FLUCTUATIONS IN CRUMP–MODE–JAGERS PROCESSES: THE LATTICE CASE

SVANTE JANSON

Abstract. Consider a supercritical Crump–Mode–Jagers process such that all births are at integer times (the lattice case). We showed in a recent paper that under a certain condition on the generating function of the intensity of the offspring process, the second-order fluctuations of the age distribution are asymptotically normal.

In the present paper we study mainly the case when this condition is not satisfied. There are two other cases; in one (boundary) case the fluctuations are still asymptotically normal, but with a larger order of the variance; in the last case, the fluctuations are even larger, but will oscillate and (except in degenerate cases) not converge in distribution. This trichotomy is similar to what has been seen in related situations, e.g. for Pólya urns.

We also add some further results in the first case, including a symbolic calculus.

1. Introduction

We consider a supercritical Crump–Mode–Jagers branching process, starting with a single individual born at time 0, and where an individual has $N \leqslant \infty$ children born at the times when the parent has age $\xi_1 \leqslant \xi_2 \leqslant \dots$. Here N and $(\xi_i)_i$ are random, and different individuals have independent copies of these random variables. Technically, it is convenient to regard $\{\xi_i\}_1^N$ as a point process Ξ on $[0, \infty)$, and give each individual x an independent copy Ξ_x of Ξ. For further details, see e.g. Jagers [7].

In the present paper, as in the companion paper [10], we consider for simplicity only the lattice case, where all ξ_i are multiples of some fixed real number d; we may without loss of generality assume that $d = 1$ and that d is maximal; thus the ξ_i are integer-valued and all births occur at integer times a.s., but there is no $d > 1$ such that all birth times a.s. are divisible by d. It would be very interesting to extend the results below to the non-lattice case; we expect similar results (under suitable assumptions), but this case seems to present new technical challenges, and we leave this as an open problem.

It is well-known that in the supercritical case that we study, the population grows to infinity, at least with positive probability; furthermore, under

Date: 6 December, 2017.
Partly supported by the Knut and Alice Wallenberg Foundation.

weak assumptions, the age distribution converges a.s. to some limit, see e.g. [7, Theorems (6.3.3), (6.8.1), (6.10.1) and Corollary (6.10.4)].

The purpose of the present paper is to continue the study started in [10] of the second-order fluctuations of the age distribution, or more generally, of the distribution of some other property, described by a random characteristic. Our main results (Theorems 2.1–2.3) show that there are three different cases depending on properties of the intensity measure $\mathbb{E}\,\Xi$ of the offspring process: in one case (the case studied in [10]) fluctuations are, after proper normalization, asymptotically normal, with only a short-range dependence between different times; in another case, there is a long-range dependence and, again after proper normalization (different this time), the fluctuations are a.s. approximated by oscillating (almost periodic) random functions of $\log n$, which furthermore essentially are determined by the initial phase of the branching process, and presumably non-normal; the third case is an intermediate boundary case. See Section 2 for precise results.

A similar trichotomy has been found in several related situations. Similar results are proved for multi-type Markov branching processes by Asmussen and Hering [1, Section VIII.3]. Their type space may be very general, so this setting includes also the single-type non-Markov case studied here (also in the non-lattice case [1, Section VIII.12]), since a Crump–Mode–Jagers branching process may be seen as a Markov process where the type of an individual is its entire life history until present. However, this will in general be a large type space, and the assumptions of [1] will in general not be satisfied; in particular, their "condition (M)" [1, p. 156] is typically not satisfied, by the same argument as in [1, p. 173] for a related situation. Hence, we can not obtain our results directly from the closely related results in [1], although there is an overlap in some special cases (for example the Galton–Watson case in Example 2.5).

Another related situation is given by multi-colour Pólya urn processes, see e.g. [9] (which uses methods and results from branching process theory). The same trichotomy appears there too, with a criterion formulated in terms of eigenvalues of a matrix that can be seen as the (expected) "offspring matrix" in that setting.

It would be interesting to find more general theorems that would include these different but obviously related results together.

Remark 1.1. Our setup includes the Galton–Watson case, where all births occur when the mother has age 1 (Example 2.5), but this case is much simpler than the general case and can be treated by simpler methods; see Jagers [7, Section 2.10], where results closely related to the ones below are given.

In general, our setting can, for example, be considered as a model for the (female) population of some animal that is fertile several years and gets one or several children once every year, with the numbers of children different years random, and possibly dependent and with different distributions depending on the age of the mother.

2. Assumptions and main results

We repeat some notation and assumptions from [10].

Let $\mu := \mathbb{E}\,\Xi$ be the intensity measure of the offspring process; thus $\mu := \sum_{k=0}^{\infty} \mu_k \delta_k$, where μ_k is the expected number of children that an individual bears at age k (and δ_k is the Dirac delta, i.e., a point mass at k). Let $N_k := \Xi\{k\}$ be the number of children born to an individual at age k. Thus $N = \sum_{k=1}^{\infty} N_k$ and $\mu_k = \mathbb{E}\,N_k$.

We make the following standing assumptions, valid throughout the paper. The first assumption (supercriticality) is essential. The assumptions (A2)–(A4) are simplifying and convenient but presumably not essential. (For (A4), this is shown in Example 9.2.)

(A1) The process is supercritical, i.e., $\mu([0,\infty]) = \sum_{k=0}^{\infty} \mu_i = \mathbb{E}\,N > 1$.
(A2) No children are born instantaneously, i.e., $\mu_0 = 0$.
(A3) $N \geqslant 1$ a.s. Thus the process a.s. survives.
(A4) There are no deaths.

Define, for all complex z such that either $z \geqslant 0$ or the sums or expectations below converge absolutely,

$$\widehat{\mu}(z) := \sum_{k=0}^{\infty} \mu_k z^k = \sum_{k=0}^{\infty} \mathbb{E}[N_k] z^k = \mathbb{E} \sum_{i=1}^{N} z^{\xi_i} \tag{2.1}$$

and the complex-valued random variable

$$\widehat{\Xi}(z) := \int_0^{\infty} z^x \,\mathrm{d}\Xi(x) = \sum_{i=1}^{N} z^{\xi_i} = \sum_{k=0}^{\infty} N_k z^k. \tag{2.2}$$

Thus $\widehat{\mu}(z) = \mathbb{E}\,\widehat{\Xi}(z)$.

We make two other standing assumptions:

(A5) $\widehat{\mu}\bigl(m^{-1}\bigr) = 1$ for some $m > 1$.

Thus $\alpha := \log m$ satisfies $\sum_{k=1}^{\infty} \mu_k e^{-k\alpha} = \widehat{\mu}(e^{-\alpha}) = 1$, so $\log m$ is the Malthusian parameter, and the population grows roughly with a factor $e^{\alpha} = m$ for each generation (see e.g. (2.7) and (2.8) below).

(A6) $\mathbb{E}[\widehat{\Xi}(r)^2] < \infty$ for some $r > m^{-1/2}$.

We fix in the sequel some $r > m^{-1/2}$ satisfying (A6). We assume for convenience $r \leqslant 1$. Note that (A6) implies

$$\widehat{\mu}(r) = \mathbb{E}\,\widehat{\Xi}(r) < \infty. \tag{2.3}$$

Hence $\widehat{\mu}(z)$ and $\widehat{\Xi}(z)$ are defined, and analytic, at least for $|z| \leqslant r$. Since $\widehat{\mu}(z)$ is a strictly increasing function on $[0,\infty)$, m^{-1} in (A5) is the unique positive root of $\widehat{\mu}(z) = 1$. However, $\widehat{\mu}(z) = 1$ may have other complex roots; we shall see that the asymptotic behaviour of the fluctuations depends crucially on the position of these roots. We define, with $D_r := \{|z| < r\}$,

$$\Gamma := \{z \in D_r : \widehat{\mu}(z) = 1\}, \qquad \Gamma_* := \Gamma \setminus \{m^{-1}\}, \tag{2.4}$$

$$\gamma_* := \inf\{|z| : z \in \Gamma_*\}, \tag{2.5}$$

$$\Gamma_{**} := \{z \in \Gamma_* : |z| = \gamma_*\}, \tag{2.6}$$

with $\gamma_* = \infty$ if $\Gamma_* = \emptyset$. (These sets may depend on the choice of r, but for our purposes this does not matter. Recall that we assume $r > m^{-1/2}$.) Since $\widehat{\mu}(z)$ is analytic, Γ is discrete and thus, if $\gamma_* < \infty$, then Γ_{**} is a finite non-empty set which we write as $\{\gamma_1, \dots, \gamma_q\}$.

Let Z_n be the total number of individuals at time n. (Which by (A2) equals the number of individuals born up to time n.) We define Z_n for all integers n by letting $Z_n := 0$ for $n < 0$. By assumption, $Z_0 = 1$.

It is well-known that the number of individuals Z_n grows asymptotically like m^n as $n \to \infty$. For example, see e.g. [7, Theorem (6.3.3)] (and remember that we here consider the lattice case),

$$\mathbb{E}\, Z_n \sim c_1 m^n, \qquad n \to \infty, \tag{2.7}$$

for some $c_1 > 0$. Moreover, if $\mathbb{E}[\widehat{\Xi}(m^{-1}) \log \widehat{\Xi}(m^{-1})] < \infty$, and in particular if $\mathbb{E}[\widehat{\Xi}(m^{-1})^2] < \infty$, which follows from our assumption (A6), then as $n \to \infty$,

$$Z_n/m^n \xrightarrow{\text{a.s.}} \mathcal{Z} \tag{2.8}$$

for some random variable $\mathcal{Z} > 0$, see e.g. Nerman [11]. In particular, it follows that for any fixed $k \geqslant 1$

$$Z_{n-k}/Z_n \xrightarrow{\text{a.s.}} m^{-k}. \tag{2.9}$$

The number of individuals of age $\geqslant k$ at time n is Z_{n-k}. For large n, we expect this to be roughly $m^{-k}Z_n$, see (2.9), and to study the fluctuations, we define, for $k = 0, 1, \dots$,

$$X_{n,k} := Z_{n-k} - m^{-k}Z_n. \tag{2.10}$$

Note that $X_{n,0} = 0$.

The main result of [10] showed asymptotic normality of $X_{n,k}$, suitably normalized, if $\gamma_* > m^{-1/2}$. We complement this in the present paper by treating also the cases $\gamma_* = m^{-1/2}$ and $\gamma_* < m^{-1/2}$. We state the main results as three separate theorems, showing different behaviours in the three different cases. Proofs are given in later sections.

By (A6) and (2.2), $\mathbb{E}\, N_k^2 < \infty$ for every $k \geqslant 1$. Let, for $j, k \geqslant 1$,

$$\sigma_{jk} := \operatorname{Cov}(N_j, N_k) \tag{2.11}$$

and, at least for $|z| < r$,

$$\Sigma(z) := \sum_{i,j} \sigma_{ij} z^i \bar{z}^j = \operatorname{Cov}\Bigl(\sum_i N_i z^i, \sum_j N_j \bar{z}^j\Bigr) = \mathbb{E}\bigl|\widehat{\Xi}(z) - \widehat{\mu}(z)\bigr|^2. \tag{2.12}$$

Also, for $R > 0$, let ℓ^2_R be the Hilbert space of infinite vectors

$$\ell^2_R := \Bigl\{(a_k)_{k=0}^\infty : \|(a_k)_0^\infty\|^2_{\ell^2_R} := \sum_{k=0}^\infty R^{2k}|a_k|^2 < \infty\Bigr\}. \tag{2.13}$$

For completeness, we begin with the case $\gamma_* > m^{-1/2}$ treated in [10].

Theorem 2.1 ([10])**.** *Assume* (A1)–(A6) *and* $\gamma_* > m^{-1/2}$*. Then, as* $n \to \infty$*,*

$$X_{n,k}/\sqrt{Z_n} \xrightarrow{\mathrm{d}} \zeta_k, \tag{2.14}$$

jointly for all $k \geqslant 0$*, for some jointly normal random variables* ζ_k *with mean* $\zeta_k = 0$ *and covariance matrix given by, for any finite sequence* $a_0, \dots, a_K$ *of real numbers,*

$$\operatorname{Var}\Bigl(\sum_k a_k\zeta_k\Bigr) = \frac{m-1}{m}\oint_{|z|=m^{-1/2}} \frac{\bigl|\sum_k a_k z^k - \sum_k a_k m^{-k}\bigr|^2}{|1-z|^2\,|1-\widehat{\mu}(z)|^2}\Sigma(z)\frac{|\mathrm{d}z|}{2\pi m^{-1/2}}. \tag{2.15}$$

Moreover, the convergence (2.14) *holds also in the stronger sense that* $(Z_n^{-1/2}X_{n,k})_k \xrightarrow{\mathrm{d}} (\zeta_k)_k$ *in the Hilbert space* ℓ^2_R*, for any* $R < m^{1/2}$*.*

The limit variables ζ_k *are non-degenerate unless* Ξ *is deterministic, i.e.,* $N_k = \mu_k$ *a.s. for each* $k \geqslant 0$*.*

Recall that joint convergence of an infinite number of variables means joint convergence of any finite set. (This is convergence in the product space $\mathbb{R}^\infty$, see [2].) Note that $\zeta_0 = 0$ (trivial but included for completeness).

The variance formula (2.15) can be interpreted as a stochastic calculus, where the limit variables are seen as stochastic integrals (in a general sense) of certain functions on the circle $|z| = m^{-1/2}$; these functions thus represent the random variables ζ_k, and therefore asymptotically $X_{n,k}$; moreover, they can be used for convenient calculations. See Section 8 for details.

Theorem 2.1 is, as said above, proved in [10]. Nevertheless, we give a new proof in Section 5. The main reason is that the new proof easily adapts to give a proof of Theorem 2.2 below, see Section 6.

We consider next the cases $\gamma_* \leqslant m^{1/2}$. Then $\Gamma_{**} = \{\gamma_1, \dots, \gamma_q\}$ is a non-empty finite set. For simplicity, we assume the condition

$$\widehat{\mu}'(\gamma) \neq 0, \qquad \gamma \in \Gamma_{**}, \tag{2.16}$$

i.e., that the points in Γ_{**} are simple roots of $\widehat{\mu}(z) = 1$; the modifications in the case with a multiple root are left to the reader. (See Remark 3.8, and note the related results for Pólya urns in [9, Theorems 3.23–3.24] and [12, Theorems 3.5–3.6].)

Theorem 2.2. *Assume* (A1)–(A6) *and* $\gamma_* = m^{-1/2}$*. Suppose further that* (2.16) *holds. Then, as* $n \to \infty$*,*

$$X_{n,k}/\sqrt{nZ_n} \xrightarrow{\mathrm{d}} \zeta_k, \tag{2.17}$$

jointly for all $k \geqslant 0$, for some jointly normal random variables ζ_k with mean $\zeta_k = 0$ and covariance matrix given by, for any finite sequence $a_0, \dots, a_K$ of real numbers,

$$\operatorname{Var}\Bigl(\sum_k a_k\zeta_k\Bigr) = (m-1)\sum_{p=1}^{q} \frac{\bigl|\sum_k a_k\gamma_p^k - \sum_k a_k m^{-k}\bigr|^2}{|1-\gamma_p|^2\,|\widehat{\mu}'(\gamma_p)|^2}\Sigma(\gamma_p). \tag{2.18}$$

Moreover, the convergence (2.14) holds also in the Hilbert space ℓ^2_R, for any $R < m^{1/2}$.

The limit variables ζ_k are non-degenerate unless $\widehat{\Xi}(\gamma_p)$ is deterministic for each $\gamma_p \in \Gamma_{**}$.

Theorem 2.3. *Assume (A1)–(A6) and $\gamma_* < m^{-1/2}$. Suppose further that (2.16) holds. Then there exist complex random variables $U_1, \dots, U_q$ and linearly independent vectors $\vec{u}_i := \bigl(\gamma_i^k - m^{-k}\bigr)_k$, $i = 1, \dots, q$, such that*

$$\gamma_*^n \vec{X}_n - \sum_{i=1}^{q} \bigl(\bar{\gamma}_i/|\gamma_i|\bigr)^n U_i\vec{u}_i \to 0 \tag{2.19}$$

a.s. and in $L^2(\ell^2_R)$, for any $R < m^{1/2}$. Furthermore, $\mathbb{E}\, U_i = 0$, and U_i is non-degenerate unless $\widehat{\Xi}\bigl(\gamma_i\bigr)$ is degenerate.

Theorems 2.1–2.3 exhibit several differences between the cases $\gamma_* < m^{-1/2}$, $\gamma_* = m^{-1/2}$ and $\gamma_* > m^{-1/2}$; cf. the similar results for Pólya urns in e.g. [9, Theorems 3.22–3.24].

- The fluctuations $X_{n,k}$, for a fixed k, are asymptotically normal when $\gamma_* \geqslant m^{-1/2}$, but (presumably) not when $\gamma_* < m^{-1/2}$.
- The fluctuations are typically of order $Z_n^{1/2} \asymp m^{n/2}$ when $\gamma_* > m^{-1/2}$, slightly larger (by a power of n) when $\gamma_* = m^{-1/2}$, and of the much larger order γ_*^{-n} when $\gamma_* < m^{-1/2}$.
- When $\gamma_* < m^{-1/2}$, the fluctuations exhibit oscillations that are periodic or almost periodic (see [3]) in $\log n$. (Note that $\gamma_i/|\gamma_i| \neq 1$ in (2.19), since m^{-1} is the only positive root in Γ.)
- When $\gamma_* < m^{-1/2}$, there is the a.s. approximation result (2.19), implying both long-range dependence as $n \to \infty$, and that the asymptotic behaviour essentially is determined by what happens in the first few generations. In contrast, the limits in (2.14) and (2.17) are mixing (see the proofs), i.e., the results holds also conditioned on the life histories of the first M individuals for any fixed M, and thus also conditioned on $Z_1, \dots, Z_K$ for any fixed K; hence, when $\gamma_* \geqslant m^{-1/2}$, the initial behaviour is eventually forgotten. Moreover for $\gamma_* > m^{-1/2}$, there is only a short-range dependence, see Example 8.1, while the case $\gamma_* = m^{-1/2}$ shows an intermediate "medium-range" dependence, see Subsection 8.2.
- When $\gamma_* > m^{-1/2}$, the limit random variables ζ_k in (2.14) are linearly independent, as a consequence of (2.15). When $\gamma_* \leqslant m^{-1/2}$,

the limits in (2.17), or the components of the sum in (2.19), span a (typically) q-dimensional space of random variables, and any $q+1$ of them are linearly dependent; see also Section 8.

Remark 2.4. We consider above $X_{n,k}$ for $k \geqslant 0$, i.e., the age distribution of the population at time n. We can define $X_{n,k}$ by (2.10) also for $k<0$; this means looking into the future and can be interpreted as predicting the future population. As shown in Section 8, (2.14)–(2.15) and (2.17)–(2.18) extend to all $k \in \mathbb{Z}$ (still jointly), and, similarly, taking the kth component in (2.19) yields a result that extends to all $k \in \mathbb{Z}$.

This enables us, for example, to obtain (by standard linear algebra) the best linear predictor of Z_{n+1} based on the observed $Z_n, \dots, Z_{n-K}$ for any fixed K.

Example 2.5 (Galton–Watson)**.** The simplest example is a Galton–Watson process, where all children are born in a single litter at age 1 of the parent, so $N_k = 0$ for $k \geqslant 2$. (Recall that all individuals live for ever in our setting. In the traditional setting, one considers the last generation, i.e., $Z_n - Z_{n-1}$.) Then, as noted in [10], $N = N_1$, $m = \mu_1$ and $\widehat{\mu}(z) = mz$; hence $\Gamma = \{m^{-1}\}$, $\Gamma_* = \emptyset$, and $\gamma_* = \infty > m^{-1/2}$. We assume $\mathbb{E}\, N^2 < \infty$, $N \geqslant 1$ a.s. and $\mathbb{P}(N > 1) > 0$; then (A1)–(A6) hold (with any r in (A6)). Thus Theorem 2.1 applies. We obtain, for example, with $\sigma^2 := \operatorname{Var}(N) = \sigma_{11}$,

$$\begin{aligned}\operatorname{Var}\bigl(\zeta_1\bigr) &= \frac{m-1}{m} \oint_{|z|=m^{-1/2}} \frac{\bigl|z-m^{-1}\bigr|^2}{|1-z|^2\,|1-mz|^2} \sigma^2 |z|^2 \frac{|\mathrm{d}z|}{2\pi m^{-1/2}} \\ &= \sigma^2 \frac{m-1}{m^4} \oint_{|z|=m^{-1/2}} \frac{1}{|1-z|^2} \frac{|\mathrm{d}z|}{2\pi m^{-1/2}} \\ &= \sigma^2 m^{-3}. \end{aligned} \tag{2.20}$$

This can be shown directly in a much simpler way; see [7, Theorem (2.10.1)], which is essentially equivalent to our Theorem 2.1 in the Galton–Watson case (but without our assumption (A3)).

Example 2.6. Suppose that all children are born when the mother has age one or two, i.e., $N_k = 0$ for $k > 2$. Then $\widehat{\mu}(z) = \mu_1 z + \mu_2 z^2$, where by assumption $\mu_1 + \mu_2 > 1$ and $\mu_1 > 0$. (A5) yields $m^2 = \mu_1 m + \mu_2$, and thus

$$m = \frac{\mu_1 + \sqrt{\mu_1^2 + 4\mu_2}}{2}. \tag{2.21}$$

The equation $\widehat{\mu}(z) = 1$ has one other root, viz. γ_1 with

$$\gamma_1^{-1} = -\frac{\sqrt{\mu_1^2 + 4\mu_2} - \mu_1}{2}. \tag{2.22}$$

The condition $\gamma_* > m^{-1/2}$ is thus equivalent to $\gamma_1^{-2} < m$, which after some elementary algebra is equivalent to, for example,

$$u_1^3 + 3u_1u_2 + u_2 - u_2^2 > 0. \tag{2.23}$$

Thus, Theorem 2.1 applies when (2.23) holds, Theorem 2.2 when there is equality in (2.23), and Theorem 2.3 when the left-hand side of (2.23) is negative. (In this example, (2.16) is trivial.)

For a simple numerical example with $\gamma_* = m^{-1/2}$, take $\mu_1 = 2$ and $\mu_2 = 8$. Then (2.21)–(2.22) yield $m = 4$ and $\gamma_1 = -\frac{1}{2}$. We obtain by (2.18), for example,

$$X_{n,1}/\sqrt{nZ_n} \xrightarrow{\mathrm{d}} \zeta_1 \sim N\Bigl(0, \frac{1}{768}\operatorname{Var}(N_2 - 2N_1)\Bigr). \tag{2.24}$$

Suppose now instead that (2.23) holds, so Theorem 2.1 applies. Let $\lambda := \gamma_1^{-1}$ be given by (2.22). Then $1 - \widehat{\mu}(z) = (1 - mz)(1 - \lambda z)$, and thus (2.15) yields, for example,

$$\begin{aligned}\operatorname{Var}\bigl(\zeta_1\bigr) &= \frac{m-1}{m}\oint_{|z|=m^{-1/2}} \frac{\bigl|z - m^{-1}\bigr|^2}{|1-z|^2\,|1-\widehat{\mu}(z)|^2}\Sigma(z)\frac{|\mathrm{d}z|}{2\pi m^{-1/2}} \\ &= \frac{m-1}{m^3}\oint_{|z|=m^{-1/2}} \frac{\sigma_{11}|z|^2 + \sigma_{12}(z+\bar{z})|z|^2 + \sigma_{22}|z|^4}{|1-z|^2\,|1-\lambda z|^2}\frac{|\mathrm{d}z|}{2\pi m^{-1/2}}.\end{aligned} \tag{2.25}$$

This integral can be evaluated by expanding $(1-z)^{-1}(1-\lambda z)^{-1}$ in a Taylor series; this yields after some calculations

$$\operatorname{Var}\bigl(\zeta_1\bigr) = \frac{(m+\lambda)(\sigma_{11} + \sigma_{22}/m) + 2(1+\lambda)\sigma_{12}}{m^2(m-\lambda)(m-\lambda^2)}. \tag{2.26}$$

Remark 2.7. The limit in (2.14) is by Theorem 2.1 degenerate only when the entire process is, and thus each $X_{n,k}$ is degenerate. In contrast, the limit in (2.17) or the approximation in (2.19) may be degenerate even in other (special) situations. For example, let N_1 be non-degenerate with $\mathbb{E}\,N_1 = 2$, let $N_2 := 2N_1 + 4$, and let $N_k := 0$ for $k > 2$. Then $\mu_1 = 2$ and $\mu_2 = 8$, and Example 2.6 shows that $\gamma_* = \frac{1}{2} = m^{-1/2}$; furthermore, (2.24) applies and yields $X_{n,k}/\sqrt{nZ_n} \xrightarrow{\mathrm{d}} 0$.

We conjecture that in this case (and similar ones with $\zeta_k = 0$ in Theorem 2.2), $X_{n,k}/\sqrt{Z_n}$ has a non-trivial normal limit in distribution; we leave this as an open problem. Simliarly, we conjecture that when each $\widehat{\Xi}(\gamma_i)$ is degenerate in Theorem 2.3, the distribution of $X_{n,k}$ is asymptotically determined by the next smallest roots in Γ_*.

2.1. **More notation.** For a random variable X in a Banach space $\mathcal{B}$, we define $\|X\|_{L^2(\mathcal{B})} := (\mathbb{E}\,\|X\|_{\mathcal{B}}^2)^{1/2}$, when $\mathcal{B} = \mathbb{R}$ or $\mathbb{C}$ abbreviated to $\|X\|_2$.

For infinite vectors $\vec{x} = (x_j)_{j=0}^\infty$ and $\vec{y} = (y_j)_{j=0}^\infty$, let $\langle \vec{x}, \vec{y}\rangle := \sum_{j=0}^\infty x_j y_j$, assuming that the sum converges absolutely.

C denotes different constants that may depend on the distribution of the branching process (i.e., on the distribution of N and (ξ_i)), but not on n and similar parameters; the constant may change from one occurrence to the next.

$O_{\text{a.s.}}(1)$ means a quantity that is bounded by a random constant that does not depend on n.

All unspecified limits are as $n \to \infty$.

3. Preliminaries

We give some further definitions and results. These are partly taken from [10]; we repeat some definitions and statements here for convenience, but refer to [10] for further details.

Let

$$B_n := Z_n - Z_{n-1} \tag{3.1}$$

be the number of individuals born at time n (with $B_0 = Z_0$). Also, let $B_{n,k}$ be the number of individuals born at time $n+k$ by parents that are themselves born at time n, and thus are of age k. Thus, recalling (A2),

$$B_n = \sum_{k=1}^{n} B_{n-k,k}, \qquad n \geqslant 1. \tag{3.2}$$

Let $\mathcal{F}_n$ be the σ-field generated by the life histories of all individuals born up to time n. Then $B_{n,k}$ is $\mathcal{F}_n$-measurable, and B_n is $\mathcal{F}_{n-1}$-measurable by (3.2). (With $\mathcal{F}_{-1}$ trivial.) Furthermore,

$$\mathbb{E}\bigl(B_{n,k} \mid \mathcal{F}_{n-1}\bigr) = \mu_k B_n, \qquad n \geqslant 0. \tag{3.3}$$

For $k \geqslant 1$, let

$$W_{n,k} := B_{n,k} - \mathbb{E}\bigl(B_{n,k} \mid \mathcal{F}_{n-1}\bigr) = B_{n,k} - \mu_k B_n. \tag{3.4}$$

(Thus $W_{n,k} = 0$ if $n < 0$.) Then $W_{n,k}$ is $\mathcal{F}_n$-measurable with

$$\mathbb{E}\bigl(W_{n,k} \mid \mathcal{F}_{n-1}\bigr) = 0. \tag{3.5}$$

Let further

$$W_n := B_n - \sum_{k=1}^{n} \mu_k B_{n-k} = B_n - \sum_{k=1}^{\infty} \mu_k B_{n-k}. \tag{3.6}$$

Thus $W_0 = B_0 = Z_0$, and for $n \geqslant 1$, by (3.6), (3.2) and (3.4),

$$W_n = \sum_{k=1}^{n} W_{n-k,k}. \tag{3.7}$$

Lemma 3.1 ([10])**.** *Assume* (A1)–(A6)*. Then, for all $n \geqslant 1$ and $k \geqslant 1$, $\mathbb{E}[W_{n,k}^2] \leqslant Cr^{-2k}m^n$ and $\mathbb{E}[W_n^2] \leqslant Cm^n$.*

Proof. See [10]. □

We use vector notation. Let $\vec{X}_n := (X_{n,k})_{k=0}^{\infty}$. Furthermore, let

$$\vec{v} := (0, m^{-1}, m^{-2}, \dots) = \bigl(m^{-k}\mathbf{1}\{k > 0\}\bigr)_{k=0}^{\infty} \tag{3.8}$$

and let

$$\Psi\bigl((y_k)_0^{\infty}\bigr) := \sum_{k=1}^{\infty} \mu_k (y_k - y_{k-1}), \tag{3.9}$$

for vectors $(y_k)_0^\infty$ such that the sum converges; finally, let S be the shift operator $S\big((y_k)_0^\infty\big) := (y_{k-1})_0^\infty$ with $y_{-1} := 0$, and let T be the linear operator

$$T(\vec{y}) := S(\vec{y}) + \Psi(\vec{y})\vec{v}. \tag{3.10}$$

Then elementary calculations yield, see [10],

$$\vec{X}_{n+1} = S(\vec{X}_n) + \big(\Psi(\vec{X}_n) - W_{n+1}\big)\vec{v} = T(\vec{X}_n) - W_{n+1}\vec{v}, \tag{3.11}$$

which leads to the following formula.

Lemma 3.2 ([10])**.** *For every $n \geqslant 0$,*

$$\vec{X}_n = -\sum_{k=0}^{n} W_{n-k} T^k(\vec{v}). \tag{3.12}$$

Proof. See [10]. □

Remark 3.3. It follows from the proofs below, that the sum in (3.12) is dominated by the first few terms in the case $\gamma_* > m^{-1/2}$, and by the last few terms in the case $\gamma_* < m^{-1/2}$, while all terms are of about the same size when $\gamma_* = m^{-1/2}$. This explains much of the different behaviours seen in Section 2.

We shall consider T defined in (3.10) as an operator on the complex Hilbert space ℓ^2_R defined in (2.13) for a suitable $R > 0$. Recall that the *spectrum* $\sigma(T)$ of a linear operator in a complex Hilbert (or Banach) space is the set of complex numbers λ such that $\lambda - T$ is *not* invertible; see e.g. [4, Section VII.3].

Lemma 3.4 ([10])**.** *Suppose that $1 \leqslant R < m$ and that $\widehat{\mu}(R^{-1}) < \infty$. Then $\vec{v} \in \ell^2_R$, Ψ is a bounded linear functional on ℓ^2_R and T is a bounded linear operator on ℓ^2_R. Furthermore, if $\lambda \in \mathbb{C}$ with $|\lambda| > R$, then $\lambda \in \sigma(T)$ if and only if $\lambda^{-1} \in \Gamma_*$, i.e., if and only if $\lambda \neq m$ and $\widehat{\mu}(\lambda^{-1}) = 1$.*

Proof. See [10]. □

Remark 3.5. It is easily seen that $\lambda \in \sigma(T)$ for every λ with $|\lambda| \leqslant R$, e.g. by taking $h = v$ in (3.21)–(3.22) and noting that $v(z)/(\lambda - z) \notin H^2_R$. Thus we have a complete description of the spectrum $\sigma(T)$ on ℓ^2_R.

Lemma 3.6 ([10])**.** *Suppose that $1 \leqslant R < m$ and that $\widehat{\mu}(R^{-1}) < \infty$. Suppose furthermore that $\widehat{\mu}(z) \neq 1$ for every complex $z \neq m^{-1}$ with $|z| < R^{-1}$. Then, for every $R_1 > R$, there exists $C = C(R_1)$ such that*

$$\|T^n\|_{\ell^2_R} \leqslant C R_1^n, \qquad n \geqslant 0. \tag{3.13}$$

Proof. See [10]. □

We shall use Lemma 3.6 when $\gamma_* > m^{-1/2}$. In the case $\gamma_* \leqslant m^{-1/2}$, we use instead the following lemma, based on a more careful spectral analysis of T. Recall the definitions (2.4)–(2.6).

Lemma 3.7. *Assume that* $R = r^{-1} \geqslant 1$, *where* $\widehat{\mu}(r) < \infty$. *Suppose furthermore that* $\Gamma_{**} = \{\gamma_1, \dots, \gamma_q\} \neq \emptyset$, *and that* (2.16) *holds. Let* $\lambda_i := \gamma_i^{-1}$. *Then there exist eigenvectors* $\vec{v}_i$ *with* $T\vec{v}_i = \lambda_i \vec{v}_i$ *and linear projections* P_i *with range* $\mathcal{R}(P_i) = \{c\vec{v}_i : c \in \mathbb{C}\}$ *(i.e., the span of* $\vec{v}_i$*),* $i = 1, \dots q$, *and furthermore a bounded operator* T_0 *in* ℓ^2_R *and a constant* $\tilde{R} < \gamma_*^{-1}$ *such that, for any* $n \geqslant 0$,

$$T^n = T_0^n + \sum_{i=1}^{q} \lambda_i^n P_i \tag{3.14}$$

and

$$\left\| T_0^n \right\|_{\ell^2_R} \leqslant C \tilde{R}^n. \tag{3.15}$$

Explicitly,

$$\vec{v}_i = P_i(\vec{v}) = \frac{1}{\gamma_i(\gamma_i - 1)\widehat{\mu}'(\gamma_i)} \big(\gamma_i^k - m^{-k}\big)_k. \tag{3.16}$$

Proof. Since the points in Γ_* are isolated, there is a number $\tilde{r} > \gamma_*$ such that $|z| > \tilde{r}$ for any $z \in \Gamma_* \setminus \Gamma_{**}$. We may assume $\tilde{r} < r$. Let $\tilde{R} := \tilde{r}^{-1} > R$. By Lemma 3.4, $\lambda_i = \gamma_i^{-1} \in \sigma(T)$ with $|\lambda_i| = \gamma_*^{-1}$, and $|\lambda| < \tilde{R} < \gamma_*^{-1}$ for any $\lambda \in \sigma(T) \setminus \{\lambda_1, \dots, \lambda_q\}$.

Since $\lambda_1, \dots, \lambda_q$ thus are isolated points in $\sigma(T)$, by standard functional calculus, see e.g. [4, Section VII.3], there exist commuting projections (not necessarily orthogonal) $P_0, \dots, P_q$ in ℓ^2_R such that $\sum_{i=0}^q P_i = 1$, T maps each subspace $E_i := P_i(\ell^2_R)$ into itself, and if $\hat{T}_i$ is the restriction of T to E_i, then $\hat{T}_i$ has spectrum $\sigma(\hat{T}_i) = \{\lambda_i\}$ for $1 \leqslant i \leqslant q$ and $\sigma(\hat{T}_0) = \sigma(T) \setminus \{\lambda_i\}_1^q$. In particular, the spectral radius $r(\hat{T}_0) < \tilde{R}$, and thus, by the spectral radius formula [4, Lemma VII.3.4],

$$\|\hat{T}_0^n\| \leqslant C\tilde{R}^n, \qquad n \geqslant 0. \tag{3.17}$$

Let $T_0 := TP_0$. Then $T_0^n = T^n P_0 = \hat{T}_0^n P_0$, and (3.15) follows.

We use, as the proof of Lemma 3.4 in [10], the fact that the mapping $(a_k)_0^\infty \mapsto \sum_{k=0}^\infty a_k z^k$ is an isometry of ℓ^2_R onto the Hardy space H^2_R consisting of all analytic functions $f(z)$ in the disc $\{z : |z| < R\}$ such that

$$\|f\|^2_{H^2_R} := \sup_{r<R} \frac{1}{2\pi} \int_0^{2\pi} |f\big(re^{i\theta}\big)|^2 \, d\theta < \infty. \tag{3.18}$$

(See e.g. [5].) In particular, $\vec{v}$ corresponds to the function

$$v(z) := \sum_{k=1}^{\infty} m^{-k} z^k = \frac{z/m}{1 - z/m} = \frac{z}{m - z}. \tag{3.19}$$

We use the same notations Ψ, S and T for the corresponding linear functional and operators on H^2_R, and note that then, on H^2_R, $Sf(z) = zf(z)$ and

$$Tf(z) = zf(z) + \Psi(f)v(z). \tag{3.20}$$

Consequently, for any $h \in H^2_R$, the equation $(\lambda - T)f = h$ is equivalent to

$$(\lambda - z)f(z) - \Psi(f)v(z) = h(z). \tag{3.21}$$

As in [10], we note that any solution to (3.21) has to be of the form

$$f(z)=c\frac{v(z)}{\lambda-z}+\frac{h(z)}{\lambda-z}, \tag{3.22}$$

where

$$c=\Psi(f)=c\Psi\Bigl(\frac{v(z)}{\lambda-z}\Bigr)+\Psi\Bigl(\frac{h(z)}{\lambda-z}\Bigr). \tag{3.23}$$

It follows that if $|\lambda|>R$, then (3.21) has a unique solution $f\in H^2_R$ if and only if $\Psi\bigl(v(z)/(\lambda-z)\bigr)\neq 1$. Furthermore, a calculation, see [10], shows that if $|\lambda|>R$ and $\lambda\neq m$, then

$$\Psi\Bigl(\frac{v(z)}{\lambda-z}\Bigr)=\frac{1}{m-\lambda}\bigl((1-\lambda)\widehat{\mu}(\lambda^{-1})+m-1\bigr), \tag{3.24}$$

which equals 1 if and only if $\widehat{\mu}(\lambda^{-1})=1$, i.e., if and only if $\lambda^{-1}\in\Gamma_*$.

For each λ_i, thus $\Psi\bigl(v(z)/(\lambda_i-z)\bigr)=1$ by (3.24), and the kernel $\mathcal{N}(\lambda_i-T)$ is one-dimensional and spanned by $v(z)/(\lambda_i-z)$, see (3.21)–(3.23). Furthermore, see again (3.21)–(3.23), the range $\mathcal{R}(\lambda_i-T)$ is given by

$$\mathcal{R}(\lambda_i-T)=\Bigl\{h\in\ell^2_R:\Psi\Bigl(\frac{h(z)}{\lambda_i-z}\Bigr)=0\Bigr\}. \tag{3.25}$$

By differentiating (3.24), we find for $|\lambda|>R$ with $\lambda^{-1}\in\Gamma_*$, i.e., $\lambda\neq m$ and $\widehat{\mu}\bigl(\lambda^{-1}\bigr)=1$,

$$\begin{aligned}\Psi\Bigl(\frac{v(z)}{(\lambda-z)^2}\Bigr)&=-\frac{\mathrm{d}}{\mathrm{d}\lambda}\Psi\Bigl(\frac{v(z)}{\lambda-z}\Bigr)=\frac{\mathrm{d}}{\mathrm{d}\lambda}\Bigl(1-\Psi\Bigl(\frac{v(z)}{\lambda-z}\Bigr)\Bigr)\\&=\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{(1-\lambda)(1-\widehat{\mu}(\lambda^{-1}))}{m-\lambda}=\frac{(1-\lambda)\widehat{\mu}'(\lambda^{-1})}{(m-\lambda)\lambda^2}.\end{aligned} \tag{3.26}$$

Thus, the assumption (2.16) implies that $\Psi\bigl(v(z)/(\lambda_i-z)^2\bigr)\neq 0$, and thus $v(z)/(\lambda_i-z)\notin\mathcal{R}(\lambda_i-T)$ by (3.25). Hence, $\mathcal{N}(\lambda_i-T)\cap\mathcal{R}(\lambda_i-T)=\{0\}$. Consequently, for every $h\in\mathcal{R}(\lambda_i-T)$, (3.21) has a unique solution $f\in\mathcal{R}(\lambda_i-T)$, i.e., the restriction of λ_i-T to $\mathcal{R}(\lambda_i-T)$ is invertible.

It follows that the projection P_i is the projection onto $\mathcal{N}(\lambda_i-T)=\{cv(z)/(\lambda_i-z)\}$ that vanishes on $\mathcal{R}(\lambda_i-T)$, which by (3.25) is given by

$$P_i(f(z))=\frac{\Psi\bigl(f(z)/(\lambda_i-z)\bigr)}{\Psi\bigl(v(z)/(\lambda_i-z)^2\bigr)}\cdot\frac{v(z)}{\lambda_i-z}. \tag{3.27}$$

In particular, since $\Psi\bigl(v(z)/(\lambda_i-z)\bigr)=1\neq 0$, $P_i(v)$ is a non-zero multiple of $v(z)/(\lambda_i-z)$. Let $\vec{v}_i:=P_i(\vec{v})$. Thus $T\vec{v}_i=\lambda_i\vec{v}_i$, and, for $n\geqslant 0$,

$$T^n=T^nP_0+\sum_{i=1}^q T^nP_i=T_0^n+\sum_{i=1}^q\lambda_i^nP_i, \tag{3.28}$$

showing (3.14).

Finally, (3.27) and (3.26) yield

$$v_i(z):=P_i\bigl(v(z)\bigr)=\frac{(m-\lambda_i)\lambda_i^2}{(1-\lambda_i)\widehat{\mu}'(\lambda_i^{-1})}\cdot\frac{v(z)}{\lambda_i-z}, \tag{3.29}$$

and (3.16) follows because $\lambda_i = \gamma_i^{-1}$ and by (3.19), for $|\lambda| > R$,

$$(m-\lambda)\frac{v(z)}{\lambda - z} = \frac{\lambda}{\lambda - z} - \frac{m}{m-z} = \sum_{k=0}^{\infty}(\lambda^{-k} - m^{-k})z^k. \tag{3.30}$$

□

Remark 3.8. It follows also that (2.16) implies that the points $\lambda_i \in \sigma(T)$ are simple poles of the resolvent $(\lambda - T)^{-1}$, and conversely. Lemma 3.7 can be extended without assuming (2.16); the general result is similar but more complicated, and is left to the reader. Cf. [4, Theorem VII.3.18].

We shall also use another similar calculation.

Lemma 3.9. *Suppose that* $1 \leqslant R < m$ *and that* $\widehat{\mu}(R^{-1}) < \infty$*. If* $|\lambda| > R$ *and* $\widehat{\mu}\big(\lambda^{-1}\big) \neq 1$*, then*

$$(\lambda - T)^{-1}(\vec{v}) = \frac{1}{(1-\lambda)(1-\widehat{\mu}(\lambda^{-1}))}\big(\lambda^{-k} - m^{-k}\big)_k. \tag{3.31}$$

Proof. Taking $h = v$ in (3.21)–(3.23), we find

$$(\lambda - T)^{-1}v(z) = f(z) = b\frac{v(z)}{\lambda - z} \tag{3.32}$$

for a constant b such that $b = \Psi(f) + 1$. This yields by (3.24)

$$b - 1 = \Psi(f) = \frac{b}{m-\lambda}\big((1-\lambda)\widehat{\mu}(\lambda^{-1}) + m - 1\big) \tag{3.33}$$

with the solution

$$b = \frac{m-\lambda}{(1-\lambda)(1-\widehat{\mu}(\lambda^{-1}))}. \tag{3.34}$$

Hence, using (3.30), for $|z| < R$,

$$f(z) = b\frac{v(z)}{\lambda - z} = \frac{1}{(1-\lambda)(1-\widehat{\mu}(\lambda^{-1}))}\sum_{\ell=0}^{\infty}(\lambda^{-k} - m^{-k})z^k. \tag{3.35}$$

□

4. A martingale

In the remaining sections, we let $R := r^{-1} < m^{1/2}$, where r is as in (A6). (We may assume that R is arbitrarily close to $m^{1/2}$ by decreasing r.) We consider as above the operator T on ℓ^2_R.

Fix a real vector $\vec{a} \in \ell^2_{R^{-1}}$ (for example any finite real vector), and write

$$\alpha_k = \alpha_k(\vec{a}) := \langle T^k(\vec{v}), \vec{a}\rangle. \tag{4.1}$$

Then (3.12) and (3.7) yield

$$\langle \vec{X}_n, \vec{a}\rangle = -\sum_{k=0}^{\infty}\sum_{j=1}^{\infty} W_{n-k-j,j}\alpha_k = -\sum_{\ell=0}^{n}\sum_{j=1}^{n-\ell} W_{\ell,j}\alpha_{n-j-\ell} \tag{4.2}$$

Define

$$\Delta M_{n,\ell} := \sum_{j=1}^{n-\ell} \alpha_{n-j-\ell} W_{\ell,j}, \tag{4.3}$$

$$M_{n,k} := \sum_{\ell=0}^{k} \Delta M_{n,\ell}. \tag{4.4}$$

Then (3.5) shows that $\mathbb{E}\bigl(\Delta M_{n,\ell} \mid \mathcal{F}_{\ell-1}\bigr) = 0$, and thus $(M_{n,k})_{k=0}^{n}$ is a martingale with respect to $(\mathcal{F}_k)_k$. Furthermore, by (4.2),

$$\langle \vec{X}_n, \vec{a} \rangle = -M_{n,n}. \tag{4.5}$$

Conditioned on $\mathcal{F}_{\ell-1}$, the vector $(W_{\ell,j})_j$ is the sum of B_ℓ independent copies of $\vec{N} - \mathbb{E}\,\vec{N}$, where $\vec{N} = (N_j)_0^\infty$, and thus, recalling (2.11),

$$\begin{aligned} Q_{n,l} := \mathbb{E}\bigl((\Delta M_{n,\ell})^2 \mid \mathcal{F}_{\ell-1}\bigr) &= B_\ell \operatorname{Var}\Bigl(\sum_{j=1}^{n-\ell} \alpha_{n-\ell-j} N_j\Bigr) \\ &= B_\ell \sum_{i,j=1}^{n-\ell} \sigma_{ij} \alpha_{n-\ell-i} \alpha_{n-\ell-j}. \end{aligned} \tag{4.6}$$

The conditional quadratic variation of the martingale $(M_{n,k})_k$ is thus

$$\begin{aligned} V_n := \sum_{\ell=0}^{n} Q_{n,\ell} &= \sum_{\ell=0}^{n} B_\ell \sum_{i,j=1}^{n-\ell} \sigma_{ij} \alpha_{n-\ell-i} \alpha_{n-\ell-j} \\ &= \sum_{\ell=0}^{n} B_{n-\ell} \sum_{i,j=1}^{\ell} \sigma_{ij} \alpha_{\ell-i} \alpha_{\ell-j}. \end{aligned} \tag{4.7}$$

By (2.2), $N_k \leqslant r^{-k} \widehat{\Xi}(r)$, and thus by (2.11) and the Cauchy–Schwarz inequality,

$$|\sigma_{ij}| \leqslant r^{-i-j}\, \mathbb{E}\, \widehat{\Xi}(r)^2 = C R^{i+j}. \tag{4.8}$$

5. Proof of Theorem 2.1

Recall that Theorem 2.1 is proved in [10]. We give here another proof, which unlike the proof in [10] is based on a martingale central limit theorem. (The proof in [10] is based on the central limit theorem for sums of i.i.d. variables, together with some approximations.) As said earlier, the main reason is that the new proof with small modifications also applies to Theorem 2.2, see Section 6, and we prefer to present it first for Theorem 2.1. (The proof in [10] does not seem to extend easily to Theorem 2.2.)

In this section we assume $\gamma_* > m^{-1/2}$, in addition to (A1)–(A6). In other words, see (2.5), each $z \in \Gamma_*$ satisfies $|z| > m^{-1/2}$. Hence, we may decrease r so that the disc D_r contains no roots of $\widehat{\mu}(z) = 1$ except m^{-1}, and still

$r > m^{-1/2}$. Thus, with $R := 1/r$ and assuming (A1)–(A6), we see that $\gamma_* > m^{-1/2}$ is equivalent to:

$$\text{There exists } R \text{ with } 1 \leqslant R < m^{1/2} \text{ such that } \widehat{\mu}(R^{-1}) < \infty \text{ and,}$$
$$\text{furthermore, } \widehat{\mu}(z) \neq 1 \text{ for every complex } z \neq m^{-1} \text{ with } |z| < R^{-1}. \tag{5.1}$$

We fix an R such that (5.1) holds, and (A6) holds with $r = 1/R$. Furthermore, we fix R_1 with $R < R_1 < m^{1/2}$. Then (5.1) and Lemma 3.6 show that (3.13) holds, i.e., $\|T^n\|_{\ell^2_R} = O\big(R_1^n\big)$.

Lemma 5.1. *Assume* (A1)–(A6) *and $\gamma_* > m^{1/2}$. If $R < m^{1/2}$, then*

$$\mathbb{E}\,\|\vec{X}_n\|^2_{\ell^2_R} \leqslant Cm^n \tag{5.2}$$

and thus

$$\mathbb{E}\,X^2_{n,k} \leqslant CR^{-2k}m^n \tag{5.3}$$

for all $n, k \geqslant 0$.

Proof. By (3.12), Lemma 3.1, (3.13) and Minkowski's inequality,

$$\begin{aligned}\|\vec{X}_n\|_{L^2(\ell^2_R)} &\leqslant \sum_{k=0}^{n} \|W_{n-k}\|_{L^2}\|T^k(\vec{v})\|_{\ell^2_R} \leqslant C\sum_{k=0}^{n} m^{(n-k)/2}R_1^k \\ &= Cm^{n/2}\sum_{k=0}^{n}(R_1/m^{1/2})^k \leqslant Cm^{n/2}.\end{aligned} \tag{5.4}$$

This yields (5.2), and (5.3) follows by (2.13). □

Similarly, (4.1) and (3.13) show that, for a fixed $\vec{a}$, with $C = C(\vec{a})$,

$$|\alpha_k| \leqslant CR_1^k. \tag{5.5}$$

Consequently, by (4.6), (4.8) and (5.5), since $R/R_1 < 1$,

$$\frac{Q_{n,\ell}}{B_\ell} = \sum_{i,j=1}^{n-\ell} \sigma_{ij}\alpha_{n-\ell-i}\alpha_{n-\ell-j} \leqslant C\sum_{i,j=1}^{\infty} R^{i+j}R_1^{2(n-\ell)-i-j} \leqslant CR_1^{2(n-\ell)}. \tag{5.6}$$

Hence, by (4.7), (4.6), (3.1) and (2.9), using dominated convergence justified by (5.6) and $R_1^2/m < 1$,

$$\begin{aligned}\frac{V_n}{Z_n} &= \sum_{\ell=0}^{n}\frac{B_{n-\ell}}{Z_n}\frac{Q_{n,n-\ell}}{B_{n-\ell}} = \sum_{\ell=0}^{n}\frac{Z_{n-\ell}-Z_{n-\ell-1}}{Z_n}\sum_{i,j=1}^{\ell}\sigma_{ij}\alpha_{\ell-i}\alpha_{\ell-j} \\ &\xrightarrow{\text{a.s.}} \sigma^2(\vec{a}) := \sum_{\ell=0}^{\infty}\big(m^{-\ell}-m^{-\ell-1}\big)\sum_{i,j=1}^{\ell}\sigma_{ij}\alpha_{\ell-i}\alpha_{\ell-j}\end{aligned} \tag{5.7}$$

We cannot use a martingale central limit theorem directly for the martingale $(M_{n,k})_k$ defined in (4.4), because the calculations above show that most of the conditional quadratic variation V_n comes from a few terms (the last ones), cf. Remark 3.3. We thus introduce another martingale.

Number the individuals $1, 2, \dots$ in order of birth, with arbitrary order at ties, and let $\mathcal{G}_\ell$ be the σ-field generated by the life histories of individuals $1, \dots, \ell$. Each Z_n is a stopping time with respect to $(\mathcal{G}_\ell)_\ell$, and $\mathcal{G}_{Z_n} = \mathcal{F}_n$.

We refine the martingale $(M_{n,k})_k$ by adding the contribution from each individual separately. Let τ_i denote the birth time of i, and $N_{i,k}$ the copy of N_k for i (i.e., the number of children i gets at age k). Let

$$\Delta\widehat{M}_{n,i} := \sum_{j=1}^{n-\tau_i} \alpha_{n-\tau_i-j}\bigl(N_{i,j} - \mu_j\bigr), \tag{5.8}$$

$$\widehat{M}_{n,k} := \sum_{i=1}^{k} \Delta\widehat{M}_{n,i}. \tag{5.9}$$

Then $(\widehat{M}_{n,k})_k$ is a $(\mathcal{G}_k)_k$-martingale with $\widehat{M}_{n,\infty} = \widehat{M}_{n,Z_n} = M_{n,n} = -\langle \vec{X}_n, \vec{a}\rangle$, see (4.3)–(4.5), and the conditional quadratic variation

$$\widehat{V}_n := \sum_i \mathbb{E}\bigl((\Delta\widehat{M}_{n,i})^2 \mid \mathcal{G}_{i-1}\bigr) = V_n \tag{5.10}$$

given by (4.7). Moreover, by (5.8) and (5.5),

$$\bigl|\Delta\widehat{M}_{n,i}\bigr| \leqslant C\sum_{j=0}^{\infty} R_1^{n-\tau_i-j}\bigl(N_{i,j} + \mu_j\bigr) = CR_1^{n-\tau_i}\bigl(\widehat{\Xi}_i(R_1^{-1}) + \widehat{\mu}(R_1^{-1})\bigr). \tag{5.11}$$

Define the random variable $U := \widehat{\Xi}(R_1^{-1}) + \widehat{\mu}(R_1^{-1})$. Then $\mathbb{E}\, U^2 < \infty$ by (A6), since $R_1^{-1} < r$. It follows from (5.11) that for some $c > 0$ and every $\varepsilon > 0$, defining $h(x) := \mathbb{E}\bigl(U^2\mathbf{1}\{U > cx\}\bigr)$,

$$\begin{aligned}\mathbb{E}\bigl(\bigl|\Delta\widehat{M}_{n,i}\bigr|^2\mathbf{1}\{\bigl|\Delta\widehat{M}_{n,i}\bigr| > \varepsilon\} \mid \mathcal{G}_{i-1}\bigr) &\leqslant CR_1^{2(n-\tau_i)}\,\mathbb{E}\bigl(U^2\mathbf{1}\{U > c\varepsilon R_1^{\tau_i-n}\}\bigr)\\ &= CR_1^{2(n-\tau_i)}h\bigl(\varepsilon R_1^{\tau_i-n}\bigr) \leqslant CR_1^{2(n-\tau_i)}h\bigl(\varepsilon R_1^{-n}\bigr),\end{aligned} \tag{5.12}$$

Thus,

$$\sum_i \mathbb{E}\bigl(\bigl|\Delta\widehat{M}_{n,i}\bigr|^2\mathbf{1}\{\bigl|\Delta\widehat{M}_{n,i}\bigr| > \varepsilon\} \mid \mathcal{G}_{i-1}\bigr) \leqslant C\sum_{k=0}^{n} B_k R_1^{2(n-k)} h\bigl(\varepsilon R_1^{-n}\bigr). \tag{5.13}$$

Finally, we normalize $\widehat{M}_{n,k}$ and define $\widetilde{M}_{n,k} := m^{-n/2}\widehat{M}_{n,k}$; this yields a martingale $(\widetilde{M}_{n,k})_k$ with conditional quadratic variation

$$\widetilde{V}_n := \sum_i \mathbb{E}\bigl((\Delta\widetilde{M}_{n,i})^2 \mid \mathcal{G}_{i-1}\bigr) = m^{-n}\widehat{V}_n \xrightarrow{\text{a.s.}} \sigma^2(\vec{a})\mathcal{Z}, \tag{5.14}$$

by (5.10), (5.7) and (2.8). Furthermore, by (5.13),

$$\sum_i \mathbb{E}\bigl(\bigl|\Delta\widetilde{M}_{n,i}\bigr|^2\mathbf{1}\{\bigl|\Delta\widetilde{M}_{n,i}\bigr| > \varepsilon\} \mid \mathcal{G}_{i-1}\bigr) \leqslant Ch\bigl(\varepsilon m^{n/2}R_1^{-n}\bigr)m^{-n}\sum_{k=0}^{n} B_k R_1^{2(n-k)}, \tag{5.15}$$

which tends to 0 a.s. as $n \to \infty$, because $(m^{1/2}R_1^{-1})^n \to \infty$ and consequently $h\bigl(\varepsilon m^{n/2}R_1^{-n}\bigr) \to 0$, and

$$m^{-n}\sum_{k=0}^{n} B_k R_1^{2(n-k)} = m^{-n}\sum_{k=0}^{n} B_{n-k}R_1^{2k} = \sum_{k=0}^{n}\frac{B_{n-k}}{m^{n-k}}\Bigl(\frac{R_1^2}{m}\Bigr)^k = O_{\mathrm{a.s.}}(1), \tag{5.16}$$

by (2.8) and $R_1^2 < m$.

The martingales $(\widetilde{M}_{n,i})_i$ thus satisfy a conditional Lindeberg condition, which together with (5.14) implies, by [6, Corollary 3.2], that, using (5.10),

$$M_{n,n}/V_n^{1/2} = \widehat{M}_{n,Z_n}/\widehat{V}_n^{1/2} = \widetilde{M}_{n,Z_n}/\widetilde{V}_n^{1/2} \xrightarrow{\mathrm{d}} N(0,1) \tag{5.17}$$

as $n \to \infty$; furthermore, the limit is mixing. (The fact that we here sum the martingale differences to a stopping time Z_n instead of a deterministic k_n as in [6] makes no difference.) By (4.5) and (5.7), this yields

$$\langle \vec{X}_n, \vec{a}\rangle / Z_n^{1/2} \xrightarrow{\mathrm{d}} N\bigl(0, \sigma^2(\vec{a})\bigr). \tag{5.18}$$

We can evaluate the asymptotic variance $\sigma^2(\vec{a})$ given in (5.7) by

$$\begin{aligned}
\frac{\sigma^2(\vec{a})}{1-m^{-1}} &= \sum_{\ell=0}^{\infty} m^{-\ell}\sum_{i,j=1}^{\ell}\sigma_{ij}\alpha_{\ell-i}\alpha_{\ell-j} \\
&= \sum_{k,p=0}^{\infty}\sum_{i,j=1}^{\ell}\sigma_{ij}\alpha_k\alpha_p \mathbf{1}\{i+k=j+p\} m^{-i-k} \\
&= \sum_{k,p,i,j}\sigma_{ij}\alpha_k\alpha_p \oint_{|z|=m^{-1/2}} z^{i+k}\bar{z}^{j+p}\frac{|\mathrm{d}z|}{2\pi m^{-1/2}} \\
&= \oint_{|z|=m^{-1/2}} \Bigl|\sum_k \alpha_k z^k\Bigr|^2 \sum_{i,j}\sigma_{ij} z^i \bar{z}^j \frac{|\mathrm{d}z|}{2\pi m^{-1/2}}.
\end{aligned} \tag{5.19}$$

Furthermore, for $|z| = m^{-1/2}$ (and any z with $|z| < R^{-1} = r$ and $\widehat{\mu}(z) \neq 1$), by (4.1) and Lemma 3.9 with $\lambda = z^{-1}$,

$$\begin{aligned}
\sum_{k=0}^{\infty}\alpha_k z^k &= \Bigl\langle \sum_{k=0}^{\infty} z^k T^k(\vec{v}), \vec{a}\Bigr\rangle = \bigl\langle (1-zT)^{-1}(\vec{v}), \vec{a}\bigr\rangle \\
&= \frac{1}{(z-1)(1-\widehat{\mu}(z))}\sum_{\ell} a_\ell\bigl(z^\ell - m^{-\ell}\bigr).
\end{aligned} \tag{5.20}$$

By (5.19)–(5.20), $\sigma^2(\vec{a})$ equals the right-hand side in (2.15). Thus, (5.18) shows convergence as in (2.14) for any finite linear combination of $Z_n^{-1/2}X_{n,k}$, and thus joint convergence in (2.14) by the Cramér–Wold device.

Convergence in $L^2(\ell^2_R)$ follows from this and Lemma 5.1 (with a slightly increased R) by a standard truncation argument; we omit the details.

By (2.15), the variable ζ_k is degenerate only if $\Sigma(z) = 0$ for every z with $|z| = m^{-1/2}$, and thus, by (2.12), $\widehat{\Xi}(z) = \widehat{\mu}(z)$ a.s. for every such z, which by (2.1)–(2.2) implies $N_k = \mu_k$ a.s. for every k. □

6. Proof of Theorem 2.2

We assume in this section that $\gamma_* = m^{-1/2}$ and that (2.16) holds. By Lemma 3.4, the spectral radius $r(T) = \gamma_*^{-1} = m^{1/2}$. Lemma 3.7 applies with $\gamma_* = m^{-1/2}$, and thus $\tilde{R} < m^{1/2}$; we may assume $\tilde{R} > R$.

Fix as in Section 4 a real vector $\vec{a} \in \ell^2_{R^{-1}}$, and define, using (3.16),

$$\beta_i = \beta_i(\vec{a}) := \langle P_i(\vec{v}), \vec{a}\rangle = \langle \vec{v}_i, \vec{a}\rangle = \frac{1}{\gamma_i(\gamma_i - 1)\widehat{\mu}'(\gamma_i)} \sum_{k=0}^{\infty} a_k\bigl(\gamma_i^k - m^{-k}\bigr). \tag{6.1}$$

Then, by (4.1) and Lemma 3.7,

$$\alpha_k = O\bigl(\tilde{R}^k\bigr) + \sum_{i=1}^{q} \lambda_i^k \langle P_i(\vec{v}), \vec{a}\rangle = \sum_{i=1}^{q} \beta_i \lambda_i^k + O\bigl(\tilde{R}^k\bigr) = O\bigl(m^{k/2}\bigr). \tag{6.2}$$

Furthermore, the O's in (6.2) hold uniformly in all $\vec{a}$ with $\|\vec{a}\|_{\ell^2_{R^{-1}}} \leqslant 1$, as does every O in this section.

Define also, for $p, t = 1, \dots, q$,

$$\sigma^*_{pt} := \sum_{i,j=1}^{\infty} \sigma_{ij} \lambda_p^{-i} \lambda_t^{-j}, \tag{6.3}$$

and note that, using (4.8), $|\lambda_p| = m^{1/2}$ and $R < m^{1/2}$,

$$\sum_{i,j=1}^{\ell} \sigma_{ij} \lambda_p^{-i} \lambda_t^{-j} = \sigma^*_{pt} + O\Bigl(\sum_{i>\ell, j\geqslant 1} R^{i+j} (m^{1/2})^{-i-j}\Bigr) = \sigma^*_{pt} + O\bigl((R/m^{1/2})^{\ell}\bigr). \tag{6.4}$$

Let

$$s_\ell := \sum_{i,j=1}^{\ell} \sigma_{ij} \alpha_{\ell-i} \alpha_{\ell-j}. \tag{6.5}$$

Then, by (6.2) and symmetry, using again (4.8) and $|\lambda_p| = m^{1/2}$, and (6.4),

$$\begin{aligned} s_\ell &:= \sum_{i,j=1}^{\ell} \sigma_{ij} \sum_{p=1}^{q} \sum_{t=1}^{q} \beta_p \lambda_p^{\ell-i} \beta_t \lambda_t^{\ell-j} + O\Bigl(\sum_{i,j=1}^{\ell} R^{i+j} m^{(\ell-i)/2} \tilde{R}^{\ell-j}\Bigr) \\ &= \sum_{p=1}^{q} \sum_{t=1}^{q} \beta_p \beta_t \lambda_p^{\ell} \lambda_t^{\ell} \sum_{i,j=1}^{\ell} \sigma_{ij} \lambda_p^{-i} \lambda_t^{-j} + O\bigl((m^{1/2}\tilde{R})^{\ell}\bigr) \\ &= \sum_{p=1}^{q} \sum_{t=1}^{q} \beta_p \beta_t \lambda_p^{\ell} \lambda_t^{\ell} \sigma^*_{pt} + O\bigl((m^{1/2}\tilde{R})^{\ell}\bigr). \end{aligned} \tag{6.6}$$

In particular,

$$s_\ell = O\bigl(m^\ell\bigr). \tag{6.7}$$

It follows by (4.7), (6.5), (2.8), (6.7) and (6.6) that, a.s.,

$$\begin{aligned}\frac{V_n}{B_n} &= \sum_{\ell=0}^{n} \frac{B_{n-\ell}}{B_n} s_\ell = \sum_{\ell=0}^{n} m^{-\ell}\bigl(1 + o(1) + O_{\mathrm{a.s.}}(1)\mathbf{1}\{n-\ell < \log n\}\bigr)s_\ell \\ &= \sum_{\ell=0}^{n} m^{-\ell} s_\ell + o(n) = \sum_{\ell=0}^{n} m^{-\ell} \sum_{p=1}^{q}\sum_{t=1}^{q} \beta_p\beta_t\lambda_p^\ell\lambda_t^\ell\sigma^*_{pt} + o(n) \\ &= \sum_{p=1}^{q}\sum_{t=1}^{q} \beta_p\beta_t\sigma^*_{pt}\sum_{\ell=0}^{n}\Bigl(\frac{\lambda_p\lambda_t}{m}\Bigr)^\ell + o(n). \end{aligned} \tag{6.8}$$

Recall that $|\lambda_p| = |\lambda_t| = m^{1/2}$, so $|\lambda_p\lambda_t/m| = 1$. Hence, if $\lambda_t = \bar\lambda_p$, then $\sum_{\ell=0}^{n}\bigl(\lambda_p\lambda_t/m\bigr)^\ell = n+1$, while if $\lambda_t \neq \bar\lambda_p$, then $\sum_{\ell=0}^{n}\bigl(\lambda_p\lambda_t/m\bigr)^\ell = O(1)$. Consequently, (6.8) yields, since $B_n/Z_n \xrightarrow{\text{a.s.}} 1 - m^{-1}$ by (3.1) and (2.8),

$$\begin{aligned}\frac{V_n}{nZ_n} \xrightarrow{\text{a.s.}} \sigma^2(\vec a) &:= \frac{m-1}{m}\sum_{p=1}^{q}\sum_{t=1}^{q}\beta_p\beta_t\sigma^*_{pt}\mathbf{1}\{\lambda_t = \bar\lambda_p\} \\ &= \frac{m-1}{m}\sum_{p=1}^{q}|\beta_p|^2\sum_{i,j=1}^{\infty}\sigma_{ij}\lambda_p^{-i}\bar\lambda_p^{-j} \\ &= \frac{m-1}{m}\sum_{p=1}^{q}|\beta_p|^2\Sigma(\gamma_p). \end{aligned} \tag{6.9}$$

We refine the martingale $(M_{n,k})_k$ to $(\widehat M_{n,k})_k$ as in Section 5, but this time we normalize it to $\widetilde M_{n,k} := (nm^n)^{-1/2}\widehat M_{n,k}$. It follows from (6.9) and (2.8) that the conditional quadratic variation $\widetilde V_n = V_n/(nm^n) \xrightarrow{\text{a.s.}} \sigma^2(\vec a)\mathcal{Z}$, i.e., (5.14) holds also in the present case. Furthermore, if we now let $R_1 := m^{1/2}$, then (5.5) and (5.11)–(5.13) hold, and it follows that (5.15) is modified to

$$\begin{aligned}\sum_i \mathbb{E}\bigl(\bigl|\Delta\widetilde M_{n,i}\bigr|^2\mathbf{1}\{\bigl|\Delta\widetilde M_{n,i}\bigr| > \varepsilon\} \mid \mathcal{G}_{i-1}\bigr) &\leqslant Ch\bigl(\varepsilon n^{1/2}\bigr)\frac{1}{nm^n}\sum_{k=0}^{n}B_k m^{n-k} \\ &= O_{\mathrm{a.s.}}\bigl(h\bigl(\varepsilon n^{1/2}\bigr)\bigr) \xrightarrow{\text{a.s.}} 0. \end{aligned} \tag{6.10}$$

Hence the conditional Lindeberg condition holds in the present case too, and (5.17) follows again by [6, Corollary 3.2], which now by (6.9) and (4.5) yields (mixing)

$$\langle \vec X_n, \vec a\rangle/(nZ_n)^{1/2} \xrightarrow{\text{d}} N\bigl(0, \sigma^2(\vec a)\bigr). \tag{6.11}$$

By (6.9) and (6.1), this proves (2.17)–(2.18).

By (2.18), the variable ζ_k is degenerate only if $\Sigma(\gamma_p) = 0$ for every p, and thus, by (2.12), $\widehat\Xi(\gamma_p) = \widehat\mu(\gamma_p)$ a.s.

As in Section 5, convergence in $L^2(\ell^2_R)$ follows by a standard truncation argument, now using the following lemma (with an increased R); we omit the details. $\square$

Lemma 6.1. *Assume* (A1)–(A6), $\gamma_* = m^{-1/2}$ *and* (2.16). *If* $R < m^{1/2}$, *then*

$$\mathbb{E}\,\|\vec{X}_n\|^2_{\ell^2_R} \leqslant Cnm^n \tag{6.12}$$

and

$$\mathbb{E}\,X^2_{n,k} \leqslant Cnm^n R^{-2k} \tag{6.13}$$

for all $n, k \geqslant 0$.

Proof. By (4.5), (4.7), (6.5), (2.7) and (6.7),

$$\mathbb{E}\langle \vec{X}_n, \vec{a}\rangle^2 = \mathbb{E}\,V_n = \mathbb{E}\sum_{\ell=0}^{n} B_{n-\ell}s_\ell \leqslant Cnm^n, \tag{6.14}$$

uniformly for $\|\vec{a}\|_{\ell^2_{R^{-1}}} \leqslant 1$. Taking $\vec{a} = R^k(\delta_{kj})_j$, we obtain (6.13).

Finally, applying (6.13) with R replaced by some R' with $R < R' < m^{1/2}$,

$$\mathbb{E}\,\|\vec{X}_n\|^2_{\ell^2_R} = \sum_{k=0}^{\infty} R^{2k}\,\mathbb{E}\,X^2_{n,k} \leqslant Cnm^n \sum_{k=0}^{\infty} (R/R')^{2k} = Cnm^n. \tag{6.15}$$

$\square$

7. Proof of Theorem 2.3

Assume now that $\gamma_* < m^{-1/2}$. By Lemma 3.4, the spectral radius $r(T) = \gamma_*^{-1} \geqslant m^{1/2}$. We apply Lemma 3.7, assuming as we may that $\tilde{R} > m^{1/2}$. (Otherwise we increase $\tilde{R}$, keeping $\tilde{R} < \gamma_*^{-1}$.) Hence, by (3.14),

$$T^k(\vec{v}) = T_0^k(\vec{v}) + \sum_{i=1}^{q} \lambda_i^k P_i(\vec{v}) = T_0^k(\vec{v}) + \sum_{i=1}^{q} \lambda_i^k \vec{v}_i. \tag{7.1}$$

Thus, by (3.12),

$$\vec{X}_n = -\sum_{k=0}^{n} W_k (TP_0)^{n-k}(\vec{v}) - \sum_{i=1}^{q}\sum_{k=0}^{n} \lambda_i^{n-k} W_k \vec{v}_i. \tag{7.2}$$

Let, recalling (3.7),

$$\check{U}_i := -\sum_{k=0}^{\infty} \gamma_i^k W_k = -\sum_{k=0}^{\infty} \lambda_i^{-k} W_k = -\sum_{\ell=0}^{\infty}\sum_{j=1}^{\infty} \lambda_i^{-\ell-j} W_{\ell,j}, \tag{7.3}$$

noting that by Lemma 3.1 and $|\gamma_i| = \gamma_* < m^{-1/2}$, the sum converges in L^2 and

$$\Big\|\check{U}_i + \sum_{k=0}^{n} \lambda_i^{-k} W_k\Big\|_2 \leqslant \sum_{k=n+1}^{\infty} C|\lambda_i|^{-k} m^{k/2} \leqslant C\big(\gamma_* m^{1/2}\big)^n. \tag{7.4}$$

Furthermore, by Lemma 3.1 and (3.15), since $\tilde{R} > m^{1/2}$,

$$\Big\|\sum_{k=0}^{n} W_k (TP_0)^{n-k}(\vec{v})\Big\|_{L^2(\ell^2_R)} \leqslant \sum_{k=0}^{n} \|W_k\|_2 \cdot \|(TP_0)^{n-k}(\vec{v})\|_{\ell^2_R} \leqslant C\sum_{k=0}^{n} m^{k/2}\tilde{R}^{n-k} \leqslant C\tilde{R}^n. \tag{7.5}$$

By (7.2), (7.5), (7.4), defining $U_i := \big(\gamma_i(\gamma_i-1)\widehat{\mu}'(\gamma_i)\big)^{-1}\check{U}_i$ so $\check{U}_i\vec{v}_i = U_i\vec{u}_i$ by (3.16),

$$\Big\|\gamma_*^n \vec{X}_n - \sum_{i=1}^{q} \big(\lambda_i/|\lambda_i|\big)^n U_i\vec{u}_i\Big\|_{L^2(\ell^2_R)} \leqslant C\gamma_*^n\tilde{R}^n + \sum_{i=1}^{q}\Big\|\sum_{k=0}^{n} \lambda_i^{-k} W_k\vec{v}_i + \check{U}_i\vec{v}_i\Big\|_{L^2(\ell^2_R)} \leqslant C(\gamma_*\tilde{R})^n + C(\gamma_* m^{1/2})^n \leqslant C(\gamma_*\tilde{R})^n. \tag{7.6}$$

Since $\gamma_*\tilde{R} < 1$, this shows convergence in (2.19) in $L^2(\ell^2_R)$; furthermore, convergence a.s. follows by (7.6) and the Borel–Cantelli lemma.

We have $\mathbb{E}\, U_i = \mathbb{E}\, \check{U}_i = 0$ by (7.3) since $\mathbb{E}\, W_k = 0$ by (3.5)–(3.7). Furthermore, $W_{0,k} = B_{0,k} - \mu_k = N_k - \mu_k$, while $\mathbb{E}\big(W_{n,k} \mid \mathcal{F}_0\big) = 0$ for $n \geqslant 1$ by (3.5); hence by (3.7), $\mathbb{E}\big(W_n \mid \mathcal{F}_0\big) = W_{0,n} = N_n - \mu_n$, and thus

$$\mathbb{E}\big(\check{U}_i \mid \mathcal{F}_0\big) = -\sum_{k=0}^{\infty} \gamma_i^k (N_k - \mu_k) = -\widehat{\Xi}(\gamma_i) + \widehat{\mu}(\gamma_i). \tag{7.7}$$

Hence, U_i is degenerate only if $\widehat{\Xi}(\gamma_i)$ is so. □

8. A stochastic integral calculus

The limit variables ζ_k in Theorems 2.1 and 2.2 can be interpreted as stochastic integrals of certain functions ("*symbols*"); which gives a useful symbolic calculus. There are also some partial related results for Theorem 2.3.

We consider the three cases in Theorems 2.1–2.3 separately.

8.1. The case $\gamma_* > m^{-1/2}$.

Assume throughout this subsection that Theorem 2.1 applies; in particular that $\gamma_* > m^{-1/2}$.

Let ν be the finite measure on the circle $|z| = m^{-1/2}$ given by

$$\mathrm{d}\nu(z) := \frac{m-1}{m}|1-z|^{-2}\,|1-\widehat{\mu}(z)|^{-2}\Sigma(z)\frac{|\mathrm{d}z|}{2\pi m^{-1/2}}, \tag{8.1}$$

and consider an isomorphism $\mathcal{I} : L^2(\nu) \to \mathcal{H}$ of the Hilbert space $L^2(\nu)$ into a Gaussian Hilbert space $\mathcal{H}$, i.e., a Hilbert space of Gaussian random variables; $\mathcal{I}$ can be interpreted as a stochastic integral, see [8, Section VII.2]. We let here $L^2(\nu)$ be the space of complex square-integrable functions, but regard it as a real Hilbert space with the inner product $\langle f, g\rangle_\nu := \operatorname{Re}\int f\bar{g}\,\mathrm{d}\nu$. Then (2.14)–(2.15) can be stated as

$$Z_n^{-1/2} X_{n,k} \xrightarrow{\mathrm{d}} \zeta_k := \mathcal{I}\big(z^k - m^{-k}\big), \tag{8.2}$$

jointly for all $k \geqslant 0$. This yields a convenient calculus for joint limits.

Example 8.1. Let $k, \ell \geqslant 0$. Then, by (2.10),

$$X_{n-\ell,k} = X_{n,k+\ell} - m^{-k} X_{n,\ell} \tag{8.3}$$

and thus, recalling (2.9), jointly for all $k, \ell \geqslant 0$,

$$\begin{aligned} Z_{n-\ell}^{-1/2} X_{n-\ell,k} &\xrightarrow{\mathrm{d}} m^{\ell/2}\bigl(\zeta_{k+\ell} - m^{-k}\zeta_\ell\bigr) = m^{\ell/2}\mathcal{I}\bigl(z^{k+\ell} - m^{-k} z^\ell\bigr) \\ &= \mathcal{I}\bigl((zm^{1/2})^\ell (z^k - m^{-k})\bigr). \end{aligned} \tag{8.4}$$

Denoting this limit by $\zeta_k^{(\ell)}$, we have of course $\zeta_k^{(\ell)} \overset{\mathrm{d}}{=} \zeta_k$, which corresponds to the fact that $|zm^{1/2}|^\ell = 1$ on the support of ν. More interesting is the joint convergence $(Z_n^{-1/2} X_{n,k}, Z_{n-\ell}^{-1/2} X_{n-\ell,k}) \xrightarrow{\mathrm{d}} (\zeta_k, \zeta_k^{(\ell)})$, with covariance

$$\begin{aligned} \operatorname{Cov}\bigl(\zeta_k, \zeta_k^{(\ell)}\bigr) &= \langle z^k - m^{-k}, (zm^{1/2})^\ell (z^k - m^{-k})\rangle_\nu \\ &= \operatorname{Re} \int_{|z| = m^{-1/2}} (zm^{1/2})^\ell |z^k - m^{-k}|^2 \,\mathrm{d}\nu. \end{aligned} \tag{8.5}$$

The measure ν is by (8.1) absolutely continuous on the circle $|z| = m^{-1/2}$. With the change of variables $z = m^{-1/2} e^{\mathrm{i}\theta}$, we have $(zm^{1/2})^\ell = e^{\mathrm{i}\ell\theta}$ and the Riemann–Lebesgue lemma shows that $\operatorname{Cov}\bigl(\zeta_k, \zeta_k^{(\ell)}\bigr) \to 0$ as $\ell \to \infty$, for fixed every k. Roughly speaking, $X_{n-\ell,k}$ and $X_{n,k}$ are thus essentially uncorrelated when ℓ is large, which justifies the claim in Section 2 that there is only a short-range dependence in this case.

Example 8.2. We can define $X_{n,k}$ by (2.10) also for $k < 0$. Then, the calculations in Example 8.1 apply to any $\ell \geqslant 0$ and any $k \geqslant -\ell$. Hence, replacing n by $n + \ell$ in (8.4), for any fixed ℓ,

$$Z_n^{-1/2} X_{n,k} \xrightarrow{\mathrm{d}} \mathcal{I}\bigl((zm^{1/2})^\ell (z^k - m^{-k})\bigr) \tag{8.6}$$

jointly for all $k \geqslant -\ell$. Since the factor $(zm^{1/2})^\ell$ does not depend on k and has absolute value 1, this means (by changing the isomorphism $\mathcal{I}$) that (8.2) holds jointly for all $k \geqslant -\ell$. Since ℓ is arbitrary, this means that (8.2) holds jointly for all $k \in \mathbb{Z}$. Hence, (2.14)–(2.15) extend to all $k \in \mathbb{Z}$, as claimed in Remark 2.4.

Example 8.3. We have, by (2.10),

$$m^{-j} Z_{n+j} - m^{-j-1} Z_{n+j+1} = m^{-j} X_{n+j+1,1}. \tag{8.7}$$

Hence, by Lemma 5.1, for $j \geqslant 0$,

$$\|m^{-j} Z_{n+j} - m^{-j-1} Z_{n+j+1}\|_2 \leqslant C m^{-j+(n+j+1)/2} = C m^{n/2 - j/2}. \tag{8.8}$$

Summing (8.8) for $j \geqslant \ell$ we obtain, recalling (2.8),

$$\|m^{-\ell} Z_{n+\ell} - m^n \mathcal{Z}\|_2 \leqslant C m^{n/2 - \ell/2} \tag{8.9}$$

for $n \geqslant 1$ and $\ell \geqslant 0$. Hence, as $\ell \to \infty$, $m^{-n/2}\bigl(m^{-\ell}Z_{n+\ell} - m^n\mathcal{Z}\bigr) \to 0$ in L^2, and thus in probability, uniformly in n. Since $Z_n/m^n \xrightarrow{\text{a.s.}} \mathcal{Z} > 0$, and thus $\sup_n m^n/Z_n < \infty$ a.s., it follows that, still uniformly in n,

$$Z_n^{-1/2}\bigl(m^{-\ell}Z_{n+\ell} - m^n\mathcal{Z}\bigr) \xrightarrow{\text{p}} 0, \qquad \ell \to \infty. \tag{8.10}$$

Define the random variables

$$Y_{n,\ell} := Z_n^{-1/2}\bigl(Z_n - m^{-\ell}Z_{n+\ell}\bigr) = -Z_n^{-1/2}m^{-\ell}X_{n,-\ell}, \qquad \ell \geqslant 0. \tag{8.11}$$

Then, by (8.2) and Example 8.2, for every fixed ℓ,

$$Y_{n,\ell} \xrightarrow{\text{d}} -m^{-\ell}\zeta_{-\ell} = \mathcal{I}\bigl(1 - m^{-\ell}z^{-\ell}\bigr), \qquad n \to \infty. \tag{8.12}$$

Furthermore, by (8.10), $Y_{n,\ell} \xrightarrow{\text{p}} Z_n^{-1/2}\bigl(Z_n - m^n\mathcal{Z}\bigr)$ as $\ell \to \infty$, uniformly in n. Finally, $|mz| = m^{1/2} > 1$ on the support of ν, and thus $1 - (mz)^{-\ell} \to 1$ in $L^2(\nu)$ as $\ell \to \infty$; hence $\mathcal{I}\bigl(1 - m^{-\ell}z^{-\ell}\bigr) \to \mathcal{I}(1)$ as $\ell \to \infty$, in L^2 and thus in distribution. It follows that we can let $\ell \to \infty$ in (8.12), see [2, Theorem 4.2], and obtain

$$Z_n^{-1/2}\bigl(Z_n - m^n\mathcal{Z}\bigr) \xrightarrow{\text{d}} \mathcal{I}\bigl(1\bigr), \qquad n \to \infty. \tag{8.13}$$

This is jointly with all (8.2), and thus, jointly for all $k \in \mathbb{Z}$,

$$Z_n^{-1/2}\bigl(Z_{n-k} - m^{n-k}\mathcal{Z}\bigr) = Z_n^{-1/2}\bigl(X_{n,k} + m^{-k}(Z_n - m^n\mathcal{Z})\bigr) \xrightarrow{\text{d}} \mathcal{I}\bigl(z^k\bigr). \tag{8.14}$$

Conversely, (8.2) follows immediately from (8.14).

In the Galton–Watson case (Example 2.5), (8.14) is equivalent to the case $q = 0$ of [7, Theorem (2.10.2)].

8.2. **The case $\gamma_* = m^{-1/2}$.** Assume now that Theorem 2.2 applies; thus $\gamma_* = m^{-1/2}$ and (2.16) holds.

In this case, let ν be the discrete measure , with support Γ_{**},

$$\nu := (m-1)\sum_{p=1}^{q} |1 - \gamma_p|^{-2}\,|\widehat{\mu}'(\gamma_p)|^{-2}\Sigma(\gamma_p)\delta_{\gamma_p}, \tag{8.15}$$

and consider an isomorphism I of $L^2(\nu)$ into a Gaussian Hilbert space as above. Then (2.17)–(2.18) can be stated as (8.2), with the normalizing factor changed from $Z_n^{-1/2}$ to $(nZ_n)^{-1/2}$.

With this change of normalization of $X_{n,k}$, all results in the preceding subsection hold, with one exception: The measure ν has finite support, and thus there exists a sequence $\ell_j \to \infty$ such that $(zm^{1/2})^{\ell_j} \to 1$ as $j \to \infty$ for every $z \in \operatorname{supp}(\nu) = \Gamma_{**}$; hence (8.5) implies $\limsup_{\ell\to\infty} \operatorname{Corr}\bigl(\zeta_k, \zeta_k^{(\ell)}\bigr) = 1$. Hence, although the convergence in (2.18) is mixing, so there is no dependence on the initial generations as in the case $\gamma_* < m^{-1/2}$, there is a dependence over longer ranges than in the case $\gamma_* > m^{-1/2}$.

Furthermore, each ζ_k now belongs to the (typically q-dimensional) space spanned by $\zeta_1, \dots, \zeta_q$, which yields the linear dependence of the limits ζ_k claimed in Section 2.

Example 8.4. In the simplest case, $\Gamma_{**} = \{-m^{1/2}\}$. (See Example 2.6 for an example.) Then $\zeta_k = \bigl((-1)^k m^{-k/2} - m^{-k}\bigr)\zeta$ for some $\zeta \sim N\bigl(0, \nu\{-m^{1/2}\}\bigr)$ and all $k \in \mathbb{Z}$.

Furthermore, $zm^{1/2} = -1$ on $\operatorname{supp}\nu$, and thus (8.4) yields $\zeta_k^{(\ell)} = (-1)^\ell \zeta_k$; in particular, $\zeta_k^{(\ell)} = \zeta_k$ for every even ℓ.

8.3. **The case $\gamma_* < m^{-1/2}$.** In this case, there is no limit, but we can argue with the components of the approximating sum in (2.19) in the same way as with ζ_k in Examples 8.1–8.2, and draw the conclusion that (2.19), interpreted component-wise, extends also to $k < 0$, as claimed in Remark 2.4. We omit the details.

9. Random characteristics

A random characteristic is a random function $\chi(t) : [0, \infty) \to \mathbb{R}$ defined on the same probability space as the prototype offspring process Ξ; we assume that each individual x has an independent copy (Ξ_x, χ_x) of (Ξ, χ), and interpret $\chi_x(t)$ as the characteristic of x at age t. We consider as above the lattice case, and define, denoting the birth time of x by τ_x,

$$Z_n^\chi := \sum_{x:\tau_x \leqslant n} \chi_x(n - \tau_x), \tag{9.1}$$

the total characteristic of all individuals at time n. See further Jagers [7].

We assume in this section (A1)–(A6), and also that there exists $R_2 < m^{1/2}$ such that for some $C < \infty$

$$\mathbb{E}[\chi(k)^2] \leqslant CR_2^{2k}, \qquad k \geqslant 0. \tag{9.2}$$

We assume below that R is chosen with $R_2 < R < m^{-1}$. We define

$$\lambda_k^\chi := \mathbb{E}\,\chi(k), \qquad k \geqslant 0, \tag{9.3}$$

and $\lambda_k^\chi := 0$ for $k < 0$, and also

$$\lambda^\chi := \sum_{k=0}^\infty \bigl(m^{-k} - m^{-k-1}\bigr)\lambda_k^\chi. \tag{9.4}$$

Note that the sum in (9.4) converges absolutely since (9.2) implies

$$|\lambda_k^\chi| = |\,\mathbb{E}\,\chi(k)| \leqslant CR_2^k. \tag{9.5}$$

We split the characteristic into its mean $\lambda_k^\chi = \mathbb{E}\,\chi(k)$ and the centered part

$$\tilde\chi(k) := \chi(k) - \mathbb{E}\,\chi(k) = \chi(k) - \lambda_k^\chi. \tag{9.6}$$

Simple calculations, see [10], yield the decomposition

$$Z_n^\chi - \lambda^\chi Z_n = Z_n^{\tilde\chi} + \sum_{k=1}^n \bigl(\lambda_k^\chi - \lambda_{k-1}^\chi\bigr) X_{n,k} = Z_n^{\tilde\chi} + \langle X_n, \Delta\vec\lambda^\chi \rangle, \tag{9.7}$$

with $\Delta\vec{\lambda}^\chi = (\lambda^\chi_k - \lambda^\chi_{k-1})_k$. Here $\Delta\vec{\lambda}^\chi \in \ell^2_{R^{-1}}$ by (9.5), and thus the asymptotic behaviour of $\langle X_n, \Delta\vec{\lambda}^\chi\rangle$ is given by Theorems 2.1–2.3.

In particular, if $\gamma_* > m^{-1/2}$, then Theorem 2.1 applies and $\langle X_n, \Delta\vec{\lambda}^\chi\rangle/\sqrt{Z_n}$ is asymptotically normal. Moreover, it is shown in [10] that in this case also $Z^{\tilde\chi}_n/\sqrt{Z_n}$ is asymptotically normal, and that the joint distribution is asymptotically normal, so $(Z^\chi_n - \lambda^\chi Z_n)/\sqrt{Z_n} \xrightarrow{\mathrm{d}} N(0,\sigma^2_\chi)$, with σ^2_χ given by an explicit but rather complicated formula in [10, Theorem 6.1].

The proof in [10] of asymptotic normality of $Z^{\tilde\chi}_n$ does not require $\gamma_* > m^{-1/2}$; it gives the following result that is valid in all three cases. (Note that the assumption $\mathbb{E}\,\chi(k) = 0$ is equivalent to $\chi = \tilde\chi$.)

Theorem 9.1. *Assume* (A1)–(A6) *and* (9.2). *If* $\mathbb{E}\,\chi(k) = 0$ *for every* $k \geqslant 0$*, then as* $n \to \infty$*,*

$$Z_n^{-1/2} Z^\chi \xrightarrow{\mathrm{d}} \zeta^\chi, \tag{9.8}$$

for some normal random variable ζ^χ *with mean* $\mathbb{E}\,\zeta^\chi = 0$ *and variance*

$$\operatorname{Var}\bigl(\zeta^\chi\bigr) = \frac{m-1}{m}\sum_{k=0}^\infty m^{-k}\operatorname{Var}\bigl(\chi(k)\bigr). \tag{9.9}$$

Proof. See [10, Section 6]. □

It remains to consider the case when $\lambda^\chi_k = \mathbb{E}\,\chi(k) \neq 0$ for some k. If $\gamma_* > m^{-1/2}$, then, as said above, asymptotic normality of $(Z^\chi_n - \lambda^\chi Z_n)/\sqrt{Z_n}$ is shown in [10]. If $\gamma_* = m^{-1/2}$ and (2.16) holds, then Theorem 2.2 shows that $\langle X_n, \Delta\vec{\lambda}^\chi\rangle/\sqrt{nZ_n} \xrightarrow{\mathrm{d}} N(0,\sigma^2)$, where σ^2 is given by (2.18) and $\sigma^2 > 0$ except in degenerate cases. Since Theorem 9.1 implies that $Z^{\tilde\chi}_n/\sqrt{nZ_n} \xrightarrow{\mathrm{p}} 0$, it follows from (9.7) that $(Z^\chi_n - \lambda^\chi Z_n)/\sqrt{nZ_n} \xrightarrow{\mathrm{d}} N(0,\sigma^2)$. Similarly, if $\gamma_* < m^{-1/2}$, then Theorem 9.1 implies $\gamma_*^n Z^{\tilde\chi}_n \xrightarrow{\mathrm{p}} 0$, and (9.7) shows that $Z^\chi_n - \lambda^\chi Z_n$ has the same (oscillating) asymptotic behaviour as $\langle X_n, \Delta\vec{\lambda}^\chi\rangle$, given by Theorem 2.3.

Summarizing, if $\gamma_* \leqslant m^{-1/2}$, then the randomness in the characteristic χ only gives an effect of smaller order than the mean $\mathbb{E}\,\chi$, and unless the mean vanishes (or the limits degenerate), Z^χ_n has the same asymptotic behaviour as if χ is replaced by the deterministic $\mathbb{E}\,\chi$, which is treated by Theorems 2.2 and 2.3.

Example 9.2. We have in the present paper for simplicity assumed (A4), that there are no deaths. Suppose now, more generally, that each individual has a random lifelength $\ell \leqslant \infty$, as usual with i.i.d. copies (Ξ_x, ℓ_x) for all individuals x. The results in Section 2 apply if we ignore deaths and let Z_n denote the number of individuals born up to time n, living or dead. Moreover, the number of living individuals at time n is Z^χ_n, for the characteristic $\chi(k) := \mathbf{1}\{\ell > k\}$.

Similarly, for example, the number of living individuals at time $n - j$ is $Z^{\chi_j}_n$ with $\chi_j(k) := \mathbf{1}\{\ell > k - j \geqslant 0\}$. The analogue of $X_{n,j}$ in (2.10) but counting only living individuals is thus given by $Z_n^{\chi_j - m^{-j}\chi}$, and results extending Theorems 2.1–2.3 without assuming (A4) follow. We leave the details to the reader.

Acknowledgement

Acknowledgement. I thank Peter Jagers and Olle Nerman for helpful comments.

References

[1] Søren Asmussen and Heinrich Hering, *Branching Processes.* Birkhäuser, Boston, 1983.

[2] Patrick Billingsley, *Convergence of Probability Measures.* Wiley, New York, 1968.

[3] Harald Bohr, Zur Theorie der Fastperiodischen Funktionen. I–III. *Acta Math.* **45** (1925), no. 1, 29–127; **46** (1925), no. 1-2, 101–214; **47** (1926), no. 3, 237–281.

[4] Nelson Dunford and Jacob T. Schwartz, *Linear Operators. I. General Theory.* Interscience Publishers, New York, 1958.

[5] Peter L. Duren, *Theory of H^p Spaces.* Academic Press, New York, 1970.

[6] P. Hall and C. C. Heyde, *Martingale Limit Theory and its Application.* Academic Press, New York, 1980.

[7] Peter Jagers, *Branching Processes with Biological Applications.* John Wiley & Sons, London, 1975.

[8] Svante Janson, *Gaussian Hilbert Spaces.* Cambridge Univ. Press, Cambridge, UK, 1997.

[9] Svante Janson, Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stoch. Process. Appl.* **110** (2004), 177–245.

[10] Svante Janson, Asymptotic normality in Crump–Mode–Jagers processes: the lattice case. Preprint, 2017. `arXiv:1711.06514`

[11] Olle Nerman, On the convergence of supercritical general (C-M-J) branching processes. *Z. Wahrsch. Verw. Gebiete* **57** (1981), no. 3, 365–395.

[12] Nicolas Pouyanne, An algebraic approach to Pólya processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** (2008), no. 2, 293–323.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

E-mail address: `svante.janson@math.uu.se`

URL: `http://www2.math.uu.se/~svante/`