Multi-type branching processes with time-dependent branching rates

D. Dolgopyat, P. Hebbar, L. Koralov, M. Perlman

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.

2010 Mathematics Subject Classification Numbers: 60J80
Keywords: Multi-type branching, extinction probability, exponential limit law, non-negative matrix product.

1 Introduction

Mathematical study of branching processes goes back to the work of Galton and Watson [16] 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 [6, 7, 8, 9], and in the study of certain chemical and nuclear reactions [14, 10]. 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. Examples of such models include:

(a) In stem cell biology ([15]), particles of a different types may represent either stem cells or progenitor cells of different types with different degrees of differentiation.
(b) When modeling cancer ([5]), particles of different types may represent cells that accumulated different numbers of mutations.

(c) In physics, cosmic ray cascades involve electrons producing photons and photons producing electrons. A cascade can be modeled by a 2-type branching process (see [12]).

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.

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 total population grows exponentially with positive probability as time goes to infinity. In the sub-critical case, 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 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 types 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 [11], 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). The complete survival criterion for the single type branching process was obtained in [3]. 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. This serves as a way to identify the analogue of the critical behavior of the process in the case of time-dependent branching rates.

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 results

Let $S = \{1, \ldots, d\}$ 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 $a = (a_1, \ldots, a_d) \in \mathbb{Z}_+^d$, $P_n(i,a)$ represents the probability that a particle of type $i$ that is alive at time $n$ is replaced in the next generation by $a_1$ 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 $Z_n = (Z_n(1), \ldots, Z_n(d))$, 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 $E(jZ_n(i))$ means the same as $E(Z_n(i) | Z_0 = e_j)$, 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 $s = (s_1, \ldots, s_d) \in [0, 1]^d$, let

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

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

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

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

where $1 = (1, \ldots, 1)$. We also define

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

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

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

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

Then,

$$m_n = \mu_0 \ldots \mu_{n-1},$$

where $\mu_n$ and $m_n$ are viewed as matrices. Also define

$$m_{k,n} = \mu_k \ldots \mu_{n-1}.$$
Let \( \| \cdot \| \) denote the following norm of a \( d \)-dimensional vector: \( \| v \| = |v_1| + \ldots + |v_d| \). 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. \( \mathbb{P}(Z_{n+1} = 0|Z_n = e_j) \geq \varepsilon_0 \).
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 1-3, 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 \( \lambda_n \) 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

\[
\mu_{n-1}v_n = \lambda_{n-1}v_{n-1}, \quad \mu_{n-1}^Tu_{n-1} = \tilde{\lambda}_{n-1}u_n.
\]

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

\[
(1 - \delta)v_n \leq \frac{m_{n,n+k}v}{\|m_{n,n+k}v\|} \leq (1 + \delta)v_n, \quad (1 - \delta)u_{n+k} \leq \frac{m_{n,n+k}^Tu}{\|m_{n,n+k}^Tu\|} \leq (1 + \delta)u_{n+k}
\]

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

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

While this proposition doesn’t appear to be new, for the sake of completeness we provide a proof in Appendix A.

**Remark:** The vectors \( v_n \) and the numbers \( \lambda_n \) are uniquely defined by the above conditions. 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).

Our first result gives a necessary and sufficient condition for the almost sure extinction of \( Z_n \).

**Theorem 2.2.** 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.
The following lemma will be derived as a simple corollary of the formulas encountered in the proof of Theorem 2.2.

Lemma 2.3. Under Assumptions 1-3, for each initial population of the branching process, there is a constant $C > 0$ such that
\[
\frac{\Lambda_n}{C} \leq \mathbb{E}\|Z_n\| \leq CA_n, \quad n \geq 1, \quad (1)
\]
\[
\frac{1}{C} \left( \sum_{k=1}^{n} \frac{1}{\Lambda_k} \right)^{-1} \leq \mathbb{P}(Z_n \neq 0) \leq C \left( \sum_{k=1}^{n} \frac{1}{\Lambda_k} \right)^{-1}, \quad n \geq 1. \quad (2)
\]

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. $\mathbb{P}(Z_n \neq 0) \to 0$ as $n \to \infty$ (equivalently, $\sum_{k=1}^{n} (1/\Lambda_k) \to \infty$, by (2)).

6. $\mathbb{E}\|Z_n\|/\mathbb{P}(Z_n \neq 0) \to \infty$ as $n \to \infty$ (equivalently, $\Lambda_n \sum_{k=1}^{n} (1/\Lambda_k) \to \infty$, by (1), (2)).

Let $\zeta_n = (\zeta_n(1), \ldots, \zeta_n(d))$ be the random vector obtained from $Z_n$ by conditioning on the event that $Z_n \neq 0$. In other words, we treat the event $Z_n \neq 0$ as a new probability space, with the measure $\mathbb{P}'$ obtained from the underlying measure $\mathbb{P}$ via $\mathbb{P}'(A) = \mathbb{P}(A)/\mathbb{P}(Z_n \neq 0)$. 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 2.4. 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
\[
\frac{\langle \zeta_n, u \rangle}{\mathbb{E}\langle \zeta_n, u \rangle} \to \xi, \quad \text{as } n \to \infty, \quad (3)
\]
where $\xi$ is an exponential random variable with parameter one. Moreover, if Assumptions 1-5 are satisfied and, for some initial population, the limit in (3) is as specified, then Assumption 6 is also satisfied.

3 Survival vs extinction

Let us denote
\[
\Xi_n = \sum_{k=0}^{n-1} \frac{1}{\Lambda_{k+1}}.
\]
These are the partial sums of the series found in Theorem 2.2, but with the index of summation shifted in order to make the arguments below more transparent.

**Proof of Theorem 2.2.** (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 \( \sigma \)-algebra generated by the branching process \( Z_n = j Z_n \). Let \( z_n = (Z_n, v_n) \). Then,

\[
\mathbb{E}(z_{n+1}|\mathcal{F}_n) = (\mathbb{E}(Z_{n+1}|Z_n), v_{n+1}) = (\mu^T Z_n, v_{n+1}) = (Z_n, \mu_n v_{n+1}) = \lambda_n z_n
\]

Accordingly, \( z_n/\Lambda_n \) is a positive martingale, and hence it converges to some random variable \( z_\infty \). Now let

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

One step analysis gives

\[
D_{n+1} = \mu^T D_n \mu_n + S_n,
\]

where

\[
S_n = \sum_{i=1}^{d} m_n(j, i) \sigma_n^2(i)
\]

and

\[
\sigma_n^2(j_1, j_2)(i) = \text{Cov}(i X_n(j_1), i X_n(j_2)).
\]

By Proposition 2.1, there exists a constant \( B \) such that, \( \|S_n\| \leq B \Lambda_n \), where \( \| \cdot \| \) is a matrix norm. Iterating (4), we get

\[
D_n = \sum_{k=0}^{n-1} m_{k+1,n}^T S_k m_{k+1,n}.
\]

Hence,

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

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 \( f_{0,n}(1) = g_0(1) = 1 \). Determining the asymptotic behavior of \( \langle 1 - f_{0,n}(s), u_0 \rangle \) will be helpful for proving the theorem and also later in the proof of (15). By the Taylor formula with respect to \( s = 1 \),

\[
\langle 1 - f_{0,n}(s), u_0 \rangle = \langle Dg_0(1)(1 - f_{1,n}(s)), u_0 \rangle - \frac{1}{2} \langle (1 - f_{1,n}(s))^T Hg_0(\eta_{1,n})(1 - f_{1,n}(s)), u_0 \rangle
\]
where $Dg_0$ is the gradient of $g_0$ and $f_{0,n}(s) \leq \eta_1(n) \leq 1$ for each component $j \in S$. Here $Hg_0$ stands for the Hessian matrix applied to each component of the vector function $g_0$ separately, then multiplied by vectors $(1 - f_{1,n}(s))$ and $(1 - f_{1,n}(s))$ 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 $\mu_0$,

$$
\langle 1 - f_{0,n}(s), u_0 \rangle = \langle (1 - f_{1,n}(s), \mu_0^T u_0) - \frac{1}{2} \langle (1 - f_{1,n}(s))^T Hg_0(\eta_1, n)(1 - f_{1,n}(s)), u_0 \rangle,
$$

Thus, for $s \neq 1$,

$$
(\langle 1 - f_{0,n}(s), u_0 \rangle)^{-1} = \left( \langle (1 - f_{1,n}(s), \tilde{\lambda}_0 u_1) - \frac{1}{2} \langle (1 - f_{1,n}(s))^T Hg_0(\eta_1, n)(1 - f_{1,n}(s)), u_0 \rangle \right)^{-1}
$$

$$
= \tilde{\lambda}_0 \langle (1 - f_{1,n}(s), u_1) \rangle^{-1} \left( 1 + \frac{\langle \frac{1}{2}(1 - f_{1,n}(s))^T Hg_0(\eta_1, n)(1 - f_{1,n}(s)), u_0 \rangle}{\tilde{\lambda}_0 \langle (1 - f_{1,n}(s), u_1) \rangle} \right)^{-1}
$$

$$
= \frac{1}{\tilde{\lambda}_0 \langle (1 - f_{1,n}(s), u_1) \rangle} + \frac{\langle \frac{1}{2}(1 - f_{1,n}(s))^T Hg_0(\eta_1, n)(1 - f_{1,n}(s)), u_0 \rangle}{\tilde{\lambda}_0 \langle (1 - f_{1,n}(s), u_1) \rangle \langle 1 - f_{0,n}(s), u_0 \rangle},
$$

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

$$
(1 - f_{0,n}(s), u_0)^{-1} = \frac{1}{\Lambda_n \langle 1 - s, u_n \rangle} + \frac{1}{2} \sum_{k=0}^{n-1} \frac{\langle (1 - f_{k+1,n}(s))^T Hg_k(\eta_{k+1,n})(1 - f_{k+1,n}(s)), u_k \rangle}{\Lambda_{k+1} \langle (1 - f_{k+1,n}(s), u_{k+1}) \rangle \langle 1 - f_{k,n}(s), u_k \rangle},
$$

(5)

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

Let

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

(6)

where 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 (5) takes the form

$$
\langle 1 - f_{0,n}(s), u_0 \rangle = \left( \frac{1}{\Lambda_n \langle 1 - s, u_n \rangle} + \alpha(n, s) \right)^{-1}.
$$

(7)

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

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

7
Proof. The statement will follow if we prove the following bounds on the terms in the sums for $\alpha(n, s)$ and $\Xi_n$: for each $0 \leq k \leq n - 1$ and $s \in [0, 1]^d \setminus \{1\}$, we have

$$\frac{1}{C\Lambda_{k+1}} \leq \frac{\langle (1 - f_{k+1,n}(s))^T H g_k(\eta_{k+1,n})(1 - f_{k+1,n}(s)), u_k \rangle}{\Lambda_{k+1} \langle (1 - f_{k+1,n}(s), u_{k+1} \rangle} \leq C \frac{1}{\Lambda_{k+1}}. \quad (8)$$

By Proposition 2.1(e), in order to prove (8), it is enough to show that there exists an $L > 0$ such that

$$\frac{1}{L} \leq \frac{\langle (1 - f_{k+1,n}(s))^T H g_k(\eta_{k+1,n})(1 - f_{k+1,n}(s)), u_k \rangle}{\langle (1 - f_{k+1,n}(s), u_{k+1} \rangle} \leq L. \quad (9)$$

Now, we know that $f^{j}_{k,n}(0) \leq f^{j}_{k,n}(s) \leq \eta_{k+1,n}(j) \leq 1$ for each $k$ and $j \in S$. Also, $f^{j}_{k,n}(0) = \mathbb{P}(Z_n = 0|Z_k = e_j) \geq \varepsilon_0$ 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 $\zeta$ 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

$$\frac{1}{c_1} \|1 - f_{k+1,n}(s)\|^2 \leq \langle (1 - f_{k+1,n}(s))^T H g_{k}(\eta_{k+1,n})(1 - f_{k+1,n}(s)), u_k \rangle \leq c_1 \|1 - f_{k+1,n}(s)\|^2. \quad (10)$$

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

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

In order to prove (9), 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 $s \in [0, 1]^d \setminus \{1\}$,

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

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

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

since $g_k$ is uniformly Lipschitz due to Assumption 3.

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

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

To prove the second inequality $\|1 - f_{k+1,n}(s)\| \leq c_2 \|1 - f_{k,n}(s)\|$, we consider the following two cases:
(CASE I) \( \|1 - f_{k+1,n}(s)\| \leq r \bar{\varepsilon}d/c_1 \). Then, from equation (10) and Proposition 2.1,
\[
\langle (1 - f_{k+1,n}(s))^T H g_k(\eta_{k+1,n})(1 - f_{k+1,n}(s)), u_k \rangle \leq c_1 \|1 - f_{k+1,n}(s)\|^2.
\]
and thus, substituting the above relation into the Taylor formula,
\[
\langle 1 - f_{k,n}(s), u_k \rangle = \langle \mu_k(1 - f_{k+1,n}(s)), u_k \rangle - \frac{1}{2} \langle (1 - f_{k+1,n}(s))^T H g_k(\eta_{k+1,n})(1 - f_{k+1,n}(s)), u_k \rangle,
\]
we get,
\[
\langle 1 - f_{k,n}(s), u_k \rangle \geq \langle \mu_k(1 - f_{k+1,n}(s)), u_k \rangle / 2,
\]
and thus,
\[
\|1 - f_{k,n}(s)\| \geq \langle 1 - f_{k,n}(s), u_k \rangle \geq \langle \mu_k(1 - f_{k+1,n}(s)), u_k \rangle / 2 \geq r \bar{\varepsilon} \|1 - f_{k+1,n}(s)\| / 2.
\]
So, for \( \tilde{c}_2 = 2/(r \bar{\varepsilon}) \), we have
\[
\|1 - f_{k+1,n}(s)\| \leq \tilde{c}_2 \|1 - f_{k,n}(s)\|.
\]

(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 \( j \in S \),
\[
g_k^j(s) = \mathbb{E}\left( \prod_{i=1}^d s_i^{Z_{k+1}^{(i)}} | Z_k = c_j \right) \leq (1 - \varepsilon_0) + \varepsilon_0 s_j^2,
\]
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 \( s \in [0, 1]^d \setminus \{1\} \),
\[
\|1 - f_{k+1,n}(s)\| \leq c_2 \|1 - f_{k,n}(s)\|. \quad \square
\]

Now we can return to the proof of Theorem 2.2. Using Lemma 3.1 in equation (7), we get
\[
\langle 1 - f_{0,n}(0), u_0 \rangle \leq \left( \frac{1}{\Lambda_n} + \frac{\Xi_n}{c} \right)^{-1}.
\]
Therefore,
\[
\langle 1 - f_{0,n}(0), u_0 \rangle \leq \frac{c}{\Xi_n}.
\]
Recall that $1 - f_{j,0}^j(0) = \mathbb{P}(Z_n \neq 0 | Z_0 = e_j)$ for each $j \in S$, and hence, if $\lim_{n \to \infty} \Xi_n = \infty$, then

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

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.

We have the following simple corollary of Theorem 2.2.

**Lemma 3.2.** Consider the process $jZ'_n$ that starts with one particle of type $j$ alive at time $k' \geq 0$ followed by branching with the matrices $P_{k'}, P_{k'+1}, \ldots$. Under Assumptions 1-3, extinction for this process occurs with probability one if and only if $\sum_{k=1}^{\infty} (1 / \Lambda_k) = \infty$.

**Proof.** It suffices to note that $\sum_{k=1}^{\infty} (1 / \Lambda_k) = \infty$ if and only if $\sum_{k=k'+1}^{\infty} (\Lambda_k / \Lambda_k) = \infty$, while the latter is equivalent to the almost sure extinction of the process $jZ'_n$ by Theorem 2.2. $\square$

**Proof of Lemma 2.3.** 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 $m_{1,n}^Tu = \mathbb{E}Z_n$,

$$c^{-1}m_{1,n}^Tu_1 \leq \mathbb{E}Z_n \leq cm_{1,n}^Tu_1.$$ 

Taking the norm and using the fact that $m_{1,n}^Tu_1 = (\Lambda_n / \Lambda_1)u_n$, we obtain (1).

From (7) with $s = 0$, using the fact that $1 - f_{0,n}^j(0) = \mathbb{P}(Z_n \neq 0 | Z_0 = e_j)$, we obtain

$$\frac{1}{C} \left( \frac{1}{\Lambda_n} + \alpha(n, s) \right)^{-1} \leq \mathbb{P}(jZ_n \neq 0) \leq C \left( \frac{1}{\Lambda_n} + \alpha(n, s) \right)^{-1}.$$ 

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

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

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). $\square$

### 4 Convergence of the process conditioned on survival

The following series will be important to our analysis,

$$\Gamma_n = \frac{1}{2} \sum_{k=0}^{n-1} \frac{1}{\lambda_k \Lambda_{k+1}} \frac{\langle v_{k+1}^T H g_k(1) v_{k+1}, u_k \rangle}{\langle v_{k+1}, u_{k+1} \rangle \langle v_k, u_k \rangle}. \quad (11)$$
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 $v_k$ 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 (11) are positive, the sequence $\Gamma_n$ is increasing. In each term in (11), each of the factors, $\lambda_k, \langle v_{k+1}^THg_k(1)v_{k+1}, u_k \rangle, \langle v_{k+1}, u_{k+1} \rangle,$ 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 estimates from Proposition 2.1, there is a positive constant $C$ such that

$$\frac{1}{C\Lambda_{k+1}} \leq \frac{1}{2\lambda_k\Lambda_{k+1}} \frac{\langle v_{k+1}^THg_k(1)v_{k+1}, u_k \rangle}{\langle v_{k+1}, u_{k+1} \rangle \langle v_k, u_k \rangle} \leq \frac{C}{\Lambda_{k+1}},$$

and consequently,

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

Assumptions 5 and 6 now can be rewritten as

$$\Gamma_n \to \infty, \quad \lambda_n\Gamma_n \to \infty \quad \text{as} \quad n \to \infty.$$

We’ll start the proof of Theorem 2.4 by formulating a seemingly weaker statement.

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

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

where $\xi$ is an exponential random variable with parameter one.

**Proof.** It is sufficient to show convergence of moment generating functions. That is, we want to prove that for each $\bar{s} \in \mathbb{R}$ we have

$$\mathbb{E}\left(\exp\left(-\bar{s}\langle jZ_n, u_n \rangle \mathbb{P}(jZ_n \neq 0)\right) | jZ_n \neq 0\right) \to \frac{1}{1 + \bar{s}} \quad \text{as} \quad n \to \infty.$$

Let us define vectors $s_j$ such that the $i$-th component of $s_j$ is

$$s_j(i) = \exp\left(-\bar{s}u_n(i) \mathbb{P}(jZ_n \neq 0)\right).$$

Then the $j$-th component of the vector $f_n(s_j)$ is equal to

$$f_n^j(s_j) = \mathbb{E}\left(\exp\left(-\bar{s}\langle jZ_n, u_n \rangle \mathbb{P}(jZ_n \neq 0)\right)\right).$$

Thus we want to show that

$$1 - \frac{1 - f_n^j(s_j)}{\mathbb{P}(jZ_n \neq 0)} \to \frac{1}{1 + \bar{s}} \quad \text{as} \quad n \to \infty.$$
In order to prove (15), it will be useful to study the asymptotic behavior of the sum on the right hand side of (5). We first find the upper and lower bounds of the sum using the upper and lower bounds for $\eta_{k,n}$. Observe that $H g_k^j(s)$ is monotonic in $s$ for each $j$ since $g_k^j$ is a polynomial with non-negative coefficients and $H g_k^j(s)$ is a matrix with entries that are mixed second derivatives of $g_k^j$. Therefore, (5) gives

\[
\left(\frac{1}{\Lambda_n(1 - s, u_n)} + \frac{1}{2} \sum_{k=0}^{n-1} \langle (1 - f_{k+1,n}(s))^T H g_k(1 - f_{k+1,n}(s)), u_k \rangle \right)^{-1} \leq (1 - f_{0,n}(s), u_0) \\
\leq \left(\frac{1}{\Lambda_n(1 - s, u_n)} + \frac{1}{2} \sum_{k=0}^{n-1} \langle (1 - f_{k+1,n}(s))^T H g_k(f_{k,n}(s))(1 - f_{k+1,n}(s)), u_{k+1} \rangle \right)^{-1}.
\]

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 1 for $k \leq K + 1$. By formally linearizing the mappings $g_k, g_{k+1}, \ldots, g_K$, we write

\[
1 - f_{k,n}(s) \approx \mu_k \mu_{k+1} \cdots \mu_K (1 - f_{K+1,n}(s)).
\]

We know that

\[
v_k = \frac{\mu_{k+1} v_{k+1} - 1}{\lambda_k},
\]

and thus

\[
v_k = \frac{\mu_k \mu_{k+1} \cdots \mu_K}{\prod_{i=k}^{K} \lambda_i}. \tag{18}
\]

Note the similarity in the expressions (17) and (18): the same product of matrices is applied, albeit to different vectors. The 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 $1 - f_{k,n}(s)$ (and $1 - f_{k+1,n}(s)$) 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 (16) 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 (16).

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,

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

**Lemma 4.2.** For each $\varepsilon' > 0$, there exist a natural number $K$ and an $\varepsilon > 0$ such that, for each $s \in [0, 1]^d \setminus \{1\}$,

\[
1 - f_{k,n}(s) = c_{k,n}(v_k + \delta_{k,n}), \tag{19}
\]

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 $n$.  

Proof. Suppose that we have (19) with \( \|\delta_{k,n}\| \leq \varepsilon'' \), but without any assumptions on \( c_{k,n} \).

Then we have, for \( 0 \leq k < J(n, \varepsilon) - K \),

\[
1 - f_{k,n}(s) = 1 - g_k(f_{k+1,n}(s)) = \mu_k(1 - f_{k+1,n}(s)) + \alpha_{k,n}\|1 - f_{k+1,n}(s)\|,
\]

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

\[
\mu_k(j, i) = \frac{\partial g^j_k}{\partial s_i}(1).
\]

The uniformity in \( k \) follows from Assumption 3. Thus

\[
1 - f_{k,n}(s) = \mu_k(c_{k+1,n}(v_{k+1} + \delta_{k+1,n})) + \alpha_{k,n}\|1 - f_{k+1,n}(s)\| = c_{k+1,n}\lambda_k v_k + c_{k+1,n}\mu_k\delta_{k+1,n} + \alpha_{k,n}\|1 - f_{k+1,n}(s)\| = c_{k+1,n}\lambda_k(v_k + \alpha'_{k,n}),
\]

where \( \|\alpha'_{k,n}\| \) can be made arbitrarily small, uniformly in \( k \), by selecting sufficiently small \( \varepsilon \) and \( \varepsilon'' \). Here we used (19) with \( k + 1 \) instead of \( k \) to estimate the contribution from the term \( \alpha_{k,n}\|1 - f_{k+1,n}(s)\| \). 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 \( \varepsilon \) and \( \varepsilon'' \) are sufficiently small. We have demonstrated, therefore, that it is sufficient to establish (19) with the estimate \( \|\delta_{k,n}\| \leq \varepsilon' \) only.

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

\[
\left(1 - \frac{\varepsilon'}{2d}\right)v_k \leq \frac{m_{k,k+k'}v}{\|m_{k,k+k'}v\|} \leq \left(1 + \frac{\varepsilon'}{2d}\right)v_k \quad \text{(20)}
\]

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

\[
m_{k,k+k'}(j, i) = \frac{\partial g^j_{k,k+k'}}{\partial s_i}(1),
\]

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

\[
\|1 - f_{k,k+k'}(1 - v) - m_{k,k+k'}v\| \leq \frac{\varepsilon'}{2d}\|m_{k,k+k'}v\|
\]

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

\[
m_{k,k+k'}v - \frac{\varepsilon'}{2d}\|m_{k,k+k'}v\|1 \leq 1 - f_{k,k+k'}(1 - v) \leq m_{k,k+k'}v + \frac{\varepsilon'}{2d}\|m_{k,k+k'}v\|1.
\]
Combined with (20), this gives
\[ \|m_{k,k'k}v\|(v_k - \frac{\varepsilon'}{d}1) \leq 1 - f_{k,k'+k'}(1 - v) \leq \|m_{k,k'k}v\|(v_k + \frac{\varepsilon'}{d}1). \]

Setting \( K = k' + 1 \), we see that the last inequality can be applied to \( v = 1 - f_{k+k',n}(s) \), provided that \( 0 \leq k \leq J(n,\varepsilon) - K \), resulting in
\[ c_{k,n}(v_k - \frac{\varepsilon'}{d}1) \leq 1 - f_{k,n}(s) \leq c_{k,n}(v_k + \frac{\varepsilon'}{d}1), \]
which gives the desired estimate. \( \square \)

Note that \( J(n,\varepsilon) \to \infty \) as \( n \to \infty \) since each component of the vector \( 1 - f_{k,n}(0) \) is
\[ 1 - f_{i,k,n}(0) = \mathbb{P}(Z_n \neq 0|Z_k = e_i) \]
and \( \mathbb{P}(Z_n \neq 0|Z_k = e_i) \to 0 \) as \( n \to \infty \) for each \( i \) and each \( k \) by Lemma 3.2.

Recall the definition of \( \alpha(n, s) \) from (6).

**Lemma 4.3.** Under Assumptions 1-6,
\[ \lim_{n \to \infty} \frac{\alpha(n, s)}{\Gamma_n} = 1, \]
uniformly in \( s \in [0,1]^d \setminus \{1\} \).

**Proof.** 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
\[ \left| \frac{\alpha(n, s) - \alpha(J(n, \varepsilon) - K - 1, s)}{\Gamma_n} + \frac{\alpha(J(n, \varepsilon) - K - 1, s) - \Gamma_{J(n, \varepsilon) - K - 1}}{\Gamma_n} + \frac{\Gamma_{J(n, \varepsilon) - K - 1} - \Gamma_n}{\Gamma_n} \right| \leq \sigma \]  \hspace{1cm} (21)
for all \( s \) and all sufficiently large \( n \).

1. We first estimate the middle term in the inequality above. By Lemma 4.2, for each \( \sigma' > 0 \), there exist a natural number \( K \) and \( \varepsilon_1 > 0 \) such that
\[ (1 - \sigma')c_{k,n}v_k \leq 1 - f_{k,n}(s) \leq (1 + \sigma')c_{k,n}v_k \]
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 \( \|Hg_k^i(1)\| \geq c > 0 \) for all \( k \geq 0, i \in S \). Thus there exists \( \varepsilon_2 > 0 \) such that the matrix norm satisfies \( \|Hg_k^i(\eta_{k+1,n}) - Hg_k^j(1)\| < \)
\[ \sigma' \cdot \| H g_k^t(1) \| \] 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

\[ \left| \frac{1}{2} J(n, \varepsilon) - K - \frac{1}{2} \sum_{k=0}^{J(n, \varepsilon) - K - 1} \frac{\langle (1 - f_{k+1,n}(s))^{T} H g_k(\eta_{k+1,n})(1 - f_{k+1,n}(s)), u_k \rangle}{\Lambda_{k+1}\langle (1 - f_{k+1,n}(s)), u_{k+1}\rangle \langle 1 - f_{k,n}(s), u_k \rangle} - \Gamma_{J(n, \varepsilon) - K - 1} \right| \]

\[ \leq \tilde{c} \sigma' \Gamma_{J(n, \varepsilon) - K - 1} \leq \tilde{c} \sigma' \Gamma_n. \]

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

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

2. Now we estimate the third term in (21). We can assume that \( K \) and \( \varepsilon \) are fixed. We first observe that we can obtain a relation similar to (5) by starting with the expression \( \langle 1 - f_{J(n, \varepsilon),n}(s), u_{J(n, \varepsilon)} \rangle \) instead of \( \langle 1 - f_{0,n}(s), u_0 \rangle \). Thus, by doing the same steps that we carried out to obtain (5), we get

\[ \langle 1 - f_{J(n, \varepsilon),n}(s), u_{J(n, \varepsilon)} \rangle = \left( \frac{\tilde{\Lambda}_{J(n, \varepsilon)}}{\tilde{\Lambda}_n} \right) + \frac{1}{2} \sum_{k=J(n, \varepsilon)}^{n-1} \left( \frac{\tilde{\Lambda}_{J(n, \varepsilon)}}{\tilde{\Lambda}_{k+1}} \right) \left( \frac{1}{\tilde{\Lambda}_n} \right) \langle 1 - s, u_n \rangle \left( \alpha(n, s) - \alpha(J(n, \varepsilon), s) \right) \left( 1 - f_{k,n}(s), u_k \right) \]

\[ \leq \frac{1}{\tilde{\Lambda}_{J(n, \varepsilon)}} \langle \alpha(n, s) - \alpha(J(n, \varepsilon), s) \rangle \leq C \frac{1}{\tilde{\Lambda}_{J(n, \varepsilon)}} \left( \Gamma_n - \Gamma_{J(n, \varepsilon)} \right), \]

where the last inequality follows from (8) and (12). From here it follows that

\[ \langle \Gamma_n - \Gamma_{J(n, \varepsilon)} \rangle \leq C \frac{1}{\tilde{\Lambda}_{J(n, \varepsilon)}} \langle 1 - f_{J(n, \varepsilon),n}(s), u_{J(n, \varepsilon)} \rangle \leq \frac{C}{\tilde{\Lambda}_{J(n, \varepsilon)}}. \]

Therefore,

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

Since \( J(n, \varepsilon) \to \infty \) as \( n \to \infty \), by Assumption 6 (see (14) and part (e) of Proposition 2.1), we have

\[ \left| \frac{C}{\tilde{\Lambda}_{J(n, \varepsilon)}} \langle \varepsilon \varepsilon \varepsilon \rangle \right| < \sigma/6. \]
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 \( J(n,\varepsilon) \to \infty \) as \( n \to \infty \), we see for all sufficiently large \( n \) and \( k \geq J(n,\varepsilon) - K - 1 \),

\[
\left| \frac{1}{\Gamma_n} \left( \frac{1}{\lambda_k \Lambda_{k+1}} \frac{1}{v_{k+1}^T H g_k(1) v_{k+1}, u_k} \langle v_{k+1}, u_{k+1} \rangle \langle v_k, u_k \rangle \right) \right| < \frac{\sigma}{3(K+1)}.
\]

Therefore,

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

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

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

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

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

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

Let us return to the proof of (15). By Lemma 4.2, when \( n \) is large, the vector \( 1 - f_n(s) = 1 - f_{0,n}(s) \) is nearly aligned to the vector \( v_0 \). Thus, in (15) we can replace the \( j \)-th component of the vector \( 1 - f_n(s_j) \) by

\[
\langle 1 - f_n(s_j), u_0 \rangle \frac{\langle v_0, e_j \rangle}{\langle v_0, u_0 \rangle}.
\]
Therefore, in order to prove (15), it is sufficient to show that

\[
1 - \frac{\langle v_0, e_j \rangle}{\langle v_0, u_0 \rangle \mathbb{P}(j Z_n \neq 0)} \left( \frac{1}{\tilde{\Lambda}_n \tilde{s}} \left( \mathbb{P}(j Z_n \neq 0) + o \left( \frac{\mathbb{P}(j Z_n \neq 0)}{\mathbb{E}(j Z_n, u_n)} \right) \right) \langle u_n, u_n \rangle + \Gamma_n \right)^{-1} \rightarrow \frac{1}{1 + \bar{s}},
\]

where we used Lemma 4.3 to transform (7) and linearized \(1 - s_j\). The LHS can be written as

\[
1 - \frac{\langle v_0, e_j \rangle}{\langle v_0, u_0 \rangle \mathbb{P}(j Z_n \neq 0) \Gamma_n} \left( \frac{1}{\tilde{\Lambda}_n \Gamma_n \tilde{s}} \left( \mathbb{P}(j Z_n \neq 0) \mathbb{E}(j Z_n, u_n) + o \left( \frac{\mathbb{P}(j Z_n \neq 0)}{\mathbb{E}(j Z_n, u_n)} \right) \right) \langle u_n, u_n \rangle + 1 \right)^{-1}.
\]  

(22)

We will need the following two lemmas.

**Lemma 4.4.** Under Assumptions 1-6,

\[
\lim_{n \to \infty} \frac{\langle v_0, u_0 \rangle \mathbb{P}(j Z_n \neq 0) \Gamma_n}{\langle v_0, e_j \rangle} = 1.
\]

*Proof.* We know that \(\mathbb{P}(j Z_n \neq 0) = 1 - f_n(0)\), and therefore for a fixed \(j \in S\), by Lemma 4.2, it is sufficient to prove that

\[
\lim_{n \to \infty} \left( c_{0,n} \langle v_0, u_0 \rangle \Gamma_n \right) = 1,
\]

(23)

where \(c_{0,n}\) is the same as in Lemma 4.2. From Lemma 4.3 and (7) we know that

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

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

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

Thus we have

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

By Assumption 6, \(\tilde{\Lambda}_n \Gamma_n \to \infty\) as \(n \to \infty\) proving (23).

**Lemma 4.5.** Under Assumptions 1-6,

\[
\lim_{n \to \infty} \tilde{\Lambda}_n \Gamma_n \frac{\mathbb{P}(j Z_n \neq 0) \langle u_n, u_n \rangle}{\mathbb{E}(j Z_n, u_n)} = 1.
\]
Proof. As in the proof of Lemma 4.4, it is sufficient to show that
\[
\lim_{n \to \infty} \frac{c_{0,n} \tilde{\Lambda}_n \Gamma_n \langle v_0, e_j \rangle \langle u_n, u_n \rangle}{\mathbb{E}(\langle j Z_n, u_n \rangle)} = 1.
\]
We observe that
\[
\mathbb{E}(\langle j Z_n, u_n \rangle) = \sum_{i=1}^{d} \mathbb{E}(Z_n(i) u_n(i) | Z_0 = e_j)
\]
\[
= \sum_{i=1}^{d} u_n(i) \mathbb{E}(Z_n(i) | Z_0 = e_j)
\]
\[
= \sum_{i=1}^{d} u_n(i)m_n(j, i)
\]
\[
= \langle m_n u_n, e_j \rangle
\]
\[
= \langle u_n, m_n^T e_j \rangle
\]
We know that \( v_0 = m_n v_n / \Lambda_n \), and thus what we want to prove is that
\[
\lim_{n \to \infty} \frac{c_{0,n} \tilde{\Lambda}_n \Gamma_n \langle v_n, m_n^T e_j \rangle \langle u_n, u_n \rangle}{\Lambda_n \langle u_n, m_n^T e_j \rangle} = 1.
\]
By (23), it is sufficient to show that
\[
\lim_{n \to \infty} \frac{\tilde{\Lambda}_n \langle v_n, m_n^T e_j \rangle \langle u_n, u_n \rangle}{\Lambda_n \langle v_0, u_0 \rangle \langle u_n, m_n^T e_j \rangle} = 1.
\]
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
\[
\langle v_n, u_n \rangle = \langle v_n, \frac{\mu_{n-1}^T u_{n-1}}{\lambda_{n-1}} \rangle = \langle \mu_{n-1} v_n, \frac{u_{n-1}}{\lambda_{n-1}} \rangle = \frac{\lambda_{n-1}}{\lambda_{n-1}} \langle v_{n-1}, u_{n-1} \rangle = \frac{\Lambda_n}{\Lambda_n} \langle v_0, u_0 \rangle,
\]
where the last equality is obtained by iterating the previous steps \( n \) times. \( \square \)

Applying the above two lemmas to transform the expression in (22), we obtain
\[
\lim_{n \to \infty} \left( 1 - \frac{\langle v_0, e_j \rangle}{\langle v_0, u_0 \rangle \mathbb{P}(j \neq 0) \Gamma_n} \left( \tilde{\Lambda}_n \Gamma_n \mathbb{E} \left( \frac{\mathbb{P}(j Z_n \neq 0)}{\mathbb{E}(j Z_n, u_n)} \right) \left( \frac{1}{\mathbb{E}(j Z_n, u_n)} + o \left( \frac{\mathbb{E}(j Z_n, u_n)}{\mathbb{E}(j Z_n, u_n)} \right) \right) \langle u_n, u_n \rangle \right)^{-1} \right)
\]
\[
= 1 - \left( \frac{1}{s} + 1 \right)^{-1} = \frac{1}{1 + s}.
\]
This completes the proof of Theorem 4.1.

Proof of Theorem 2.4. First, let Assumptions 1-6 be satisfied. Let $P : v \to v/\|v\|$ be the projection onto the unit sphere, with the convention that $P(0) = 0$. We claim that
\[ \lim_{n \to \infty} \|P(E_j \zeta_n) - u_n\| = 0, \quad \lim_{n \to \infty} P(\|P_j \zeta_n\| > \varepsilon) = 0 \] (24)
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
\[ (1 - \delta)u_{n+k'} \leq \frac{m_{n+k'}^{\ast}}{m_{n+k'}^{\ast} u} \leq (1 + \delta)u_{n+k'} \] (25)
whenever $u$ is a non-zero vector with non-negative components. Let $j\zeta_n$ be the random vector obtained by taking $j\zeta_n$ as the initial population of a branching process, then branching for $k'$ steps using the matrices $P_n, ..., P_{n+k'-1}$, 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 $\mathbb{P}(\|j\zeta_n\| > A) \to 1$ as $n \to \infty$, we obtain
\[ \lim_{n \to \infty} (\mathbb{P}(\|P_j \zeta_n^{k'} - u_{n+k'}\| > \varepsilon) - \mathbb{P}(\|P_j \zeta_n^{k'} - u_{n+k'}\| > \varepsilon)) = 0. \]
Also note that $\lim_{n \to \infty}(\mathbb{P}(E_j \zeta_n^{k'}) - \mathbb{P}(E_j \zeta_{n+k'})) = 0$. Therefore, since $\delta > 0$ was arbitrarily small, (24) will follow if we show that
\[ \|P(E_j \zeta_n^{k'}) - u_{n+k'}\| \leq \delta \] (26)
for all sufficiently large $n$ and
\[ \lim_{n \to \infty} \mathbb{P}(\|P_j \zeta_n^{k'} - u_{n+k'}\| > \varepsilon) = 0. \] (27)
(26) immediately follows from (25). (27) is a consequence of
\[ \lim_{n \to \infty} \mathbb{P}(\|P_j \zeta_n^{k'} - P(E_j \zeta_n^{k'})\| > \varepsilon - \delta) = 0, \]
which can be derived from the Chebyshev inequality since for each $A > 0$ we have $\mathbb{P}(\|j\zeta_n\| > A) \to 1$ as $n \to \infty$. Thus we have (24).

Theorem 4.1 and (24) clearly imply (3) with $j\zeta_n$ replaced by $j\zeta_n$. If the initial population of the process $Z_n$ is not necessarily $\zeta_n$, then we can consider a new process $Z'_n$ for which the transition distribution $P_n'$ is such that $jZ_n'$ coincides in distribution with $Z_n$. We also define $P_n' = P_n$ for $n \geq 1$. Then $\langle j\zeta_n, u \rangle/E\langle \zeta_n, u \rangle$ is equal, in distribution, to $\langle j\zeta_n, u \rangle/E\langle j\zeta_n, u \rangle$ when $n \geq 1$, and therefore (3) holds for every initial population.

Finally, suppose that Assumptions 1-5 are satisfied. If Assumption 6 fails, then $E\|\zeta_n\| = E\|Z_n\|/\mathbb{P}(Z_n \neq 0)$ is bounded along a subsequence for every initial population. Then (3) does not hold since $\zeta_n$ is integer-valued, which gives a contradiction. \qed
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 $jX_t$ 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.

1. $\rho_t(j) \leq K_0.$
2. $P(jX_t(i) \geq 2) \geq \varepsilon_0.$
3. $P(jX_t = 0) \geq \varepsilon_0.$
4. $E(\|jX_t\|^2) \leq K_0.$

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 $a_1 + \ldots + a_d$ particles: $a_1$ particles of type one, $a_2$ particles of type two, etc., during the time interval $[t, t + \Delta]$ with probability $\rho_t(j)P_t(j, a)(\Delta + o(\Delta)).$ Moreover, $Z_n, n \in \mathbb{N}, n \geq 0,$ is a discrete time branching process that satisfies Assumptions 1-3 (with different $\varepsilon_0$ and $K_0$).

The above claims can be justified by first defining a sequence of discrete time processes with small time steps, and then letting the time step go to zero to obtain the limiting continuous time process with the desired properties. Its first moment $M(t) = E Z_t$ satisfies

$$M'(t) = A(t)M(t),$$  (28)

where $A(t)_{ij} = \rho_t(j)(E(jX_t(i)) - \delta_{ij}).$ Similarly, if $E(\|jX_t\|^p)$ exists and depends continuously on $t,$ then the moments of $Z_t$ of order $p$ satisfy inhomogenous linear equations, and if $E(\|jX_t\|^p)$ 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 $\|jX_t\|$ (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,

$$E Z_n = m_n^T Z_0.$$ 

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

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

From (28) it follows that there is a positive constant $c$ such that

$$\frac{1}{c} ||E Z_n|| \leq ||E Z_t|| \leq c ||E Z_n||, \quad n \leq t \leq n + 1.$$
Therefore, the condition $\sum_{k=1}^{\infty} (1/\Lambda_k) = \infty$ used in Theorem 2.2 is equivalent to the following:

$$\int_0^\infty \frac{1}{\|EZ_t\|} dt = \infty. \quad (29)$$

Thus we have the following continuous time analogue of Theorem 2.2.

**Theorem 5.1.** Under Assumptions 0'-3', if extinction of the process $Z_t$ occurs with probability one for some initial population, then (29) holds. If (29) 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'. $E(\|jX_t\|^3) \leq K_0$ for some $K_0 > 0$.

5'. $P(Z_t \neq 0) \to 0$ as $t \to \infty$.

6'. $E\|Z_t\|/P(Z_t \neq 0) \to \infty$ as $t \to \infty$.

Note that if Assumptions 0'-6' are satisfied, then Assumptions 1-6 are satisfied by the discrete time process $Z_n$. Let $\zeta_t = (\zeta_t(1), ..., \zeta_t(d))$ be the random vector obtained from $Z_t$ by conditioning on the event that $Z_t \neq 0$. The following theorem is an easy consequence of Theorem 2.4. 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

$$\frac{\langle \zeta_t, u \rangle}{E\langle \zeta_t, u \rangle} \to \xi, \quad \text{as } t \to \infty, \quad (30)$$

where $\xi$ is an exponential random variable with parameter one. Moreover, if Assumptions 0'-5' are satisfied and, for some initial population, the limit in (30) is as specified, then Assumption 6' is also satisfied.

### 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., [13, 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 [4, Theorem XVI.3.3] or [13, Theorem 1.1]) If \( \mu \) is a linear operator that maps \( \mathcal{K} \) into itself so that \( \mu(\mathcal{K}) \) has finite diameter \( \Delta \) with respect to the Hilbert metric, then for each \( u, v \in \mathcal{K} \)
\[
\frac{d(\mu u, \mu v)}{d(u,v)} \leq \tanh \left( \frac{\Delta}{4} \right) < 1.
\]

Proof of Proposition 2.1. Assumptions 1-3 imply that \( \mu_n(\mathcal{K}) \subset \bar{\mathcal{K}}(R) \), 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 then \( (2 \ln R)(\tanh(\ln R/2))^{n-k-1} \). Hence Lemma A.1 shows that \( \cap_{n>k} \mathbb{K}_{k,n} \) is a single point, which we call \( v_k \). Since \( \mu_{k-1}(\cap_{n>k} \mathbb{K}_{k,n}) = \cap_{n>k-1} \mathbb{K}_{k-1,n} \), it follows that \( \mu_{k-1} v_k = \lambda_{k-1} v_{k-1} \) for some \( \lambda_{k-1} > 0 \).

Next, let \( u_0 \) be an arbitrary vector with \( \|u_0\| = 1, \ u_0(i) > \varepsilon_0 \) for each \( i \).

Let \( u_n = m_n^T u_0/\|m_n^T u_0\|, \ \lambda_n = \|\mu_n^T u_n\| \). 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 \( \mu_n \) is between \( \varepsilon_0 \) and \( \sqrt{K_0} \).

We prove the first inequality of part (d), the second is similar. By Lemma A.2,
\[
d(m_{n,n+k} v_n, v_n) \leq \varepsilon_k := 2(\ln R) \left( \tanh \frac{\ln R}{2} \right)^{k-1}.
\]
Note that $\varepsilon_k$ 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,

$$\langle u_n, v_n \rangle = \langle m^T_{k,n}u_k, v_n \rangle = \frac{1}{\Lambda_{k,n}} \langle u_k, m_{k,n}v_n \rangle = \frac{\Lambda_{k,n}}{\Lambda_{k,n}} \langle u_k, v_k \rangle.$$ 

Due to parts (a) and (b) proved above, $\langle u_j, v_j \rangle$ are uniformly bounded from above and below, i.e., $\varepsilon_0d \leq \langle u_j, v_j \rangle \leq 1$, 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

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

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. 

\textbf{Acknowledgments:} 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. Koralov was supported by NSF grant DMS-1309084.

\textbf{References}


