Chapter 3

Branching Processes I: Supercritical growth and population structure

The fundamental characteristic of biological populations is that individuals undergo birth and death and that individuals carry information passed on from their parents at birth. Furthermore there is a randomness in this process in that the number of births that an individual gives rise to is in general not deterministic but random. Branching processes model this process under simplifying assumptions but nevertheless provide the starting point for the modelling and analysis of such populations. In this chapter we present some of the central ideas and key results in the theory of branching processes.

3.1 Basic Concepts and Results on Branching Processes

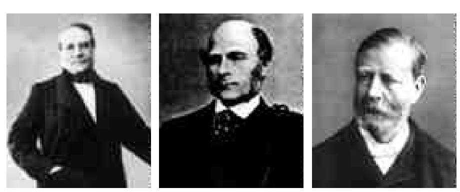

Figure 3.1: Bienamyé, Galton and Watson

3.1.1 Bienamyé-Galton-Watson processes

The Bienamyé-Galton-Watson branching process (BGW process) is a Markov chain on $\mathbb{N}_0 := \{0, 1, 2, \dots\}$. The discrete time parameter is interpreted as the generation number and X_n denotes the number of individuals alive in the n'th generation. Generation $(n+1)$ consists of the offspring of the nth generation as follows:

- each individual i in the nth generation produces a random number ξ_i with distribution
$$p_k = P[\xi_i = k],\ k \in \mathbb{N}_0$$
- $\xi_1, \xi_2, \ldots, \xi_{X_n}$ are independent.

Let $X_0 = 1$. Then for $n \geq 0$

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_i, \quad \{\xi_i\} \text{ independent}$$

We assume that the mean number of offspring

$$m = \sum_{i=1}^{\infty} i p_i < \infty.$$

The BGW process is said to be *subcritical* if $m < 1$, *critical* if $m = 1$ and *supercritical* if $m > 1$.

A basic tool in the study of branching processes is the *generating function*

$$(3.1)\quad f(s) = E[s^{\xi}] = \sum_{k=0}^{\infty} p_k s^k, \quad 0 \leq s \leq 1.$$

Then

$$(3.2)\quad f'(1) = m, \quad f''(1) = E[\xi(\xi-1)] \geq 0.$$

Let

$$f_n(s) = E[s^{X_n}],\ n \in \mathbb{N}.$$

Then conditioned on X_n, and using the independence of the $\{\xi_i\}$,

$$f_{n+1}(s) = E[s^{\sum_{i=1}^{X_n} \xi_i}] = E[f(s)^{X_n}] = f_n(f(s)) = f(f_n(s)).$$

Note that $f(0) = P[\xi = 0] = p_0$ and

$$P[X_{n+1} = 0] = f(f_n(0)) = f(P[X_n = 0])$$

Then if $m > 1,\ p_0 > 0,\ P[X_n = 0] = f_n(0) \uparrow q$ where q is the smallest nonnegative root of

$$f(s) = s,$$

and if $m \leq 1$, $P[X_n = 0] \uparrow 1$. Note that 1 and q are the only roots of $f(s) = s$.

Since $E[X_{n+1}|X_n] = mX_n$,

$$(3.3)\quad W_n := \frac{X_n}{m^n} \text{ is a martingale and } lim_{n\to\infty} W_n = W \text{ exists } a.s.$$

Proposition 3.1 *We have $P[W=0]=q$ or 1, that is, conditioned on nonextinction either $W=0$ a.s. or $W>0$ a.s.*

Proof. It suffices to show that $P[W=0]$ is a root of $f(s)=s$. The ith individual of the first generation has a descendant family with a martingale limit which we denote by $W^{(i)}$. Then $\{W^{(i)}\}_{i=1,\dots,X_1}$ are independent and have the same distribution as W. Therefore

$$(3.4)\quad W=\frac{1}{m}\sum_{i=1}^{X_1}W^{(i)}$$

and therefore $W=0$ if and only if for all $i\le X_1$, $W^{(i)}=0$. Conditioning on X_1 implies that

$$(3.5)\quad P[W=0]=E(P(W^{(i)}=0)^{X_1})=f(P[W=0]).$$

Therefore $P[W=0]$ is a root of $f(s)=s$. ■

Remark 3.2 *In the case* $\mathrm{Var(X_1)}=\sigma^2<\infty$ *we can show by induction that*

$$(3.6)\quad \mathrm{Var(X_n)}=\begin{array}{l}\dfrac{\sigma^2m^n(m^n-1)}{m^2-m},\quad m\neq 1,\\ =n\sigma^2,\quad m=1\end{array}$$

Then if $m>1$ the martingale $\frac{X_n}{m^n}$ is uniformly integrable and $E(W)=1$. Moreover $\frac{X_n}{m^n}\to W$ in L^2 and

$$(3.7)\quad \mathrm{Var(W)}=\frac{\sigma^2}{\mathrm{m^2-m}}>0\qquad\text{(see Harris [292] Theorem 8.1).}$$

If $m>1$, $\sigma^2=\infty$, a basic question concerns the nature of the random variable W and the question whether or not $\frac{X_n}{m^n}\to W$ in L^1. The question was settled by a celebrated result of Kesten and Stigum which we present in Theorem 3.6 below. We first introduce some further basic notions.

Bienamyé-Galton-Watson process with immigration (BGWI)

The Bienamyé-Galton-Watson process with offspring distribution $\{p_k\}$ and immigration process $\{Y_n\}_{n\in\mathbb{N}_0}$ satisfies

$$(3.8)\quad X_{n+1}=\sum_{i=1}^{X_n}\xi_i+Y_{n+1},$$

where the ξ_i are iid with distribution $\{p_k\}$.

Let $\mathcal{F}^Y$ be the σ-field generated by $\{Y_k : k \geq 1\}$ and $X_{n,k}$ be the number of descendants at generation n of the individuals who immigrated in generation k. Then the total number of individuals in generation n is $X_n = \sum_{k=1}^n X_{n,k}$.

For $k < n$ the random variable $W_{n,k} = X_{n,k}/m^{n-k}$ is the has the same law as $\widetilde{X}_{n-k}/m^{n-k}$ where $\widetilde{X}_n$ is the BGW process with Y_k initial particles. Therefore

$$(3.9)\quad E[\frac{X_{n,k}}{m^{n-k}}] = Y_k.$$

Now consider the subcritical case $m < 1$. If $\{Y_i\}$ are i.i.d. with $E[Y_i] < \infty$, then the Markov chain X_n has a stationary measure with mean $\frac{E[Y]}{1-m}$.

Next consider the supercritical case $m > 1$. Then

$$(3.10)\quad E[\frac{X_n}{m^n}|\mathcal{F}^Y] = E[\frac{1}{m^n}\sum_{k=1}^n X_{n,k}|\mathcal{F}^Y] = \sum_{k=1}^n \frac{1}{m^k}E[\frac{X_{n,k}}{m^{n-k}}|\mathcal{F}^Y] = \sum_{k=1}^n \frac{Y_k}{m^k}.$$

If $\sup_k E[Y_k] < \infty$, then

$$(3.11)\quad \lim_{n\to\infty} \frac{E[X_n]}{m^n} = \sum_{k=1}^\infty \frac{E[Y_k]}{m^k} < \infty.$$

A dichotomy in the more subtle case $E[Y_k] = \infty$ is provided by the following theorem of Seneta.

Theorem 3.3 *(Seneta (1970) [557]) Let X_n denote the BGW process with mean offspring $m > 1$, $X_0 = 0$ and with i.i.d. immigration process Y_n.*
(a) If $E[\log^+ Y_1] < \infty$, then $\lim \frac{X_n}{m^n}$ exists and is finite a.s.
(b) If $E[\log^+ Y_1] = \infty$, then $\limsup \frac{X_n}{c^n} = \infty$ for every constant $c > 0$.

Proof. The theorem is a consequence of the following elementary result.

Lemma 3.4 *Let $Y, Y_1, Y_2, \ldots$ be nonnegative iid rv. Then a.s.*

$$(3.12)\quad \limsup_{n\to\infty} \frac{1}{n}Y_n = \begin{cases} 0, & \text{if } E[Y] < \infty \\ \infty, & \text{if } E[Y] = \infty \end{cases}$$

Proof. Recall that $E[Y] = \int_0^\infty P(Y > x)dx$. This gives $\sum_n P(\frac{Y}{n} > c) < \infty$ for any $c > 0$ if $E[Y] < \infty$ and the result follows by Borel-Cantelli. If $E[Y] = \infty$, then $\sum P(\frac{Y}{n} > c) = \infty$ for any $c > 0$ and the result follows by the second Borel-Cantelli Lemma since the Y_n are independent. ■

Proof of (a). By (3.10)

$$(3.13)\quad E[\frac{X_n}{m^n}|\mathcal{F}^Y] = \sum_{k=1}^n \frac{Y_k}{m^k}.$$

Since here we assume $E[\log^+ Y_1] < \infty$, Lemma 3.4 gives $\limsup_{k\to\infty} \frac{Y_k}{c^k} < \infty$ for any $c > 0$. Therefore the series given by the last expression in (3.13) converges a.s. and therefore $\lim_{n\to\infty} E[\frac{X_n}{m^n}|\mathcal{F}^Y]$ exists and is finite a.s. This implies (a)

Proof of (b). If $E[\log^+ Y_1] = \infty$, then by Lemma 3.4 $\limsup_{n\to\infty} \frac{\log^+ Y_n}{n} = \infty$ a.s. Therefore for any $c > 0$

$$(3.14)\quad \limsup_{n\to\infty} \frac{Y_n}{c^n} = \infty$$

a.s. Since $X_n \geq Y_n$, (b) follows.

■

3.1.2 Bienamyé-Galton-Watson trees

In addition to the keeping track of the total population of generation $n+1$ in a BGW process it is useful to incorporate genealogical information, for example, which individuals in generation $n+1$ have the same parent in generation n. This leads to a natural family tree structure which was introduced in the papers of Joffe and Waugh (1982), (1985), [347], [348] in their determination of the distribution of kin numbers and developed in the papers of Chauvin (1986) [72] and Neveu (1986) [482].

A convenient representation of the BGW random tree is as follows. Let $u = (i_1, \ldots, i_n)$ denote an individual in generation n who is the i_nth child of the i_{n-1}-th child of ... of the i_1-th child of the ancestor, denoted by $\emptyset$. The *space of individuals* (vertices) is given by

$$(3.15)\quad \mathcal{I} = \{\emptyset\} \cup \cup_{n=1}^{\infty} \mathbb{N}^n.$$

Given $u = (u_1, \ldots, u_m),\ v = (v_1, \ldots, v_n)\ \in \mathcal{I}$, we denote the composition by $uv := (u_1, \ldots, u_m, v_1, \ldots, v_n)$

A *plane rooted tree* $\mathcal{T}$ with root $\emptyset$ is a subset of $\mathcal{I}$ such that

1. $\emptyset \in \mathcal{T}$,
2. If $v \in \mathcal{T}$ and $v = uj$ for some $u \in \mathcal{I}$ and $j \in \mathcal{I}$, then $u \in \mathcal{T}$,
3. For every $u \in \mathcal{T}$, there exists a number $k_u(\mathcal{T}) \geq 0$, such that $uj \in \mathcal{T}$ if and only if $1 \leq j \leq k_u(\mathcal{T})$.

A plane tree can be given the structure of a graph in which a parent is connected by an edge to each of its offspring.

Let $\mathbf{T}$ be the set of all plane trees. If $t \in \mathbf{T}$ let $[t]_n$ be the set of rooted trees whose first n levels agree with those of t. Let $\mathbf{V}$ denote the set of connected sequences in $\mathcal{I}$, $\varnothing, v_1, v_2, \ldots,$ which do not backtrack. Given $t \in \mathbf{T}$, let $V(t)$ denote the set of paths in t. If v_n is a vertex at the nth level, let $[t; v]_n$ denote the set of trees with distinguished paths such that the tree is in $[t]_n$ $v \in V(t)$ and the path goes through v_n.

Given a finite plane tree $\mathcal{T}$ the *height* $h(\mathcal{T})$ is the maximal generation of a vertex in $\mathcal{T}$ and $\#(\mathcal{T})$ denotes the number of vertices in $\mathcal{T}$. Let $\mathbf{T}_n$ be the set of trees of height n.

A *random tree* is given by a probability measure on $\mathbf{T}$. Given an offspring distribution $\mathcal{L}(\xi) = \{p_k\}_{k\in\mathbb{N}}$, the corresponding BGW tree is constructed as follows:

Let the initial individual be labelled $\emptyset$. Give it a random number of children denoted $1, 2, \ldots, \xi_\emptyset$.

Then each of these has a random number of children, for example i has children denoted $(i, 1), \ldots, (i, \xi_i)$ etc. Each of these has children, for example (i, j) has $\xi_{i,j}$ children labelled $(i, j, 1), \ldots, (i, j, \xi_{i,j})$, etc. Then considering the first n generations in this way we obtain a probability measure P_n^{BGW} on $\mathbf{T}_n$.

The probability measures, P_n^{BGW} form a consistent family and induce a probability measure P^{BGW} on $\mathbf{T}$, the law the BGW random tree.

Let

$$Z_n = \text{ number of vertices in the tree at level } n. \tag{3.16}$$

Then by the construction it follows that Z_n is a version of the BGW process and we can think of the BGW tree as an enriched version of the BGW process.

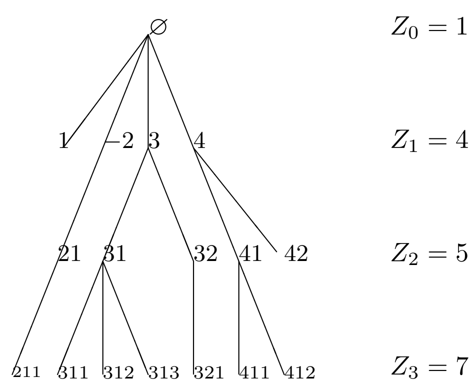

Figure 3.2: BGW Tree

The size-biased BGW tree

The fundamental notion of size-biasing has many applications. It will be used below in the proof of Lyons, Pemantle and Peres (1995) [438] of some basic results on Bienamyé-Galton-Watson processes (see Theorem 3.6 below).

To exploit this notion for branching processes we consider the *size-biased offspring distribution*

(3.17) $\widehat{p}_k = \frac{kp_k}{m}, \quad k = 1, 2, \dots.$

We denote by $\widehat{\xi}$ a random variable having the size biased offspring distribution. The size-biased BGW tree $\widehat{T}$ is constructed as follows:

- the initial individual is labelled $\varnothing$; $\varnothing$ has a random number $\widehat{\xi}_\varnothing$ of children (with the size-biased offspring distribution) $\widehat{p}$,
- one of the children of $\varnothing$ is selected at random and denoted v_1 and given an independent size-biased number $\widehat{\xi}_{v_1}$ of children,
- the other children of $\varnothing$ are independently assigned ordinary BGW descendant trees with offspring number ξ,
- again one of the children of v_1 is selected at random and denoted v_2 and given an independent size-biased number $\widehat{\xi}_{v_2}$ of children,
- this process is continued and produces the size-biased BGW tree $\widehat{T}$ which is *immortal* and infinite distinguished path v which we call the *backbone*.

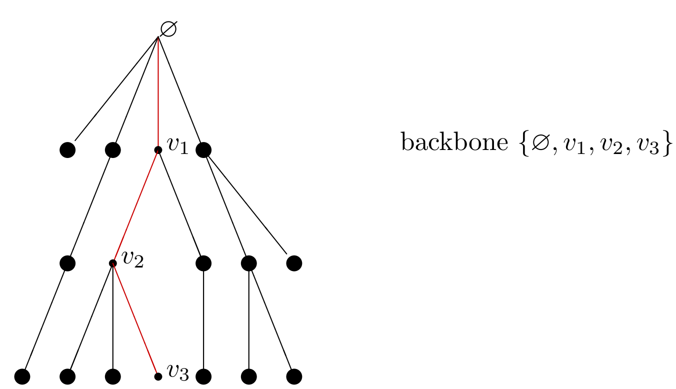

Figure 3.3: Size-biased BGW Tree

Define the measure $\bar{P}_*^{BGW} \in \mathcal{P}(\mathbf{T} \times \mathbf{V})$ to be the joint distribution of the random tree $\widehat{T}$ and backbone $\{v_0, v_1, v_2, \dots\}$. Let $\bar{P}^{BGW}$ denote the marginal distribution of $\widehat{T}$. We can view the vertices off the backbone $(v_0, v_1, \dots)$ of the size-biased tree as a branching process with immigration in which the immigrants are the siblings of the individuals on the backbone. The distribution of the number of immigrants at generation n, Y_n, is given by the law $\widehat{\xi} - 1$.

Given a tree t let $[t]_n$ denote the tree restricted to generations $1, \dots, n$. Let $Z_n(t)$ denote the number of vertices in the tree at the nth level (generation) and $\mathcal{F}_n = \sigma([t]_n)$. Let

$$(3.18)\quad W_n(t) := \frac{Z_n(t)}{m^n}$$

denote the martingale associated to a tree t with $Z_n(t)$ vertices at generation n.

Lemma 3.5 *(a) The Radon-Nikodym derivative of the marginal distribution $\bar{P}^{BGW}|_{\mathcal{F}_n}$ of $\widehat{T}$ with respect to $P^{PGW}|_{\mathcal{F}_n}$ is given by*

$$(3.19)\quad \frac{d\bar{P}_n^{BGW}}{dP_n^{BGW}}(t) = W_n(t).$$

(b) Under the measure $\bar{P}_*^{BGW}$, the vertex v_n at the nth level of the tree $\widehat{T}$ in the random path $(v_0, v_1, \dots)$ is uniformly distributed on the vertices at the nth level of $\widehat{T}$.

Proof.

We will verify that

$$(3.20)\quad \bar{P}_*^{BGW}[t, v]_n = \frac{1}{m^n} P^{BGW}[t]_n$$

and therefore

$$(3.21)\quad \bar{P}^{BGW}[t]_n = W_n(t) P^{BGW}[t]_n.$$

First observe that the

$$(3.22)\quad \begin{aligned} \bar{P}_*^{BGW}(Z_1 = k, v_1 = i) &= \frac{kp_k}{m} \cdot \frac{1}{k} \\ &= \frac{p_k}{m} = \frac{1}{m} P(\xi = k), \text{ for } i = 1, \dots, k. \end{aligned}$$

since v_1 is randomly chosen from the offspring $(1, \dots, \widehat{\xi}_\varnothing)$.

Now consider $[\widehat{T}, v]_{n+1}$. We can construct this by first selecting $\widehat{\xi}_0$ and v_1 and then following the next n generations of the resulting descendant tree and backbone as well as the BGW descendant trees of the remaining $\widehat{\xi}_0 - 1$ vertices in the first generation. If $\widehat{\xi}_\varnothing(t) = k$ we denote the resulting descendant trees by $t^{(1)}, t^{(2)}, \dots, t^{(k)}$.

Let $v_{n+1}(t)$ be a vertex (determined by a position in the lexicographic order) at level $n + 1$. It determines $v_1(t)$ and the descendant tree $t^{(v_1)}$ that it belongs to. If $\widehat{\xi}_\varnothing(t) = k,\ v_1(t) = i$, then we obtain

$$\text{(3.23)}\quad \bar{P}_*^{BGW}[t;v]_{n+1} = \frac{p_k}{m} \cdot \bar{P}_*^{BGW}[t^{(i)};v_{n+1}]_n \cdot \prod_{j=1,j\neq i}^{k} P^{BGW}[t^{(j)}]_n.$$

Then by induction for each n

$$\text{(3.24)}\quad \bar{P}_*^{BGW}[t;v]_n = \frac{1}{m^n} P^{BGW}[t]_n$$

for each of the $Z_n(t)$ positions v in the lexicographic order at level n and $[t]_n$. Consequently we have obtained the *martingale change of measure*

$$\text{(3.25)}\quad \bar{P}_*^{BGW}[t]_n = \frac{Z_n(t)}{m^n} P^{BGW}[t]_n$$

and

$$\text{(3.26)}\quad \bar{P}_*^{BGW}[v = i|t] = \frac{1}{Z_n(t)} \text{ for } i = 1, \dots, Z_n(t).$$

■

For an infinite tree t we define

$$\text{(3.27)}\quad W(t) := \limsup_{n\to\infty} W_n(t).$$

Note that in the critical and subcritical cases the measures P^{BGW} and $\bar{P}^{BGW}$ are singular since the P^{BGW}- probability of nonextinction is zero. The question as to whether or not they are singular in the supercritical case will be the focus of the next subsection.

3.1.3 Supercritical branching

As mentioned above if $0 < m < \infty$, then under P^{BGW}

$$\text{(3.28)}\quad W_n = \frac{Z_n}{m^n}$$

is a martingale and converges to a random variable W a.s. as $n \to \infty$. The characterization of the limit W in the supercritical case, $m > 1$, under minimal conditions was obtained in the following theorem of Kesten and Stigum (1966) [365]. The proof given below follows the "conceptual proof" of Lyons, Pemantle and Peres (1995) [438].

Theorem 3.6 *(Kesten-Stigum (1966) [365]) Consider the BGW process with offspring ξ and mean offspring size m. If $1 < m < \infty$, the following are equivalent*

(i) $P^{BGW}[W = 0] = q$

(ii) $E^{BGW}[W] = 1$

(iii) $E[\xi \log^+ \xi] < \infty$

Proof. By Lemma 3.5

$$\frac{d\bar{P}_n^{BGW}}{dP_n^{BGW}}(t) = W_n(t) \tag{3.29}$$

where the left side denotes the Radon-Nikodym derivative wrt $\mathcal{F}_n = \sigma([t]_n)$.

Note that $P^{BGW}(W = 0) \geq q$ where $q = P^{BGW}(E_0)$ where $E_0 := \{Z_n = 0 \text{ for some } n < \infty\}$ (extinction probability). Moreover, since $\mathcal{F}_n \uparrow \mathcal{F} = \sigma(t)$, we have the Radon-Nikodym dichotomy (see Theorem 16.12)

$$W = 0, \quad P^{BGW} - a.s. \quad \Leftrightarrow P^{BGW} \perp \bar{P}^{BGW} \quad \Leftrightarrow W = \infty \quad \bar{P}^{BGW} - a.s. \tag{3.30}$$

and

$$\int W dP^{BGW} = 1 \quad \Leftrightarrow \bar{P}^{BGW} \ll P^{BGW} \quad \Leftrightarrow W < \infty \quad \bar{P}^{BGW} - a.s. \tag{3.31}$$

Now recall (3.13) that the size-biased tree can be represented as a branching process with immigration in which the distribution of the number of immigrants at generation n, Y_n, is given by the law $\widehat{\xi} - 1$, that is

$$E[Z_n|\mathcal{Y}] = \sum_{k=1}^{n} \frac{Y_k}{m^k}. \tag{3.32}$$

If $E[\log^+ \widehat{\xi}] = E[\xi \log^+ \xi] = \sum_{k=1}^{\infty} kp_k \log k = \infty$, then

$$W = \lim_{n\to\infty} \frac{Z_n}{m^n} = \infty, \quad \bar{P}^{BGW} a.s. \tag{3.33}$$

by Theorem 3.3 (b). Therefore $P^{BGW}(W = 0) = 1$ by (3.30).

If $E[\xi \log^+ \xi] < \infty$, then $E[\log^+ \widehat{\xi}] = \sum_{k=1}^{\infty} kp_k \log k < \infty$. and by Theorem 3.3(a)

$$\lim_{n\to\infty} E(\frac{Z_n}{m^n}|\mathcal{Y}) = \sum_{k=1}^{\infty} \frac{Y_k}{m^k} < \infty, \quad \bar{P}^{BGW} a.s. \tag{3.34}$$

and therefore

$$W = \lim_{n\to\infty} \frac{Z_n}{m^n} < \infty, \bar{P}^{BGW} - a.s. \tag{3.35}$$

Then $E^{BGW}[W] = \int W dP^{BGW} = 1$ by (3.31).

Finally, since by Proposition 3.1 $P^{BGW}(W = 0) = q$ or 0, we obtain (i).

■

Remark 3.7 *The supercritical branching model is the basic model for a growing population with unlimited resources. A more realistic model is a spatial model in which resources are locally limited but the population can grow by spreading spatially. A simple deterministic model of this type is the Fisher-KPP equation. We will consider the analogous spatial stochastic models in a later chapter.*

3.1.4 The general branching model of Crump-Mode-Jagers

We now consider a far-reaching generalization of the Bienamyé-Galton-Watson process known as a Crump-Mode Jagers (CMJ) process ([112], [338]). This is a process with time parameter set $[0, \infty)$ consisting of finitely many individuals at each time.

With each individual x we denote its birth time τ_x, lifetime λ_x and reproduction process ξ_x. The latter is a point process which gives the sequence of birth times of individuals. $\xi_x(t)$ is the number of offspring produced (during its lifetime) by an individual x born at time 0 during $[0, t]$. The intensity of ξ_x, called the *reproduction function* is defined by

(3.36) $\mu(t) = E[\xi(t)]$.

The lifetime distribution is defined by

(3.37) $L(u) = P[\lambda \leq u]$.

We begin with one individual $\emptyset$ which we assume is born at time $\tau_\emptyset = 0$. The reproduction processes ξ_x of different individuals are iid copies of ξ.

The basic probability space is

(3.38) $$(\Omega_{\mathcal{I}}, \mathcal{B}_{\mathcal{I}}, P_{\mathcal{I}}) = \prod_{x \in \mathcal{I}} (\Omega_x, \mathcal{B}_x, P_x)$$

where $\mathcal{I}$ is given as in (3.15) and (ξ_x, λ_x) are random variables defined on $(\Omega_x, \mathcal{B}_x, P_x)$ with distribution as above.

We then determine the birth times $\{\tau_x,\ x \in \mathcal{I}\}$ as follows:

(3.39) $$\tau_\emptyset = 0, \qquad \tau_{(x',i)} = \tau_{x'} + \inf\{u : \xi_{x'}(u) \geq i\}.$$

Note that for individuals never born $\tau_x = \infty$.

Let

(3.40) $$Z_t = \sum_{x \in \mathcal{I}} 1_{\tau_x \leq t < \lambda_x}, \quad T_t = \sum_{x \in \mathcal{I}} 1_{\tau_x \leq t}$$

that is, the *number of individuals alive at time* t and *total number of births before time* t, respectively.

For $\lambda > 0$ we define

(3.41) $$\xi^\lambda(t) := \int_0^t e^{-\lambda t} \xi(dt).$$

The *Malthusian parameter* α is defined by the equation

(3.42) $E[\xi^\alpha(\infty)] = 1$

that is,

$$(3.43)\quad \int_0^\infty e^{-\alpha t}\mu(dt) = 1.$$

The *stable average age of child-bearing* is defined as

$$(3.44)\quad \beta = \int_0^\infty t\widetilde{\mu}(dt) \text{ where } \widetilde{\mu}(dt) = e^{-\alpha t}\mu(dt).$$

Example 3.8 *Consider a population in which individuals have an internal state space, say* $\mathbb{N}$. *Assume that the individual starts in state* 0 *at its time of birth and and its internal state changes according to a Markov transition mechanism. Finally assume that when it is in state* i *it produces offspring at rate* λ_i.

Definition 3.9 Characteristics of an individual A characteristic *of an individual is given by a process* $\phi : \mathbb{R}\times\Omega \to \mathbb{R}_+$ *which is given by a* $\mathcal{B}(\mathbb{R})\times\sigma(\xi)$*-measurable non-negative function satisfying* $\phi(t) = 0$ *for* $t < 0$, *let*

$$(3.45)\quad Z_t^\phi = \sum_{x\in\mathcal{I}} \phi_x(t-\tau_x)$$

denote the process counted with characteristic ϕ.

Example 3.10 *If* $\phi^a(t) = 1_{[0,\inf(a,\lambda))}(t)$, *then* $Z_t^{\phi^a}$ *counts the number of individuals alive at time* t *whose ages are less than* a.

The following fundamental generalization of the Kesten-Stigum theorem was developed in papers of Doney (1972),(1976) [176], [177], and Nerman (1981) [475].

Theorem 3.11 *Consider a CMJ process with malthusian parameter* α *and assume that* $\beta < \infty$.

(a) [176] Then as $t\to\infty$, $e^{-\alpha t}Z_t$ *converges in distribution to* mW_∞ *where*

$$(3.46)\quad m = \frac{\int_0^\infty e^{-\alpha s}(1-L(s))ds}{\beta}$$

and W_∞ *is a random variable (see Proposition 3.13) and*

(b) The following are equivalent:

$$(3.47)\quad E[\xi^\alpha(\infty)\log^+\xi^\alpha(\infty)] < \infty$$

$$(3.48)\quad E[W] > 0$$

$$(3.49)\quad E[e^{-\alpha t}Z_t] \to E[W] \quad \text{as } t\to\infty$$

$$(3.50)\quad W > 0 \text{ a.s. on } \{T_t\to\infty\}.$$

(c) [475] Under the condition that there exists a non-increasing integrable function g *such that*

(3.51) $E[\sup_t \frac{(\xi^\alpha(\infty) - \xi^\alpha(t))}{g(t)}] < \infty,$

then $e^{-\alpha t} Z_t$ *converges a.s. as* $t \to \infty$*.*

Remark 3.12 *A sufficient condition is the existence of non-increasing integrable function* g *such that*

(3.52) $\int_0^\infty \frac{1}{g(t)} e^{-\alpha t} \mu(dt) < \infty.$

(See Nerman [475] (5.4)).

Comments on Proofs

(b) The equivalence statements can be proved in this general case following the same lines as that of Lyons, Pemantle and Peres - see Olofsson (1996) [492].

(a) - convergence in distribution was proved by Doney (1972) [176]. However the almost sure convergence required some basic new ideas since we can no longer directly use the martingale convergence theorem since Z_t is not a martingale in the general case. The a.s. convergence was proved by Nerman [475]. We will not give Nerman's long detailed technical proof of this result but will now introduce the key tool used in its proof and which is of independent interest, namely, an underlying intrinsic martingale W_t discovered by Nerman [475] and then give an intuitive idea of the remainder of the proof.

Denote the mother of x by $m(x)$ and let

(3.53) $\mathcal{I}_t = \{x \in \mathcal{I} : \tau_{m(x)} \leq t < \tau_x < \infty\},$

the set of individuals whose mothers are born before time t but who themselves are born after t

Consider the individuals ordered by their times of birth

(3.54) $0 = \tau_{x_1} \leq \tau_{x_2} \leq \dots$

Define $\mathcal{A}_n$ = σ-algebra generated by $\{(\tau_{x_i}, \xi_{x_i}, \lambda_{x_i}) : i = 1, \dots, n\}$ Recall (3.40) and let $\mathcal{F}_t = \mathcal{A}_{T_t}$.

Define

(3.55) $W_t := \sum_{x \in \mathcal{I}_t} e^{-\alpha \tau_x}.$

Proposition 3.13 *(Nerman (1981) [475]) (a) The process* $\{W_t, \mathcal{F}_t\}$ *is a non-negative martingale with* $E[W_t] = 1$*.*

(b) There exists a random variable $W_\infty < \infty$ *such that* $W_t \to W_\infty$ *a.s. as* $t \to \infty$*.*

Proof. Define

$$R_0 = 1, \tag{3.56}$$

$$R_n = 1 + \sum_{i=1}^{n} e^{-\alpha\tau_{x_i}}(\xi^{\alpha}_{x_i}(\infty) - 1), \quad n = 1, 2, \ldots$$

Equivalently, letting $\tau_{(x_i,k)}$ denote the time of birth of the kth offspring of x_i,

$$R_n = 1 + \sum_{i=1}^{n} \sum_{k=1}^{\xi_{x_i}(\infty)} e^{-\alpha\tau_{(x_i,k)}} - \sum_{i=1}^{n} e^{-\alpha\tau_{x_i}} \tag{3.57}$$

so that R_n is a weighted (weights $e^{-\alpha\tau_x}$) sum of children of the first n individuals.

We next show that $(R_n, \mathcal{A}_n)$ is a non-negative martingale. R_n and $\tau_{x_{n+1}}$ are $\mathcal{A}_n$-measurable and $\xi^{\alpha}_{x_{n+1}}$ is independent of $\mathcal{A}_n$ and

$$E[\xi^{\alpha}_{x_{n+1}}(\infty)] = \mu_\alpha(\infty) = 1. \tag{3.58}$$

Therefore

$$E[R_{n+1} - R_n] = e^{-\alpha\tau_{x_{n+1}}} E[\xi^{\alpha}_{x_{n+1}} - 1] = 0. \tag{3.59}$$

Next we observe that since $\mathcal{I}(t)$ consists of exactly the children of the first T_t individuals to be born after t, it follows that $W_t = R_{T_t}$.

Note that for fixed t, $\{T_t \le k\} = \{\tau_{x_n} \le t\} \in \mathcal{A}_n$ and therefore T_t is an increasing family of integer-valued stopping times with respect to $\{\mathcal{A}_n\}$. Therefore $\{W_t\}$ is a supermartingale with respect to the filtration $\{\mathcal{A}_{T_t}\}$.

Since $E[T_t] < \infty$ and

$$E[|R_{n+1} - R_n| \, |\mathcal{A}_n] = e^{-\alpha\tau_{x_{n+1}}} E[|\xi^{\alpha}(\infty) - 1|] \le 2. \tag{3.60}$$

a standard argument (e.g. Breiman [34] Prop. 5.33) implies that $E[W_t] = E[R_{T_t}] = 1$ and $\{W_t\}$ is actually a martingale.

(b) This follows from (a) and the martingale convergence theorem.

■

Remark 3.14 *We now sketch an intuitive explanation for the proof of the a.s. convergence of $e^{-\alpha t}Z_t$ using Proposition 3.13. This is based on the relation between W_t and Z_t which is somewhat is somewhat indirect. To give some idea of this, let*

$$W_{t,c} = \sum_{x \in \mathcal{I}_{t,c}} e^{-\alpha\tau_x}, \tag{3.61}$$

where

$$\mathcal{I}_{t,c} = \{x = (x', i) : \tau_{x'} \le t,\ t + c < \tau_x < \infty\}. \tag{3.62}$$

Note that if we consider the characteristic χ^c defined by

(3.63) $\chi^c(s) = (\xi^\alpha(\infty) - \xi^\alpha(s+c))e^{\alpha s}$ *for* $s \geq 0$,

then

(3.64) $W_{t,c} = e^{-\alpha t} Z_t^{\chi^c}$

where

(3.65) $Z_t^{\chi^c} = \sum_{x \in \mathcal{I}} \chi_x^c(t - \tau_x), \quad \chi_x^c(s) = (\xi_x^\alpha(\infty) - \xi_x^\alpha(s+c))e^{\alpha s}.$

Note that $\lim_{c\to 0} W_{t,c} = W_t$ *and* $\lim_{c\to 0} Z_t^{\chi^c} = Z_t^\chi$ *where*

(3.66) $\chi(s) = \int_s^\infty e^{-\alpha(u-s)}\xi(du).$

Then $Z_t^\chi \to m_\chi W_\infty$, *a.s. where*

(3.67) $m_\chi = \frac{\int_0^\infty e^{-\alpha t}(1 - L(t))dt}{\beta}.$

In the special case where ξ *is stationary then the distribution of* $\chi(s)$ *does not depend on* s. *Then* Z_t^χ *is a sum of* Z_t *i.i.d. random variables and therefore as* $t \to \infty$, Z_t^χ *should approach a constant times* Z_t *by the law of large numbers.*

Stable age distribution

The notion of the stable age distribution of a population is a basic concept in demography going back to Euler. The stable age distribution in the deterministic setting of the Euler-Lotka equation (2.2) is

(3.68) $U(\infty, ds) = \frac{(1 - L(s))e^{-\alpha s}ds}{\int_0^\infty (1 - L(s))e^{-\alpha s}ds}.$

It was introduced into the study of branching processes by Athreya and Kaplan (1976) [10]. Let Z_t^a denote the number of individuals of age $\leq a$. The normalized age distribution at time t is defined by

(3.69) $U(t, [0,a)) := \frac{Z_t^a}{Z_t}, \quad a \geq 0.$

Theorem 3.15 *(Nerman [475] Theorem 6.3 - Convergence to stable age distribution) Assume that* ξ *satisfies the conditions of Theorem 3.11. Then on the event* $T_t \to \infty$,

(3.70) $U(t, [0,a)) \to \frac{\int_0^a (1 - L(u))e^{-\alpha u}du}{\int_0^\infty (1 - L(u))e^{-\alpha u}du}$ *a.s. as* $t \to \infty$.

3.1.5 Multitype branching

A central idea in evolutionary biology is the differential growth rates of different types of individuals. Multitype branching processes provide a starting point for our discussion of this basic topic.

Consider a multitype BGW process with K types. Let $\xi^{(i,j)}$ be a random variable representing the number of particles of type K produced by one type i particle in one generation.

Let $Z^{(j)}$ be the number of particles of type j in generation n and $\mathbf{Z}_n := (Z_n^{(1)}, \ldots, Z_n^K)$.

For $\mathbf{k} = (k_1, \ldots, k_K)$, let $p_{\mathbf{k}}^{(i)} = P[\xi^{(i,j)} = k_j,\ j = 1, \ldots, K]$. Assume that

$$(3.71)\quad \mathbf{M} = (m_{(i,j)})_{i,j=1,\ldots,K}, \qquad m^{(i,j)} = E[\xi^{(i,j)}] < \infty \quad \forall\, i, j.$$

Then

$$(3.72)\quad E(\mathbf{Z}_{m+n}|\mathbf{Z}_m) = \mathbf{Z}_m\mathbf{M}^n, \quad m, n \in \mathbb{N}.$$

The behaviour of $E[\mathbf{Z}_n]$ as $n \to \infty$ is then obtained from the classical Perron-Frobenius Theorem:

Theorem 3.16 *(Perron-Frobenius) Let $\mathbf{M}$ be a nonnegative $K \times K$ matrix. Assume that $\mathbf{M}^n$ is strictly positive for some $n \in \mathbb{N}$. Then $\mathbf{M}$ has a largest positive eigenvalue ρ which is a simple eigenvalue with positive right and left normalized eigenvectors $\mathbf{u} = (u_i)$ ($\sum u_i = 1$) and $\mathbf{v} = (v_i)$ which are the only nonnegative eigenvectors. Moreover*

$$(3.73)\quad \mathbf{M}^n = \rho^n\mathbf{M}_1 + \mathbf{M}_2^n$$

where $\mathbf{M}_1 = (u_iv_j)_{i,j\in\{1,\ldots,K\}}$ normalized by $\sum_i, ju_iv_j = 1$. Moreover $\mathbf{M}_1\mathbf{M}_2 = \mathbf{M}_2\mathbf{M}_1 = 0$, $\mathbf{M}_1^n = \mathbf{M}_1$.

Finally,

$$(3.74)\quad |M_2^n| = O(\alpha^n)$$

for some $0 < \alpha < \rho$.

The analogue of the Kesten-Stigum theorem stated above is given as follows.

Theorem 3.17 *(Kesten-Stigum (1966) [365]), (Kurtz, Lyons, Pemantle and Peres (1997) [408])*
(a) There is a scalar random variable W such that

$$(3.75)\quad \lim_{n\to\infty} \frac{\mathbf{Z}_n}{\rho^n} = W\mathbf{u} \ \ a.s.$$

and $P[W > 0] > 0$ *iff*

$$E[\sum_{i,j=1}^{J} \xi^{(i,j)} \log^{+} \xi^{(i,j)}] < \infty. \tag{3.76}$$

(b) Almost surely, conditioned on nonextinction,

$$\lim_{n\to\infty} \frac{\mathbf{Z}_n}{|\mathbf{Z}_n|} = \mathbf{u}. \tag{3.77}$$

3.2 Multilevel branching

Consider a *host-parasite population* in which the individuals in the host population reproduce by BGW branching and the population of parasite on a given host also develop by an independent BGW branching. This is an example of a *multilevel branching system*.

A *multilevel population system* is a hierarchically structured collection of objects at different levels as follows:
E_0 denotes the set of possible types of level 1 object,
for $n \geq 1$ each level $(n+1)$ object is given by a collection of level n object including their their multiplicities.

Multilevel branching dynamics
Consider a continuous time branching process such that

- for $n \geq 1$, when a level n object branches, all its offspring are copies of it
- if $n \geq 2$, then the offspring contains a copy of the set of level-$n-1$ objects contained in the parent level n object.
- let γ_n the level n branching rate and by $f_n(s)$ the level n offspring generating function.

Then the questions of extinction, classification into critical, subcritical and supercritical case and growth asymptotics in the supercritical case are more complex than the single level branching case. See for example, Dawson and Wu (1996) [154].