0}) \rightarrow 0$ as $n \rightarrow \infty$ (equivalently, $\sum_{k=1}^n (1/\Lambda_k) \rightarrow \infty$, by (2.2)).
6. $\mathbb{E}\|Z_n\|/\mathbb{P}(Z_n \neq \mathbf{0}) \rightarrow \infty$ as $n \rightarrow \infty$ (equivalently, $\Lambda_n \sum_{k=1}^n (1/\Lambda_k) \rightarrow \infty$, by (2.1), (2.2)).

Let $\zeta_n = (\zeta_n(1), \dots, \zeta_n(d))$ be the random vector obtained from Z_n by conditioning on the event that $Z_n \neq \mathbf{0}$. In other words, we treat the event $Z_n \neq \mathbf{0}$ as a new probability space, with the measure \mathbb{P}' obtained from the underlying measure \mathbb{P} via $\mathbb{P}'(A) = \mathbb{P}(A)/\mathbb{P}(Z_n \neq \mathbf{0})$. When we write $j\zeta_n$, we mean that the initial population for the branching process is specified as e_j .

We will prove exponential limit for the multi-type random variable under the assumptions listed above.

Theorem 2.2. *Under Assumptions 1-6, for each initial population of the branching process and each vector u with positive components, we have the following limit in distribution*

$$\frac{\langle \zeta_n, u \rangle}{\mathbb{E}\langle \zeta_n, u \rangle} \rightarrow \xi, \quad \text{as } n \rightarrow \infty, \quad (2.3)$$

where ξ is an exponential random variable with parameter one. Moreover, if Assumptions 1-5 are satisfied and, for some initial population, the limit in (2.3) is as specified, then Assumption 6 is also satisfied.

We say that a process is *uniformly critical* if it satisfies Assumptions 1-4 and there is a constant b such that for each n, k, i, j , we have

$$\frac{1}{b} \leq M_{n,n+k}(j, i) \leq b. \quad (2.4)$$

For uniformly critical processes, Λ_k are uniformly bounded from above and below, so we have

$$\sum_{k=1}^{\infty} \frac{1}{\Lambda_k} = \infty, \quad \lim_{n \rightarrow \infty} \Lambda_n \left(\sum_{k=1}^n \frac{1}{\Lambda_k} \right) = \infty.$$

Therefore, uniformly critical processes become extinct with probability 1 and the distribution of the appropriately scaled number of particles at time n , conditioned on survival, converges to exponential.

The next proposition and the lemma that follows will be helpful for comparing our results to those of [15]. An important part of Proposition 2.2 (part (d)) shows that, under (2.4) (or even under a weaker condition (2.5)), Assumption 2 almost follows from Assumption 1, in the sense that Assumption 2 is satisfied for an appropriate subprocess. Given l , let \tilde{P}_n be the transition probability of the process $\{\tilde{Z}_n\}$, where $\tilde{Z}_n = Z_{nl}$. That is, $\tilde{P}_n(i, a)$ represents the probability that a particle of type i that is alive at time nl is replaced in generation $(n+1)l$ by $a_1 + \dots + a_d$ particles: a_1 particles of type one, a_2 particles of type two, etc.

Proposition 2.2. (a) If P_n satisfies Assumption 1, then \tilde{P}_n satisfies Assumption 1 for each l .

(b) If P_n satisfies Assumption 3, then \tilde{P}_n satisfies Assumption 3 for each l .

(c) If P_n satisfies Assumption 4, then \tilde{P}_n satisfies Assumption 4 for each l .

(d) If P_n satisfies Assumption 1, and there is a constant \mathbf{b} such that for each n, k, j

$$\mathbb{E}(|Z_{n+k}| | Z_n = e_j) \leq \mathbf{b}, \quad (2.5)$$

then there exist $l = l(\varepsilon_0, \mathbf{b})$ and $\varepsilon_1 = \varepsilon_1(\varepsilon_0, \mathbf{b})$ such that for each n and j

$$\tilde{P}_n(\tilde{Z}_{n+1} = \mathbf{0} | \tilde{Z}_n = e_j) \geq \varepsilon_1.$$

The above Proposition is proved in Appendix B. The following lemma is proved in Section 4.

Lemma 2.3. If $\{Z_n\}$ satisfies Assumptions 1 and 4, and (2.4) holds, then extinction happens with probability one and (2.3) holds.

For single type branching processes, Lemma 2.3 is helpful in showing that our results imply Theorem 5 of [15]. In fact, the assumptions of Theorem 5 of [15] (generalized to the multi-type case) are: (a) our Assumption 4, (b) that (2.4) holds, and (c) that there is $\bar{\varepsilon}_0 > 0$ such that for each n, i, j

$$\mathbb{E}(Z_{n+1}^2(i) - Z_{n+1}(i) | Z_n = e_j) \geq \bar{\varepsilon}_0. \quad (2.6)$$

We claim that under Assumption 4, (2.6) is equivalent to Assumption 1. On the one hand,

$$\mathbb{E}(Z_{n+1}^2(i) - Z_{n+1}(i) | Z_n = e_j) \geq 2\mathbb{P}(Z_{n+1}(i) \geq 2 | Z_n = e_j).$$

On the other hand, by Assumption 4, we can take $N \geq 2$ such that

$$\mathbb{E}(Z_{n+1}^2(i) \chi_{Z_{n+1}(i) \geq N} | Z_n = e_j) \leq \frac{\bar{\varepsilon}_0}{2}$$

for all n , where $\chi_{Z_{n+1}(i) \geq N}$ is the indicator function of the event $\{Z_{n+1}(i) \geq N\}$. Then

$$\mathbb{E}(Z_{n+1}^2(i) - Z_{n+1}(i) | Z_n = e_j) \leq \frac{\bar{\varepsilon}_0}{2} + (N^2 - N)\mathbb{P}(Z_{n+1}(i) \geq 2 | Z_n = e_j).$$

Thus if (2.6) holds, then

$$\mathbb{P}(Z_{n+1}(i) \geq 2 | Z_n = e_j) \geq \frac{\bar{\varepsilon}_0}{2(N^2 - N)},$$

proving that Assumption 1 is equivalent to (2.6).

The results of Theorem 5 of [15] are: our Lemma 4.3 (in the single type case) and our formula (2.3) (in the single type case). The latter holds by Lemma 2.3. We prove Lemma 4.3 in Section 4 under Assumptions 1-6. However, under Assumptions 1, 4, and (2.4), the conclusion of the lemma still holds (the argument is similar to that in the proof of Lemma 2.3).

3. Survival vs extinction

Proof of Theorem 2.1. (PART I) $\sum_{k=1}^{\infty} (1/\Lambda_k) < \infty$ implies positive probability of survival.

Let us fix $Z_0 = e_j$ with an arbitrary $j \in S$. Let \mathcal{F}_n be the σ -algebra generated by the branching process $\{Z_n\} = \{jZ_n\}$. Let $z_n = \langle Z_n, v_n \rangle$. Then,

$$\mathbb{E}(z_{n+1} | \mathcal{F}_n) = \langle \mathbb{E}(Z_{n+1} | Z_n), v_{n+1} \rangle = \langle A_n^T Z_n, v_{n+1} \rangle = \langle Z_n, A_n v_{n+1} \rangle = \lambda_n z_n$$

Accordingly, $\{z_n/\Lambda_n\}$ is a positive martingale, and hence it converges to some random variable z_∞ . Now let

$$D_n(j_1, j_2) = \text{Cov}(Z_n(j_1), Z_n(j_2)).$$

One step analysis gives

$$D_{n+1} = A_n^T D_n A_n + S_n, \tag{3.1}$$

where

$$S_n = \sum_{i=1}^d M_n(j, i) \sigma_n^2(i)$$

and

$$\sigma_n^2(j_1, j_2)(i) = \text{Cov}(iX_n(j_1), iX_n(j_2)).$$

By Proposition 2.1, there exists a constant B such that, $\|S_n\| \leq B\Lambda_n$, where $\|\cdot\|$ is a matrix norm. Iterating (3.1), we get

$$D_n = \sum_{k=0}^{n-1} M_{k+1,n}^T S_k M_{k+1,n}.$$

Hence,

$$\|D_n\| \leq B_1 \sum_{k=0}^{n-1} \left(\frac{\Lambda_n}{\Lambda_{k+1}} \right)^2 \Lambda_k \leq B_2 \Lambda_n^2 \sum_{k=0}^{n-1} \frac{1}{\Lambda_k}$$

with some constants B_1, B_2 .

Thus $\|D_n\| \leq \tilde{B} \Lambda_n^2$, and so the martingale $\{z_n/\Lambda_n\}$ is uniformly bounded in L^2 . Therefore, $\mathbb{E}(z_\infty) = \mathbb{E}(z_0) > 0$, and hence $\mathbb{P}(z_\infty > 0) > 0$, implying that the probability of survival of the branching process starting with a single particle of type j is positive. Therefore, the probability of survival is positive for every initial population.

(PART II) $\sum_{k=1}^{\infty} (1/\Lambda_k) = \infty$ implies that extinction occurs with probability one.

Recall that $f_{0,n}(s) = g_0(f_{1,n}(s))$ and $f_{0,n}(\mathbf{1}) = g_0(\mathbf{1}) = \mathbf{1}$. Determining the asymptotic behavior of $\langle \mathbf{1} - f_{0,n}(s), u_0 \rangle$ will be helpful for proving the theorem and also later in the proof of (4.5). By the Taylor formula with respect to $s = \mathbf{1}$,

$$\langle \mathbf{1} - f_{0,n}(s), u_0 \rangle = \langle Dg_0(\mathbf{1})(\mathbf{1} - f_{1,n}(s)), u_0 \rangle - \frac{1}{2} \langle (\mathbf{1} - f_{1,n}(s))^T Hg_0(\eta_{1,n})(\mathbf{1} - f_{1,n}(s)), u_0 \rangle$$

$$= \langle A_0(\mathbf{1} - f_{1,n}(s)), u_0 \rangle - \frac{1}{2} \langle (\mathbf{1} - f_{1,n}(s))^T Hg_0(\eta_{1,n})(\mathbf{1} - f_{1,n}(s)), u_0 \rangle,$$

where Dg_0 is the gradient of g_0 and $\eta_{1,n} = \eta_{1,n}(j, s)$ satisfies $f_{0,n}^j(s) \leq \eta_{1,n} \leq 1$ for each component $j \in S$ and $s \in [0, 1]^d$. Here Hg_0 stands for the Hessian matrix applied to each component of the vector function g_0 separately, then multiplied by vectors $(\mathbf{1} - f_{1,n}(s))^T$ and $(\mathbf{1} - f_{1,n}(s))$ to get scalars, which are then multiplied by the corresponding components of u_0 to form the scalar product. Therefore, by taking the transpose of A_0 ,

$$\begin{aligned} \langle \mathbf{1} - f_{0,n}(s), u_0 \rangle &= \langle (\mathbf{1} - f_{1,n}(s)), A_0^T u_0 \rangle - \frac{1}{2} \langle (\mathbf{1} - f_{1,n}(s))^T Hg_0(\eta_{1,n})(\mathbf{1} - f_{1,n}(s)), u_0 \rangle \\ &= \langle (\mathbf{1} - f_{1,n}(s)), \tilde{\lambda}_0 u_1 \rangle - \frac{1}{2} \langle (\mathbf{1} - f_{1,n}(s))^T Hg_0(\eta_{1,n})(\mathbf{1} - f_{1,n}(s)), u_0 \rangle. \end{aligned}$$

Thus, for $s \neq \mathbf{1}$,

$$\begin{aligned} \langle \mathbf{1} - f_{0,n}(s), u_0 \rangle^{-1} &= \left(\langle (\mathbf{1} - f_{1,n}(s)), \tilde{\lambda}_0 u_1 \rangle - \frac{1}{2} \langle (\mathbf{1} - f_{1,n}(s))^T Hg_0(\eta_{1,n})(\mathbf{1} - f_{1,n}(s)), u_0 \rangle \right)^{-1} \\ &= (\tilde{\lambda}_0 \langle (\mathbf{1} - f_{1,n}(s)), u_1 \rangle)^{-1} \left(1 + \frac{\langle -\frac{1}{2}(\mathbf{1} - f_{1,n}(s))^T Hg_0(\eta_{1,n})(\mathbf{1} - f_{1,n}(s)), u_0 \rangle}{\tilde{\lambda}_0 \langle (\mathbf{1} - f_{1,n}(s)), u_1 \rangle} \right)^{-1} \\ &= \frac{1}{\tilde{\lambda}_0 \langle (\mathbf{1} - f_{1,n}(s)), u_1 \rangle} + \frac{\langle \frac{1}{2}(\mathbf{1} - f_{1,n}(s))^T Hg_0(\eta_{1,n})(\mathbf{1} - f_{1,n}(s)), u_0 \rangle}{\tilde{\lambda}_0 \langle (\mathbf{1} - f_{1,n}(s)), u_1 \rangle \langle \mathbf{1} - f_{0,n}(s), u_0 \rangle}, \end{aligned}$$

where the last equality follows from the simple relation

$$\frac{1}{a} = \frac{1}{b} \left(1 - \frac{c}{b} \right)^{-1} \implies \frac{1}{a} = \frac{1}{b} + \frac{c}{ba}.$$

By iterating the previous equality n times, we get

$$\langle \mathbf{1} - f_{0,n}(s), u_0 \rangle^{-1} = \frac{1}{\tilde{\Lambda}_n \langle \mathbf{1} - s, u_n \rangle} + \frac{1}{2} \sum_{k=0}^{n-1} \frac{\langle (\mathbf{1} - f_{k+1,n}(s))^T Hg_k(\eta_{k+1,n})(\mathbf{1} - f_{k+1,n}(s)), u_k \rangle}{\tilde{\Lambda}_{k+1} \langle (\mathbf{1} - f_{k+1,n}(s)), u_{k+1} \rangle \langle \mathbf{1} - f_{k,n}(s), u_k \rangle}, \quad (3.2)$$

where $f_{k,n}^j(s) \leq \eta_{k+1,n}(j, s) \leq 1$ for each $k \geq 0$ and $j \in S$.

Let

$$\alpha(n, s) = \frac{1}{2} \sum_{k=0}^{n-1} \frac{\langle (\mathbf{1} - f_{k+1,n}(s))^T Hg_k(\eta_{k+1,n})(\mathbf{1} - f_{k+1,n}(s)), u_k \rangle}{\tilde{\Lambda}_{k+1} \langle (\mathbf{1} - f_{k+1,n}(s)), u_{k+1} \rangle \langle \mathbf{1} - f_{k,n}(s), u_k \rangle}, \quad (3.3)$$

where we note again that the dependence on s also lies in the vector $\eta_{k+1,n}$ since the components of $\eta_{k+1,n}$ satisfy $f_{k,n}(s)(i) \leq \eta_{k+1,n}(i) \leq 1$. Then (3.2) takes the form

$$\langle \mathbf{1} - f_{0,n}(s), u_0 \rangle = \left(\frac{1}{\tilde{\Lambda}_n \langle \mathbf{1} - s, u_n \rangle} + \alpha(n, s) \right)^{-1}. \quad (3.4)$$

We will need the following lemma which will be proved after we complete the proof of this Theorem.

Let us denote

$$\Xi_n = \sum_{k=0}^{n-1} \frac{1}{\tilde{\Lambda}_{k+1}}.$$

These are the partial sums of the series found in Theorem 2.1, but with the index of summation shifted in order to make the arguments below more transparent.

Lemma 3.1. *Under Assumptions 1-3, there exists $C > 1$ such that for each n and each $s \in [0, 1]^d \setminus \{\mathbf{1}\}$ we have*

$$\frac{1}{C} \leq \frac{\alpha(n, s)}{\Xi_n} \leq C.$$

Using Lemma 3.1 in equation (3.4), we get

$$\langle \mathbf{1} - f_{0,n}(\mathbf{0}), u_0 \rangle \leq \left(\frac{1}{\tilde{\Lambda}_n} + \frac{\Xi_n}{C} \right)^{-1}.$$

Therefore,

$$\langle \mathbf{1} - f_{0,n}(\mathbf{0}), u_0 \rangle \leq \frac{C}{\Xi_n}.$$

We note that that $1 - f_{0,n}^j(\mathbf{0}) = \mathbb{P}(Z_n \neq \mathbf{0} | Z_0 = e_j)$ for each $j \in S$, and hence, if $\lim_{n \rightarrow \infty} \Xi_n = \infty$, then

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n \neq \mathbf{0} | Z_0 = e_j) = 0.$$

Thus, extinction occurs with probability one if the initial population is e_j . Therefore, since j was arbitrary, extinction occurs with probability one for every initial population. \square

Proof of Lemma 3.1. The statement will follow if we prove the following bounds on the terms in the sums for $\alpha(n, s)$ and Ξ_n : for each $0 \leq k \leq n-1$ and $s \in [0, 1]^d \setminus \{\mathbf{1}\}$, we have

$$\frac{1}{C\Lambda_{k+1}} \leq \frac{\langle (\mathbf{1} - f_{k+1,n}(s))^T H g_k(\eta_{k+1,n})(\mathbf{1} - f_{k+1,n}(s)), u_k \rangle}{\tilde{\Lambda}_{k+1} \langle (\mathbf{1} - f_{k+1,n}(s)), u_{k+1} \rangle \langle \mathbf{1} - f_{k,n}(s), u_k \rangle} \leq \frac{C}{\Lambda_{k+1}}. \quad (3.5)$$

By Proposition 2.1(e), in order to prove (3.5), it is enough to show that there exists an $L > 0$ such that

$$\frac{1}{L} \leq \frac{\langle (\mathbf{1} - f_{k+1,n}(s))^T H g_k(\eta_{k+1,n})(\mathbf{1} - f_{k+1,n}(s)), u_k \rangle}{\langle \mathbf{1} - f_{k+1,n}(s), u_{k+1} \rangle \langle \mathbf{1} - f_{k,n}(s), u_k \rangle} \leq L. \quad (3.6)$$

Now, we know that $f_{k,n}^j(\mathbf{0}) \leq f_{k,n}^j(s) \leq \eta_{k+1,n}(j) \leq 1$ for each k and $j \in S$. Also, $f_{k,n}^j(\mathbf{0}) = \mathbb{P}(Z_n = \mathbf{0} | Z_k = e_j) \geq \varepsilon_0$ for each $k \leq n-1$, and thus $\varepsilon_0 \leq \eta_{k+1,n}(j) \leq 1$ for each $k \leq n-1$ and $j \in S$. Thus, by Assumptions 1-3, there exists a constant $c_1 > 0$ such that for each vector ζ with non-negative components, we have

$$\frac{1}{c_1} \|\zeta\|^2 \leq \langle \zeta^T H g_k(\eta_{k+1,n}) \zeta, u_k \rangle \leq c_1 \|\zeta\|^2.$$

In particular, we have

$$\frac{1}{c_1} \|\mathbf{1} - f_{k+1,n}(s)\|^2 \leq \langle (\mathbf{1} - f_{k+1,n}(s))^T H g_k(\eta_{k+1,n})(\mathbf{1} - f_{k+1,n}(s)), u_k \rangle \leq c_1 \|\mathbf{1} - f_{k+1,n}(s)\|^2. \quad (3.7)$$

By Proposition 2.1, for each $0 \leq k \leq n-1$,

$$\bar{\varepsilon} \|\mathbf{1} - f_{k,n}(s)\| \leq \langle \mathbf{1} - f_{k,n}(s), u_k \rangle \leq \|\mathbf{1} - f_{k,n}(s)\|.$$

In order to prove (3.6), it is sufficient to prove that there exists a constant $c_2 > 0$ such that for each $0 \leq k \leq n-1$ and each $s \in [0, 1]^d \setminus \{\mathbf{1}\}$,

$$\frac{1}{c_2} \leq \frac{\|\mathbf{1} - f_{k+1,n}(s)\|}{\|\mathbf{1} - f_{k,n}(s)\|} \leq c_2.$$

The first inequality $\|\mathbf{1} - f_{k,n}(s)\| \leq c_2 \|\mathbf{1} - f_{k+1,n}(s)\|$ follows from the fact that

$$\|\mathbf{1} - f_{k,n}(s)\| = \|g_k(\mathbf{1}) - g_k(f_{k+1,n}(s))\| \leq c_2 \|\mathbf{1} - f_{k+1,n}(s)\|,$$

since g_k is uniformly Lipschitz due to Assumption 3.

We observe that by Assumptions 1-3, each entry of the matrix A_k is uniformly bounded from above and below, i.e., there exist positive constants r and R such that, for each $i, j \in S$,

$$r \leq A_k(i, j) \leq R.$$

To prove the second inequality $\|\mathbf{1} - f_{k+1,n}(s)\| \leq c_2 \|\mathbf{1} - f_{k,n}(s)\|$, we consider the following two cases:

(CASE I) $\|\mathbf{1} - f_{k+1,n}(s)\| \leq r\bar{\varepsilon}d/c_1$. Then, from equation (3.7) and Proposition 2.1,

$$\begin{aligned} \langle (\mathbf{1} - f_{k+1,n}(s))^T H g_k(\eta_{k+1,n})(\mathbf{1} - f_{k+1,n}(s)), u_k \rangle &\leq c_1 \|\mathbf{1} - f_{k+1,n}(s)\|^2 \\ &\leq r\bar{\varepsilon}d \|\mathbf{1} - f_{k+1,n}(s)\| \leq \langle A_k(\mathbf{1} - f_{k+1,n}(s)), u_k \rangle, \end{aligned}$$

and thus, substituting the above relation into the Taylor formula,

$$\langle \mathbf{1} - f_{k,n}(s), u_k \rangle = \langle A_k(\mathbf{1} - f_{k+1,n}(s)), u_k \rangle - \frac{1}{2} \langle (\mathbf{1} - f_{k+1,n}(s))^T H g_k(\eta_{k+1,n})(\mathbf{1} - f_{k+1,n}(s)), u_k \rangle,$$

we get,

$$\langle \mathbf{1} - f_{k,n}(s), u_k \rangle \geq \langle A_k(\mathbf{1} - f_{k+1,n}(s)), u_k \rangle / 2,$$

and thus,

$$\|\mathbf{1} - f_{k,n}(s)\| \geq \langle \mathbf{1} - f_{k,n}(s), u_k \rangle \geq \langle A_k(\mathbf{1} - f_{k+1,n}(s)), u_k \rangle / 2 \geq r\bar{\varepsilon} \|\mathbf{1} - f_{k+1,n}(s)\| / 2.$$

So, for $\tilde{c}_2 = 2/(r\bar{\varepsilon})$, we have

$$\|\mathbf{1} - f_{k+1,n}(s)\| \leq \tilde{c}_2 \|\mathbf{1} - f_{k,n}(s)\|.$$

(CASE II) Now suppose that $1 - f_{k+1,n}^j(s) > r\bar{\varepsilon}/c_1$ for some $j \in S$. We want to prove that there exists a $\gamma > 0$ such that $1 - f_{k,n}^j(s) \geq \gamma$. From Assumptions 1-2, for each $j \in S$,

$$g_k^j(s) = \mathbb{E} \left(\prod_{i=1}^d s_i^{Z_k^{k+1}(i)} \mid Z_k = e_j \right) \leq (1 - \varepsilon_0) + \varepsilon_0 s_j^2,$$

and thus, since $f_{k+1,n}^j(s) < 1 - r\bar{\varepsilon}/c_1$,

$$f_{k,n}^j(s) = g_k^j(f_{k+1,n}(s)) \leq (1 - \varepsilon_0) + \varepsilon_0 (1 - r\bar{\varepsilon}/c_1)^2 < 1,$$

where the last inequality holds since $0 < r\bar{\varepsilon}/c_1 < 1$. Setting $\gamma = \varepsilon_0 - \varepsilon_0(1 - r\bar{\varepsilon}/c_1)^2$, we obtain

$$1 - f_{k,n}^j(s) > \gamma,$$

which is the required inequality.

So, from the two cases above, we can define $c_2 = \max(\tilde{c}_2, d/\gamma)$ to get, for each $0 \leq k \leq n-1$ and $s \in [0, 1]^d \setminus \{\mathbf{1}\}$,

$$\|\mathbf{1} - f_{k+1,n}(s)\| \leq c_2 \|\mathbf{1} - f_{k,n}(s)\|.$$

□

Proof of Lemma 2.2. Let $\bar{u} = \mathbb{E}Z_1$. By Assumptions 1-3, for every initial population, there is a constant $c > 0$ such that $c^{-1}u_1 \leq \bar{u} \leq cu_1$, where inequality between vectors is understood as the inequality between their components. Then, since $M_{1,n}^T \bar{u} = \mathbb{E}Z_n$,

$$c^{-1}M_{1,n}^T u_1 \leq \mathbb{E}Z_n \leq cM_{1,n}^T u_1.$$

Taking the norm and using the fact that $M_{1,n}^T u_1 = (\Lambda_n/\Lambda_1)u_n$, we obtain (2.1).

From (3.4) with $s = \mathbf{0}$, using the fact that $1 - f_{0,n}^j(\mathbf{0}) = \mathbb{P}(Z_n \neq \mathbf{0} | Z_0 = e_j)$, we obtain

$$\frac{1}{C} \left(\frac{1}{\tilde{\Lambda}_n} + \alpha(n, s) \right)^{-1} \leq \mathbb{P}(Z_n \neq \mathbf{0}) \leq C \left(\frac{1}{\tilde{\Lambda}_n} + \alpha(n, s) \right)^{-1}.$$

Using Lemma 3.1 and the first estimate in part (e) of Proposition 2.1, we obtain, for a different constant C ,

$$\frac{1}{C} \left(\frac{1}{\Lambda_n} + \Xi_n \right)^{-1} \leq \mathbb{P}(Z_n \neq \mathbf{0}) \leq C \left(\frac{1}{\Lambda_n} + \Xi_n \right)^{-1}.$$

Since this is valid for every j , we have the same inequality for an arbitrary initial population (with a constant C that depends on the initial population). Since $\Lambda_n \Xi_n \geq 1$, this implies (2.2). \square

4. Convergence of the process conditioned on survival

The following series will be important to our analysis,

$$\Gamma_n = \frac{1}{2} \sum_{k=0}^{n-1} \frac{1}{\lambda_k \tilde{\Lambda}_{k+1}} \frac{\langle v_{k+1}^T H g_k(\mathbf{1}) v_{k+1}, u_k \rangle}{\langle v_{k+1}, u_{k+1} \rangle \langle v_k, u_k \rangle}. \quad (4.1)$$

Here H denotes the Hessian matrix. It is applied to each component of g_k separately, then multiplied by vectors v_{k+1}^T and v_{k+1} to get scalars, which are then multiplied by the corresponding components of u_k to form the scalar product in the numerator. Since all terms in the right side of (4.1) are positive, the sequence Γ_n is increasing. In each term in (4.1), each of the factors, λ_k , $\langle v_{k+1}^T H g_k(\mathbf{1}) v_{k+1}, u_k \rangle$, $\langle v_{k+1}, u_{k+1} \rangle$, and $\langle v_k, u_k \rangle$, is bounded from above and below uniformly in k by Assumptions 1-3 and Proposition 2.1. Therefore, by Proposition 2.1, there is a positive constant C such that

$$\frac{1}{C \Lambda_{k+1}} \leq \frac{1}{2} \frac{1}{\lambda_k \tilde{\Lambda}_{k+1}} \frac{\langle v_{k+1}^T H g_k(\mathbf{1}) v_{k+1}, u_k \rangle}{\langle v_{k+1}, u_{k+1} \rangle \langle v_k, u_k \rangle} \leq \frac{C}{\Lambda_{k+1}}, \quad (4.2)$$

and consequently,

$$\frac{\Xi_n}{C} = \frac{1}{C} \sum_{k=0}^{n-1} \frac{1}{\Lambda_{k+1}} \leq \Gamma_n \leq C \sum_{k=0}^{n-1} \frac{1}{\Lambda_{k+1}} = C \Xi_n. \quad (4.3)$$

Assumptions 5 and 6 now can be rewritten as

$$\Gamma_n \rightarrow \infty, \quad \Lambda_n \Gamma_n \rightarrow \infty \quad \text{as } n \rightarrow \infty. \quad (4.4)$$

The proof of Theorem 2.2 will rely on the following seemingly weaker statement.

Theorem 4.1. *Under Assumptions 1-6, for each $j \in S$, we have the following limit in distribution*

$$\frac{\langle j \zeta_n, u_n \rangle}{\mathbb{E} \langle j \zeta_n, u_n \rangle} \rightarrow \xi \quad \text{as } n \rightarrow \infty,$$

where ξ is an exponential random variable with parameter one.

Proof. The proof will rely on several lemmas, which will be formulated as needed. The proofs of these lemmas will be given in the end of this section. It is sufficient to show convergence of moment generating functions. That is, we want to prove that for each $\varkappa \in [0, \infty)$

$$\mathbb{E} \left(\exp \left(\frac{-\varkappa \langle j Z_n, u_n \rangle \mathbb{P}(Z_n \neq \mathbf{0})}{\mathbb{E} \langle j Z_n, u_n \rangle} \right) \middle| Z_n \neq \mathbf{0} \right) \rightarrow \frac{1}{1 + \varkappa} \quad \text{as } n \rightarrow \infty.$$

Let us define vectors \bar{s}_j such that the i -th component of \bar{s}_j is

$$\bar{s}_j(i) = \exp\left(\frac{-\varkappa u_n(i) \mathbb{P}(j Z_n \neq \mathbf{0})}{\mathbb{E}\langle j Z_n, u_n \rangle}\right).$$

Then the j -th component of the vector $f_n(\bar{s}_j)$ is equal to

$$f_n^j(\bar{s}_j) = \mathbb{E}\left(\exp\left(\frac{-\varkappa \langle j Z_n, u_n \rangle \mathbb{P}(j Z_n \neq \mathbf{0})}{\mathbb{E}\langle j Z_n, u_n \rangle}\right)\right).$$

Thus we want to show that

$$1 - \frac{1 - f_n^j(\bar{s}_j)}{\mathbb{P}(j Z_n \neq \mathbf{0})} \rightarrow \frac{1}{1 + \varkappa} \text{ as } n \rightarrow \infty. \quad (4.5)$$

In order to prove (4.5), it will be useful to study the asymptotic behavior of the sum on the right hand side of (3.2). We first find the upper and lower bounds of the sum using the upper and lower bounds for $\eta_{k,n}$. Observe that $Hg_k^j(s)$ is monotonic in s for each j since g_k^j is a polynomial with non-negative coefficients and $Hg_k^j(s)$ is a matrix with entries that are mixed second derivatives of g_k^j . Therefore, (3.2) gives

$$\begin{aligned} & \left(\frac{1}{\tilde{\Lambda}_n \langle \mathbf{1} - s, u_n \rangle} + \frac{1}{2} \sum_{k=0}^{n-1} \frac{\langle (\mathbf{1} - f_{k+1,n}(s))^T Hg_k(\mathbf{1})(\mathbf{1} - f_{k+1,n}(s)), u_k \rangle}{\tilde{\Lambda}_{k+1} \langle (\mathbf{1} - f_{k+1,n}(s)), u_{k+1} \rangle \langle \mathbf{1} - f_{k,n}(s), u_k \rangle} \right)^{-1} \\ & \leq \langle \mathbf{1} - f_{0,n}(s), u_0 \rangle \\ & \leq \left(\frac{1}{\tilde{\Lambda}_n \langle \mathbf{1} - s, u_n \rangle} + \frac{1}{2} \sum_{k=0}^{n-1} \frac{\langle (\mathbf{1} - f_{k+1,n}(s))^T Hg_k(f_{k,n}(s))(\mathbf{1} - f_{k+1,n}(s)), u_k \rangle}{\tilde{\Lambda}_{k+1} \langle (\mathbf{1} - f_{k+1,n}(s)), u_{k+1} \rangle \langle \mathbf{1} - f_{k,n}(s), u_k \rangle} \right)^{-1}. \end{aligned} \quad (4.6)$$

Let us briefly explain the idea for the next step. Assume that K is such that $n - K$ is large and $f_{k,n}(s)$ is close to $\mathbf{1}$ for $k \leq K + 1$. By formally linearizing the mappings g_k, g_{k+1}, \dots, g_K , we write

$$\mathbf{1} - f_{k,n}(s) \approx A_k A_{k+1} \cdots A_K (\mathbf{1} - f_{K+1,n}(s)). \quad (4.7)$$

We know that

$$v_k = A_k \frac{v_{k+1}}{\lambda_k},$$

and thus

$$v_k = A_k A_{k+1} \cdots A_K \frac{v_{K+1}}{\prod_{i=k}^K \lambda_i}. \quad (4.8)$$

Note the similarity in the expressions (4.7) and (4.8): the same product of matrices is applied, albeit to different vectors. Proposition 2.1(d) (contractive property of the matrices) implies that the resulting expressions will be aligned in the same direction if $K - k$ is sufficiently large. That is we can replace $\mathbf{1} - f_{k,n}(s)$ (and $\mathbf{1} - f_{k+1,n}(s)$) by the vectors $c_{k,n} v_k$ (and $c_{k+1,n} v_{k+1}$) in each of the terms in the sums in (4.6) for all k that are sufficiently far away from n , where $c_{k,n}$ satisfy the relation $c_{k,n}/c_{k+1,n} = \lambda_k$. This will simplify (4.6).

Now let us make the above arguments rigorous. For a given $\varepsilon > 0$ and a positive integer n , we define $J(n, \varepsilon)$ as follows,

$$J(n, \varepsilon) = \min\{k : 1 - f_{k,n}^i(\mathbf{0}) > \varepsilon \text{ for some } i \in S\}.$$

Lemma 4.1. *For each $\varepsilon' > 0$, there exist a natural number K and an $\varepsilon > 0$ such that, for each $s \in [0, 1]^d \setminus \{\mathbf{1}\}$,*

$$\mathbf{1} - f_{k,n}(s) = c_{k,n}(v_k + \delta_{k,n}), \quad (4.9)$$

where $\delta_{k,n}$ and $c_{k,n}$ depend on s and satisfy $\|\delta_{k,n}\| \leq \varepsilon'$ and $|(c_{k,n}/c_{k+1,n}) - \lambda_k| \leq \varepsilon'$ for each $0 \leq k \leq J(n, \varepsilon) - K$ and each n .

Note that $J(n, \varepsilon) \rightarrow \infty$ as $n \rightarrow \infty$ since each component of the vector $\mathbf{1} - f_{k,n}(\mathbf{0})$ is

$$1 - f_{k,n}^i(\mathbf{0}) = \mathbb{P}(Z_n \neq \mathbf{0} | Z_k = e_i)$$

and $\mathbb{P}(Z_n \neq \mathbf{0} | Z_k = e_i) \rightarrow 0$ as $n \rightarrow \infty$ for each i and each k by Lemma 2.1.

Recall the definition of $\alpha(n, s)$ from (3.3).

Lemma 4.2. *Under Assumptions 1-6,*

$$\lim_{n \rightarrow \infty} \frac{\alpha(n, s)}{\Gamma_n} = 1,$$

uniformly in $s \in [0, 1]^d \setminus \{\mathbf{1}\}$.

Let us return to the proof of (4.5). By Lemma 4.1, when n is large, the vector $\mathbf{1} - f_n(s) = \mathbf{1} - f_{0,n}(s)$ is nearly aligned to the vector v_0 . Thus, in (4.5) we can replace the j -th component of the vector $\mathbf{1} - f_n(\bar{s}_j)$ by

$$\langle \mathbf{1} - f_n(\bar{s}_j), u_0 \rangle \frac{\langle v_0, e_j \rangle}{\langle v_0, u_0 \rangle}.$$

Therefore, in order to prove (4.5), it is sufficient to show that

$$1 - \frac{\langle v_0, e_j \rangle}{\langle v_0, u_0 \rangle \mathbb{P}(Z_n \neq \mathbf{0})} \left(\frac{1}{\tilde{\Lambda}_n \varkappa \left(\frac{\mathbb{P}(Z_n \neq \mathbf{0})}{\mathbb{E}(\langle Z_n, u_n \rangle)} + o\left(\frac{\mathbb{P}(Z_n \neq \mathbf{0})}{\mathbb{E}(\langle Z_n, u_n \rangle)}\right)\right)} + \Gamma_n \right)^{-1} \rightarrow \frac{1}{1 + \varkappa},$$

where we used Lemma 4.2 to transform (3.4) and linearized $\mathbf{1} - \bar{s}_j$. The LHS can be written as

$$1 - \frac{\langle v_0, e_j \rangle}{\langle v_0, u_0 \rangle \mathbb{P}(Z_n \neq \mathbf{0}) \Gamma_n} \left(\frac{1}{\tilde{\Lambda}_n \Gamma_n \varkappa \left(\frac{\mathbb{P}(Z_n \neq \mathbf{0})}{\mathbb{E}(\langle Z_n, u_n \rangle)} + o\left(\frac{\mathbb{P}(Z_n \neq \mathbf{0})}{\mathbb{E}(\langle Z_n, u_n \rangle)}\right)\right)} + 1 \right)^{-1}. \quad (4.10)$$

We will need the following two lemmas.

Lemma 4.3. *Under Assumptions 1-6,*

$$\lim_{n \rightarrow \infty} \frac{\langle v_0, u_0 \rangle \mathbb{P}(Z_n \neq \mathbf{0}) \Gamma_n}{\langle v_0, e_j \rangle} = 1.$$

Lemma 4.4. *Under Assumptions 1-6,*

$$\lim_{n \rightarrow \infty} \tilde{\Lambda}_n \Gamma_n \frac{\mathbb{P}(Z_n \neq \mathbf{0}) \langle u_n, u_n \rangle}{\mathbb{E}(\langle Z_n, u_n \rangle)} = 1.$$

Applying the above two lemmas to transform the expression in (4.10), we obtain

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 - \frac{\langle v_0, e_j \rangle}{\langle v_0, u_0 \rangle \mathbb{P}(Z_n \neq \mathbf{0}) \Gamma_n} \left(\frac{1}{\tilde{\Lambda}_n \Gamma_n \varkappa \left(\frac{\mathbb{P}(Z_n \neq \mathbf{0})}{\mathbb{E}(\langle Z_n, u_n \rangle)} + o\left(\frac{\mathbb{P}(Z_n \neq \mathbf{0})}{\mathbb{E}(\langle Z_n, u_n \rangle)}\right)\right)} + 1 \right)^{-1} \right) \\ = 1 - \left(\frac{1}{\varkappa} + 1 \right)^{-1} = \frac{1}{1 + \varkappa}. \end{aligned}$$

This completes the proof of Theorem 4.1. \square

Proof of Theorem 2.2. First, let Assumptions 1–6 be satisfied. Let $\mathcal{P} : v \rightarrow v/\|v\|$ be the projection onto the unit sphere, with the convention that $\mathcal{P}(\mathbf{0}) = \mathbf{0}$. We claim that

$$\lim_{n \rightarrow \infty} \|\mathcal{P}(\mathbb{E}_j \zeta_n) - u_n\| = 0, \quad (4.11)$$

$$\lim_{n \rightarrow \infty} \mathbb{P}(\|\mathcal{P}(j \zeta_n) - u_n\| > \varepsilon) = 0 \quad (4.12)$$

for each j and each $\varepsilon > 0$. Let us fix $\delta \in (0, \varepsilon)$. By Proposition 2.1, we can find $k' \in \mathbb{N}$ such that

$$(1 - \delta)u_{n+k'} \leq \frac{M_{n,n+k'}^T u}{\|M_{n,n+k'}^T u\|} \leq (1 + \delta)u_{n+k'} \quad (4.13)$$

whenever u is a non-zero vector with non-negative components. Let ${}_j \zeta_n^{k'}$ be the random vector obtained by taking ${}_j \zeta_n$ as the initial population of a branching process, then branching for k' steps using our original branching distributions $P_n, \dots, P_{n+k'-1}$ and evaluating the resulting population. Note that ${}_j \zeta_n^{k'}$ is different from ${}_j \zeta_{n+k'}$, the latter can be obtained from ${}_j \zeta_n^{k'}$ by conditioning on the event of non-extinction. Since the extinction of a large initial population in k' steps occurs with small probability and since, by Theorem 4.1, for each $a > 0$ we have $\mathbb{P}(\|{}_j \zeta_n\| > a) \rightarrow 1$ as $n \rightarrow \infty$, we obtain

$$\lim_{n \rightarrow \infty} (\mathbb{P}(\|\mathcal{P}({}_j \zeta_n^{k'}) - u_{n+k'}\| > \varepsilon) - \mathbb{P}(\|\mathcal{P}({}_j \zeta_{n+k'}) - u_{n+k'}\| > \varepsilon)) = 0.$$

Also note that $\lim_{n \rightarrow \infty} (\mathcal{P}(\mathbb{E}_j \zeta_n^{k'}) - \mathcal{P}(\mathbb{E}_j \zeta_{n+k'})) = \mathbf{0}$. Therefore, since $\delta > 0$ was arbitrarily small, (4.11) and (4.12) will follow if we show that

$$\|\mathcal{P}(\mathbb{E}_j \zeta_n^{k'}) - u_{n+k'}\| \leq \delta \quad (4.14)$$

for all sufficiently large n and

$$\lim_{n \rightarrow \infty} \mathbb{P}(\|\mathcal{P}({}_j \zeta_n^{k'}) - u_{n+k'}\| > \varepsilon) = 0. \quad (4.15)$$

(4.14) immediately follows from (4.13). (4.15) is a consequence of

$$\lim_{n \rightarrow \infty} \mathbb{P}(\|\mathcal{P}({}_j \zeta_n^{k'}) - \mathcal{P}(\mathbb{E}_j \zeta_n^{k'})\| > \varepsilon - \delta) = 0,$$

which can be derived from the Chebyshev inequality since for each $a > 0$ we have $\mathbb{P}(\|{}_j \zeta_n\| > a) \rightarrow 1$ as $n \rightarrow \infty$. Thus we have (4.11) and (4.12).

Next, let us show that (4.11) and (4.12), along with Theorem 4.1, imply (2.3) with ${}_j \zeta_n$ in place of ζ_n . By Theorem 4.1, it is sufficient to show that we have the following limit in probability

$$\frac{\langle {}_j \zeta_n, u_n \rangle}{\mathbb{E} \langle {}_j \zeta_n, u_n \rangle} - \frac{\langle {}_j \zeta_n, u \rangle}{\mathbb{E} \langle {}_j \zeta_n, u \rangle} \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (4.16)$$

From (4.11) we know that

$$\lim_{n \rightarrow \infty} (\langle \mathcal{P}(\mathbb{E}_j \zeta_n), u_n \rangle - \langle u_n, u_n \rangle) = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} (\langle \mathcal{P}(\mathbb{E}_j \zeta_n), u \rangle - \langle u_n, u \rangle) = 0.$$

By (4.12) the two following limits hold in probability,

$$\lim_{n \rightarrow \infty} (\langle \mathcal{P}({}_j \zeta_n), u_n \rangle - \langle u_n, u_n \rangle) = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} (\langle \mathcal{P}({}_j \zeta_n), u \rangle - \langle u_n, u \rangle) = 0.$$

Therefore, the right hand side in the following equality

$$\frac{\langle {}_j \zeta_n, u_n \rangle}{\mathbb{E} \langle {}_j \zeta_n, u_n \rangle} - \frac{\langle {}_j \zeta_n, u \rangle}{\mathbb{E} \langle {}_j \zeta_n, u \rangle} = \frac{\|{}_j \zeta_n\|}{\|\mathbb{E}({}_j \zeta_n)\|} \left(\frac{\langle \mathcal{P}({}_j \zeta_n), u_n \rangle}{\langle \mathcal{P}(\mathbb{E}_j \zeta_n), u_n \rangle} - \frac{\langle \mathcal{P}({}_j \zeta_n), u \rangle}{\langle \mathcal{P}(\mathbb{E}_j \zeta_n), u \rangle} \right)$$

tends to zero in probability (the factor in the brackets tends to zero in probability while the first factor is bounded in L^1). This justifies (4.16) and therefore (2.3) with ${}_j\zeta_n$ in place of ζ_n .

If the initial population of the process $\{Z_n\}$ is such that we have more than one particle at time zero, then we can consider a new process $\{Z'_n\}$ for which $Z'_0 = e_1$ and the transition distribution P'_0 is such that ${}_jZ'_1$ coincides in distribution with Z_1 . We also define $P'_n = P_n$ for $n \geq 1$. It is easy to see that the modified process satisfies Assumptions 1-4 with possibly different values of ε_0 and K_0 . On the other hand, $\langle \zeta_n, u \rangle / \mathbb{E}\langle \zeta_n, u \rangle$ is equal, in distribution, to $\langle {}_j\zeta'_n, u \rangle / \mathbb{E}\langle {}_j\zeta'_n, u \rangle$ when $n \geq 1$, and therefore (2.3) holds for every initial population.

Finally, suppose that Assumptions 1-5 are satisfied. If Assumption 6 fails, then $\mathbb{E}\|\zeta_n\| = \mathbb{E}\|Z_n\| / \mathbb{P}(Z_n \neq \mathbf{0})$ is bounded along a subsequence for every initial population. Then (2.3) does not hold since ζ_n is integer-valued, which gives a contradiction. \square

Proof of Lemma 2.3. From Proposition 2.2 it follows that if $\{Z_n\}$ satisfies the assumptions of Lemma 2.3, then there exists l such that $\{\tilde{Z}_n\} = \{Z_{nl}\}$ satisfies assumptions 1-3. By Theorem 2.1, which can be applied due to (2.4), $\{\tilde{Z}_n\}$ goes extinct almost surely. This implies the almost sure extinction of $\{Z_n\}$.

Similarly, from Theorem 2.2 (which can be applied since Assumptions 5 and 6 are met by the process $\{Z_{nl}\}$ due to (2.4)), it follows that (2.3) holds along a subsequence nl . To show that (2.3) holds (without restriction to a subsequence) let $n = Nl + r$ with $0 \leq r < l$. We claim that for every u and every $0 \leq r < l$,

$$\lim_{N \rightarrow \infty} \left(\frac{\langle \zeta_{Nl+r}, u \rangle}{\langle \zeta_{Nl}, u_{Nl} \rangle} - \frac{\Lambda_{Nl+r}(u_{Nl+r}, u)}{\Lambda_{Nl}} \right) = 0 \quad \text{in probability.} \quad (4.17)$$

Lemma 2.3 follows directly from (4.17) and Theorem 2.2 applied to $\{Z_{Nl}\}$. The proof of (4.17) is similar to the proof of (4.12) and (4.16), so we leave it to the reader. \square

It still remains to prove Lemmas 4.1–4.4.

Proof of Lemma 4.1. Suppose that we have (4.9) with $\|\delta_{k,n}\| \leq \varepsilon''$, but without any assumptions on $c_{k,n}$. Then we have, for $0 \leq k < J(n, \varepsilon) - K$,

$$\mathbf{1} - f_{k,n}(s) = \mathbf{1} - g_k(f_{k+1,n}(s)) = A_k(\mathbf{1} - f_{k+1,n}(s)) + \alpha_{k,n} \|\mathbf{1} - f_{k+1,n}(s)\|,$$

where $\|\alpha_{k,n}\|$ can be made arbitrarily small, uniformly in k , by selecting sufficiently small ε . The latter statement about $\alpha_{k,n}$ follows from the assumption that $1 - f_{k+1,n}^i(s) \leq \varepsilon$ for all i (from definition of $J(n, \varepsilon)$) and the fact that

$$A_k(j, i) = \frac{\partial g_k^j}{\partial s_i}(\mathbf{1}).$$

The uniformity in k follows from Assumption 3. Thus

$$\mathbf{1} - f_{k,n}(s) = A_k(c_{k+1,n}(v_{k+1} + \delta_{k+1,n})) + \alpha_{k,n} \|\mathbf{1} - f_{k+1,n}(s)\| =$$

$$c_{k+1,n} \lambda_k v_k + c_{k+1,n} A_k \delta_{k+1,n} + \alpha_{k,n} \|\mathbf{1} - f_{k+1,n}(s)\| = c_{k+1,n} \lambda_k (v_k + \alpha'_{k,n}),$$

where $\|\alpha'_{k,n}\|$ can be made arbitrarily small, uniformly in k , by selecting sufficiently small ε and ε'' . Here we used (4.9) with $k+1$ instead of k to estimate the contribution from the term $\alpha_{k,n} \|\mathbf{1} - f_{k+1,n}(s)\|$. Thus

$$c_{k+1,n} \lambda_k (v_k + \alpha'_{k,n}) = c_{k,n} (v_k + \delta_{k,n}),$$

which implies that $|(c_{k,n}/c_{k+1,n}) - \lambda_k| \leq \varepsilon'$ holds for $0 \leq k < J(n, \varepsilon) - K$, provided that ε and ε'' are sufficiently small. We have demonstrated, therefore, that it is sufficient to establish (4.9) with the estimate $\|\delta_{k,n}\| \leq \varepsilon'$ only.

By part (d) of Proposition 2.1, there is $k' \in \mathbb{N}$ such that

$$\left(1 - \frac{\varepsilon'}{2d}\right) v_k \leq \frac{M_{k,k+k'} v}{\|M_{k,k+k'} v\|} \leq \left(1 + \frac{\varepsilon'}{2d}\right) v_k \quad (4.18)$$

for each k and each non-zero vector v with non-negative components. Since

$$M_{k,k+k'}(j, i) = \frac{\partial f_{k,k+k'}^j}{\partial s_i}(\mathbf{1}),$$

we can linearize the mapping $\mathbf{1} - f_{k,k+k'}(s)$ at $s = \mathbf{1}$ and obtain that there is ε such that

$$\|\mathbf{1} - f_{k,k+k'}(\mathbf{1} - v) - M_{k,k+k'}v\| \leq \frac{\varepsilon'}{2d} \|M_{k,k+k'}v\|$$

whenever $0 < \|v\| \leq \varepsilon d$. (We have used here that $M_{k,k+k'}$ is bounded uniformly in k .) Therefore,

$$M_{k,k+k'}v - \frac{\varepsilon'}{2d} \|M_{k,k+k'}v\| \mathbf{1} \leq \mathbf{1} - f_{k,k+k'}(\mathbf{1} - v) \leq M_{k,k+k'}v + \frac{\varepsilon'}{2d} \|M_{k,k+k'}v\| \mathbf{1}.$$

Combined with (4.18), this gives

$$\|M_{k,k+k'}v\|(v_k - \frac{\varepsilon'}{d} \mathbf{1}) \leq \mathbf{1} - f_{k,k+k'}(\mathbf{1} - v) \leq \|M_{k,k+k'}v\|(v_k + \frac{\varepsilon'}{d} \mathbf{1}).$$

Setting $K = k' + 1$, we see that the last inequality can be applied to $v = \mathbf{1} - f_{k+k',n}(s)$, provided that $0 \leq k \leq J(n, \varepsilon) - K$, resulting in

$$c_{k,n}(v_k - \frac{\varepsilon'}{d} \mathbf{1}) \leq \mathbf{1} - f_{k,n}(s) \leq c_{k,n}(v_k + \frac{\varepsilon'}{d} \mathbf{1}),$$

which gives the desired estimate. \square

Proof of Lemma 4.2. We break up the difference $(\alpha(n, s)/\Gamma_n) - 1$ into three parts. So we want to prove that for each $\sigma > 0$ there is $\varepsilon > 0$ such that

$$\begin{aligned} & \left| \frac{\alpha(n, s) - \alpha(J(n, \varepsilon) - K - 1, s)}{\Gamma_n} + \frac{\alpha(J(n, \varepsilon) - K - 1, s) - \Gamma_{J(n, \varepsilon) - K - 1}}{\Gamma_n} \right. \\ & \left. + \frac{\Gamma_{J(n, \varepsilon) - K - 1} - \Gamma_n}{\Gamma_n} \right| \leq \sigma \end{aligned} \quad (4.19)$$

for all s and all sufficiently large n .

1. We first estimate the middle term in the inequality above. By Lemma 4.1, for each $\sigma' > 0$, there exist a natural number K and $\varepsilon_1 > 0$ such that

$$(1 - \sigma')c_{k,n}v_k \leq \mathbf{1} - f_{k,n}(s) \leq (1 + \sigma')c_{k,n}v_k$$

for each $k < J(n, \varepsilon_1) - K$.

By Assumption 4, $\|X_n\|^2$ are uniformly integrable, and thus the matrices $Hg_k^i(s)$, $k \geq 0$, $i \in S$, are equicontinuous in s . Note also that $\|Hg_k^i(\mathbf{1})\| \geq c > 0$ for all $k \geq 0$, $i \in S$. Thus there exists $\varepsilon_2 > 0$ such that the matrix norm satisfies $\|Hg_k^i(\eta_{k+1,n}) - Hg_k^i(\mathbf{1})\| < \sigma' \|Hg_k^i(\mathbf{1})\|$ for each $k < J(n, \varepsilon_2) - K$. Choosing $\varepsilon = \min(\varepsilon_1, \varepsilon_2)$, we see that there is a constant \tilde{c} independent of $\sigma' > 0$ such that

$$\begin{aligned} & \left| \frac{1}{2} \sum_{k=0}^{J(n, \varepsilon) - K - 1} \frac{\langle (\mathbf{1} - f_{k+1,n}(s))^T Hg_k(\eta_{k+1,n})(\mathbf{1} - f_{k+1,n}(s)), u_k \rangle}{\tilde{\Lambda}_{k+1} \langle (\mathbf{1} - f_{k+1,n}(s)), u_{k+1} \rangle \langle \mathbf{1} - f_{k,n}(s), u_k \rangle} - \Gamma_{J(n, \varepsilon) - K - 1} \right| \\ & \leq \tilde{c} \sigma' \Gamma_{J(n, \varepsilon) - K - 1} \leq \tilde{c} \sigma' \Gamma_n. \end{aligned}$$

(In essence, there a small relative error, linear in σ' , in the factors in each of the terms of the sum, and thus the total relative error is small.) By choosing $\sigma' \leq \sigma/(3\bar{c})$, we obtain

$$\left| \frac{\alpha(J(n, \varepsilon) - K - 1, s) - \Gamma_{J(n, \varepsilon) - K - 1}}{\Gamma_n} \right| < \frac{\sigma}{3}.$$

2. Now we estimate the third term in (4.19). We can assume that K and ε are fixed. We first observe that we can obtain a relation similar to (3.2) by starting with the expression $\langle \mathbf{1} - f_{J(n, \varepsilon), n}(s), u_{J(n, \varepsilon)} \rangle$ instead of $\langle \mathbf{1} - f_{0, n}(s), u_0 \rangle$. Thus, by doing the same steps that we carried out to obtain (3.2), we get

$$\begin{aligned} & \langle \mathbf{1} - f_{J(n, \varepsilon), n}(s), u_{J(n, \varepsilon)} \rangle \\ &= \left(\frac{\tilde{\Lambda}_{J(n, \varepsilon)}}{\tilde{\Lambda}_n \langle \mathbf{1} - s, u_n \rangle} + \frac{1}{2} \sum_{k=J(n, \varepsilon)}^{n-1} \frac{\tilde{\Lambda}_{J(n, \varepsilon)} \langle (\mathbf{1} - f_{k+1, n}(s))^T H g_k(\eta_{k+1, n})(\mathbf{1} - f_{k+1, n}(s)), u_k \rangle}{\tilde{\Lambda}_{k+1} \langle (\mathbf{1} - f_{k+1, n}(s)), u_{k+1} \rangle \langle \mathbf{1} - f_{k, n}(s), u_k \rangle} \right)^{-1} \\ &= \frac{1}{\tilde{\Lambda}_{J(n, \varepsilon)}} \left(\frac{1}{\tilde{\Lambda}_n \langle \mathbf{1} - s, u_n \rangle} + \frac{1}{2} \sum_{k=J(n, \varepsilon)}^{n-1} \frac{\langle (\mathbf{1} - f_{k+1, n}(s))^T H g_k(\eta_{k+1, n})(\mathbf{1} - f_{k+1, n}(s)), u_k \rangle}{\tilde{\Lambda}_{k+1} \langle (\mathbf{1} - f_{k+1, n}(s)), u_{k+1} \rangle \langle \mathbf{1} - f_{k, n}(s), u_k \rangle} \right)^{-1} \\ &= \frac{1}{\tilde{\Lambda}_{J(n, \varepsilon)}} \left(\frac{1}{\tilde{\Lambda}_n \langle \mathbf{1} - s, u_n \rangle} + (\alpha(n, s) - \alpha(J(n, \varepsilon), s)) \right)^{-1} \\ &\leq \frac{1}{\tilde{\Lambda}_{J(n, \varepsilon)} (\alpha(n, s) - \alpha(J(n, \varepsilon), s))} \leq C \frac{1}{\tilde{\Lambda}_{J(n, \varepsilon)} (\Gamma_n - \Gamma_{J(n, \varepsilon)})}, \end{aligned}$$

where the last inequality follows from (3.5) and (4.2). From here it follows that

$$(\Gamma_n - \Gamma_{J(n, \varepsilon)}) \leq C \frac{1}{\tilde{\Lambda}_{J(n, \varepsilon)} \langle \mathbf{1} - f_{J(n, \varepsilon), n}(\mathbf{0}), u_{J(n, \varepsilon)} \rangle} \leq \frac{C}{\tilde{\Lambda}_{J(n, \varepsilon)} \varepsilon \bar{\varepsilon}}.$$

Therefore,

$$\frac{\Gamma_n - \Gamma_{J(n, \varepsilon)}}{\Gamma_n} \leq \frac{C}{\tilde{\Lambda}_{J(n, \varepsilon)} \Gamma_n \varepsilon \bar{\varepsilon}} \leq \frac{C}{\tilde{\Lambda}_{J(n, \varepsilon)} \Gamma_{J(n, \varepsilon)} \varepsilon \bar{\varepsilon}}.$$

Since $J(n, \varepsilon) \rightarrow \infty$ as $n \rightarrow \infty$, by Assumption 6 (see (4.4) and part (e) of Proposition 2.1), we have

$$\left| \frac{C}{\tilde{\Lambda}_{J(n, \varepsilon)} \Gamma_{J(n, \varepsilon)} \varepsilon \bar{\varepsilon}} \right| < \frac{\sigma}{6}$$

for all sufficiently large n . Now,

$$\frac{\Gamma_n - \Gamma_{J(n, \varepsilon) - K - 1}}{\Gamma_n} = \frac{\Gamma_n - \Gamma_{J(n, \varepsilon)}}{\Gamma_n} + \frac{\Gamma_{J(n, \varepsilon)} - \Gamma_{J(n, \varepsilon) - K - 1}}{\Gamma_n}.$$

For a fixed K , using Assumption 6 and the fact that $J(n, \varepsilon) \rightarrow \infty$ as $n \rightarrow \infty$, we see for all sufficiently large n and $k \geq J(n, \varepsilon) - K - 1$,

$$\left| \frac{1}{\Gamma_n} \left(\frac{1}{\lambda_k \tilde{\Lambda}_{k+1}} \frac{\langle v_{k+1}^T H g_k(\mathbf{1}) v_{k+1}, u_k \rangle}{\langle v_{k+1}, u_{k+1} \rangle \langle v_k, u_k \rangle} \right) \right| < \frac{\sigma}{3(K+1)}.$$

Therefore,

$$\left| \frac{1}{\Gamma_n} \left(\frac{1}{2} \sum_{k=J(n, \varepsilon) - K - 1}^{J(n, \varepsilon)} \frac{1}{\lambda_k \tilde{\Lambda}_{k+1}} \frac{\langle v_{k+1}^T H g_k(\mathbf{1}) v_{k+1}, u_k \rangle}{\langle v_{k+1}, u_{k+1} \rangle \langle v_k, u_k \rangle} \right) \right| < \frac{\sigma}{6}.$$

Thus,

$$\frac{\Gamma_{J(n,\varepsilon)} - \Gamma_{J(n,\varepsilon)-K-1}}{\Gamma_n} < \frac{\sigma}{6},$$

and, therefore, for all sufficiently large n we have

$$\left| \frac{\Gamma_n - \Gamma_{J(n,\varepsilon)}}{\Gamma_n} \right| < \frac{\sigma}{3}.$$

3. We know that

$$\frac{\alpha(n, s) - \alpha(J(n, \varepsilon) - K - 1, s)}{\Gamma_n} \leq C \frac{\Gamma_n - \Gamma_{J(n, \varepsilon) - K - 1}}{\Gamma_n},$$

and by the same arguments as above, for all sufficiently large n we have

$$\left| C \frac{\Gamma_n - \Gamma_{J(n, \varepsilon) - K - 1}}{\Gamma_n} \right| < \frac{\sigma}{3}.$$

By the estimates from steps 1-3, for all sufficiently large n and all $s \in [0, 1]^d \setminus \{\mathbf{1}\}$ we have

$$\left| \frac{\alpha(n, s)}{\Gamma_n} - 1 \right| < \sigma.$$

Since $\sigma > 0$ was arbitrary, the proof is complete. \square

Proof of Lemma 4.3. We know that $\mathbb{P}(Z_n \neq \mathbf{0}) = 1 - f_n^j(\mathbf{0})$, and therefore for a fixed $j \in S$, by Lemma 4.1, it is sufficient to prove that

$$\lim_{n \rightarrow \infty} (c_{0,n} \langle v_0, u_0 \rangle \Gamma_n) = 1, \quad (4.20)$$

where $c_{0,n}$ is the same as in Lemma 4.1. From Lemma 4.2 and (3.4) we know that

$$\langle \mathbf{1} - f_{0,n}(s), u_0 \rangle \sim \left(\frac{1}{\tilde{\Lambda}_n \langle \mathbf{1} - s, u_n \rangle} + \Gamma_n \right)^{-1} \quad \text{as } n \rightarrow \infty.$$

By plugging in $s = \mathbf{0}$ and by replacing $\langle \mathbf{1} - f_{0,n}(\mathbf{0}), u_0 \rangle$ by $c_{0,n} \langle v_0, u_0 \rangle$, we get

$$\lim_{n \rightarrow \infty} c_{0,n} \langle v_0, u_0 \rangle \left(\frac{1}{\tilde{\Lambda}_n \langle \mathbf{1}, u_n \rangle} + \Gamma_n \right) = 1.$$

Thus we have

$$\lim_{n \rightarrow \infty} c_{0,n} \Gamma_n \langle v_0, u_0 \rangle \left(\frac{1}{\tilde{\Lambda}_n \Gamma_n \langle \mathbf{1}, u_n \rangle} + 1 \right) = 1.$$

By Assumption 6, $\tilde{\Lambda}_n \Gamma_n \rightarrow \infty$ as $n \rightarrow \infty$ proving (4.20). \square

Proof of Lemma 4.4. As in the proof of Lemma 4.3, it is sufficient to show that

$$\lim_{n \rightarrow \infty} \frac{c_{0,n} \tilde{\Lambda}_n \Gamma_n \langle v_0, e_j \rangle \langle u_n, u_n \rangle}{\mathbb{E}(\langle Z_n, u_n \rangle)} = 1.$$

We observe that

$$\begin{aligned} \mathbb{E}(\langle Z_n, u_n \rangle) &= \sum_{i=1}^d \mathbb{E}(Z_n(i) u_n(i) | Z_0 = e_j) \\ &= \sum_{i=1}^d u_n(i) \mathbb{E}(Z_n(i) | Z_0 = e_j) \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^d u_n(i) M_n(j, i) \\
&= \langle M_n u_n, e_j \rangle \\
&= \langle u_n, M_n^T e_j \rangle
\end{aligned}$$

We know that $v_0 = M_n v_n / \Lambda_n$, and thus what we want to prove is that

$$\lim_{n \rightarrow \infty} \frac{c_{0,n} \tilde{\Lambda}_n \Gamma_n \langle v_n, M_n^T e_j \rangle \langle u_n, u_n \rangle}{\Lambda_n \langle u_n, M_n^T e_j \rangle} = 1.$$

By (4.20), it is sufficient to show that

$$\lim_{n \rightarrow \infty} \frac{\tilde{\Lambda}_n \langle v_n, M_n^T e_j \rangle \langle u_n, u_n \rangle}{\Lambda_n \langle v_0, u_0 \rangle \langle u_n, M_n^T e_j \rangle} = 1.$$

By Proposition 2.1 (part (d)), the vectors $M_n^T e_j$ align with the vectors u_n . Therefore, it remains to prove that

$$\frac{\tilde{\Lambda}_n \langle v_n, u_n \rangle}{\Lambda_n \langle v_0, u_0 \rangle} = 1.$$

But this is true because

$$\langle v_n, u_n \rangle = \langle v_n, \frac{A_{n-1}^T u_{n-1}}{\tilde{\lambda}_{n-1}} \rangle = \langle A_{n-1} v_n, \frac{u_{n-1}}{\tilde{\lambda}_{n-1}} \rangle = \frac{\lambda_{n-1}}{\tilde{\lambda}_{n-1}} \langle v_{n-1}, u_{n-1} \rangle = \frac{\Lambda_n}{\tilde{\Lambda}_n} \langle v_0, u_0 \rangle,$$

where the last equality is obtained by iterating the previous steps n times. \square

5. Branching processes with continuous time

In this section, we provide an application of our results to continuous time branching. Let $\rho_t(j)$, $1 \leq j \leq d$, be continuous functions and $P_t(j, \cdot)$ be transition distributions on \mathbb{Z}_+^d such that $P_t(j, a)$ is continuous for each $a \in \mathbb{Z}_+^d$.

Let ${}_j X_t$ be a random vector with values in \mathbb{Z}_+^d , whose distribution is given by $P_t(j, \cdot)$. We assume that there are $\varepsilon_0, K_0 > 0$ such that for all $i, j \in S$ the following bounds hold.

- 0'. $\varepsilon_0 \leq \rho_t(j) \leq K_0$.
- 1'. $\mathbb{P}({}_j X_t(i) \geq 2) \geq \varepsilon_0$.
- 2'. $\mathbb{P}({}_j X_t = \mathbf{0}) \geq \varepsilon_0$.
- 3'. $\mathbb{E}(\|{}_j X_t\|^2) \leq K_0$.

Assuming that we start with a finite number of particles and that the above bounds hold, the transition rates $\rho_t(j)$ and the transition distributions $P_t(j, \cdot)$ define a continuous time branching process $\{Z_t\}$ with particles of d different types. Namely, each particle of type j alive at time t undergoes transformation into $a_1 + \dots + a_d$ particles: a_1 particles of type one, a_2 particles of type two, etc., during the time interval $[t, t + \Delta]$ with probability $\rho_t(j) P_t(j, a)(\Delta + o(\Delta))$.

Observe that $\{Z_n\}$, $n \in \mathbb{N}$, $n \geq 0$, is a discrete time branching process that satisfies Assumptions 1-3 (with different ε_0 and K_0). The fact that it satisfies Assumptions 1 and 2 is clear. The first moment $M(t) = \mathbb{E}Z_t$ satisfies

$$M'(t) = B^T(t)M(t), \tag{5.1}$$

where $B(t)_{ji} = \rho_i(j)(\mathbb{E}(X_t(i)) - \delta_{ij})$. Similarly, if $\mathbb{E}(\|X_t\|^p)$ exists and depends continuously on t , then the moments of $\{Z_t\}$ of order p satisfy inhomogeneous linear equations, and if $\mathbb{E}(\|X_t\|^p)$ is uniformly bounded in both t and j , then the coefficients of those equations are uniformly bounded. In particular, Assumption 3' implies that Assumption 3 is satisfied, while a bound on the third moment of $\|X_t\|$ (see Assumption 4' below) would imply that Assumption 4 is satisfied.

Recall that, in the notation of Section 2 applied to the process observed at integer time points,

$$\mathbb{E}Z_n = M_n^T Z_0.$$

Therefore, by part (e) of Proposition 2.1, for each initial population, there is a positive constant C such that

$$\frac{1}{C}\Lambda_n \leq \|\mathbb{E}Z_n\| \leq C\Lambda_n.$$

From (5.1) it follows that there is a positive constant c such that

$$\frac{1}{c}\|\mathbb{E}Z_n\| \leq \|\mathbb{E}Z_t\| \leq c\|\mathbb{E}Z_n\|, \quad n \leq t \leq n+1.$$

Therefore, the condition $\sum_{k=1}^{\infty} (1/\Lambda_k) = \infty$ used in Theorem 2.1 is equivalent to the following:

$$\int_0^{\infty} \frac{1}{\|\mathbb{E}Z_t\|} dt = \infty. \quad (5.2)$$

Thus we have the following continuous time analogue of Theorem 2.1.

Theorem 5.1. *Under Assumptions 0'-3', if extinction of the process $\{Z_t\}$ occurs with probability one for some initial population, then (5.2) holds. If (5.2) holds, then extinction with probability one occurs for every initial population.*

To formulate the next theorem, we will make use of the following assumptions:

$$4'. \mathbb{E}(\|X_t\|^3) \leq K_0 \text{ for some } K_0 > 0.$$

$$5'. \mathbb{P}(Z_t \neq \mathbf{0}) \rightarrow 0 \text{ as } t \rightarrow \infty.$$

$$6'. \mathbb{E}\|Z_t\|/\mathbb{P}(Z_t \neq \mathbf{0}) \rightarrow \infty \text{ as } t \rightarrow \infty.$$

Note that if Assumptions 0'-6' are satisfied, then Assumptions 1-6 are satisfied by the discrete time process $\{Z_n\}$. Let $\zeta_t = (\zeta_t(1), \dots, \zeta_t(d))$ be the random vector obtained from Z_t by conditioning on the event that $Z_t \neq \mathbf{0}$. The following theorem is an easy consequence of Theorem 2.2. The proof is left to the reader.

Theorem 5.2. *Under Assumptions 0'-6', for each initial population of the branching process and each vector u with positive components, we have the following limit in distribution*

$$\frac{\langle \zeta_t, u \rangle}{\mathbb{E}\langle \zeta_t, u \rangle} \rightarrow \xi, \quad \text{as } t \rightarrow \infty, \quad (5.3)$$

where ξ is an exponential random variable with parameter one. Moreover, if Assumptions 0'-5' are satisfied and, for some initial population, the limit in (5.3) is as specified, then Assumption 6' is also satisfied.

Appendix A. Proof of Proposition 2.1

Let \mathcal{K} be the cone of positive vectors. Given $u, v \in \mathcal{K}$, their Hilbert metric distance is defined by

$$d(u, v) = \ln \frac{\beta(u, v)}{\alpha(u, v)}, \quad \text{where } \beta(u, v) = \max_i \frac{v(i)}{u(i)}, \quad \alpha(u, v) = \min_i \frac{v(i)}{u(i)}.$$

Note that d defines the distance on the space of lines in \mathcal{K} in the sense that

$$d(au, bv) = d(u, v), \quad d(u, cu) = 0.$$

Moreover the following estimate holds.

Lemma A.1. (see, e.g., [20, Lemma 1.3]) *If $\|u\| = \|v\| = 1$, then*

$$\|u - v\| \leq e^{d(u, v)} - 1.$$

We will also use the following result of G. Birkhoff.

Lemma A.2. (see [5, Theorem XVI.3.3] or [20, Theorem 1.1]) *If A is a linear operator that maps \mathcal{K} into itself so that $A(\mathcal{K})$ has finite diameter Δ with respect to the Hilbert metric, then for each $u, v \in \mathcal{K}$*

$$\frac{d(Au, Av)}{d(u, v)} \leq \tanh\left(\frac{\Delta}{4}\right) < 1.$$

Proof of Proposition 2.1. Assumptions 1-3 imply that $A_n(\mathcal{K}) \subset \bar{\mathcal{K}}(R)$, where $R = \sqrt{K_0}/\varepsilon_0$ and

$$\bar{\mathcal{K}}(R) := \{u : u(i) > 0 \text{ for each } i \text{ and } \max_i u(i) \leq R \min_i u(i)\}.$$

Note that if $u, v \in \bar{\mathcal{K}}(R)$, then multiplying these vectors by $c_u = (\max_i u(i))^{-1}$ and $c_v = (\max_i v(i))^{-1}$, respectively, we get

$$\beta(u, v) \leq R, \quad \alpha(u, v) \leq \frac{1}{R}$$

and so $\text{diam}(\bar{\mathcal{K}}(R)) \leq 2 \ln R$.

Now let $\mathcal{K}_{k,n} = M_{k,n}\mathcal{K}$ and let $\mathbb{K}_{k,n}$ denote the set of elements of $\mathcal{K}_{k,n}$ with unit norm. Then, for each fixed k , $\mathbb{K}_{k,n}$ is a nested sequence of compact sets, and Lemma A.2 shows that the diameter of $\mathbb{K}_{k,n}$ with respect to the Hilbert metric is less than $(2 \ln R)(\tanh(\ln R/2))^{n-k-1}$. Hence Lemma A.1 shows that $\bigcap_{n>k} \mathbb{K}_{k,n}$ is a single point, which we call v_k . Since $A_{k-1}(\bigcap_{n>k} \mathbb{K}_{k,n}) = \bigcap_{n>k-1} \mathbb{K}_{k-1,n}$, it follows that $A_{k-1}v_k = \lambda_{k-1}v_{k-1}$ for some $\lambda_{k-1} > 0$.

Next, let u_0 be an arbitrary vector with

$$\|u_0\| = 1, \quad u_0(i) > \varepsilon_0 \text{ for each } i.$$

Let $u_n = M_n^T u_0 / \|M_n^T u_0\|$, $\tilde{\lambda}_n = \|A_n^T u_n\|$. Note that $u_n \in \bar{\mathcal{K}}(R)$.

Then $\{u_n\}$ and $\{v_n\}$ satisfy statements (a)–(e) of Proposition 2.1. Indeed, (a) holds by construction. (b) holds since for each vector w in $\bar{\mathcal{K}}(R)$ of unit norm

$$\min_i w(i) \geq \frac{\max_i w(i)}{R} \geq \frac{1}{dR}.$$

(c) holds since each entry of $u_n(i)$ and $v_n(i)$ is squeezed between $1/R$ and 1 while each entry of A_n is between ε_0 and $\sqrt{K_0}$.

We prove the first inequality of part (d), the second is similar. By Lemma A.2,

$$d(M_{n,n+k}v, v_n) \leq \varepsilon_k := 2(\ln R) \left(\tanh \frac{\ln R}{2} \right)^{k-1}.$$

Note that ε_k can be made as close to 0 as we wish by taking k large. By the definition of the Hilbert metric, there is a number $a_{n,k}$ such that

$$a_{n,k}v_n \leq \frac{M_{n,n+k}v}{\|M_{n,n+k}v\|} \leq a_{n,k}e^{\varepsilon_k}v_n.$$

Taking the norm, we see that $e^{-\varepsilon_k} \leq a_{n,k} \leq 1$. This proves part (d) for k' such that $e^{\varepsilon_{k'}} \leq 1 + \delta$.

Next,

$$\langle u_n, v_n \rangle = \left\langle \frac{M_{k,n}^T u_k}{\tilde{\Lambda}_n / \tilde{\Lambda}_k}, v_n \right\rangle = \frac{1}{\tilde{\Lambda}_n / \tilde{\Lambda}_k} \langle u_k, M_{k,n} v_n \rangle = \frac{\Lambda_n / \Lambda_k}{\tilde{\Lambda}_n / \tilde{\Lambda}_k} \langle u_k, v_k \rangle.$$

Due to parts (a) and (b) proved above, $\langle u_j, v_j \rangle$ are uniformly bounded from above and below, i.e., $\varepsilon_0 d \leq \langle u_j, v_j \rangle \leq 1$, proving the first inequality of part (e). To prove the second inequality, we note that by the foregoing discussion there is a constant L such that for each j and n we have

$$\frac{1}{L}v_{n-1} \leq A_{n-1}e_j \leq Lv_{n-1}.$$

Applying $M_{k,n-1}$ to this inequality and using that $M_{k,n-1}v_{n-1} = \Lambda_{n-1}/\Lambda_k v_k$, we get

$$\frac{v_k(i)}{L} \leq \frac{M_{k,n}(i,j)}{\Lambda_{n-1}/\Lambda_k} \leq Lv_k(i).$$

Combining this with parts (b) and (c) established above, we obtain the second inequality of part (e). The proof of Proposition 2.1 is complete. \square

Appendix B. Skipping generations

Proof of Proposition 2.2. (a) If Assumption 1 is satisfied, then the probability to survive till time $l(n+1) - 1$ starting from a single particle at time ln is bounded from below. One of the surviving particles will have two or more offspring of type i with probability bounded from below.

To prove parts (b) and (c), let us consider $l = 2$. Then the result for larger l follows similarly by induction since particles of generation $l+1$ are children of particles of generation l . To prove part (b), it suffices to show that $\mathbb{E}((Z_{n+2}(i;k))^2 | Z_n = e_j) \leq \bar{K}$, where $Z_{n+2}(i;k)$ is the number of particles of type i at time $n+2$ whose parents have type k . In other words, it suffices to bound $\mathbb{E}(Y^2)$, where $Y = \sum_{m=1}^N X_m$, X_m are independent, have common distribution \mathcal{X} , and are independent of the random variable N , where also $\mathbb{E}(\mathcal{X}^2) \leq K_1$, $\mathbb{E}(N^2) \leq K_2$. Note that

$$\mathbb{E}(Y^2) = \mathbb{E}(N\mathbb{E}(\mathcal{X}^2) + (N(N-1)/2)(\mathbb{E}(\mathcal{X}))^2),$$

which gives the desired bound.

Likewise, to prove (c) it suffices to show that the random variables Y_n^2 are uniformly integrable, where $Y_n = \sum_{m=1}^{N_n} X_{m,n}$, $X_{m,n}$ are independent, have common distribution \mathcal{X}_n and are independent of the random variable N_n , and \mathcal{X}_n^2, N_n^2 are uniformly integrable. We have

$$Y_n^2 = Y_n^2 \chi_{\{N_n > M\}} + Y_n^2 \chi_{\{N_n \leq M\}}.$$

The expectation of the first term equals

$$\mathbb{E}(Y_n^2 \chi_{\{N_n > M\}}) = \mathbb{E}\left(N_n \mathbb{E}(\mathcal{X}_n^2) \chi_{\{N_n > M\}} + \frac{N_n(N_n-1)}{2} \chi_{\{N_n > M\}} (\mathbb{E}(\mathcal{X}_n))^2\right).$$

This expression can be made arbitrarily small by choosing a sufficiently large M since $\mathbb{E}(\mathcal{X}_n^2)$ are uniformly bounded and N_n^2 are uniformly integrable. For the second term we have

$$Y_n^2 \mathcal{X}_{\{N_n \leq M\}} \leq \left(\sum_{m=1}^M X_{m,n} \right)^2,$$

which is uniformly integrable due to the uniform integrability of \mathcal{X}_n^2 .

It remains to establish (d). Choose l so that

$$\left(1 + \frac{\varepsilon_0}{2}\right)^l > \mathfrak{b}. \quad (\text{B.1})$$

It suffices to show that for each j

$$\mathbb{P}(Z_l = \mathbf{0}) \geq \varepsilon_1, \quad (\text{B.2})$$

where ε_1 depends only on ε_0 and \mathfrak{b} .

Given $j \in S$ and $0 \leq n \leq l$, we will say that (n, j) is 1-unstable if $n < l$ and

$$\mathbb{P}(Z_{n+1} = \mathbf{0} | Z_n = e_j) \geq \frac{\varepsilon_0}{2}.$$

Otherwise we will say that (n, j) is 1-stable.

For $p > 0$, we say that (n, j) is $(p+1)$ -unstable if it is either p -unstable or $n < l-1$ and

$$\mathbb{P}(Z_{n+1}(m) = 0 \quad \forall m : (n+1, m) \text{ is } p\text{-stable} | Z_n = e_j) \geq \frac{\varepsilon_0}{2}.$$

Otherwise we will say that (n, j) is $(p+1)$ -stable. For example (n, j) is 2-unstable if either it is 1-unstable, or, with probability which is not too small, all its children are 1-unstable.

We call l -stable pairs simply stable. A particle from generation n of type j will be called stable if the pair (n, j) is stable. We claim that all generation 0 particles are unstable. Indeed, by definition, each stable particle has at least one stable child with probability at least $1 - \frac{\varepsilon_0}{2}$, and, by Assumption 1, it has at least two stable children with probability at least ε_0 . Accordingly, for each stable particle, the expected number of its stable children is at least $1 + \frac{\varepsilon_0}{2}$. Hence, had $(0, j)$ been stable, the expected number of its (stable) decedents after l generations would have been greater than $(1 + \frac{\varepsilon_0}{2})^l \geq \mathfrak{b}$ contradicting (2.5).

Set $M = 4\mathfrak{b}/\varepsilon_0$. Then, from (2.5), it follows that for each $j \in S$ and $n \geq 0$,

$$\mathbb{P}(|Z_{n+1}| \geq M | Z_n = e_j) \leq \frac{\varepsilon_0}{4}.$$

Define η_p as follows

$$\eta_p = \inf \mathbb{P}(Z_{n+p} = \mathbf{0} | Z_n = e_j),$$

where the infimum is taken over all (n, j) which are p -unstable. Note that

$$\eta_1 \geq \frac{\varepsilon_0}{2} \quad \text{and} \quad \eta_p \geq \frac{\varepsilon_0}{4} \eta_{p-1}^M,$$

where the factor $\varepsilon_0/4$ represents the probability that the original particle had fewer than M children all of which are $(p-1)$ -unstable and η_{p-1}^M is the probability that all these children leave no descendants after $p-1$ steps. This proves (B.2) with l given by (B.1) and $\varepsilon_1 = \eta_l$. \square

Acknowledgements

While working on this article, all authors were partially supported by UMD REU grant DMS-1359307. In addition, D. Dolgopyat was supported by NSF grant DMS-1362064 and L. Korolov was supported by NSF grant DMS-1309084 and ARO grant W911NF1710419.

References

- [1] AGRESTI, A. (1975). On the extinction times of varying and random environment branching processes. *J. Appl. Prob.* **12**, 39–46.
- [2] ATHREYA K.B., NEY P.E. (2004). *Branching Processes*, Dover Publications, Mineola, NY.
- [3] BIGGINS J. D., COHN H., NERMAN O. (1999) Multi-type branching in varying environment. *Stoch. Proc. Appl.* **83**, 357–400.
- [4] BHATTACHARYA N., PERLMAN M. (2017). Time inhomogeneous branching processes conditioned on non extinction. Preprint arXiv:1703.00337.
- [5] BIRKHOFF G. (1967). *Lattice Theory*, 3d edition. AMS Colloquium Publications, **XXV**, AMS, Providence, R.I., vi+418 pp.
- [6] BOROVKOV, K., DAY, R., RICE, T. (2013). High host density favors greater virulence: a model of parasite-host dynamics based on multi-type branching processes. *J. Math. Biol.* **66**, 1123–1153.
- [7] COHN H., WANG Q. (2003). Multitype branching limit behavior. *Ann. Appl. Prob.* **13**, 490–500.
- [8] DURRETT R. (2015). *Branching Process Models of Cancer*, Mathematical Biosciences Institute Lecture Series Volume 1: Stochastics in Biological Systems, Springer.
- [9] FISCHER R. A. (1922). On the dominance ratio. *Proc. R. Soc. Edinburgh* **42**, 321–341.
- [10] FISCHER R. A. (1930). *The Genetical Theory of Natural Selection*, Oxford University Press.
- [11] FISCHER R. A. (1930). The distribution of gene ratios for rare mutations. *Proc. R. Soc. Edinburgh* **50**, 204–219.
- [12] HACCOU P., JAGERS P., VATUTIN V.A. (2005). *Branching Processes: Variation, Growth, and Extinction of Populations*, Cambridge University Press, Cambridge.
- [13] HALDANE J. B. S. (1927). A mathematical theory of natural and artificial selection, part V: Selection and mutation. *Proc. Camb. Phil. Soc.* **23**, 838–844.
- [14] HAWKINS D., ULAM S. (1944). Theory of multiplicative processes I. Los Alamos Scientific Laboratory, LADS–265.
- [15] JAGERS P. (1974). Galton-Watson processes in varying environments. *J. Appl. Prob.* **11**, 174–178.
- [16] JONES O. D. (1997). On the convergence of multitype branching processes with varying environments. *Ann. Appl. Prob.* **7**, 772–801.
- [17] MODE C.J., (1971). *Multitype Branching Processes: Theory and Applications*, Elsevier, New York.
- [18] KERSTING G. (2017). A unifying approach to branching processes in varying environments. Preprint arXiv:1703.01960.
- [19] KIMMEL M., AXELROD D. E. (2015). *Branching Processes in Biology*, 2d edition. Interdisciplinary Applied Mathematics, 19. Springer, New York, xx+280 pp.
- [20] LIVERANI C. (1995). Decay of correlations. *Ann. Math.* **142**, 239–301.
- [21] SEMENOV N. N. (1934). *Chain Reactions*, (in Russian), Goshimizdat.
- [22] SENETA E. (2006). *Non-negative Matrices and Markov Chains*, Revised Printing, Springer, New York.
- [23] WATSON H.W., GALTON F. (1874) On the probability of the extinction of families. *J. Anthropol. Inst. Great Britain and Ireland* **4**, 138–144.