J. Appl. Prob. **55,** 701–727 (2018)
*doi:*10.1017/jpr.2018.46

MULTI-TYPE BRANCHING PROCESSES WITH TIME-DEPENDENT BRANCHING RATES

D. DOLGOPYAT,* **
P. HEBBAR * AND
L. KORALOV,* *Univeristy of Maryland*
M. PERLMAN,*** *Stanford Univeristy*

Abstract

Under mild nondegeneracy 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 ∞.

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

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

1. Introduction

The mathematical study of branching processes goes back to the work of Galton and Watson [23] and their interest in the probabilities of the long-term survival of aristocratic family names. Later it was realized that similar mathematical models could be used to describe the evolution of a variety of biological populations, for example, in genetics [9]–[11], [13], and in the study of certain chemical and nuclear reactions [14], [21]. Branching processes are central in the study of the evolution of various populations such as bacteria, cancer cells, carriers of a particular form of a gene, 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, a multi-type continuous-time Markov branching process may be used to describe the dynamics of the spread of two types of parasite 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 a different number 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 [20]. In addition, a vast number of applications of multi-type branching processes in biology can be found in [12] and [18].

In this paper we are concerned with the long-time behavior of multi-type branching processes with time-dependent branching rates. We stress that the temporal inhomogeneity is due to the

Received 1 March 2017; revision received 19 May 2018.
* Postal address: Department of Mathematics, University of Maryland, College Park, MD 20742, USA.
** Email address: dmitry@math.umd.edu
*** Postal address: Department of Mathematics, Stanford University, Stanford, CA 94305, USA.

dependence of the branching rates over time (this dependence may model a varying environment for the entire process) and not on the age of the particles (which is a well-studied model). We believe that the methods presented in this 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 the subject of future work.

For multi-type processes with constant branching rates, according to classical results (see [2, Chapter 5] and the references therein), three different cases can be distinguished.

The supercritical case. The expectation of the total population size grows exponentially, and the total population grows exponentially with positive probability as time goes to ∞.

The subcritical case. The expectation of the total population size decays exponentially, and the population goes extinct with overwhelming probability, that is, the probability that the population at time n is nonzero decays exponentially in n.

The critical case. The population also goes extinct, but the expectation of the total population size remains bounded away from 0 and ∞, 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 supercritical, subcritical, 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 yields a criterion for almost sure extinction of the process in terms of the asymptotic behavior of the branching rates, under mild nondegeneracy 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 the particle number conditioned on survival in terms of the branching rates was obtained independently in [3] and [17]. Our second result yields 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 nondegeneracy 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) become extinct with probability 1, 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) become extinct with probability 1, 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.

We 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 the 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 [4]. Sufficient conditions for the continuity of the limiting distribution function were stated in [7].

On the other hand, in the time-independent case, a subcritical process conditioned on survival tends to a random limit. Now, our processes in the third category do not necessarily have this property (for example, 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 $n \ge 0$, 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 $a_1 + \cdots + a_d$ particles: a_1 particles of type 1, a_2 particles of type 2, and so on. A d-type branching process $\{Z_n\}$ is obtained by starting with a positive finite number of particles at time 0, and then replacing each particle of each type i, $i \in S$, that is alive at time n, $n \ge 0$, 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 $\mathbb{E}(Z_n(i) \mid Z_0 = e_j)$, where e_j is the unit vector in the jth 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

$$f_n^j(s) = \mathbb{E}\biggl(\prod_{i=1}^d s_i^{Z_n(i)} \biggm| Z_0 = e_j\biggr), \qquad g_n^j(s) = \mathbb{E}\biggl(\prod_{i=1}^d s_i^{Z_{n+1}(i)} \biggm| Z_n = e_j\biggr).$$

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

$$f_n(s) = f_{n-1}(g_{n-1}(s)) = (g_0 \circ g_1 \circ \cdots \circ g_{n-1})(s), \qquad f_n(\mathbf{1}) = \mathbf{1},$$

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

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

Thus, $f_{0,n} = f_n$. Denote

$$M_n(j,i) = \frac{\partial f_n^j}{\partial s_i}(\mathbf{1}) = \mathbb{E}(Z_n(i) \mid Z_0 = e_j),$$
$$A_n(j,i) = \frac{\partial g_n^j}{\partial s_i}(\mathbf{1}) = \mathbb{E}(Z_{n+1}(i) \mid Z_n = e_j),$$

Then

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

where A_n and M_n are viewed as matrices. We also define

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

Let $\|\cdot\|$ denote the following norm of a d-dimensional vector: $\|v\| = |v_1| + \cdots + |v_d|$. We will use certain nondegeneracy 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:

(A1) $\mathbb{P}(Z_{n+1}(i) \geq 2 \mid Z_n = e_j) \geq \varepsilon_0$;

(A2) $\mathbb{P}(Z_{n+1} = \mathbf{0} \mid Z_n = e_j) \geq \varepsilon_0$, where $\mathbf{0} = (0, 0, \dots, 0) \in \mathbb{R}^d$;

(A3) $\mathbb{E}(\|Z_{n+1}\|^2 \mid Z_n = e_j) \leq K_0$.

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 (A1) and (A3), 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*

$$A_{n-1} v_n = \lambda_{n-1} v_{n-1}, \qquad A_{n-1}^{\top} u_{n-1} = \tilde{\lambda}_{n-1} u_n;$$

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

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

whenever $k \geq k'$, v *and* u *are nonzero vectors with nonnegative components, and the inequality between vectors is understood as the inequality between their components;*

(e) *there is a* $K > 0$ *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*

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

This proposition can be derived from, for example, the results of [22, Chapter 3]. Indeed, from our assumptions (A1)–(A3), it follows that the matrices A_n have Birkhoff's contraction coefficient (in the terminology of [22]) uniformly bounded away from 1. This implies that the conditions of [22, Lemma 3.4.] are fulfilled (which, in particular, implies that the family $M_{k,n}$ is weakly ergodic; see [22]). This lemma and [22, Exercise 3.5] easily imply the existence of the 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 the vectors u_n and v_n is as follows. The vector u_n yields the asymptotic proportions of different particles in the population provided that Z_n is large (see (4.10) and (4.11) 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

$$\lim_{N\to\infty} \frac{\mathbb{E}(z_N^* \mid Z_n = u')}{\mathbb{E}(z_N^* \mid Z_n = u'')} = \frac{\langle v_n, u'\rangle}{\langle v_n, u''\rangle} \quad \text{for each } u', u'' \in \mathbb{Z}^d,\ n \in \mathbb{Z}_+.$$

Thus, v_n controls the expected future size of the population.

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

Theorem 2.1. *Under assumptions (A1)–(A3), if extinction of the process $\{Z_n\}$ occurs with probability 1 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 1 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 [17] for recent results in the case of single-type branching processes.

The next lemma easily follows from Theorem 2.1.

Lemma 2.1. *Suppose assumptions (A1)–(A3) hold.*

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

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

Remark 2.4. The divergence of $\sum_{k=1}^{\infty}(1/\Lambda_{lk})$ 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. (a) It suffices to note that $\sum_{k=1}^{\infty}(1/\Lambda_k) = \infty$ if and only if $\sum_{k=l+1}^{\infty}(\Lambda_l/\Lambda_k) = \infty$, while the latter is equivalent to the almost sure extinction of the process $\{{}_jZ'_n\}$ by Theorem 2.1.

(b) We observe that, under assumption (A3), there exists a constant C such that for each k and each $lk \le n < l(k+1)$, we have $\Lambda_{lk}/C \le \Lambda_n \le C\Lambda_{lk}$. □

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

Lemma 2.2. *Under assumptions (A1)–(A3), for each initial population of the branching process, there is a constant $C > 0$ such that*

$$\frac{\Lambda_n}{C} \le \mathbb{E}\|Z_n\| \le C\Lambda_n, \qquad n \ge 1, \tag{2.1}$$

$$\frac{1}{C}\left(\sum_{k=1}^{n}\frac{1}{\Lambda_k}\right)^{-1} \le \mathbb{P}(Z_n \ne \mathbf{0}) \le C\left(\sum_{k=1}^{n}\frac{1}{\Lambda_k}\right)^{-1}, \qquad n \ge 1. \tag{2.2}$$

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

(A4) the random variables $\|{}_jX_n\|^2$, $j \in S$, $n \ge 0$, are uniformly integrable;

(A5) $\mathbb{P}(Z_n \ne \mathbf{0}) \to 0$ as $n \to \infty$ (equivalently, $\sum_{k=1}^{n}(1/\Lambda_k) \to \infty$, by (2.2));

(A6) $\mathbb{E}\|Z_n\|/\mathbb{P}(Z_n \ne \mathbf{0}) \to \infty$ as $n \to \infty$ (equivalently, $\Lambda_n\sum_{k=1}^{n}(1/\Lambda_k) \to \infty$, by (2.1) and (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 $Z_n \ne \mathbf{0}$. In other words, we treat the event $Z_n \ne \mathbf{0}$ as a new probability space, with the measure $\mathbb{P}'$ obtained from the underlying measure $\mathbb{P}$ via $\mathbb{P}'(A) = \mathbb{P}(A)/\mathbb{P}(Z_n \ne \mathbf{0})$. When we write ${}_j\zeta_n$, we mean that the initial population for the branching process is specified as e_j.

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

Theorem 2.2. *Under assumptions (A1)–(A6), for each initial population of the branching process and each vector u with positive components, we have the following limit in distribution:*

$$\frac{\langle \zeta_n, u\rangle}{\mathbb{E}\langle \zeta_n, u\rangle} \xrightarrow{\text{D}} \xi \quad \textit{as } n \to \infty, \tag{2.3}$$

where ξ is an exponential random variable with parameter 1. Moreover, if assumptions (A1)–(A5) are satisfied and, for some initial population, the limit in (2.3) is as specified, then assumption (A6) is also satisfied.

We say that a process is *uniformly critical* if it satisfies assumptions (A1)–(A4) and there is a constant b such that for each n, k, i, j, we have

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

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

$$\sum_{k=1}^{\infty} \frac{1}{\Lambda_k} = \infty, \qquad \lim_{n\to\infty} \Lambda_n \bigg(\sum_{k=1}^{n} \frac{1}{\Lambda_k} \bigg) = \infty.$$

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 an exponential.

The next proposition and the subsequent lemma will be helpful for comparing our results to those of [15]. Note that Proposition 2.2(d) can be used to show that, under (2.4) (or even under a weaker condition (2.5)), assumption (A2) almost follows from assumption (A1) in the sense that assumption (A2) 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 $a_1 + \cdots + a_d$ particles: a_1 particles of type 1, a_2 particles of type 2, and so on.

Proposition 2.2. (a) *If P_n satisfies assumption (A1) then $\tilde{P}_n$ satisfies assumption (A1) for each l.*

(b) *If P_n satisfies assumption (A3) then $\tilde{P}_n$ satisfies assumption (A3) for each l.*

(c) *If P_n satisfies assumption (A4) then $\tilde{P}_n$ satisfies assumption (A4) for each l.*

(d) *If P_n satisfies assumption (A1) and there is a constant $\mathfrak{b}$ such that, for each n, k, j,*

$$\mathbb{E}(|Z_{n+k}| \mid Z_n = e_j) \le \mathfrak{b}, \tag{2.5}$$

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

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

Proof. See Appendix B. □

Lemma 2.3. *If $\{Z_n\}$ satisfies assumptions (A1) and (A4), and (2.4) holds, then extinction happens with probability 1 and (2.3) holds.*

Proof. See Section 4. □

For single-type branching processes, Lemma 2.3 is helpful in showing that our results imply [15, Theorem 5]. In fact, the assumptions of [15, Theorem 5] (generalized to the multi-type case) are:

- our assumption (A4);
- that (2.4) holds;
- that there is $\bar{\varepsilon}_0 > 0$ such that, for each n, i, j,

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

We claim that under assumption (A4), (2.6) is equivalent to assumption (A1). On the one hand,

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

On the other hand, by assumption (A4), we can take $N \ge 2$ such that

$$\mathbb{E}(Z^2_{n+1}(i)\,\mathbf{1}_{\{Z_{n+1}(i)\ge N\}} \mid Z_n = e_j) \le \tfrac{1}{2}\bar{\varepsilon}_0 \quad \text{for all } n,$$

where $\mathbf{1}_{\{Z_{n+1}(i)\ge N\}}$ is the indicator function of the event $\{Z_{n+1}(i) \ge N\}$. Then

$$\mathbb{E}(Z^2_{n+1}(i) - Z_{n+1}(i) \mid Z_n = e_j) \le \tfrac{1}{2}\bar{\varepsilon}_0 + (N^2 - N)\mathbb{P}(Z_{n+1}(i) \ge 2 \mid Z_n = e_j).$$

Thus, if (2.6) holds then

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

proving that assumption (A1) is equivalent to (2.6).

The results of [15, Theorem 5] are: our Lemma 4.3 (in the single-type case) and our (2.3) (in the single-type case). The latter holds by Lemma 2.3. We prove Lemma 4.3 in Section 4 under assumptions (A1)–(A6). However, under assumptions (A1) and (A4), 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 versus extinction

Proof of Theorem 2.1. We split the proof into two parts.

Part 1: $\sum_{k=1}^{\infty}(1/\Lambda_k) < \infty$ implies a positive probability of survival. 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

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

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

$$D_n(j_1, j_2) = \operatorname{cov}(Z_n(j_1), Z_n(j_2)).$$

One-step analysis yields

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

where

$$S_n = \sum_{i=1}^{d} M_n(j,i)\sigma_n^2(i) \quad \text{and} \quad \sigma_n^2(j_1, j_2)(i) = \operatorname{cov}({}_iX_n(j_1), {}_iX_n(j_2)).$$

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

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

Hence,

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

with some constants B_1, B_2.

Thus, $\|D_n\| \le \tilde{B}\Lambda_n^2$, 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 2: $\sum_{k=1}^{\infty}(1/\Lambda_k) = \infty$ *implies that extinction occurs with probability 1.* 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 (4.4). By the Taylor formula with respect to $s = \mathbf{1}$,

$$\begin{aligned}\langle \mathbf{1} - f_{0,n}(s), u_0\rangle &\\ &= \langle Dg_0(\mathbf{1})(\mathbf{1} - f_{1,n}(s)), u_0\rangle - \tfrac{1}{2}\langle(\mathbf{1} - f_{1,n}(s))^\top Hg_0(\eta_{1,n})(\mathbf{1} - f_{1,n}(s)), u_0\rangle\\ &= \langle A_0(\mathbf{1} - f_{1,n}(s)), u_0\rangle - \tfrac{1}{2}\langle(\mathbf{1} - f_{1,n}(s))^\top Hg_0(\eta_{1,n})(\mathbf{1} - f_{1,n}(s)), u_0\rangle,\end{aligned}$$

where Dg_0 is the gradient of g_0 and $\eta_{1,n} = \eta_{1,n}(j, s)$ satisfies $f_{0,n}^j(s) \le \eta_{1,n} \le 1$ for each component $j \in S$ and $s \in [0, 1]^d$. Here Hg_0 denotes the Hessian matrix applied to each component of the vector function g_0 separately, then multiplied by vectors $(\mathbf{1} - f_{1,n}(s))^\top$ and $(\mathbf{1} - f_{1,n}(s))$ yielding 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,

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

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

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

where the last equality follows from the simple relation

$$\frac{1}{a} = \frac{1}{b}\Big(1 - \frac{c}{b}\Big)^{-1} \quad\Longrightarrow\quad \frac{1}{a} = \frac{1}{b} + \frac{c}{ba}.$$

By iterating the previous equality n times, we obtain

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

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

Let

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

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 $f_{k,n}(s)(i) \le \eta_{k+1,n}(i) \le 1$. Then (3.2) takes the form

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

We will need the following lemma, the proof of which follows the proof of this theorem. Denote

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

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 (A1)–(A3), there exists $C > 1$ such that for each n and each $s \in [0, 1]^d \setminus \{\mathbf{1}\}$, we have*

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

Using Lemma 3.1 in (3.4), we obtain

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

Therefore,

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

Note that $1 - f^j_{0,n}(\mathbf{0}) = \mathbb{P}(Z_n \ne \mathbf{0} \mid Z_0 = e_j)$ for each $j \in S$ and, hence, if $\lim_{n\to\infty} \Xi_n = \infty$ then

$$\lim_{n\to\infty} \mathbb{P}(Z_n \ne \mathbf{0} \mid Z_0 = e_j) = 0.$$

Thus, extinction occurs with probability 1 if the initial population is e_j. Therefore, since j was arbitrary, extinction occurs with probability 1 for every initial population. □

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 $0 \le k \le n - 1$ and $s \in [0, 1]^d \setminus \{\mathbf{1}\}$, we have

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

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

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

Now, we know that $f^j_{k,n}(\mathbf{0}) \le f^j_{k,n}(s) \le \eta_{k+1,n}(j) \le 1$ for each k and $j \in S$. Also, $f^j_{k,n}(\mathbf{0}) = \mathbb{P}(Z_n = \mathbf{0} \mid Z_k = e_j) \ge \varepsilon_0$ for each $k \le n - 1$ and, thus, $\varepsilon_0 \le \eta_{k+1,n}(j) \le 1$ for each $k \le n - 1$ and $j \in S$. Thus, by assumptions (A1)–(A3), there exists a constant $c_1 > 0$ such that for each vector ζ with nonnegative components, we have

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

In particular, we have

$$\frac{1}{c_1}\|\mathbf{1}-f_{k+1,n}(s)\|^2 \le \langle(\mathbf{1}-f_{k+1,n}(s))^T Hg_k(\eta_{k+1,n})(\mathbf{1}-f_{k+1,n}(s)), u_k\rangle \le c_1\|\mathbf{1}-f_{k+1,n}(s)\|^2. \tag{3.7}$$

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

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

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

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

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

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

since g_k is uniformly Lipschitz due to assumption (A3).

We observe that by assumptions (A1)–(A3), each entry of the matrix A_k is uniformly bounded from above and below, that is, there exist positive constants r and R such that, for each $i, j \in S$,

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

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

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

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

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

$$\begin{aligned}&\langle \mathbf{1}-f_{k,n}(s), u_k\rangle\\ &\quad= \langle A_k(\mathbf{1}-f_{k+1,n}(s)), u_k\rangle - \tfrac{1}{2}\langle(\mathbf{1}-f_{k+1,n}(s))^\top Hg_k(\eta_{k+1,n})(\mathbf{1}-f_{k+1,n}(s)), u_k\rangle,\end{aligned}$$

we obtain

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

thus,

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

So, for $\tilde{c_2} = 2/(r\bar{\varepsilon})$, we have

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

Case 2: $1 - f^j_{k+1,n}(s) > r\bar{\varepsilon}/c_1$ *for some* $j \in S$. We want to prove that there exists a $\gamma > 0$ such that $1 - f^j_{k,n}(s) \geq \gamma$. From assumptions (A1) and (A2), for each $j \in S$,

$$g^j_k(s) = \mathbb{E}\bigg(\prod_{i=1}^{d} s_i^{Z_{k+1}(i)} \;\bigg|\; Z_k = e_j\bigg) \leq (1 - \varepsilon_0) + \varepsilon_0 s_j^2,$$

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

$$f^j_{k,n}(s) = g^j_k(f_{k+1,n}(s)) \leq (1 - \varepsilon_0) + \varepsilon_0\bigg(1 - \frac{r\bar{\varepsilon}}{c_1}\bigg)^2 < 1,$$

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

$$1 - f^j_{k,n}(s) > \gamma,$$

which is the required inequality.

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

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

Proof of Lemma 2.2. Let $\bar{u} = \mathbb{E}Z_1$. By assumptions (A1)–(A3), for every initial population, there is a constant $c > 0$ such that $c^{-1}u_1 \leq \bar{u} \leq cu_1$, where the inequality between vectors is understood as the inequality between their components. Then, since $M^\top_{1,n}\bar{u} = \mathbb{E}Z_n$,

$$c^{-1}M^\top_{1,n}u_1 \leq \mathbb{E}Z_n \leq cM^\top_{1,n}u_1.$$

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

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

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

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

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

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:

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

Here H denotes the Hessian matrix. It is applied to each component of g_k separately, then multiplied by vectors $v^\top_{k+1}$ and v_{k+1} to obtain 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-hand side of (4.1) are positive, the sequence Γ_n is increasing. In each term in (4.1), each of the factors λ_k, $\langle v_{k+1}^\top H g_k(\mathbf{1}) v_{k+1}, u_k\rangle$, $\langle v_{k+1}, u_{k+1}\rangle$, and $\langle v_k, u_k\rangle$ is bounded from above and below uniformly in k by assumptions (A1)–(A3) and Proposition 2.1. Therefore, by Proposition 2.1, there is a positive constant C such that

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

and, consequently,

$$\frac{\Xi_n}{C} = \frac{1}{C}\sum_{k=0}^{n-1}\frac{1}{\Lambda_{k+1}} \le \Gamma_n \le C\sum_{k=0}^{n-1}\frac{1}{\Lambda_{k+1}} = C\Xi_n.$$

Assumptions (A5) and (A6) can now be written as

$$\Gamma_n \to \infty, \qquad \Lambda_n\Gamma_n \to \infty \quad \text{as } n \to \infty. \tag{4.3}$$

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

Theorem 4.1. *Under assumptions (A1)–(A6), for each $j \in S$, we have the limit in distribution*

$$\frac{\langle {}_j\zeta_n, u_n\rangle}{\mathbb{E}\langle {}_j\zeta_n, u_n\rangle} \xrightarrow{\text{D}} \xi \quad \text{as } n \to \infty,$$

where ξ is an exponential random variable with parameter 1.

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

$$\mathbb{E}\biggl(\exp\biggl(\frac{-\varkappa\langle {}_jZ_n, u_n\rangle\mathbb{P}({}_jZ_n \ne \mathbf{0})}{\mathbb{E}(\langle {}_jZ_n, u_n\rangle)}\biggr) \biggm| {}_jZ_n \ne \mathbf{0}\biggr) \to \frac{1}{1+\varkappa} \quad \text{as } n \to \infty.$$

Define vectors $\bar{s}_j$ such that the ith component of $\bar{s}_j$ is

$$\bar{s}_j(i) = \exp\biggl(-\frac{\varkappa u_n(i)\mathbb{P}({}_jZ_n \ne \mathbf{0})}{\mathbb{E}(\langle {}_jZ_n, u_n\rangle)}\biggr).$$

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

$$f_n^j(\bar{s}_j) = \mathbb{E}\biggl(\exp\biggl(-\frac{\varkappa\langle {}_jZ_n, u_n\rangle\mathbb{P}({}_jZ_n \ne \mathbf{0})}{\mathbb{E}(\langle {}_jZ_n, u_n\rangle)}\biggr)\biggr).$$

Thus, we want to show that

$$1 - \frac{1 - f_n^j(\bar{s}_j)}{\mathbb{P}({}_jZ_n \ne \mathbf{0})} \to \frac{1}{1+\varkappa} \quad \text{as } n \to \infty. \tag{4.4}$$

In order to prove (4.4), it will be useful to study the asymptotic behavior of the sum on the right-hand side of (3.2). We first obtain 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 nonnegative coefficients and $Hg_k^j(s)$ is a matrix with entries that are mixed second derivatives of g_k^j. Therefore, (3.2) yields

$$\begin{aligned}&\left(\frac{1}{\tilde{\Lambda}_n\langle \mathbf{1}-s,u_n\rangle}+\frac{1}{2}\sum_{k=0}^{n-1}\frac{\langle(\mathbf{1}-f_{k+1,n}(s))^\top Hg_k(\mathbf{1})(\mathbf{1}-f_{k+1,n}(s)),u_k\rangle}{\tilde{\Lambda}_{k+1}\langle(\mathbf{1}-f_{k+1,n}(s)),u_{k+1}\rangle\langle\mathbf{1}-f_{k,n}(s),u_k\rangle}\right)^{-1}\\&\quad\le\langle\mathbf{1}-f_{0,n}(s),u_0\rangle\\&\quad\le\left(\frac{1}{\tilde{\Lambda}_n\langle \mathbf{1}-s,u_n\rangle}+\frac{1}{2}\sum_{k=0}^{n-1}\frac{\langle(\mathbf{1}-f_{k+1,n}(s))^\top Hg_k(f_{k,n}(s))(\mathbf{1}-f_{k+1,n}(s)),u_k\rangle}{\tilde{\Lambda}_{k+1}\langle(\mathbf{1}-f_{k+1,n}(s)),u_{k+1}\rangle\langle\mathbf{1}-f_{k,n}(s),u_k\rangle}\right)^{-1}.\end{aligned}\tag{4.5}$$

We 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 $k\le K+1$. By formally linearizing the mappings $g_k, g_{k+1},\dots,g_K$, we write

$$\mathbf{1}-f_{k,n}(s)\approx A_kA_{k+1}\cdots A_K(\mathbf{1}-f_{K+1,n}(s)).\tag{4.6}$$

We know that

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

and, thus,

$$v_k=A_kA_{k+1}\cdots A_K\frac{v_{K+1}}{\prod_{i=k}^K\lambda_i}.\tag{4.7}$$

Note the similarity between (4.6) and (4.7): 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 (4.5) 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 allow us to simplify (4.5).

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

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

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

$$\mathbf{1}-f_{k,n}(s)=c_{k,n}(v_k+\delta_{k,n}),\tag{4.8}$$

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

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

$$1-f_{k,n}^i(\mathbf{0})=\mathbb{P}(Z_n\ne\mathbf{0}\mid Z_k=e_i)\quad\text{and}\quad\mathbb{P}(Z_n\ne\mathbf{0}\mid Z_k=e_i)\to 0\quad\text{as }\to\infty$$

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 (A1)–(A6),*

$$\lim_{n\to\infty} \frac{\alpha(n,s)}{\Gamma_n} = 1 \quad \textit{uniformly in } s \in [0,1]^d \setminus \{\mathbf{1}\}.$$

We return to the proof of (4.4). 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 (4.4) we can replace the jth 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 (4.4), it is sufficient to show that

$$1 - \frac{\langle v_0, e_j \rangle}{\langle v_0, u_0 \rangle \mathbb{P}({}_jZ_n \neq \mathbf{0})} \bigg(\bigg[\tilde{\Lambda}_n \varkappa \bigg(\frac{\mathbb{P}({}_jZ_n \neq \mathbf{0})}{\mathbb{E}(\langle {}_jZ_n, u_n \rangle)} + o\bigg(\frac{\mathbb{P}({}_jZ_n \neq \mathbf{0})}{\mathbb{E}(\langle {}_jZ_n, u_n \rangle)} \bigg) \bigg) \langle u_n, u_n \rangle \bigg]^{-1} + \Gamma_n \bigg)^{-1} \to \frac{1}{1+\varkappa},$$

where we use Lemma 4.2 to transform (3.4) and linearize $\mathbf{1} - \bar{s}_j$. The left-hand side can be written as

$$1 - \frac{\langle v_0, e_j \rangle}{\langle v_0, u_0 \rangle \mathbb{P}({}_jZ_n \neq \mathbf{0}) \Gamma_n} \times \bigg(\bigg[\tilde{\Lambda}_n \Gamma_n \varkappa \bigg(\frac{\mathbb{P}({}_jZ_n \neq \mathbf{0})}{\mathbb{E}(\langle {}_jZ_n, u_n \rangle)} + o\bigg(\frac{\mathbb{P}({}_jZ_n \neq \mathbf{0})}{\mathbb{E}(\langle {}_jZ_n, u_n \rangle)} \bigg) \bigg) \langle u_n, u_n \rangle \bigg]^{-1} + 1 \bigg)^{-1}. \tag{4.9}$$

We will need the next two lemmas.

Lemma 4.3. *Under assumptions (A1)–(A6),*

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

Lemma 4.4. *Under assumptions (A1)–(A6),*

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

Applying the above two lemmas to transform the expression in (4.9), we obtain

$$\begin{aligned} &\lim_{n\to\infty} \bigg(1 - \frac{\langle v_0, e_j \rangle}{\langle v_0, u_0 \rangle \mathbb{P}({}_jZ_n \neq \mathbf{0}) \Gamma_n} \\ &\qquad \times \bigg(\bigg[\tilde{\Lambda}_n \Gamma_n \varkappa \bigg(\frac{\mathbb{P}({}_jZ_n \neq \mathbf{0})}{\mathbb{E}(\langle {}_jZ_n, u_n \rangle)} + o\bigg(\frac{\mathbb{P}({}_jZ_n \neq \mathbf{0})}{\mathbb{E}(\langle {}_jZ_n, u_n \rangle)} \bigg) \bigg) \langle u_n, u_n \rangle \bigg]^{-1} + 1 \bigg)^{-1} \bigg) \\ &= 1 - \bigg(\frac{1}{\varkappa} + 1 \bigg)^{-1} \\ &= \frac{1}{1+\varkappa}. \end{aligned}$$

This completes the proof of Theorem 4.1. □

Proof of Theorem 2.2. First, let assumptions (A1)–(A6) be satisfied. Let $\mathcal{P}\colon v \to v/\|v\|$ be the projection onto the unit sphere with the convention that $\mathcal{P}(\mathbf{0}) = \mathbf{0}$. We claim that

$$\lim_{n\to\infty} \|\mathcal{P}(\mathbb{E}_j\zeta_n) - u_n\| = 0, \tag{4.10}$$

$$\lim_{n\to\infty} \mathbb{P}(\|\mathcal{P}\left({}_j\zeta_n\right) - u_n\| > \varepsilon) = 0 \tag{4.11}$$

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

$$(1-\delta)u_{n+k'} \le \frac{M^\top_{n,n+k'}u}{\|M^\top_{n,n+k'}u\|} \le (1+\delta)u_{n+k'} \tag{4.12}$$

whenever u is a nonzero vector with nonnegative components. Let ${}_j\zeta^{k'}_n$ 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 ${}_j\zeta^{k'}_n$ is different from ${}_j\zeta_{n+k'}$, the latter can be obtained from ${}_j\zeta^{k'}_n$ by conditioning on the event of nonextinction. Since the extinction of a large initial population in k' steps occurs with a small probability and since, by Theorem 4.1, for each $a > 0$ we have $\mathbb{P}(\|{}_j\zeta_n\| > a) \to 1$ as $n \to \infty$, we then obtain

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

Also note that $\lim_{n\to\infty} (\mathcal{P}(\mathbb{E}_j\zeta^{k'}_n) - \mathcal{P}(\mathbb{E}_j\zeta_{n+k'})) = 0$. Therefore, since $\delta > 0$ was arbitrarily small, (4.10) and (4.11) will follow if we show that

$$\|\mathcal{P}(\mathbb{E}_j\zeta^{k'}_n) - u_{n+k'}\| \le \delta \quad \text{for all sufficiently large } n \tag{4.13}$$

and

$$\lim_{n\to\infty} \mathbb{P}(\|\mathcal{P}({}_j\zeta^{k'}_n) - u_{n+k'}\| > \varepsilon) = 0. \tag{4.14}$$

Equation (4.13) immediately follows from (4.12). Equation (4.14) is a consequence of

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

which can be derived from the Chebyshev inequality since, for each $a > 0$, we have

$$\mathbb{P}(\|{}_j\zeta_n\| > a) \to 1 \quad \text{as } n \to \infty.$$

Thus, we have (4.10) and (4.11).

Next, we show that (4.10) and (4.11), 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 limit in probability

$$\frac{\langle {}_j\zeta_n, u_n\rangle}{\mathbb{E}\langle {}_j\zeta_n, u_n\rangle} - \frac{\langle {}_j\zeta_n, u\rangle}{\mathbb{E}\langle {}_j\zeta_n, u\rangle} \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \to \infty. \tag{4.15}$$

From (4.10), we know that

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

From (4.11), the following two limits hold in probability:

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

Therefore, the right-hand side of

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

tends to 0 in probability (the factor in the brackets tends to 0 in probability while the first factor is bounded in L^1). This justifies (4.15) 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 0, 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 (A1)–(A4) 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 (A1)–(A5) are satisfied. If assumption (A6) 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 yields a contradiction. □

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 (A1)–(A3). By Theorem 2.1, which can be applied due to (2.4), $\{\tilde{Z}_n\}$ becomes extinct almost surely. This implies the almost sure extinction of $\{Z_n\}$.

Similarly, from Theorem 2.2 (which can be applied since assumptions (A5) and (A6) 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 $0 \leq r < l$,

$$\lim_{N\to\infty}\Big(\frac{\langle\zeta_{Nl+r}, u\rangle}{\langle\zeta_{Nl}, u_{Nl}\rangle} - \frac{\Lambda_{Nl+r}}{\Lambda_{Nl}}\langle u_{Nl+r}, u\rangle\Big) = 0 \quad \text{in probability.} \tag{4.16}$$

Lemma 2.3 follows directly from (4.16) and Theorem 2.2 applied to $\{Z_{Nl}\}$. The proof of (4.16) is similar to the proof of (4.11) and (4.15), so we leave it to the reader. □

It still remains to prove Lemmas 4.1–4.4.

Proof of Lemma 4.1. Suppose that we have (4.8) 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 $1 - f^i_{k+1,n}(s) \leq \varepsilon$ for all i (from definition of $J(n, \varepsilon)$) and the fact that

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

The uniformity in k follows from assumption (A3). Thus,

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

where $\|\alpha'_{k,n}\|$ can be made arbitrarily small, uniformly in k, by selecting sufficiently small ε and ε''. Here we use (4.8) 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 $|(c_{k,n}/c_{k+1,n})-\lambda_k|\le\varepsilon'$ holds for $0\le k<J(n,\varepsilon)-K$, provided that ε and ε'' are sufficiently small. We have demonstrated, therefore, that it is sufficient to establish (4.8) with the estimate $\|\delta_{k,n}\|\le\varepsilon'$ only.

From Proposition 2.1(d), there exists $k'\in\mathbb{N}$ such that

$$\left(1-\frac{\varepsilon'}{2d}\right)v_k\le\frac{M_{k,k+k'}v}{\|M_{k,k+k'}v\|}\le\left(1+\frac{\varepsilon'}{2d}\right)v_k \tag{4.17}$$

for each k and each nonzero vector v with nonnegative components. Since

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

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

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

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

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

Combined with (4.17), this yields

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

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

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

which yields the desired estimate. □

Proof of Lemma 4.2. We split 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

$$\left|\frac{\alpha(n,s)-\alpha(J(n,\varepsilon)-K-1,s)}{\Gamma_n}+\frac{\alpha(J(n,\varepsilon)-K-1,s)-\Gamma_{J(n,\varepsilon)-K-1}}{\Gamma_n}\right.$$
$$\left.+\frac{\Gamma_{J(n,\varepsilon)-K-1}-\Gamma_n}{\Gamma_n}\right|$$
$$\le\sigma\quad\text{for all } s \text{ and all sufficiently large } n. \tag{4.18}$$

Part 1. We first estimate the middle term in (4.18). By Lemma 4.1, for each $\sigma' > 0$, there exist a natural number K and $\varepsilon_1 > 0$ such that

$$(1-\sigma')c_{k,n}v_k \le \mathbf{1} - f_{k,n}(s) \le (1+\sigma')c_{k,n}v_k \quad \text{for each } k < J(n,\varepsilon_1) - K.$$

By assumption (A4), $\|{}_iX_n\|^2$ are uniformly integrable, and, thus, the matrices $Hg_k^i(s)$, $k \ge 0$, $i \in S$, are equicontinuous in s. Note also that $\|Hg_k^i(\mathbf{1})\| \ge c > 0$ for all $k \ge 0$, $i \in S$. 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

$$\begin{aligned}
&\left|\frac{1}{2}\sum_{k=0}^{J(n,\varepsilon)-K-1} \frac{\langle(\mathbf{1}-f_{k+1,n}(s))^\top Hg_k(\eta_{k+1,n})(\mathbf{1}-f_{k+1,n}(s)), u_k\rangle}{\tilde{\Lambda}_{k+1}\langle(\mathbf{1}-f_{k+1,n}(s)), u_{k+1}\rangle\langle\mathbf{1}-f_{k,n}(s), u_k\rangle} - \Gamma_{J(n,\varepsilon)-K-1}\right| \\
&\quad \le \tilde{c}\sigma'\Gamma_{J(n,\varepsilon)-K-1} \\
&\quad \le \tilde{c}\sigma'\Gamma_n.
\end{aligned}$$

(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 $\sigma' \le \sigma/3\tilde{c}$, we obtain

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

Part 2. Now we estimate the third term in (4.18). 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 carrying out the same steps as used to obtain (3.2), we have

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

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

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

Therefore,

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

Since $J(n, \varepsilon) \to \infty$ as $n \to \infty$, by assumption (A6) (see (4.3) and Proposition 2.1(e)), we have

$$\left| \frac{C}{\tilde{\Lambda}_{J(n,\varepsilon)} \Gamma_{J(n,\varepsilon)} \varepsilon \bar{\varepsilon}} \right| < \frac{\sigma}{6} \quad \text{for all sufficiently large } n.$$

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 (A6) and the fact that $J(n, \varepsilon) \to \infty$ as $n \to \infty$, we see that for all sufficiently large n and $k \ge J(n, \varepsilon) - K - 1$,

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

Therefore,

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

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

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

Part 3. We know that

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

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 $\sigma > 0$ is arbitrary, the proof is complete. □

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

$$\lim_{n\to\infty}(c_{0,n}\langle v_0, u_0\rangle\Gamma_n) = 1, \tag{4.19}$$

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

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

By substituting 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 obtain

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

Thus, we have

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

By assumption (A6), $\tilde{\Lambda}_n\Gamma_n \to \infty$ as $n \to \infty$, proving (4.19). □

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

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

We observe that

$$\begin{aligned}
\mathbb{E}(\langle {}_jZ_n, u_n\rangle) &= \sum_{i=1}^{d}\mathbb{E}(Z_n(i)u_n(i) \mid Z_0 = e_j)\\
&= \sum_{i=1}^{d} u_n(i)\mathbb{E}(Z_n(i) \mid Z_0 = e_j)\\
&= \sum_{i=1}^{d} u_n(i)M_n(j, i)\\
&= \langle M_nu_n, e_j\rangle\\
&= \langle u_n, M_n^T e_j\rangle.
\end{aligned}$$

We know that $v_0 = M_nv_n/\Lambda_n$, and, thus, we want to prove that

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

By (4.19), it is sufficient to show that

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

By Proposition 2.1(d), the vectors $M_n^\top e_j$ 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 holds since

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

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

5. Continuous-time branching processes

In this section we provide an application of our results to continuous-time branching processes. Let $\rho_t(j)$, $1 \le j \le d$, 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 ${}_jX_t$ 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:

(A0′) $\varepsilon_0 \le \rho_t(j) \le K_0$;

(A1′) $\mathbb{P}({}_jX_t(i) \ge 2) \ge \varepsilon_0$;

(A2′) $\mathbb{P}({}_jX_t = \mathbf{0}) \ge \varepsilon_0$;

(A3′) $\mathbb{E}(\|{}_jX_t\|^2) \le K_0$.

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 $a_1 + \cdots + a_d$ particles: a_1 particles of type 1, a_2 particles of type 2, and so on, during the time interval $[t, t+\Delta]$ with probability $\rho_t(j)P_t(j, a)(\Delta + o(\Delta))$.

Observe that $\{Z_n\}$, $n \in \mathbb{N}$, $n \ge 0$, is a discrete-time branching process that satisfies assumptions (A1)–(A3) (with different ε_0 and K_0). The fact that it satisfies assumptions (A1) and (A2) is clear. The first moment $M(t) = \mathbb{E}Z_t$ satisfies

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

where $B(t)_{ji} = \rho_t(j)(\mathbb{E}({}_jX_t(i)) - \delta_{ij})$. Similarly, if $\mathbb{E}(||{}_jX_t||^p)$ exists and depends continuously on t, then the moments of $\{Z_t\}$ of order p satisfy inhomogeneous linear equations, and if $\mathbb{E}(||{}_jX_t||^p)$ is uniformly bounded in both t and j, then the coefficients of those equations are uniformly bounded. In particular, assumption (A3′) implies that (A3) is satisfied, while a bound on the third moment of $\|{}_jX_t\|$ (see assumption (A4′) below) would imply that (A4) is satisfied.

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

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

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

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

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

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

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

$$\int_0^{\infty} \frac{1}{\|\mathbb{E}Z_t\|}\, \mathrm{d}t = \infty. \tag{5.2}$$

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

Theorem 5.1. *Under assumptions (A0′)–(A3′), if extinction of the process* $\{Z_t\}$ *occurs with probability 1 for some initial population, then (5.2) holds. If (5.2) holds then extinction with probability 1 occurs for every initial population.*

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

(A4′) $\mathbb{E}(\|{}_jX_t\|^3) \leq K_0$ for some $K_0 > 0$;

(A5′) $\mathbb{P}(Z_t \neq \mathbf{0}) \to 0$ as $t \to \infty$;

(A6′) $\mathbb{E}\|Z_t\|/\mathbb{P}(Z_t \neq \mathbf{0}) \to \infty$ as $t \to \infty$.

Note that if (A0′)–(A6′) are satisfied then (A1)–(A6) 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 (A0′)–(A6′), for each initial population of the branching process and each vector* u *with positive components, we have the limit in distribution*

$$\frac{\langle \zeta_t, u\rangle}{\mathbb{E}\langle \zeta_t, u\rangle} \xrightarrow{\mathrm{D}} \xi \quad \text{as } t \to \infty, \tag{5.3}$$

where ξ *is an exponential random variable with parameter 1. Moreover, if (A0′)–(A5′) are satisfied and, for some initial population, the limit in (5.3) is as specified, then (A6′) 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

$$d(u, v) = \ln \frac{\beta(u, v)}{\alpha(u, v)},$$

where $\beta(u, v) = \max_i v(i)/u(i)$ and $\alpha(u, v) = \min_i v(i)/u(i)$. Note that d defines the distance on the space of lines in $\mathcal{K}$ in the sense that

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

Moreover, our next estimate holds.

Lemma A.1. ([19, Lemma 1.3].) *If* $\|u\| = \|v\| = 1$ *then*

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

We will also use the next result due to Birkhoff [5].

Lemma A.2. ([5, Theorem XVI.3.3] or [19, 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 (A1)–(A3) imply that $A_n(\mathcal{K}) \subset \bar{\mathcal{K}}(R)$, where $R = \sqrt{K_0}/\varepsilon_0$ and

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

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, leads to

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

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

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 from Lemma A.2 we see that the diameter of $\mathbb{K}_{k,n}$ with respect to the Hilbert metric is less then $(2 \ln R)(\tanh(\ln R/2))^{n-k-1}$. Hence, Lemma A.1 can be used to show 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

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

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

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

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

Proposition 2.1(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 Proposition 2.1(d), the second is similar. By Lemma A.2,

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

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

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

Taking the norm, we see that $\mathrm{e}^{-\varepsilon_k} \leq a_{n,k} \leq 1$. This proves Proposition 2.1(d) for k' such that $\mathrm{e}^{\varepsilon_{k'}} \leq 1 + \delta$.

Next,

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

Due to Proposition 2.1(a) and (b) proved above, $\langle u_j, v_j \rangle$ are uniformly bounded from above and below, that is, $\varepsilon_0 d \le \langle u_j, v_j \rangle \le 1$, proving the first inequality of Proposition 2.1(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

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

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

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

Combining this with (b) and (c) established above, we obtain the second inequality in Proposition 2.1(e). The proof is complete. □

Appendix B. Skipping generations

Proof of Proposition 2.2. (a) If assumption (A1) 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), we 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.

(b) It suffices to show that $\mathbb{E}((Z_{n+2}(i;k))^2 \mid Z_n = e_j) \le \bar{K}$, 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 $\mathbb{E}(\mathcal{X}^2) \le K_1$, $\mathbb{E}(N^2) \le K_2$. Note that

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

which yields the desired bound.

(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

$$Y_n^2 = Y_n^2 \mathbf{1}_{\{N_n > M\}} + Y_n^2 \mathbf{1}_{\{N_n \le M\}}.$$

The expectation of the first term is equal to

$$\mathbb{E}(Y_n^2 \mathbf{1}_{\{N_n > M\}}) = \mathbb{E}\big(N_n \mathbb{E}(\mathcal{X}_n^2) \mathbf{1}_{\{N_n > M\}} + \tfrac{1}{2} N_n(N_n - 1) \mathbf{1}_{\{N_n > M\}} (\mathbb{E}(\mathcal{X}_n))^2\big).$$

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

$$Y_n^2 \mathbf{1}_{\{N_n \le M\}} \le \bigg(\sum_{m=1}^{M} X_{m,n}\bigg)^2,$$

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

(d) Choose l so that

$$\left(1 + \tfrac{1}{2}\varepsilon_0\right)^l > \mathfrak{b}. \tag{B.1}$$

It suffices to show that, for each j,

$$\mathbb{P}({}_jZ_l = \mathbf{0}) \geq \varepsilon_1, \tag{B.2}$$

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

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

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

Otherwise, we 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

$$\mathbb{P}(Z_{n+1}(m) = 0 \text{ for all } m\colon (n+1, m) \text{ is } p\text{-stable} \mid Z_n = e_j) \geq \tfrac{1}{2}\varepsilon_0.$$

Otherwise, we say that (n, j) is $(p + 1)$-stable. For example, (n, j) is 2-unstable if it is either 1-unstable or, with a probability that 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 $1 - \frac{1}{2}\varepsilon_0$, and, by assumption (A1), 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 $1 + \frac{1}{2}\varepsilon_0$. Hence, for $(0, j)$ to be stable, the expected number of its (stable) descedents after l generations would need to be greater than $(1 + \frac{1}{2}\varepsilon_0)^l \geq \mathfrak{b}$, contradicting (2.5).

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

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

Define η_p as

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

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

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

where the factor $\frac{1}{4}\varepsilon_0$ 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 $\varepsilon_1 = \eta_l$. □

Acknowledgements

The authors were partially supported by the National Science Foundation (NSF) through the Research Experiences for Undergraduates at the University of Maryland (grant number DMS-1359307). In addition, D. Dolgopyat was supported by NSF grant DMS-1362064 and L. Koralov was supported by NSF grant DMS-1309084 and the Army Research Office (grant number 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. and Ney, P. E. (2004). *Branching Processes.* Dover, Mineola, NY.

[3] Bhattacharya, N. and Perlman, M. (2017). Time inhomogeneous branching processes conditioned on non-extinction. Preprint. Available at https://arxiv.org/abs/1703.00337.

[4] Biggins, J. D., Cohn, H. and Nerman, O. (1999). Multi-type branching in varying environment. *Stoch. Process. Appl.* **83,** 357–400.

[5] Birkhoff, G. (1967). *Lattice Theory*, 3rd edn. AMS, Providence, RI.

[6] Borovkov, K., Day, R. and 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. and Wang, Q. (2003). Multitype branching limit behavior. *Ann. Appl. Prob.* **13,** 490–500.

[8] Durrett, R. (2015). *Branching Process Models of Cancer*. Springer, Cham.

[9] Fisher, R. A. (1923). On the dominance ratio. *Proc. R. Soc. Edinburgh* **42,** 321–341.

[10] Fisher, R. A. (1930). *The Genetical Theory of Natural Selection.* Oxford University Press.

[11] Fisher, R. A. (1931). The distribution of gene ratios for rare mutations. *Proc. R. Soc. Edinburgh* **50,** 204–219.

[12] Haccou, P., Jagers, P. and Vatutin, V. A. (2007). *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge University Press.

[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. and Ulam, S. (1946). Theory of multiplicative processes I. Los Alamos Scientific Laboratory.

[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] Kersting, G. (2017). A unifying approach to branching processes in varying environments. Preprint. Available at https://arxiv.org/abs/1703.01960.

[18] Kimmel, M. and Axelrod, D. E. (2015). *Branching Processes in Biology,* 2nd edn. Springer, New York.

[19] Liverani, C. (1995). Decay of correlations. *Ann. Math.* **142,** 239–301.

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

[21] Semenov, N. N. (1934). *Chain Reactions*. Goshimizdat, Leningrad (in Russian).

[22] Seneta, E. (2006). *Non-Negative Matrices and Markov Chains*. Springer, New York.

[23] Watson, H. W. and Galton, F. (1875). On the probability of the extinction of families. *J. R. Anthropol. Inst. G.* **4,** 138–144.