The Annals of Applied Probability
1992, Vol. 2, No. 3, 575–596

THE POPULATION COMPOSITION OF A MULTITYPE BRANCHING RANDOM WALK

BY M. BRAMSON,[1] P. NEY[2] AND J. TAO[2]

University of Wisconsin—Madison

We consider a multitype branching process in which particles move according to probability laws that depend on the type of a particle. The composition of the population, namely the proportions of numbers of particles of different types, varies over the region of motion. In this paper, we study the growth rate and population composition of the process throughout its domain of supercriticality.

1. Introduction. A branching random walk (BRW) is a model of a population multiplying in accordance with a Galton–Watson (GW) branching process and moving in some space X (typically Euclidean) according to a random walk. The process $\{Z^n(\Gamma) =$ the number of particles in the set Γ at time n; $n = 0, 1, \dots\}$ has been extensively studied (see, e.g., [1]–[3], [6], [7], [9], [16], [23]), as have such quantities as the range of the process (the convex hull of the locations of the particles, see, e.g., [8], [10], [11]). For a brief historical survey, see [20].

In the multitype BRW, there are $d > 1$ types of particles multiplying in accordance with a multitype GW process and moving according to probability laws that depend on the particle type. The object of study is now the distribution of the number of particles of the various types throughout the space X at time n. The total number of particles in a particular set grows (or decays) at an exponential rate depending on the set, and the *population composition*, namely the proportion of particles of the various types, also varies throughout the space. This population composition is the principal objective of our study.

To focus on the principal ideas in a simple setting, we limit ourselves to motion on the one-dimensional lattice $I = \{0, \pm 1, \pm 2, \dots\}$. We also assume an aperiodicity condition to be specified later. More general spaces will be considered in [22]. Thus for any $x \in \mathbb{R}$, we consider

$Z_{ij}^n([x])$ = the number of type j particles at the point "integer part of x" at time n, descended from a type i ancestor at time 0.

We typically abbreviate $Z_{ij}^n([x])$ as $Z_{ij}^n[x]$.

Received May 1990; revised April 1991.
[1]Supported in part by NSF Grant DMS-89-01545.
[2]Supported in part by NSF Grant DMS-89-01464.
AMS 1980 *subject classifications*. Primary 60J80; secondary 60F10.
Key words and phrases. Branching process, random walk, large deviations.

Our main result says that subject to moment and regularity assumptions,

$$b_i^n \sqrt{n}\, e^{\Lambda^*(a)n} Z_{ij}^n[an] \to u_j(\theta_a) W_i(\theta_a) \quad \text{a.s. as } n \to \infty, \tag{1.1}$$

where $\Lambda^*(a)$ and $(u_1(\theta_a), \dots, u_d(\theta_a))$ are a rate function and eigenvector coming out of a large deviation analysis and $W_i(\theta_a)$ is a random variable which is a martingale limit. The term b_i^n is bounded above and away from zero and incorporates the slight dependence in n due to taking integer parts. Of course an immediate corollary is that on the set $\{W_i(\theta_a) > 0\}$,

$$\frac{Z_{ij}^n[an]}{Z_{ik}^n[an]} \to \frac{u_j(\theta_a)}{u_k(\theta_a)} \quad \text{a.s.,} \tag{1.2}$$

which describes the above-mentioned population composition.

In the case of a single particle type, the above result reduces to

$$b_i^n \sqrt{n}\, e^{\Lambda^*(a)n} Z^n[an] \to W(\theta_a). \tag{1.3}$$

This was proved by Biggins [9] under a logarithmic moment condition which is a natural extension of the familiar hypothesis $EZ^1 \log Z^1 < \infty$ in branching theory. A similar result for nonlattice $\mathbb{R}^d$-valued random walk was proved by Uchiyama [23]. A related result for the first- and last-birth problem for a multitype age-dependent branching process was also studied by Biggins [5].

Here is an outline of the paper: Section 2 introduces notation, definitions and hypotheses and gives a summary of results. Section 3 gives a little background on positive kernels which will be needed later. Section 4 introduces some martingales that play a central role in our proofs and also proves some technical lemmas. Section 5 proves the main result.

2. Notation, definitions, hypotheses, results. The process under study is defined in terms of a family of point measures on I equal to the integers, namely

$$\{Z_{ij}(x); i, j = 1, \dots, d; x \in I\}.$$

The interpretation is that

$Z_{ij}(x)$ = the number of first generation offspring of type j, located at $x \in I$, which are produced by a "parent" of type i located at the origin 0.

An nth generation type i particle located at the point $\gamma \in I$ produces $(n+1)$st generation type j offspring according to a random measure equivalent to $Z_{ij}(\cdot)$ translated by γ, independently of the history of the process up to generation (time) n and of other particles existing at time n. Thus the process is spatially homogeneous and the history of motion of a single particle consists of independent but not identically distributed increments.

We introduce the following notation:

$Z^n_{ij}(x)$ = the number of type j particles located at $x \in I$ at time n, given an initial type i particle located at 0 at time 0,

$$Z^n_{ij}(\Gamma) = \sum_{x \in \Gamma \cap I} Z^n_{ij}(x) \quad \text{for any set } \Gamma \subset \mathbb{R},$$

$$Z_{ij}(\Gamma) = Z^1_{ij}(\Gamma),$$

$$Z_{ij} = Z_{ij}(\mathbb{R}), \qquad Z^n_{ij} = Z^n_{ij}(\mathbb{R}).$$

Fix $a \in \mathbb{R}$ throughout. Since we will frequently take $x = [an]$, we abbreviate

$$X^n_{ij} = Z^n_{ij}[an].$$

Define the expectations

$$m^n_{ij}(\Gamma) = EZ^n_{ij}(\Gamma), \qquad \Gamma \subset \mathbb{R}.$$

For each $n = 1, 2, \ldots$, $\{m^n_{ij}(\cdot)\}$ is a matrix of measures. Define $m^n_{ij}(x)$ and $m^n_{ij}[x]$ analogously. Also let

$$m^n_{ij} = EZ^n_{ij}$$

and write

$$m_{ij} = m^1_{ij} \quad \text{and} \quad M = \text{the matrix } \{m_{ij}\}.$$

Note that $m^n_{ij} = (M^n)_{i,j}$ is the (i, j) entry of M^n. For a matrix $\{a_{ij};\ i, j = 1, \ldots, d\}$, we will denote the ith row by

$$a_i = (a_{i1}, \ldots, a_{id}).$$

We will say that the sequence of matrices of measures $\{m^n_{ij}(\cdot);\ i, j = 1, \ldots, d;\ n = 1, 2, \ldots\}$ is *irreducible aperiodic* if

(2.0a) $\qquad M = \{m_{ij}\}$ is an irreducible, aperiodic matrix

and

(2.0b) $\qquad$ the smallest lattice supporting the measures $m^n_{ii}(\cdot)$ for all i and n is the integer lattice.

We will assume that:

(H1). *$\{m^n_{ij}(\cdot)\}$ is an irreducible aperiodic sequence of matrix-valued measures.*

Note that (H1) excludes the cases where $m^n_{ij}(\cdot)$ is periodic or nonlattice. This assumption is made to simplify notation when employing the local central limit theorem in Lemma 3.1. Analogous versions of Lemma 3.1 and the theorem hold in the periodic and nonlattice cases as well, after the usual modifications.

In the proofs we will have to argue on the states of the process at suitable intermediate times between 0 and n. This necessitates a regrettably complicated notation. We set

$$Z_{ej}^{ln}(k;x) = \text{the number of type } j\text{'s located at } x \text{ at time } n, \text{ descended from the } k\text{th type } e \text{ at time } l.$$

When $x = [an]$, we again abbreviate

$$X_{ej}^{ln}(k) = Z_{ej}^{ln}(k;[an]).$$

Also set

$$Z_{ej}^{ln}(k) = \sum_{x \in I} Z_{ej}^{ln}(k;x)$$

and

$$Y_{ij}^{n}(h) = \text{the location of the } h\text{th type } j \text{ at time } n, \text{ descended from the type } i \text{ ancestor at time } 0 \ (h = 1, 2, \dots),$$

$$Y_{ej}^{ln}(k,h) = \text{the location of the } h\text{th type } j \text{ at time } n, \text{ descended from the } k\text{th type } e \text{ at time } l.$$

In all the above notation, superscripts denote time, subscripts denote type, quantities on the main line denote location (e.g., x) and counting variables (e.g., k, h).

Let

$$\mathscr{F}_n = \text{the } \sigma\text{-field generated by the process up to time } n$$

so that all the above quantities are measurable.

We write $\mathscr{F} = (\mathscr{F}_n,\ n = 0, 1, 2, \dots)$. Also, we introduce the following transforms:

$$\hat{Z}_{ij}^{n}(\theta) = \sum_{x \in I} e^{\theta x} Z_{ij}^{n}(x), \qquad \hat{Z}_{ij}(\theta) = \hat{Z}_{ij}^{1}(\theta),$$

$$\hat{m}_{ij}^{n}(\theta) = \sum_{x \in I} e^{\theta x} m_{ij}^{n}(x), \qquad \hat{m}_{ij}(\theta) = \hat{m}_{ij}^{1}(\theta),$$

$$\hat{M}(\theta) = \text{the matrix } \{\hat{m}_{ij}(\theta)\}.$$

Let $\mathscr{D} = \{\theta \in \mathbb{R}: \hat{m}_{ij}(\theta) < \infty \text{ for } i, j = 1, \dots, d\}$.

(H1) implies that $\hat{M}(\theta)$ is also irreducible and aperiodic for $\theta \in \mathscr{D}$. It then has a maximal (real) eigenvalue, which we denote by $\lambda(\theta)$. Let $\Lambda(\theta) = \log \lambda(\theta)$, $\rho = \lambda(0)$ and $\mu = \lambda'(0)/\rho = \Lambda'(0)$. Note that ρ is the maximal eigenvalue of M. We will also see that μ can be interpreted as the "mean drift" of the process. Associated with the eigenvalue $\lambda(\theta)$, the matrix $\hat{M}(\theta)$ will have left and right eigenvectors

$$u(\theta) = (u_1(\theta), \dots, u_d(\theta)) \quad \text{and} \quad v(\theta) = (v_1(\theta), \dots, v_d(\theta)),$$

normalized so that $\sum_i u_i(\theta) v_i(\theta) = 1$.

We assume the following further hypotheses. Recall that a is (so far) an arbitrary fixed point in $\mathbb{R}$.

(H2). $\rho > 1$.

(H3). *$\mathscr{D}$ contains a neighborhood of* 0.

(H4). $\Lambda'(\theta) = a$ *has a solution* $\theta_a \in \operatorname{In} \mathscr{D}$. *(This solution will necessarily be unique.)*

(H5). $E(\hat{Z}_{ij}(\theta_a))^{p_0} < \infty$ *for some* $p_0 > 1$ *for* $i, j = 1, \dots, d$.

Sometimes we will write

$$\lambda = \lambda(\theta_a), \qquad u_i = u_i(\theta_a), \qquad v_i = v_i(\theta_a).$$

We set

$$\Lambda^*(x) = \sup_\theta (x\theta - \Lambda(\theta)).$$

It is easy to check that $\Lambda^*(\cdot)$ is convex and that $\Lambda^*(a) = a\theta_a - \Lambda(\theta_a)$. From hypothesis (H3) one can show that the level set $\{x\colon \Lambda^*(x) \le 0\}$ is a closed bounded interval, say $[\alpha_0, a_0]$. Thus $\Lambda^*(x) < 0$ for $x \in (\alpha_0, a_0)$ and $\Lambda^*(x) > 0$ for $x \in [\alpha_0, a_0]^c$, and since $\rho > 1$, $\mu \in (\alpha_0, a_0)$. It is well known (see, e.g., [17], page 1269) that

$$\lim_{n\to\infty} \frac{1}{n} \log m_{ij}^n[an] = -\Lambda^*(a) \tag{2.1}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \log m_{ij}^n([an, \infty)) = -\Lambda^*(a) \quad \text{for } a > \mu. \tag{2.2}$$

On account of (2.1), $m_{ij}^n[an]$ grows exponentially with rate $-\Lambda^*(a) > 0$ for $a \in (\alpha_0, a_0)$. Hence we call the interval (α_0, a_0) the domain of supercriticality of the process.

We will assume that:

(H6). *a is such that $\Lambda^*(a) < 0$, that is, a is in the range of supercriticality of the process. For definiteness we also take $a > \mu$. (This implies $\theta_a > 0$.)*

We define

$$b_i^n = \sqrt{2\pi}\,\sigma_i(\theta_a)\exp\{\theta_a([an] - an)\}/v_i(\theta_a), \tag{2.3}$$

where $\sigma_i(\theta_a) > 0$ will be specified later [in (3.23)]. (The $\exp\{\cdot\}$ term arises from evaluation at $[an]$ rather than an for $Z_{ij}^n[an]$ below.)

Our main result is:

THEOREM. *Assume* (H1)–(H6). *Then for* $i, j = 1, \dots, d$,

$$\lim_{n \to \infty} b_i^n \sqrt{n}\, e^{\Lambda^*(a)n} Z_{ij}^n[an] = u_j(\theta_a) W_i(\theta_a) \quad a.s. \tag{2.4}$$

Here $W_i(\theta_a) = \lim_{n \to \infty} W_i^n(\theta_a)$ *is the limit of a martingale sequence to be defined later, with* $W_i(\theta_a) > 0$ *on the set of nonextinction of the branching process* Z^n.

An immediate consequence is:

COROLLARY. *On the set* $\{Z^n \nrightarrow 0\}$,

$$\frac{Z_{ij}^n[an]}{Z_{ik}^n[an]} \to \frac{u_j(\theta_a)}{u_k(\theta_a)} \quad a.s., \tag{2.5}$$

for $i, j, k = 1, \dots, d$.

3. The expectation matrix. We have defined $m_{ij}(x) = EZ_{ij}(x)$. Also recall that

$m_{ij}(\Gamma) =$ the expected number of type j's in a set $\Gamma \subset \mathbb{R}$, produced by a type i parent.

For each $i, j = 1, \dots, d$, $m_{ij}(\cdot)$ is a measure on $\mathbb{R}$ and $M(\Gamma) = \{m_{ij}(\Gamma)\}$ is a matrix of measures. Such objects are central to the study of Markov-additive (MA) processes and much has been written about them (e.g., [13], [17], [18], [21]). In this section we summarize some facts about the asymptotic behavior of $m_{ij}^n(\cdot)$.

An MA-process is a Markov chain $\{(X_n, S_n); n = 0, 1, \dots\}$, where X_n is an ordinary Markov chain (with state space $\{1, \dots, d\}$ in our case) and S_n is an "additive" component whose increments $(S_{n+1} - S_n)$ have a distribution depending on X_n and X_{n+1}. We assume $S_0 = 0$. The distribution of this process is determined by a so-called MA-kernel, which is just a matrix of measures

$$\{p_{ij}(\cdot); i, j = 1, \dots, d\}, \tag{3.1}$$

with $\{p_{ij}(R)\}$ a stochastic matrix. We will assume that $\{p_{ij}(\cdot)\}$ has support on the lattice $I - a$, $a \in \mathbb{R}$. This is done to compensate for translation by a, while centering our MA-kernel. Define

$$p_{ij}^2(x) = \sum_{k=1}^{d} \sum_{y \in I - a} p_{ik}(y) p_{kj}(x - y) \tag{3.2}$$

for $x \in I - 2a$ and then define $p_{ij}^n(\cdot)$ by induction. This has the interpretation

$$p_{ij}^n(x) = P\{X_n = j, S_n = x | X_0 = i\}. \tag{3.3}$$

When the support of $\{p_{ij}(\cdot)\} \subset I - a$, the support of $\{p_{ij}^n(\cdot)\} \subset I_n = I - an$. In analogy to (2.0) we will say that $\{p_{ij}^n(\cdot)\}$ is *irreducible aperiodic* if

(3.4a) $\{p_{ij}(\mathbb{R}); i, j = 1, \dots, d\}$ is an irreducible aperiodic matrix

and

(3.4b) for $p_{ij}(\cdot)$ supported on $I - a$ ($a = 0$ is allowed), I is the smallest lattice J such that $J - an$ supports $p_{ii}^n(\cdot)$ for all i and n.

The following result is a useful local limit theorem about $p_{ij}^n(x)$.

LEMMA 3.1 (Nagaev [18]). *Suppose that $\{p_{ij}(\cdot)\}$ is supported on $I - a$ and is irreducible aperiodic. Assume that*

$$\sum_{x \in \mathbf{I}_1} |x|^3 p_{ij}(x) < \infty \quad \text{for } i, j = 1, \dots, d, \tag{3.5}$$

and

$$\{p_{ij}(\mathbb{R})\} \text{ has invariant probability measure } \pi = (\pi_1, \dots, \pi_d),$$

which is centered so that

$$\sum_{i=1}^{d} \sum_{j=1}^{d} \sum_{x \in I_1} \pi_i x p_{ij}(x) = 0. \tag{3.6}$$

Then

$$\lim_{n \to \infty} \sup_{x \in I_n} \left| \sigma\sqrt{n}\, p_{ij}^n(x) - \pi_j \varphi\left(\frac{x}{\sigma\sqrt{n}} \right) \right| = 0, \tag{3.7}$$

where $\varphi(\cdot)$ is the standard normal density. Here $\sigma > 0$ is an appropriate second moment determined by $\{p_{ij}(\cdot)\}$.

Nagaev's proof uses a spectral argument on the matrix of Fourier transforms $p_{ij}(t) = \int p_{ij}(dx) e^{itx}$. It is similar in spirit to that of the standard local limit theorem for sums of i.i.d. random variables. When $\{p_{ij}(\cdot)\}$ is periodic or is nonlattice, there are corresponding limits in the usual manner of local limit theorems.

We will need a large deviation version of this result for $\{p_{ij}(\cdot)\}$ supported on I. Define the transform matrix

$$\{\hat{p}_{ij}(\theta)\} = \left\{ \sum_{x \in I} e^{\theta x} p_{ij}(x) \right\},$$

and let $\rho(\theta)$, $l(\theta)$ and $r(\theta)$ denote, respectively, its maximal eigenvalue and associated left and right eigenvectors. Let

$$p_{ij}(x; \theta) = \frac{e^{\theta x} p_{ij}(x) r_j(\theta)}{\rho(\theta) r_i(\theta)}, \quad i, j = 1, \dots, d, \qquad \theta \in \mathbb{R}. \tag{3.8}$$

Then this is another MA-kernel and $p_{ij}(\mathbb{R};\theta)$ is a stochastic matrix with invariant measure

$$\{l_i(\theta)r_i(\theta);\ i = 1,\dots,d\}.$$

Let $\pi_i(\theta) = l_i(\theta)r_i(\theta)$ and normalize so that $\sum \pi_i(\theta) = 1$. [Note that $\pi_i(0) = \pi_i$.] Iteration of (3.8) gives

$$p_{ij}^n(x;\theta) = \frac{e^{\theta x}p_{ij}^n(x)r_j(\theta)}{\rho^n(\theta)r_i(\theta)}, \qquad i,j = 1,\dots,d,\ \theta \in \mathbb{R}, \tag{3.9}$$

$n = 1,2,\dots$.

Clearly $\rho(\theta) < \infty$ if and only if $\hat{p}_{ij}(\theta) < \infty$ for all $i,j = 1,\dots,d$. Let $\mathscr{D}(\rho) = \{\theta\colon \rho(\theta) < \infty\}$. Fix a point $a \in \mathbb{R}$ and assume that

$$\rho'(\theta) = a\rho(\theta) \quad \text{has a solution } \theta_a \in \operatorname{In} \mathscr{D}(\rho). \tag{3.10}$$

It is easy to check that

$$\sum_{j=1}^{d}\sum_{i=1}^{d}\sum_{x\in I} \pi_i(\theta_a)xp_{ij}(x;\theta_a) = a. \tag{3.11}$$

Let

$$\begin{aligned}\bar{p}_{ij}(x;\theta_a) &= p_{ij}(a+x;\theta_a)\\ &= \frac{e^{\theta_a(a+x)}p_{ij}(a+x)r_j(\theta_a)}{\rho(\theta_a)r_i(\theta_a)}, \qquad i,j = 1,\dots,d.\end{aligned} \tag{3.12}$$

This is a centered MA-kernel [in the sense of (3.6)] on I. Iteration gives

$$\begin{aligned}\bar{p}_{ij}^n(x;\theta_a) &= \frac{e^{\theta_a(an+x)}p_{ij}^n(an+x)r_j(\theta_a)}{\rho^n(\theta_a)r_i(\theta_a)}\\ &= e^{nI(a)+\theta_a x}\frac{r_j(\theta_a)}{r_i(\theta_a)}p_{ij}^n(an+x),\end{aligned} \tag{3.13}$$

where

$$I(a) = \sup_\theta[\theta a - \log\rho(\theta)] = \theta_a a - \log\rho(\theta_a). \tag{3.14}$$

It can also be shown that $\theta_a = I'(a)$.

Now applying (3.7) to $\bar{p}_{ij}^n(x;\theta_a)$ and noting (3.11), we have the following:

LEMMA 3.2 (*LD* local limit theorem). *Assume that* $\{p_{ij}(\cdot)\}$ *is supported on* I *and that* (3.4) *and* (3.10) *hold. Then for all* $i,j = 1,\dots,d$, *we have*

$$\begin{aligned}\lim_{n\to\infty}\sup_{x\in I_n}\Big|&\sigma_i\sqrt{n}\,e^{nI(a)+xI'(a)}p_{ij}^n(an+x)\\ &-r_i(\theta_a)l_j(\theta_a)\varphi\Big(\frac{x}{\sigma_i\sqrt{n}}\Big)\Big| = 0.\end{aligned} \tag{3.15}$$

Here $\sigma_i = \sigma_i(\theta_a) > 0$ *is again a second moment parameter.*

The above results on the stochastic MA-kernels $\{p_{ij}^n(\cdot)\}$ can easily be converted into corresponding statements about general nonnegative kernels. We do this for $\{m_{ij}^n(\cdot)\}$ in Lemma 3.3, which is the analog of Lemma 3.2.

Recall that

$$\hat{m}_{ij}(\theta) = \sum_{x \in I} e^{\theta x} m_{ij}(x), \qquad i, j = 1, \dots, d,$$

with maximal eigenvalue $\lambda(\theta)$ and eigenvectors $u(\theta), v(\theta)$, where $\{m_{ij}(\cdot)\}$ is supported on I. Let

$$m_{ij}(x;\theta) = \frac{e^{\theta x} m_{ij}(x) v_j(\theta)}{\lambda(\theta) v_i(\theta)}, \qquad i, j = 1, \dots, d. \tag{3.16}$$

From the definition of the eigenvalue $\lambda(\theta)$,

$$\sum_j \sum_x m_{ij}(x;\theta) = \sum_j \frac{\hat{m}_{ij}(\theta) v_j(\theta)}{\lambda(\theta) v_i(\theta)} = 1,$$

namely $m_{ij}(x;\theta)$ is a stochastic MA-kernel even if $m_{ij}(x)$ is not. Letting $\theta = \theta_a$ as in (H4), we can define the centered stochastic kernel

$$\overline{m}_{ij}^n(x;\theta_a) = \frac{e^{\theta_a(an+x)} m_{ij}^n(an+x) v_j(\theta_a)}{\lambda^n(\theta_a) v_i(\theta_a)}. \tag{3.17}$$

This is the analog of $\overline{p}_{ij}^n(x;\theta_a)$ in (3.13). Recall that $\pi(\theta) = \{\pi_i(\theta);\ i = 1, \dots, d\} = \{u_i(\theta) v_i(\theta)\}$ is assumed normalized to $\sum \pi_i = 1$ and is then the invariant probability measure of $\{\hat{m}_{ij}(\theta)\}$. We can substitute $\overline{m}_{ij}^n(x;\theta_a)$ into Lemma 3.1 to obtain the following analog of Lemma 3.2.

LEMMA 3.3. *If* $\{m_{ij}(\cdot)\}$ *satisfies* (H1)–(H4), *then*

$$\lim_{n\to\infty} \sup_{x \in I_n} \left| \sigma_i \sqrt{n}\, \overline{m}_{ij}^n(x;\theta_a) - \pi_j(\theta_a) \varphi\left(\frac{x}{\sigma_i \sqrt{n}}\right) \right| = 0. \tag{3.18}$$

Substituting (3.17), *this says that*

$$\lim_{n\to\infty} \sup_{x \in I_n} \left| \sigma_i \sqrt{n}\, e^{\Lambda^*(a)n + \theta_a x} m_{ij}^n(an+x) - u_j(\theta_a) v_i(\theta_a) \varphi\left(\frac{x}{\sigma_i \sqrt{n}}\right) \right| = 0. \tag{3.19}$$

Here, $\sigma_i = \sigma_i(\theta_a) > 0$ is a second moment determined by $m_{ij}(\cdot;\theta_a)$. Taking $x = [an] - an$, it also follows that

$$\lim_{n\to\infty} b_i^n \sqrt{n}\, e^{\Lambda^*(a)n} m_{ij}^n[an] = u_j(\theta_a), \tag{3.20}$$

where b_i^n is defined in (2.3). For our purposes, b_i^n can be thought of as being "almost constant in n."

4. Martingales and preliminary lemmas. We first introduce

$$W_i^n(\theta) = \frac{\sum_j \hat{Z}_{ij}^n(\theta)v_j(\theta)}{\lambda^n(\theta)v_i(\theta)}. \tag{4.1}$$

LEMMA 4.1. *If $\theta \in \mathscr{D}$, then $\{W_i^n(\theta);\ n = 0, 1, \dots\}$ is a martingale with respect to $\mathscr{F}$.*

The proof is analogous to that of the well-known single type case. Just decompose $\hat{Z}_{il}^{n+k}(\theta)$ according to the number, type and positions of particles. Then evaluate $E(\hat{Z}_{il}^{n+k}(\theta)|\mathscr{F}_n)$, and apply the definition of eigenvector.

COROLLARY. *If $\theta \in \mathscr{D}$,*

$$\lim_{n\to\infty} W_i^n(\theta) = W_i(\theta) \quad \textit{exists a.s. for } i = 1, \dots, d. \tag{4.2}$$

$W_i(\theta_a)$ is the limiting quantity appearing in the theorem.

We next demonstrate:

LEMMA 4.2. *Assume* (H1)–(H6). *Then for all $i = 1, \dots, d$,*

$$\sup_n E|W_i^n(\theta_a)|^p < \infty \tag{4.3}$$

for some $p > 1$.

We will use Lemma 4.2 in the proof of Lemma 4.3 and Proposition 5.1. The following corollary will also be used.

COROLLARY. $EW_i(\theta_a) = 1$.

Note also that it is immediate from (4.3) that

$$E\left(\hat{Z}_{ij}^n(\theta_a)\right)^p \le c\beta^n \tag{4.4}$$

for appropriate $c < \infty$, $\beta < \infty$ and all $i, j = 1, \dots, d$, $n = 1, 2, \dots$.

Before starting the proof of Lemma 4.2, we will express $W_i^{n+1}(\theta)$ in a convenient form. Note that

$$W_i^n(\theta) = \frac{\sum_{j=1}^d \sum_{h=1}^{Z_{ij}^n} e^{\theta Y_{ij}^n(h)} v_j(\theta)}{\lambda^n(\theta)v_i(\theta)}.$$

So one has

$$W_i^{n+1}(\theta) = \sum_{e=1}^{d} \sum_{j=1}^{d} \sum_{k=1}^{Z_{ie}^n} \sum_{h=1}^{Z_{ej}^{n,n+1}(k)} e^{\theta Y_{ej}^{n,n+1}(k,h)} \frac{v_j(\theta)}{\lambda^{n+1}(\theta) v_i(\theta)}$$

$$= \sum_e \sum_k \left[\frac{v_e(\theta) e^{\theta Y_{ie}^n(k)}}{\lambda^n(\theta) v_i(\theta)} \right] \left[\sum_j \sum_h \frac{e^{\theta(Y_{ej}^{n,n+1}(k,h) - Y_{ie}^n(k))} v_j(\theta)}{\lambda(\theta) v_e(\theta)} \right].$$

Denoting the terms in brackets by $A_e(k)$ and $\zeta_e(k)$, respectively, we can write

$$W_i^{n+1}(\theta) = \sum_e \sum_k A_e(k) \zeta_e(k),$$

where

$$A_e(k) \quad \text{are } \mathscr{F}_n \text{ measurable r.v.'s}$$

and

$$\zeta_e(k) \quad \text{are conditionally i.i.d., given } \mathscr{F}_n.$$

PROOF OF LEMMA 4.2. The following proof is an adaptation of its single type version in Neveu [19] and Joffe [14]. Define $\operatorname{var}_p X = EX^p - (EX)^p$ for $X \geq 0$. Then by the lemma in Section 3 of Neveu [19], one has for X, Y nonnegative and independent, that

$$\operatorname{var}_p(X+Y) \leq \operatorname{var}_p X + \operatorname{var}_p Y, \qquad 1 < p \leq 2. \tag{4.5}$$

Applying (4.5) yields

$$\begin{aligned}\operatorname{var}_p[W_i^{n+1}(\theta) | \mathscr{F}_n] &= \operatorname{var}_p\left[\sum_e \sum_k A_e(k) \zeta_e(k) \middle| \mathscr{F}_n \right] \\ &\leq \sum_{e,k} \operatorname{var}_p(A_e(k) \zeta_e(k) | \mathscr{F}_n) \\ &= \sum_{e,k} (A_e(k))^p \operatorname{var}_p(\zeta_e(k) | \mathscr{F}_n).\end{aligned}$$

Let $\operatorname{var}_p(\zeta_e(k) | \mathscr{F}_n) = c_e$. This is independent of k. By (H5), $c_e < \infty$ for $\theta = \theta_a$ and $p \leq p_0$. Thus by the martingale property and the definition of var_p,

$$\begin{aligned}E\left[(W_i^{n+1})^p | \mathscr{F}_n\right] - (W_i^n)^p &= E\left[(W_i^{n+1})^p | \mathscr{F}_n\right] - \left[E(W_i^{n+1} | \mathscr{F}_n)\right]^p \\ &\leq \sum_e c_e \sum_k (A_e(k))^p.\end{aligned} \tag{4.6}$$

Taking expectations of both sides gives

$$
\begin{aligned}
E(W_i^{n+1})^p - E(W_i^n)^p &\le \sum_e c_e E\left[\sum_k (A_e(k))^p\right] \\
&= \sum_e c_e E\left[\sum_k \frac{e^{p\theta Y_{ie}^n(k)} v_e^p(\theta)}{\lambda^{pn}(\theta) v_i^p(\theta)}\right] \\
&= \sum_e c_e \frac{\hat{m}_{ie}^n(p\theta) v_e^p(\theta)}{\lambda^{pn}(\theta) v_i^p(\theta)} \qquad (4.7)\\
&= \sum_e c_e \left[\frac{\lambda(p\theta)}{\lambda^p(\theta)}\right]^n \left(\frac{v_e^p(\theta)}{v_i^p(\theta)} \frac{v_i(p\theta)}{v_e(p\theta)}\right) \frac{\hat{m}_{ie}^n(p\theta) v_e(p\theta)}{\lambda^n(p\theta) v_i(p\theta)}.
\end{aligned}
$$

Letting

$$c = \max\left[c_e \frac{v_e^p(\theta) v_i(p\theta)}{v_i^p(\theta) v_e(p\theta)}; e = 1, \dots, d\right],$$

this is at most

$$c\left[\frac{\lambda(p\theta)}{\lambda^p(\theta)}\right]^n \sum_e \frac{\hat{m}_{ie}^n(p\theta) v_e(p\theta)}{\lambda^n(p\theta) v_i(p\theta)}.$$

For $\theta = \theta_a$, $p \le p_0$, one has $c < \infty$.

The sum in this last expression equals 1 since $\lambda(p\theta_a)$ and $v(p\theta_a)$ are an eigenvalue and eigenvector for $\{\hat{m}_{ij}(p\theta_a)\}$. Thus we conclude from (4.7) that

$$E(W_i^{n+1}(\theta_a))^p - E(W_i^n(\theta_a))^p \le c\left[\frac{\lambda(p\theta_a)}{\lambda^p(\theta_a)}\right]^n. \qquad (4.8)$$

We wish to show that

$$\lambda(p\theta_a) < \lambda^p(\theta_a) \quad \text{for some } p > 1. \qquad (4.9)$$

Then (4.3) will follow upon summing (4.8) over n. To check (4.9), write

$$f(\alpha) = \lambda(\alpha\theta_a) - \lambda^\alpha(\theta_a).$$

Since $f(1) = 0$, it is sufficient to show that $f'(1) < 0$. But

$$
\begin{aligned}
f'(1) &= [\theta_a \lambda'(\alpha\theta_a) - \lambda^\alpha(\theta_a) \log \lambda(\theta_a)]_{\alpha=1} \\
&= \lambda(\theta_a)\left[\theta_a \cdot \frac{\lambda'(\theta_a)}{\lambda(\theta_a)} - \log \lambda(\theta_a)\right] \\
&= \lambda(\theta_a)\Lambda^*(a) < 0,
\end{aligned}
$$

since $\Lambda^*(a) < 0$ by (H6). This proves Lemma 4.2. □

Note that in the proof of Lemma 4.2, one actually needs to use p-variances, rather than just variances [even if one strengthens (H5)], on account of (4.9).

CONVENTION. From now on, we take $p < p_0$ sufficiently close to 1 so that (4.3) holds.

The following inequality will be used in the proof of Proposition 5.1 and is an immediate consequence of Lemma 4.2.

LEMMA 4.3. *Assume* (H1)–(H6) *and suppose there are at least k type r particles at time l. Then*

$$E\Big[\big|X_{rj}^{ln}(k)\big|^p\Big|\mathscr{F}_l\Big] \le ce^{-\Lambda^*(a)pn}\lambda^{-pl}(\theta_a)e^{p\theta_a Y_{ir}^l(k)}, \qquad c<\infty. \tag{4.10}$$

PROOF. Abbreviate $\lambda(\theta_a) = \lambda$. Since $X_{rj}^{ln}(k)$ counts only particles at an, we see that

$$\begin{aligned}
&E\Big[\big|e^{\theta_a an}X_{rj}^{ln}(k)\big|^p\Big|\mathscr{F}_l\Big]\\
&\le E\left[\left|\sum_{h=1}^{Z_{rj}^{ln}(k)} e^{\theta_a Y_{rj}^{ln}(k,h)}\right|^p\middle|\mathscr{F}_l\right] \qquad (4.11)\\
&= e^{p\theta_a Y_{ir}^l(k)}\lambda^{(n-l)p}E\left[\left|\sum_h \frac{e^{\theta_a(Y_{rj}^{ln}(k,h)-Y_{ir}^l(k))}}{\lambda^{n-l}}\right|^p\middle|\mathscr{F}_l\right].
\end{aligned}$$

Now, the term under the expectation in the last expression is bounded in distribution by a constant times $|W_r^{n-l}(\theta_a)|^p$. Hence (4.11) is

$$\le c' e^{p\theta_a Y_{ir}^l(k)}\lambda^{(n-l)p}E\big|W_r^{n-l}(\theta_a)\big|^p$$

for some $c' < \infty$. By Lemma 4.2 this is

$$\begin{aligned}
&\le ce^{p\theta_a Y_{ir}^l(k)}\lambda^{(n-l)p}\\
&= ce^{p\theta_a Y_{ir}^l(k)}e^{-pn\Lambda^*(a)}\lambda^{-lp}e^{npa\theta_a}
\end{aligned} \tag{4.12}$$

for some c. Cancelling $e^{npa\theta_a}$ from (4.12) and the left side of (4.11) yields (4.10). □

Let q_i denote the extinction probability of the Galton–Watson process $Z_{ij}^n = \sum_{x\in I} Z_{ij}^n(x)$.

LEMMA 4.4. $P(W_i(\theta_a) > 0) = 1 - q_i$.

PROOF. The reasoning is essentially the same as for nonextinction in the Galton–Watson case. Decomposition based on the processes emanating from the first generation shows that

$$\hat{Z}_{ij}^n(\theta_a) =_D \sum_{e=1}^{d}\sum_{k=1}^{Z_{ie}^1} e^{\theta_a Y_{ie}^1(k)}\hat{Z}_{ej}^{n-1}(\theta_a;k),$$

where $\hat{Z}_{ej}^{n-1}(\theta_a, k)$ are i.i.d. copies of $\hat{Z}_{ej}^{n-1}(\theta_a)$. Multiplying through by $\lambda^{-n}(\theta_a)v_j(\theta_a)/v_i(\theta_a)$ and letting $n \to \infty$ yields

$$(4.13) \qquad W_i(\theta_a) =_D \lambda^{-1}(\theta_a) \sum_{e=1}^{d} \frac{v_e(\theta_a)}{v_i(\theta_a)} \sum_{k=1}^{Z_{ie}^1} e^{\theta_a Y_{ie}^1(k)} W_e(\theta_a; k),$$

where $W_e(\theta_a, k)$ are i.i.d. copies of $W_e(\theta_a)$. Thus $W_i(\theta_a) = 0$ corresponds to $W_e(\theta_a; k) = 0$ for all e and k. Letting $\mathbf{f}(\cdot)$ be the generating function of Z_1, $r_i = P(W_i(\theta_a) = 0)$ and $\mathbf{r} = (r_1, \dots, r_d)$, we conclude from (4.13) that $\mathbf{r}$ satisfies $\mathbf{r} = \mathbf{f}(\mathbf{r})$. Since $r_i < 1$ by Lemma 4.2, $\mathbf{r} = \mathbf{q}$. □

Finally, we will need the following lemma in the proof of Proposition 5.1.

LEMMA 4.5. *For any fixed* $\delta > 0$,

$$(4.14) \qquad P\left\{ \bigcup_{j=1}^{d} \bigcup_{h=1}^{Z_{ij}^n} \{Y_{ij}^n(h) \geq (a_0 + \delta)n\} \right\} \leq e^{-\tilde{\delta} n}$$

for some $\tilde{\delta} > 0$. [*Namely, the probability that any particles are located to the right of* $(a_0 + \delta)n$ *is small.*]

PROOF. The left side of (4.14) is

$$\leq \sum_j P\left\{ \bigcup_h Y_{ij}^n(h) \geq (a_0 + \delta)n \right\}$$

$$= E\left[\sum_{j=1}^{d} Z_{ij}^n P\{Y_{ij}^n(1) \geq (a_0 + \delta)n\} \right]$$

$$= \sum_j m_{ij}^n([(a_0 + \delta)n, \infty)).$$

By (2.2), this is

$$\leq ce^{-(\Lambda^*(a_0+\delta)-\delta')n}$$

for any $\delta' > 0$ and appropriate c. Since $\Lambda^*(a_0 + \delta) > 0$ by definition of a_0, the conclusion follows. □

5. The asymptotic behavior of $Z_{ij}^n[an]$. In this section, we prove our main result:

THEOREM. *Assume* (H1)–(H6). *Then for* $i, j = 1, \dots, d$,

$$(5.1) \qquad \lim_{n \to \infty} b_i^n \sqrt{n}\, e^{\Lambda^*(a)n} Z_{ij}^n[an] = u_j(\theta_a) W_i(\theta_a) \quad a.s.,$$

where b_i^n *is as in* (2.3) *and* $W_i(\theta_a)$ *is as in* (4.2).

Note that on account of Lemma 4.4, $W_i(\theta_a) > 0$ on the set of nonextinction of $\{Z^n\}$.

In the proof, we will sometimes abbreviate frequently used notation. Namely, we write

$$u_i = u_i(\theta_a), \qquad v_i = v_i(\theta_a), \qquad i = 1, \dots, d; \qquad \lambda = \lambda(\theta_a),$$

and

$$X_{ij}^n = Z_{ij}^n[an], \qquad X_{ij}^{ln}(k) = Z_{ij}^{ln}(k; [an]). \tag{5.2}$$

The proof of the theorem is divided into two propositions. The idea is to first show that X_{ij}^n is close to its conditional expectation with respect to $\mathscr{F}_l$ at a suitable time $l < n$, then to show that this conditional expectation converges to the asserted limit as $l, n \to \infty$ appropriately. The choice of l will be as follows: take $0 < \alpha < \frac{1}{2}$ and $t = 0, 1, \dots,$ and then define

$$l = l(n) = [t^\alpha] \quad \text{when } t^3 \le n < (t+1)^3. \tag{5.3}$$

We now state Propositions 5.1 and 5.2. The proof of Proposition 5.2 below follows closely the lines of the argument for the single type case in Biggins [9], where the choice of $l(n)$ in (5.3) is employed.

Proposition 5.1. *Assume* (H1)–(H6). *Then*

$$\sqrt{n}\, e^{\Lambda^*(a)n} \left| X_{ij}^n - E\left(X_{ij}^n | \mathscr{F}_l\right) \right| \to 0 \quad a.s., \tag{5.4}$$

as $l, n \to \infty$, with l as in (5.3).

Proposition 5.2. *Assume* (H1)–(H6). *Then*

$$\left| b_i^n \sqrt{n}\, e^{\Lambda^*(a)n} E\left(X_{ij}^n | \mathscr{F}_l\right) - u_j(\theta_a) W_i^l(\theta_a) \right| \to 0 \quad a.s., \tag{5.5}$$

as $l, n \to \infty$, with l as in (5.3).

Combining these two facts yields

$$\left| b_i^n \sqrt{n}\, e^{\Lambda^*(a)n} X_{ij}^n - u_j(\theta_a) W_i^l(\theta_a) \right| \to 0 \quad \text{a.s.} \tag{5.6}$$

But $W_i^l(\theta_a) \to W_i(\theta_a)$ a.s. by (4.2), and hence the theorem follows from the two propositions.

Proof of Proposition 5.1. Let

$$\sigma^p\left(X_{ij}^n | \mathscr{F}_l\right) = E\left(\left| X_{ij}^n - E\left(X_{ij}^n | \mathscr{F}_l\right)\right|^p \middle| \mathscr{F}_l\right), \qquad 1 < p < 2.$$

We first show:

Part 1. There exist constants $c < \infty$ and $\eta > 0$ such that

$$P\left\{e^{\Lambda^*(a)pn} \sigma^p\left(X_{ij}^n | \mathscr{F}_l\right) \ge c e^{-\eta l}\right\} \le c e^{-\eta l}. \tag{5.7}$$

We will then use (5.7) to prove:

PART 2. For sufficiently small $\eta' > 0$, there is an $\eta'' > 0$ such that

$$P\left\{e^{\Lambda^*(a)n}\left|X_{ij}^n - E\left(X_{ij}^n|\mathscr{F}_l\right)\right| \geq e^{-\eta' l}\right\} \leq ce^{-\eta'' l}. \tag{5.8}$$

The conclusion of Proposition 5.1 follows from (5.8). Namely, by adjusting the constant η' in (5.8), one obtains

$$P\left\{\sqrt{n}\, e^{\Lambda^*(a)n}\left|X_{ij}^n - E\left(X_{ij}^n|\mathscr{F}_l\right)\right| \geq ce^{-\eta' l}\right\} \leq ce^{-\eta'' l}. \tag{5.9}$$

Since $l = l(n) \sim n^{\alpha/3}$, $0 < \alpha < \frac{1}{2}$, application of the Borel–Cantelli lemma to (5.9) then yields Proposition 5.1.

PROOF OF PART 1. The idea behind the proof is to control the growth of the p-variance of X_{ij}^n as $n \to \infty$. The p-variance should be well-behaved, since large numbers of independent random variables $X_{rj}^{ln}(k)$, $k = 1, \dots, Z_{ir}^l$, are being summed in the branching process. One needs to be careful with large individual $Y_{ir}^l(k)$ though, which can increase the p-variance. It is easy to show that large $Y_{ir}^l(k)$, as given by $\mathscr{S}^c$ [(5.11)], only occur with small probability [(5.12)]. The contribution of the smaller terms given by $\mathscr{S}$ can be controlled using Lemma 4.3 to give a bound in terms of $W_i^l(\theta_a)$ [(5.13)]. Application of Lemma 4.2 reduces this to (5.7).

We first rewrite $\sigma^p(X_{ij}^n|\mathscr{F}_l)$ as

$$\begin{aligned}\sigma^p\left(X_{ij}^n|\mathscr{F}_l\right) &= E\left(\left|X_{ij}^n - E\left(X_{ij}^n|\mathscr{F}_l\right)\right|^p\Big|\mathscr{F}_l\right)\\ &= E\left\{\left|\sum_{r=1}^{d}\sum_{k=1}^{Z_{ir}^l}\left[X_{rj}^{ln}(k) - E\left(X_{rj}^{ln}(k)|\mathscr{F}_l\right)\right]\right|^p\Bigg|\mathscr{F}_l\right\}.\end{aligned}$$

We apply the von Bahr–Esseen inequality [4], which states that if $\xi_1, \dots, \xi_n$ are independent random variables with $E[\xi_k] = 0$ and $S_n = \sum_{k=1}^n \xi_k$, then

$$E\left[|S_n|^p\right] \leq c\sum_{k=1}^{n} E\left[|\xi_k|^p\right]$$

for a suitable constant c, $1 \leq p \leq 2$. Applying this gives the upper bound

$$c\sum_r\sum_k E\left\{\left|X_{rj}^{ln}(k) - E\left(X_{rj}^{ln}(k)|\mathscr{F}_l\right)\right|^p\Big|\mathscr{F}_l\right\}. \tag{5.10}$$

Decompose this sum into $A_l + B_l$, where A_l is the sum (5.10) over the set

$$\mathscr{S} = \left\{(k, r)\colon Y_{ir}^l(k) \leq (a_0 + \delta)l\right\}, \tag{5.11}$$

$\delta > 0$ to be chosen later, and B_l is the sum over $\mathscr{S}^c$. An upper bound for B_l is immediate from Lemma 4.5, namely

$$P\{|B_l| > 0\} \leq e^{-\tilde{\delta} l} \quad \text{for some } \tilde{\delta} > 0. \tag{5.12}$$

We proceed to bound A_l. Note that $|a+b|^p \le 2|a|^p + 2|b|^p$ and apply Jensen's inequality to obtain

$$E\Big(\big|X_{rj}^{ln}(k) - E\big(X_{rj}^{ln}(k)|\mathscr{F}_l\big)\big|^p\Big|\mathscr{F}_l\Big) \le 4E\Big(\big|X_{rj}^{ln}(k)\big|^p\Big|\mathscr{F}_l\Big).$$

By Lemma 4.3, this is

$$\le ce^{-\Lambda^*(a)pn}\lambda^{-pl}e^{p\theta_a Y_{ir}^l(k)}$$

for another choice of c. We therefore have

$$\begin{aligned} A_l e^{\Lambda^*(a)pn} &\le c\sum_{\mathscr{S}} \lambda^{-pl}e^{p\theta_a Y_{ir}^l(k)} \\ &= c\sum_{\mathscr{S}} \big(\lambda^{-l}e^{\theta_a Y_{ir}^l(k)}\big)^{p-1}\big(\lambda^{-l}e^{\theta_a Y_{ir}^l(k)}\big), \end{aligned}$$

which, since $\theta_a > 0$,

$$\le c\big(\lambda^{-l}e^{\theta_a(a_0+\delta)l}\big)^{p-1}\sum_r \frac{v_r}{v_i}\sum_k \lambda^{-l}e^{\theta_a Y_{ir}^l(k)}$$

for another appropriate c. This

$$= c\exp\{(p-1)(\theta_a\delta l + [\theta_a a_0 - \Lambda(\theta_a)]l)\}W_i^l(\theta_a).$$

Now note that

$$f(\theta) = \theta a_0 - \Lambda(\theta)$$

is increasing in θ at $\theta = \theta_a$. So for appropriate $\eta_1 > 0$, $\eta_2 > 0$,

$$\begin{aligned} \theta_a a_0 - \Lambda(\theta_a) &\le (\theta_a + \eta_1)(a_0 - \eta_2) - \Lambda(\theta_a + \eta_1) \\ &\le \sup_\theta\big[\theta(a_0 - \eta_2) - \Lambda(\theta)\big] \\ &= \Lambda^*(a_0 - \eta_2) < 0, \end{aligned}$$

the last inequality following from the definition of a_0. Plugging this in above gives

$$A_l e^{\Lambda^*(a)pn} \le ce^{-(p-1)\gamma l}W_i^l(\theta_a) \tag{5.13}$$

for sufficiently small γ.

Let $\gamma_1 = (p-1)\gamma/2$. Then by (5.13),

$$\begin{aligned} P\{A_l e^{\Lambda^*(a)pn} \ge e^{-\gamma_1 l}\} &\le P\{ce^{-(p-1)\gamma l}W_i^l(\theta_a) \ge e^{-\gamma_1 l}\} \\ &= P\left\{W_i^l(\theta_a) \ge \frac{e^{\gamma_1 l}}{c}\right\}. \end{aligned} \tag{5.14}$$

But

$$P\{W_i^n(\theta_a) \ge M\} \le \frac{E|W_i^n(\theta_a)|^p}{M^p} \le \frac{c}{M^p}$$

for suitable c by the Chebyshev inequality and Lemma 4.2. This implies

$$P\{W_i^l(\theta_a) \geq e^{\gamma_1 l}/c\} \leq c' e^{-\gamma_1 p l}$$

for appropriate c'. Therefore

$$P\{A_l e^{\Lambda^*(a)pn} \geq e^{-\gamma_1 l}\} \leq c' e^{-\gamma_1 p l}.$$

Combining this with (5.12), we see that there exist constants $c < \infty$ and $\eta > 0$ such that for all l, n,

$$P\{e^{\Lambda^*(a)pn}(A_l + B_l) \geq c e^{-\eta l}\} \leq c e^{-\eta l}.$$

Together with (5.10) this implies (5.7) and Part 1.

PROOF OF PART 2. Define the sets

$$F = \{e^{\Lambda^*(a)pn} \sigma^p(X_{ij}^n | \mathscr{F}_l) < c e^{-\eta l}\},$$

$$G = \{e^{\Lambda^*(a)n} |X_{ij}^n - E(X_{ij}^n | \mathscr{F}_l)| \geq e^{-\eta' l}\},$$

and decompose $P(G) = P(GF) + P(GF^c)$. By (5.7),

$$P(GF^c) \leq P(F^c) \leq c e^{-\eta l}. \tag{5.15}$$

Note that since $F \in \mathscr{F}_l$,

$$P(GF | \mathscr{F}_l) = E(1_G 1_F | \mathscr{F}_l) = 1_F E(1_G | \mathscr{F}_l) = 1_F P(G | \mathscr{F}_l).$$

By the Chebyshev inequality, this is

$$\leq 1_F \frac{\sigma^p(X_{ij}^n | \mathscr{F}_l)}{(e^{-\Lambda^*(a)n} e^{-\eta' l})^p} \leq c e^{(p\eta' - \eta) l}, \tag{5.16}$$

the last inequality following from the definition of F. Now choosing η' sufficiently small, (5.15) and (5.16) imply (5.8) and Part 2. This completes the proof of Proposition 5.1. □

PROOF OF PROPOSITION 5.2. In Proposition 5.2, we wish to show that $E(X_{ij}^n | \mathscr{F}_l)$, after suitable renormalization, is approximated by the martingale $W_i^l(\theta_a)$ as $n \to \infty$ with $l \sim n^{\alpha/3}$. The approach is to first decompose $E(X_{ij}^n | \mathscr{F}_l)$ as in (5.18) according to its behavior before and after time l. The terms corresponding to the interval $[l, n]$ are centered here by subtracting $a(n - l)$. These terms are then compared to appropriately centered normal densities [(5.20)]; by Lemma 3.3, the approximation will be close. After some further computation, the randomness in the centering is shown to become insignificant as $n \to \infty$ [(5.23)]. Getting rid of this randomness produces the martingale $W_i^l(\theta_a)$.

PART 1. Recall the centered stochastic kernel defined in (3.17):

$$\overline{m}_{ij}^n(x; \theta_a) = e^{\theta_a(an + x)} \frac{m_{ij}^n(an + x) v_j(\theta_a)}{\lambda^n(\theta_a) v_i(\theta_a)}. \tag{5.17}$$

We first express $E(X_{ij}^n|\mathscr{F}_l)$ in terms of this kernel. Namely,

$$\begin{aligned} E(X_{ij}^n|\mathscr{F}_l) &= \sum_{r=1}^{d}\sum_{k=1}^{Z_{ir}^l} E\big(X_{rj}^{ln}(k)\big|\mathscr{F}_l\big) \\ &= \sum_r \sum_k m_{rj}^{n-l}([an] - Y_{ir}^l(k)) \\ &= \beta_n \sum_r \sum_k e^{-\theta_a(an - Y_{ir}^l(k))}\lambda^{n-l}(\theta_a)\frac{v_r}{v_j}\overline{m}_{rj}^{n-l}\big(y^l - Y_{ir}^l(k);\theta_a\big), \end{aligned}$$

where $y^l = [an] - a(n-l) \in I_{n-l}$ and $\beta_n = e^{\theta_a(an-[an])}$. (When $an \in I$, one has $y^l = al$ and $\beta_n = 1$.) This

$$(5.18)\qquad = \frac{\sqrt{2\pi}\,\sigma_i}{b_i^n v_j}e^{-\Lambda^*(a)n}\sum_r\sum_k \frac{v_r}{v_i}\frac{e^{\theta_a Y_{ir}^l(k)}}{\lambda^l(\theta_a)}\overline{m}_{rj}^{n-l}\big(y^l - Y_{ir}^l(k);\theta_a\big),$$

where b_i^n is given in (2.3) and $\sigma_i = \sigma_i(\theta_a)$ in (3.18).

PART 2. We next show that

$$(5.19)\qquad \frac{1}{\sqrt{2\pi}}\left| b_i^n\sqrt{n-l}\,e^{\Lambda^*(a)n}E\big(X_{ij}^n|\mathscr{F}_l\big) - u_j\sum_r\sum_k \frac{v_r}{v_i}\frac{e^{\theta Y_{ir}^l(k)}}{\lambda^l(\theta_a)}\right| \to 0 \quad \text{a.s.},$$

as $l, n \to \infty$, where l is as in (5.3). Since the sum in (5.19) is just the martingale $W_i^l(\theta_a)$, this implies

$$\left|b_i^n\sqrt{n}\,e^{\Lambda^*(a)n}E\big(X_{ij}^n|\mathscr{F}_l\big) - u_jW_i^l(\theta_a)\right| \to 0 \quad \text{a.s.},$$

which is Proposition 5.2.

To prove (5.19), note that by (5.18) and the triangle inequality, the left side of (5.19) is

$$(5.20)\qquad \begin{aligned} &\le \left|\frac{\sigma_i}{v_j}\sqrt{n-l}\sum_r\sum_k \frac{v_r}{v_i}\frac{e^{\theta_a Y_{ir}^l(k)}}{\lambda^l(\theta_a)}\overline{m}_{rj}^{n-l}\big(y^l - Y_{ir}^l(k);\theta_a\big) - \sum_r\sum_k \frac{v_r}{v_i}\frac{e^{\theta_a Y_{ir}^l(k)}}{\lambda^l(\theta_a)}u_j\varphi\left(\frac{y^l - Y_{ir}^l(k)}{\sigma_i\sqrt{n-l}}\right)\right| \\ &\quad + \left|\sum_r\sum_k \frac{v_r}{v_i}\frac{e^{\theta_a Y_{ir}^l(k)}}{\lambda^l(\theta_a)}u_j\left[\varphi\left(\frac{y^l - Y_{ir}^l(k)}{\sigma_i\sqrt{n-l}}\right) - \frac{1}{\sqrt{2\pi}}\right]\right|. \end{aligned}$$

Now Lemma 3.3 yields

$$\sup_{y \in I_{n-l}} \left| \frac{\sigma_i}{v_j} \sqrt{n}\, \overline{m}_{ij}^{n}(y;\theta_a) - u_j \varphi\left(\frac{y}{\sigma_i \sqrt{n}} \right) \right| \to 0,$$

and hence

$$\left| \frac{\sigma_i}{v_j} \sqrt{n-l}\, \overline{m}_{ij}^{n-l}(y^l - Y_{ir}^l(k);\theta_a) - u_j \varphi\left(\frac{y^l - Y_{ir}^l(k)}{\sigma_i \sqrt{n-l}} \right) \right| \le A_{n-l}, \tag{5.21}$$

where $A_k \to 0$ as $k \to \infty$. Also note that $|\varphi(x) - \varphi(0)| \le cx^2$. Applying this fact and (5.21) to (5.20), we see that the LHS of (5.19) is

$$\le c \sum_r \sum_k \frac{v_r}{v_i} \frac{e^{\theta_a Y_{ir}^l(k)}}{\lambda^l(\theta_a)} \left[A_{n-l} + \frac{(y^l - Y_{ir}^l(k))^2}{(n-l)\sigma_i^2} \right] \tag{5.22}$$

with another choice of c. Let

$$R_1(n,l) = cA_{n-l} \sum_r \sum_k \frac{v_r}{v_i} \frac{e^{\theta_a Y_{ir}^l(k)}}{\lambda^l(\theta_a)}$$

and

$$R_2(n,l) = \frac{c}{(n-l)\sigma_i^2} \sum_r \sum_k \frac{v_r}{v_i} \frac{e^{\theta_a Y_{ir}^l(k)}}{\lambda^l(\theta_a)} (y^l - Y_{ir}^l(k))^2.$$

Thus (5.19) $\le R_1(n,l) + R_2(n,l)$. Note that $R_1(n,l) = cA_{n-l} W_i^l(\theta_a)$. Recalling that $l \sim n^{\alpha/3}$, that $A_k \to 0$ as $k \to \infty$ and the corollary of Lemma 4.1, we see that

$$R_1(n,l) \to 0 \quad \text{a.s. as } n \to \infty.$$

It remains to show that $R_2(n,l) \to 0$ a.s. To this end, note that

$$\begin{aligned} ER_2(n,l) &= \frac{c}{(n-l)\sigma_i^2} \sum_{r=1}^{d} \frac{v_r}{v_i} \int \frac{e^{\theta_a y}}{\lambda^l(\theta_a)} (y - y^l)^2 m_{ir}^l(dy) \\ &= \frac{c}{(n-l)\sigma_i^2} \sum_{r=1}^{d} \int (y - \alpha_n)^2 \overline{m}_{ir}^l(dy;\theta_a), \end{aligned} \tag{5.23}$$

where $\alpha_n = [an] - an$. This is

$$\le \frac{c'l}{n-l}$$

for appropriate c', since the above second moment is finite. Thus

$$\sum_{t=1}^{\infty} ER_2(t^3, [t^\alpha]) \le c' \sum_{t=1}^{\infty} \frac{[t^\alpha]}{t^3 - [t^\alpha]} < \infty. \tag{5.24}$$

Also, note that $R_2(n, l)$ is decreasing in the first argument. So for $n_0 = (t_0)^3$,

$$P\left[\sup_{n \ge n_0} R_2(n, l) > \varepsilon\right] \le P\left[\sup_{t \ge t_0} R_2(t^3, [t^\alpha]) > \varepsilon\right],$$

and it suffices to show the right side approaches 0 for any $\varepsilon > 0$ as $t_0 \to \infty$. This, however, follows from (5.24), in conjunction with the Chebyshev inequality and the Borel–Cantelli lemma. Consequently,

$$R_2(n, l) \to 0 \quad \text{a.s. as } n \to \infty.$$

This proves (5.19) and completes the proof of Proposition 5.2. □

Acknowledgment. We would like to thank the referee for an observation leading to substantial simplification of the paper.

REFERENCES

[1] Asmussen, S. and Herring, H. (1983). *Branching Processes*. Birkhäuser, Boston.
[2] Asmussen, S. and Kaplan, N. (1976). Branching random walks I and II. *J. Stochastic Process. Appl.* **4** 1–32.
[3] Athreya, K. B. and Kaplan, N. (1978). The additive property and its application in branching processes. *Adv. in Probab.* **5** 27–60.
[4] von Bahr, B. and Essen, C.-G. (1965). Inequalities for the rth absolute moment of a sum of random variables, $1 \le r \le 2$. *Ann. Math. Statist.* **36** 299–303.
[5] Biggins, J. D. (1976). The first- and last-birth problems for a multitype age-dependent branching process. *Adv. in Appl. Probab.* **8** 446–459.
[6] Biggins, J. D. (1977). Chernoff's theorem in branching random walk. *J. Appl. Probab.* **14** 630–636.
[7] Biggins, J. D. (1977). Martingale convergence in the branching random walk. *J. Appl. Probab.* **14** 25–37.
[8] Biggins, J. D. (1978). The asymptotic shape of the branching random walk. *Adv. in Appl. Probab.* **10** 62–84.
[9] Biggins, J. D. (1979). Growth rates in the branching random walk. *Z. Wahrsch. Verw. Gebiete* **48** 17–34.
[10] Bramson, M. (1978). Maximal displacement of branching Brownian motion. *Comm. Pure Appl. Math.* **31** 531–581.
[11] Bramson, M. (1978). Minimal displacement of branching random walk. *Z. Wahrsch. Verw. Gebiete* **45** 89–108.
[12] Iltis, M. (1991). Asymptotics of large deviations for Markov chains. Ph.D. dissertation, Univ. Wisconsin.
[13] Iscoe, I., Ney, P. and Nummelin, E. (1985). Large deviations of uniformly recurrent Markov additive processes. *Adv. in Appl. Math.* **6** 373–412.
[14] Joffe, A. (1992). Martingale methods in branching processes. *Proc. 1989 Vilnius Conf.* To appear.
[15] Joffe, A., Le Cam, L. and Neveu, J. (1973). Sur la loi des grands nombres pour des variables aléatoires de Bernoulli attachées à un arbre dyadique. *C. R. Acad. Sci. Paris Ser. A–B* **277** 963–964.
[16] Kingman, J. F. C. (1975). The first birth problem for an age-dependent branching process. *Ann. Probab.* **3** 790–801.
[17] Miller, H. D. (1961). A convexity property in the theory of random variables on a finite Markov chain. *Ann. Math. Statist.* **32** 1260–1270.

[18] NAGAEV, S. V. (1957). Some limit theorems for stationary Markov chains. *Theory Probab. Appl.* **2** 378–406.
[19] NEVEU, J. (1988). *Seminar on Stochastic Processes 1987* 223–273. Birkhäuser, Boston.
[20] NEY, P. (1992). Branching random walk. In *Spatial Stochastic Processes, Progress in Probability* **19**. Birkhäuser, Boston. To appear.
[21] NEY, P. and NUMMELIN, E. (1987). Markov additive processes I and II. *Ann. Probab.* **15** 561–592, 593–609.
[22] TAO, J. (1992). Large deviations of multi-type branching random walk. Ph.D. dissertation, Univ. Wisconsin.
[23] UCHIYAMA, K. (1982). Spatial growth of a branching process of particles living in R^d. *Ann. Probab.* **10** 896–918.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WISCONSIN—MADISON
480 LINCOLN DRIVE
VAN VLECK HALL
MADISON, WISCONSIN 53706