

## MULTITYPE BIENAYMÉ–GALTON–WATSON PROCESSES ESCAPING EXTINCTION

SERIK SAGITOV\* AND

MARIA CONCEIÇÃO SERRA,\*\* *Chalmers University of Technology and Göteborg University*

### Abstract

In the framework of a multitype Bienaymé–Galton–Watson (BGW) process, the event that the daughter’s type differs from the mother’s type can be viewed as a mutation event. Assuming that mutations are rare, we study a situation where all types except one produce on average less than one offspring. We establish a neat asymptotic structure for the BGW process escaping extinction due to a sequence of mutations toward the supercritical type. Our asymptotic analysis is performed by letting mutation probabilities tend to 0. The limit process, conditional on escaping extinction, is another BGW process with an enriched set of types, allowing us to delineate a stem lineage of particles that leads toward the escape event. The stem lineage can be described by a simple Markov chain on the set of particle types. The total time to escape becomes a sum of a random number of independent, geometrically distributed times spent at intermediate types.

*Keywords:* Bienaymé–Galton–Watson process; decomposable; escape from extinction; multitype; wild-type branching process

2000 Mathematics Subject Classification: Primary 60J80  
Secondary 92D25

### 1. Introduction

Recently, Iwasa *et al.* [5], [6] studied the probability that a virus placed in a hostile environment escapes extinction via a chain of mutations. As a population model, the authors suggested using a multitype branching process with either continuous or discrete time. Branching processes naturally arise in this context because viruses reproduce asexually and the assumption of independence among coexisting particles is realistic provided that subcritical reproduction conditions are assumed. Particle types, corresponding to various genome sequences of the virus, are conveniently coded by vectors of 0s and 1s of length  $L$ , so that the set of  $2^L$  possible types can be depicted as a graph, illustrated in Figure 1 (see the remark at the end of this section). In the multitype setting, a mutation is a reproduction outcome when the daughter’s type differs from the mother’s type [10].

A discrete-time version of the key result in [6] can be stated as follows. Consider a Bienaymé–Galton–Watson (BGW) process with  $2^L$  types of particle labeled by 0-1 vectors of length  $L$ . Let all the types with exactly  $i$  1s in their vector labels have the same mean offspring number  $m_i$ . If we further assume that  $0 < m_i < 1$  for  $i = 1, \dots, L$  while  $m_0 > 1$ , and that the BGW process starts from the subcritical form  $(1, \dots, 1)$ , then, obviously, this reproduction process is doomed to become extinct unless a sequence of mutations results in the supercritical form

---

Received 21 August 2007; revision received 29 December 2008.

\* Postal address: School of Mathematical Sciences, Chalmers University of Technology and Göteborg University, SE-412 96 Göteborg, Sweden. Email address: serik@math.chalmers.se

\*\* Current address: Department of Mathematics, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal.

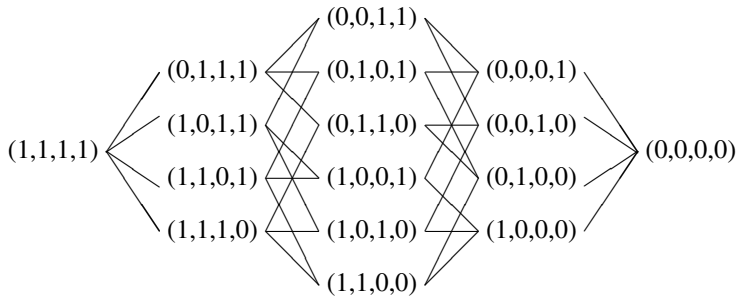


FIGURE 1: The network of 0-1 sequences of length  $L = 4$ . Here the edges represent single point mutations of probability  $\mu(1 - \mu)^{L-1}$ . More generally, mutation between two given sequences which differ at  $i$  sites has probability  $\mu^i(1 - \mu)^{L-i}$ .

$(0, \dots, 0)$ . Given that all point mutations have the same probability  $\mu$  per site per generation, what is the probability for such a BGW process to escape the extinction? According to [6], the escape probability is asymptotically equal to a positive constant (which can be expressed in terms of  $m_1, \dots, m_L$ ) times  $(1 - q)\mu^L$  as  $\mu \rightarrow 0$ . Here,  $q \in (0, 1)$  stands for the extinction probability of the branching process formed by the particles of the supercritical type in the absence of mutations. The asserted order of smallness,  $O(\mu^L)$ , implies that, as  $\mu \rightarrow 0$ , we can disregard the possibility of backward mutations, which would lead to terms of order  $o(\mu^L)$ .

Usually, asymptotic analysis of the BGW processes  $\{Z(n)\}_{n \geq 0}$  is performed as the time of observation  $n$  goes to  $\infty$ , addressing the issue of long-term demographic patterns (see, for example, [4] or [9]). On the contrary, the asymptotic approach as  $\mu \rightarrow 0$ , suggested by Iwasa *et al.*, turns to short-term demographic patterns, under the biologically relevant assumption that mutations are rare events. In this paper we study a wide class of multitype BGW processes with rare mutations conditioned on producing a certain type of particle. We obtain an asymptotic BGW process with an enriched set of types, which allows us to delineate a sequence of successful mutations leading to the desired type.

Our results are presented in the following manner. First, we consider the case in which  $L = 1$ , where the BGW process has just two types of particle, 0 and 1. Sections 2 and 3 are devoted to a general two-type BGW process and there we introduce the main tool of our approach—decompositions of so-called *wild-type BGW processes*. For a given multitype BGW process starting from a single particle, the wild-type BGW process is a one-type BGW process formed by particles of the original type. Whenever a *mutant*, a particle of another type, is born, the mutant itself and all its forthcoming descendants are excluded from the wild-type process. (Note that, throughout this paper, by *descendants* of a particle we mean not only its direct offspring (daughters) but also its granddaughters, grand-granddaughters, and so on.) Such a wild-type process can be viewed as a ‘first wave’ of the whole BGW process followed by consecutive ‘waves’ of the wild-type process. This representation of the multitype BGW processes appears to be very useful in the current context.

In Section 2 we demonstrate that the wild-type process can be viewed as a decomposable two-type BGW process. These two subtypes are meant to distinguish between the original type particles with different fates—those which have at least one mutant among their descendants and those which do not. This should be compared with a refined decomposition of Section 3, where the type-1 particles with mutant descendants are further classified into those which eventually lead to nonextinction events and those which do not.

We use such decompositions as a framework for obtaining limit structures, as  $\mu \rightarrow 0$ , of the BGW processes conditioned on escaping extinction. Note that in our limit theorems the mutation rate parameter  $\mu$  regulates different mutation probabilities in a rather flexible way, allowing mutation probabilities to depend on the corresponding offspring number. It can be biologically relevant to assume, for example, that particles with many siblings have higher probability of undergoing mutation. In particular, in the two-type case we assume that the forward mutation probability from type 1 to type 0 is asymptotically equal to  $\mu$  times a constant  $a_{10}(k)$  which may depend on the corresponding family size  $k$ .

In Section 4 we present our first limit theorem (its proof is given in Section 5) which deals with a two-type BGW process stemming from a type-1 particle with  $m_1 \in (0, 1)$ . We show that its conditional limit can be regarded as a BGW process with immigration. The limit process is effectively the same whether we condition just on a mutation event or we condition on a successful mutation resulting in escape from extinction. The immigration source, corresponding to a stem lineage leading to a desired mutation event, is turned on during a geometric time  $T_1$  with mean  $E(T_1) = (1 - m_1)^{-1}$ . The numbers of immigrants have the size-biased distribution. (In an earlier paper [8], the geometric time to escape was obtained for a simpler version of the two-type model using a more straightforward asymptotic analysis.)

Our careful treatment of the two-type case in Sections 2–5 confirms that the backward mutations on the path to escape play a negligible role. Therefore, in the remaining sections we disregard the possibility of backward mutations to simplify analysis of the BGW processes with an arbitrary number of types. In Section 6 we introduce a sequential mutation model and describe the corresponding wild-type BGW process. In the sequential mutation model, mutations may occur along an interval of types starting from type  $L$  and ending at type 0. With this model, the probability that a given daughter of a type- $i$  particle mutates to type  $j < i$  is assumed to be asymptotically equal to  $\mu^{i-j} a_{ij}(k)$  as  $\mu \rightarrow 0$ , where  $k$  is the offspring number of the mother in question. This is a natural intermediate step between the two-type case towards the so-called network model described in Figure 1. Indeed, if in the network case we treat each subset of 0-1 vectors with exactly  $i$  1s as a single type  $i$ , then we arrive at a sequential mutation model with the mutation probability between types  $i$  and  $j$  being asymptotically equivalent to  $\binom{i}{j} \mu^{i-j}$ , given that  $j < i$ .

The asymptotic results in Section 7 extend the theorem in Section 4 with  $L = 1$  to sequential mutation models with arbitrary  $L \geq 1$ . In Section 8 we discuss the asymptotic distribution of the total time to escape and its expected value. Finally, in Section 9 we apply the results of Section 7 to the network mutation model.

**Remark.** A more realistic picture of various viral DNA sequences would distinguish among the four bases  $A$ ,  $C$ ,  $T$ , and  $G$  at each of the  $L$  sites. The 0 and 1 coding system suggested by Iwasa *et al.* [5], [6] describes a simplified view of the way the DNA sequences change over time. It is well known that transitions  $A \leftrightarrow G$  and  $C \leftrightarrow T$  are much more frequent than the transversions  $A \leftrightarrow C$ ,  $A \leftrightarrow T$ ,  $G \leftrightarrow C$ , and  $G \leftrightarrow T$ . If we assume that all point mutations are transitions then it would be sufficient with the binary coding. For example, given the initial sequence, say  $ACGT$ , and its binary code 1111, it is straightforward to recover the mutant DNA sequence  $ATAT$  from the binary sequence 1001.

## 2. The wild-type BGW process in the two-type case

Consider a BGW process with two types of particle, labeled by 0 and 1, which starts from a single particle to be called the *progenitor*. If  $Z_i(n)$  is the number of type- $i$  particles in

generation  $n$  then the vector  $\{Z_0(n), Z_1(n)\}_{n \geq 0}$  forms a Markov chain describing the population size and type structure evolving generation wise. The transition probabilities of this Markov chain are determined by a pair of generating functions,

$$f_i(s_0, s_1) = E_i(s_0^{Z_0(1)} s_1^{Z_1(1)}), \quad (s_0, s_1) \in [0, 1] \times [0, 1], \quad i = 0, 1,$$

defining the reproduction laws for the particles of types 0 and 1.

Proposition 2.1, below, deals with the corresponding wild-type process involving only those descendants of the progenitor which did not undergo mutation. The proposition says that the wild-type process can be viewed as a decomposable BGW process recognizing two subtypes of the progenitor’s type 1. This construction was inspired by the well-known decomposition of a supercritical single-type BGW process into a two-type BGW process distinguishing between particles with infinite and finite lines of descent (see [1, p. 47]).

There are two possibilities from which to choose the type of progenitor:

$$\begin{aligned} B_0 &= \{Z_0(0) = 1, Z_1(0) = 0\}, \\ B_1 &= \{Z_0(0) = 0, Z_1(0) = 1\}. \end{aligned}$$

The event  $B_1$ , in turn, is the union of two disjoint events:  $B_{1_m}$ , the event that at least one mutation occurs, and  $B_{1_u}$ , the event that the progenitor’s type will be inherited unchanged by all of its descendants. This suggests a refined classification of type-1 particles into two subtypes,  $1_m$  and  $1_u$ , leading to the following relations:

$$\begin{aligned} B_{1_m} &= \{Z_{1_m}(0) = 1, Z_{1_u}(0) = 0\}, \\ B_{1_u} &= \{Z_{1_m}(0) = 0, Z_{1_u}(0) = 1\}. \end{aligned}$$

We will denote by  $P_0, P_1, P_{1_m}$ , and  $P_{1_u}$  the probabilities conditioned on the type or subtype of the progenitor and by  $E_0, E_1, E_{1_m}$ , and  $E_{1_u}$  the respective expectation operators. Set

$$Q_{1_m} = P_1(B_{1_m}), \quad Q_{1_u} = P_1(B_{1_u}) = 1 - Q_{1_m}.$$

Assume that the progenitor of type 1 has produced  $k = k_0 + k_1$  daughters, with  $Z_0(1) = k_0$  among them being of type 0 and  $Z_1(1) = k_1$  among them being of type 1. We will label the type-1 daughters by the numbers  $1, \dots, k_1$  and the type-0 daughters by  $k_1 + 1, \dots, k$ . Within the particles of each type, the labeling is done uniformly at random. For an event  $A$  concerning the original BGW process, we will denote by  $A^{(i)}$  its analog associated with the BGW process stemming from the progenitor’s daughter number  $i$ . For example, the event  $B_{1_m}^{(i)}$  simply means that the  $i$ th daughter of the progenitor has subtype  $1_m$ .

The next lemma is analogous to Lemma 2.1 of [3]. It concerns a random variable  $R$  taking values  $1, 2, \dots, \infty$ . If among the  $k_1$  progenitor’s daughters of type 1 there is at least one particle of subtype  $1_m$ , then  $R$  is defined as the smallest number  $j$  labeling a  $1_m$ -daughter. Otherwise, we set  $R = \infty$ . That is, for  $j = 1, 2, \dots$ ,

$$\{R = j\} = \{Z_1(1) \geq j\} \cap \bigcap_{i=1}^{j-1} B_{1_u}^{(i)} \cap B_{1_m}^{(j)}$$

and

$$\{R = \infty\} = B_{1_u} = \left\{ \bigcap_{i=1}^{Z_1(1)} B_{1_u}^{(i)} \cap \{Z_1(1) > 0\} \right\} \cup \{Z_1(1) = 0\}.$$

**Lemma 2.1.** Consider the random variable  $R$  defined above. Its conditional distribution is given, for  $1 \leq j \leq \infty$ ,  $k_0 \geq 0$ , and  $k_1 \geq 0$ , by

$$P_1(R = j \mid Z_0(1) = k_0, Z_1(1) = k_1) = \begin{cases} Q_{1_u}^{j-1} Q_{1_m}, & 1 \leq j \leq k_1, \\ Q_{1_u}^{k_1}, & j = \infty, k_1 \geq 1, \\ 1, & j = \infty, k_1 = 0, \end{cases} \tag{2.1}$$

implying that

$$E(s^R \mathbf{1}_{\{R < \infty\}}) = \frac{s Q_{1_m}}{1 - s Q_{1_u}} (1 - f_1(1, s Q_{1_u})), \quad s \in [0, 1),$$

and

$$Q_{1_u} = f_1(0, Q_{1_u}). \tag{2.2}$$

*Proof.* Observe that

$$\{R = j, Z_0(1) = k_0, Z_1(1) = k_1\} = \{Z_0(1) = k_0, Z_1(1) = k_1, C_i^{(i)}, 1 \leq i \leq k\},$$

where  $k = k_0 + k_1$ ,  $j \in [1, k_1]$ , and

$$C_i = \begin{cases} B_{1_u}, & 1 \leq i \leq j - 1, \\ B_{1_m}, & i = j, \\ B_1, & j + 1 \leq i \leq k_1, \\ B_0, & k_1 + 1 \leq i \leq k. \end{cases} \tag{2.3}$$

According to the key branching property, the offspring of the progenitor generate independent trees. Therefore,

$$\begin{aligned} P_1(Z_0(1) = k_0, Z_1(1) = k_1, C_i^{(i)}, 1 \leq i \leq k) \\ = P_1(Z_0(1) = k_0, Z_1(1) = k_1) \prod_{i=1}^{k_1} P_1(C_i) \prod_{i=k_1+1}^k P_0(C_i), \end{aligned} \tag{2.4}$$

implying that

$$P_1(R = j \mid Z_0(1) = k_0, Z_1(1) = k_1) = Q_{1_u}^{j-1} Q_{1_m}. \tag{2.5}$$

Now it remains only to note that

$$\begin{aligned} P_1(R = \infty \mid Z_0(1) = k_0, Z_1(1) = k_1) &= Q_{1_u}^{k_1} \quad \text{for } k_1 \geq 1, \\ P_1(R = \infty \mid Z_0(1) = k_0, Z_1(1) = 0) &= 1, \end{aligned}$$

and that (2.2) follows from

$$B_{1_u} = B_1 \cap \{Z_0(1) = 0, R = \infty\}. \tag{2.6}$$

**Lemma 2.2.** Let  $\{A_i\}_{i=1}^k$  be random events concerning the BGW process generated by a single progenitor particle, when this particle has  $k$  daughters. The daughter versions of these events,

$\{A_i^{(i)}\}_{i=1}^k$ , are conditionally independent given that  $R = j$ ,  $Z_0(1) = k_0$ ,  $Z_1(1) = k_1$ , and  $k_0 + k_1 = k$  with

$$P_1(A_i^{(i)} \mid R = j, Z_0(1) = k_0, Z_1(1) = k_1) = \begin{cases} P_{1_u}(A_i), & 1 \leq i \leq j - 1 \wedge k_1, \\ P_{1_m}(A_i), & i = j, \\ P_1(A_i), & j + 1 \leq i \leq k_1, \\ P_0(A_i), & k_1 + 1 \leq i \leq k. \end{cases}$$

*Proof.* Let  $j \in [1, k_1]$ . In terms of the system of events (2.3),

$$\begin{aligned} &\{R = j, Z_0(1) = k_0, Z_1(1) = k_1, A_i^{(i)}, 1 \leq i \leq k\} \\ &= \{Z_0(1) = k_0, Z_1(1) = k_1, A_i^{(i)} \cap C_i^{(i)}, 1 \leq i \leq k\}. \end{aligned}$$

The stated conditional independence now follows from (2.4), with events  $C_i^{(i)}$  replaced by  $A_i^{(i)} \cap C_i^{(i)}$ , and (2.5):

$$\begin{aligned} &P_1(A_i^{(i)}, 1 \leq i \leq k, R = j \mid Z_0(1) = k_0, Z_1(1) = k_1) \\ &= \prod_{i=1}^{j-1} P_1(A_i \cap B_{1_u}) P_1(A_j \cap B_{1_m}) \prod_{i=j+1}^{k_1} P_1(A_i) \prod_{i=k_1+1}^k P_0(A_i) \\ &= P_1(R = j \mid Z_0(1) = k_0, Z_1(1) = k_1) \\ &\quad \times \prod_{i=1}^{j-1} P_{1_u}(A_i) P_{1_m}(A_j) \prod_{i=j+1}^{k_1} P_1(A_i) \prod_{i=k_1+1}^k P_0(A_i). \end{aligned}$$

**Proposition 2.1.** Consider a two-type BGW process starting from a single type-1 particle. The corresponding wild-type process can be represented as a decomposable BGW process  $\{Z_{1_m}(n), Z_{1_u}(n)\}_{n \geq 0}$  with two types of particle:  $1_m$  and  $1_u$ . The progenitor’s type has distribution  $(Q_{1_m}, Q_{1_u})$  satisfying (2.2) and  $Q_{1_m} + Q_{1_u} = 1$ . The two-type reproduction law of the  $\{Z_{1_m}(n), Z_{1_u}(n)\}_{n \geq 0}$  process has probability generating functions

$$E_{1_u}(s^{Z_{1_m}(1)} t^{Z_{1_u}(1)}) = \frac{f_1(0, tQ_{1_u})}{Q_{1_u}}, \tag{2.7}$$

$$E_{1_m}(s^{Z_{1_m}(1)} t^{Z_{1_u}(1)}) = \frac{f_1(1, sQ_{1_m} + tQ_{1_u}) - f_1(0, tQ_{1_u})}{Q_{1_m}}. \tag{2.8}$$

*Proof.* Since

$$E_{1_u}(s^{Z_{1_m}(1)} t^{Z_{1_u}(1)})Q_{1_u} + E_{1_m}(s^{Z_{1_m}(1)} t^{Z_{1_u}(1)})Q_{1_m} = E_1(s^{Z_{1_m}(1)} t^{Z_{1_u}(1)}),$$

and due to the basic branching property,

$$E_1(s^{Z_{1_m}(1)} t^{Z_{1_u}(1)}) = f_1(1, sQ_{1_m} + tQ_{1_u}),$$

to prove (2.7) and (2.8), it suffices to verify that

$$E_1(s^{Z_{1_m}(1)} t^{Z_{1_u}(1)} \mathbf{1}_{B_{1_u}}) = f_1(0, tQ_{1_u}).$$

But this readily follows from (2.1) and (2.6):

$$\begin{aligned} E_1(s^{Z_{1_m}(1)}t^{Z_{1_u}(1)}\mathbf{1}_{\{Z_0(1)=0, R=\infty\}}) &= E_1(t^{Z_1(1)}\mathbf{1}_{\{Z_0(1)=0, R=\infty\}}) \\ &= f_1(0, tQ_{1_u}). \end{aligned}$$

Finally, the claimed independence of particles in the framework of the new two-type system follows from Lemma 2.2.

### 3. A refined decomposition of the wild-type BGW process

With the same two-type BGW process as in the previous section, let us now distinguish between two kinds of type-0 particles which will be labeled as  $0_e$  and  $0_d$ , depending on whether the BGW process stemming from a given 0-particle escapes extinction ( $0_e$ ) or dies out ( $0_d$ ). We wish to refine our earlier classification of type-1 particles and divide the subtype  $1_m$  into sub-subtypes  $1_e$  and  $1_d$ . A  $1_e$ -particle is by definition a  $1_m$ -particle which has at least one  $0_e$ -descendant. Similarly, a  $1_d$ -particle is a  $1_m$ -particle which has no  $0_e$ -descendants, i.e. only  $0_d$ -descendants. Let  $Z_{1_e}(n)$  and  $Z_{1_d}(n)$  stand for the numbers of particles of subtypes  $1_e$  and  $1_d$  in generation  $n$ . Set

$$\begin{aligned} Q_{1_e} &= P_1(Z_{1_e}(0) = 1, Z_{1_d}(0) = 0, Z_{1_u}(0) = 0), \\ Q_{1_d} &= P_1(Z_{1_e}(0) = 0, Z_{1_d}(0) = 1, Z_{1_u}(0) = 0), \end{aligned}$$

so that  $Q_{1_e} + Q_{1_d} = Q_{1_m}$ .

It is well known that the extinction probabilities

$$q_i = P_i\left(\lim_{n \rightarrow \infty} (Z_0(n) + Z_1(n)) = 0\right)$$

satisfy the following pair of equations (see [1, p. 186]):

$$q_0 = f_0(q_0, q_1), \quad q_1 = f_1(q_0, q_1).$$

Observe that  $Q_{1_e} = 1 - q_1$ , given that there is a possibility for type-1 particles to produce particles of type 0 and that  $m_0 > 1$ . Indeed, if a BGW process escapes extinction, it grows exponentially and, therefore, produces lots of type-0 particles, some of which will be classified as  $0_e$ -particles. On the other hand, if the type-1 progenitor is classified as a subtype  $1_e$ -particle, it will produce a  $0_e$ -descendant guaranteeing escape from extinction. This observation yields the following equation to be used in Section 5:

$$1 - Q_{1_e} = f_1(q_0, 1 - Q_{1_e}). \tag{3.1}$$

**Proposition 3.1.** *Consider a BGW process with types 0 and 1 particles starting from a single type-1 particle. The corresponding wild-type process can be represented as a decomposable BGW process  $\{Z_{1_e}(n), Z_{1_d}(n), Z_{1_u}(n)\}_{n \geq 0}$  with the three types of particle,  $1_e$ ,  $1_d$ , and  $1_u$ , described above.*

The progenitor's type in this representation has distribution  $(Q_{1_e}, Q_{1_d}, Q_{1_u})$  and the new three-type reproduction law is defined by

$$E_{1_u}(s^{Z_{1_e}(1)}v^{Z_{1_d}(1)}t^{Z_{1_u}(1)}) = \frac{f_1(0, tQ_{1_u})}{Q_{1_u}}, \tag{3.2}$$

$$E_{1_d}(s^{Z_{1_e}(1)}v^{Z_{1_d}(1)}t^{Z_{1_u}(1)}) = \frac{f_1(q_0, vQ_{1_d} + tQ_{1_u}) - f_1(0, tQ_{1_u})}{Q_{1_d}}, \tag{3.3}$$

$$E_{1_e}(s^{Z_{1_e}(1)}v^{Z_{1_d}(1)}t^{Z_{1_u}(1)}) = \frac{f_1(1, sQ_{1_e} + vQ_{1_d} + tQ_{1_u}) - f_1(q_0, vQ_{1_d} + tQ_{1_u})}{Q_{1_e}}. \tag{3.4}$$

*Proof.* It is easy to adjust Lemma 2.1 and Lemma 2.2 to verify the branching property of the three-type process, and it directly follows from Proposition 2.1 that (3.2) holds. Now, in view of the branching property of the three-type process,

$$E_1(s^{Z_{1_e}(1)}v^{Z_{1_d}(1)}t^{Z_{1_u}(1)}) = f_1(1, sQ_{1_e} + vQ_{1_d} + tQ_{1_u}),$$

and because of

$$\begin{aligned} E_1(s^{Z_{1_e}(1)}v^{Z_{1_d}(1)}t^{Z_{1_u}(1)}) &= Q_{1_e} E_{1_e}(s^{Z_{1_e}(1)}v^{Z_{1_d}(1)}t^{Z_{1_u}(1)}) \\ &\quad + Q_{1_d} E_{1_d}(s^{Z_{1_e}(1)}v^{Z_{1_d}(1)}t^{Z_{1_u}(1)}) \\ &\quad + Q_{1_u} E_{1_u}(s^{Z_{1_e}(1)}v^{Z_{1_d}(1)}t^{Z_{1_u}(1)}), \end{aligned}$$

to prove (3.3) and (3.4), it suffices to show that the sum of the last two terms equals  $f_1(q_0, vQ_{1_d} + tQ_{1_u})$ . But this is equivalent to

$$\begin{aligned} E_1(s^{Z_{1_e}(1)}v^{Z_{1_d}(1)}t^{Z_{1_u}(1)} \mathbf{1}_{\{\lim_{n \rightarrow \infty} Z_0(n)=0\}}) \\ &= E_1(v^{Z_{1_d}(1)}t^{Z_{1_u}(1)} \mathbf{1}_{\{\text{all 0-daughter processes die out}\}} \mathbf{1}_{\{Z_{1_e}(1)=0\}}) \\ &= f_1(q_0, vQ_{1_d} + tQ_{1_u}). \end{aligned}$$

### 4. Limit theorem in the two-type case

In the previous two sections we considered a two-type BGW process with a general reproduction law described by a pair of generating functions,  $f_0(s_0, s_1)$  and  $f_1(s_0, s_1)$ . In this section we deal with a family of two-type BGW processes labeled by a parameter  $0 < \mu < 1$  regulating communication rates between types 0 and 1. We will assume a particular kind of reproduction law for the type-1 particles:

$$\begin{aligned} f_1^{(\mu)}(s_0, s_1) &= \sum_{k=0}^{\infty} p_1(k)(s_1(1 - \mu a_{10}^{(\mu)}(k)) + s_0 \mu a_{10}^{(\mu)}(k))^k \\ &= \sum_{k=0}^{\infty} p_1(k)(s_1 + (s_0 - s_1)\mu a_{10}^{(\mu)}(k))^k. \end{aligned} \tag{4.1}$$

Here  $\{p_1(k)\}_{k=0}^{\infty}$  is the distribution of the total offspring number for a type-1 particle. According to (4.1), each out of  $k$  offspring independently chooses its type: type 1 with probability  $(1 - \mu a_{10}^{(\mu)}(k))$  or type 0 (mutation event) with probability  $\mu a_{10}^{(\mu)}(k)$ . Note that the offspring number  $k$  is independent of the mutation rate  $\mu$ , while, on the other hand, mutation probabilities may depend on the offspring number.



In our asymptotic analysis,  $\mu$  goes to 0, making mutations rare events. We will assume uniform convergence, i.e.

$$\sup_{k \geq 0} |a_{10}^{(\mu)}(k) - a_{10}(k)| \rightarrow 0, \quad \mu \rightarrow 0, \tag{4.2}$$

where the limit sequence is uniformly bounded, i.e.

$$C \equiv \sup_{k \geq 0} a_{10}(k) < \infty. \tag{4.3}$$

Obviously, as  $\mu \rightarrow 0$ ,  $f_1^{(\mu)}(s_0, s_1) \rightarrow \phi_1(s_1)$ , where

$$\phi_1(s) = \sum_{k=0}^{\infty} p_1(k) s^k.$$

We assume a similar convergence for the offspring numbers of the type-0 particles:

$$f_0^{(\mu)}(s_0, s_1) \rightarrow \phi_0(s_0), \tag{4.4}$$

where the limit generating function  $\phi_0(s)$  describes the limit reproduction regime of type 0 with no mutation to type 1.

The mean of the total offspring numbers  $m_i = \phi_i'(1)$ ,  $i = 0, 1$ , is always supposed to be positive and finite. Condition (4.4) implies that  $q_0^{(\mu)} \rightarrow q$ , where  $q = 1$  if  $m_0 \leq 1$  (unless  $\phi_0(s) \equiv s$ ), and  $q \in [0, 1)$ ,  $q = \phi_0(q)$  if  $m_0 > 1$ . In terms of the generating function,

$$\psi_{10}(s) = \sum_{k=1}^{\infty} k p_1(k) a_{10}(k) s^{k-1}, \quad s \in [0, 1],$$

condition (4.3) ensures that  $\psi_{10}(1) < \infty$ .

**Theorem 4.1.** *Consider the  $\mu$ -labeled two-type BGW process stemming from a type-1 particle which satisfies conditions (4.1)–(4.4). If  $m_1 < 1$  and  $\psi_{10}(1) > 0$ , then the probability of at least one mutation event  $Q_{1m}^{(\mu)}$  has the asymptotic behavior*

$$\lim_{\mu \rightarrow 0} Q_{1m}^{(\mu)} \mu^{-1} = \psi_{10}(1)(1 - m_1)^{-1}, \tag{4.5}$$

and, conditioned on at least one mutation, the process  $\{Z_{1m}(n), Z_{1u}(n)\}_{n \geq 0}$  converges in distribution to a limit process  $\{X_{1m}^*(n), X_{1u}^*(n)\}_{n \geq 0}$ , which is a decomposable two-type BGW process described below.

Furthermore, if  $m_0 > 1$  then the probability of escaping from extinction  $Q_{1e}^{(\mu)}$  has the asymptotic behavior

$$\lim_{\mu \rightarrow 0} Q_{1e}^{(\mu)} \mu^{-1} = (1 - q)\psi_{10}(1)(1 - m_1)^{-1}, \tag{4.6}$$

and, conditioned on the escape event, the process  $\{Z_{1e}(n), Z_{1d}(n), Z_{1u}(n)\}_{n \geq 0}$  converges in distribution to  $\{X_{1m}^*(n), 0, X_{1u}^*(n)\}_{n \geq 0}$ .

Because the stated weak convergence results concern Markov chains with discrete time and the set  $\{0, 1, 2, \dots\}$  as the state space, it suffices to verify convergence of the corresponding transition probabilities (see [2]). Thus, in view of Propositions 2.1 and 3.1, this theorem is a

consequence of three relations:

$$E_{1_u}^{(\mu)}(s^{Z_{1m}(1)}t^{Z_{1u}(1)}) \rightarrow \phi_1(t), \tag{4.7}$$

$$E_{1_m}^{(\mu)}(s^{Z_{1m}(1)}t^{Z_{1u}(1)}) \rightarrow m_1s \frac{\phi_1'(t)}{\phi_1'(1)} + (1 - m_1) \frac{\psi_{10}(t)}{\psi_{10}(1)}, \tag{4.8}$$

$$E_{1_e}^{(\mu)}(s^{Z_{1e}(1)}v^{Z_{1d}(1)}t^{Z_{1u}(1)}) \rightarrow m_1s \frac{\phi_1'(t)}{\phi_1'(1)} + (1 - m_1) \frac{\psi_{10}(t)}{\psi_{10}(1)}, \tag{4.9}$$

as  $\mu \rightarrow 0$ , proved in the next section. The limit generating functions in (4.7)–(4.9) imply the following reproduction rules in the limit process.

The limit  $\{X_{1_m^*}(n), X_{1_u^*}(n)\}_{n \geq 0}$  is a BGW process with types  $1_m^*$  and  $1_u^*$  corresponding to the escape (stem) lineage and extinct (side) lineages, respectively. It starts with a single  $1_m^*$ -particle, whose reproduction law is described by the limit generating function in (4.8)–(4.9). At the time of death, this particle either produces one  $1_m^*$ -particle with probability  $m_1$  or zero  $1_m^*$ -particles with probability  $(1 - m_1)$ . In both cases, it also produces a random number of  $1_u^*$ -particles: in the former case, the generating function for the number of  $1_u^*$ -offspring is  $\phi_1'(s)/m_1$  and, in the latter case, it is  $\psi_{10}(s)/\psi_{10}(1)$ . It follows that, asymptotically, the stem lineage stays alive for a geometric time  $T_1$  with mean  $E(T_1) = (1 - m_1)^{-1}$  (cf. [8]).

Relation (4.7) says that the  $1_u^*$ -particles reproduce themselves according to the generating function  $\phi_1(s)$ . Therefore, the process  $X_{1_u^*}(n)$  can be viewed as the number of particles in a BGW process with a stopped immigration. Think of the stem lineage described above as the immigration source, with every immigrant initiating an independent BGW process with the offspring generating function  $\phi_1(s)$ . At times  $1, \dots, T_1 - 1$ , the independent numbers of immigrants have a common distribution with the generating function  $\phi_1'(s)/m_1$ . At the time  $T_1$ , when the stem lineage stops, the number of immigrants has a possibly different distribution with the generating function  $\psi_{10}(s)/\psi_{10}(1)$ .

Note that if the mutation probability is independent of the family size, i.e.  $a_{10}(k) \equiv c$ , then  $\psi_{10}(s) = c\phi_1'(s)$  and  $\psi_{10}(1) = cm_1$ , so that even the last number of immigrants has the generating function  $\phi_1'(s)/m_1$ . Observe that this generating function corresponds to the so-called size-biased version of the offspring distribution  $\phi_1(s)$ ; see, for example, [7]. In this case,  $\{X_{1_m^*}(n), X_{1_u^*}(n)\}_{n \geq 0}$  becomes a size-biased version of the single-type BGW process with the offspring generating function  $\phi_1(s)$ , whose distinguished line is stopped at the geometric time  $T_1$ .

The three-dimensional limit process  $\{X_{1_m^*}(n), 0, X_{1_u^*}(n)\}_{n \geq 0}$  in the second part of the statement describes conditioning on the escape event when the particle system manages not only to produce a type-0 particle but also to escape extinction due to the supercritical reproduction rate of type-0 particles ( $m_0 > 1$ ). The fact that the first component of the limit process is the same as in the first part of the assertion confirms a common sense expectation that conditioning on the mutation event is asymptotically equivalent to conditioning on the escape event. The zero second component simply means that there are no particles of type  $1_e^*$ . In other words, the following scenario is impossible: a type-1 particle manages to produce at least one type-0 particle, but all of the mutant lineages die out.

### 5. Proof of Theorem 4.1

Throughout this section, we assume that (4.1)–(4.4) hold and that  $m_1 < 1$ . We prove (4.5)–(4.9) (where relations (4.6) and (4.9) additionally require that  $m_0 > 1$ ) using the following lemma.

**Lemma 5.1.** *As  $\mu \rightarrow 0$  uniformly over  $(s_0, s_1) \in [0, 1] \times [0, 1)$ ,*

$$f_1^{(\mu)}(s_0, s_1) = \phi_1(s_1) + \mu(s_0 - s_1)\psi_{10}(s_1) + o(\mu(s_0 - s_1)) + O\left(\mu^2(s_0 - s_1)^2 \frac{\eta(s_1)}{1 - s_1}\right),$$

where  $\eta(s) = m_1 - \phi_1'(s)$  is such that  $\eta(s) \searrow 0$  as  $s \rightarrow 1$ .

*Proof.* If  $a \neq 1$  and  $0 \leq a \leq a + b \leq 1$  then, for any  $k \in \mathbb{N}$ ,

$$0 \leq (a + b)^k - a^k - ka^{k-1}b \leq b^2 \sum_{i=1}^{k-1} (k - i)a^{i-1} \leq kb^2 \frac{1 - a^{k-1}}{1 - a}. \tag{5.1}$$

These inequalities become obvious in the light of the representation

$$\begin{aligned} (a + b)^k - a^k - ka^{k-1}b &= b \sum_{i=1}^{k-1} ((a + b)^{k-i} - a^{k-i})a^{i-1} \\ &= b^2 \sum_{i=1}^{k-1} a^{i-1} \sum_{j=1}^{k-i} (a + b)^{k-i-j} a^{j-1}. \end{aligned}$$

It follows from (5.1) that, for any  $(s_0, s_1) \in [0, 1] \times [0, 1)$ ,

$$\begin{aligned} &\left| \sum_{k=0}^{\infty} p_1(k)[s_1 + \mu(s_0 - s_1)a_{10}(k)]^k - \phi_1(s_1) - \mu(s_0 - s_1)\psi_{10}(s_1) \right| \\ &\leq \sum_{k=1}^{\infty} p_1(k) | [s_1 + \mu(s_0 - s_1)a_{10}(k)]^k - s_1^k - k\mu a_{10}(k)(s_0 - s_1)s_1^{k-1} | \\ &\leq \sum_{k=1}^{\infty} p_1(k) \left( \mu^2(s_0 - s_1)^2 a_{10}^2(k) \sum_{i=1}^{k-1} (k - i)s_1^{i-1} \right) \\ &\leq C^2 \mu^2(s_0 - s_1)^2 \sum_{k=1}^{\infty} p_1(k) k \frac{1 - s_1^{k-1}}{1 - s_1} \\ &= C^2 \mu^2(s_0 - s_1)^2 \frac{m_1 - \phi_1'(s_1)}{1 - s_1}, \end{aligned}$$

where  $C$  is as in (4.3). On the other hand, (4.1) implies that

$$\begin{aligned} &\left| f_1^{(\mu)}(s_0, s_1) - \sum_{k=0}^{\infty} p_1(k)[s_1 + \mu(s_0 - s_1)a_{10}(k)]^k \right| \\ &\leq \mu |s_0 - s_1| \sum_{k=1}^{\infty} p_1(k) k |a_{10}^{(\mu)}(k) - a_{10}(k)|, \end{aligned}$$

which gives the  $o(\mu(s_0 - s_1))$  term due to the uniform convergence condition.

*Proof of (4.5)–(4.6).* The probability  $Q_{1m}^{(\mu)}$  that a 1-particle will have at least one type-0 descendant is bounded from below by

$$p_1(k)k\mu a_{10}^{(\mu)}(k)[1 - \mu a_{10}^{(\mu)}(k)]^{k-1}$$

for all  $k = 1, 2, 3, \dots$ . Since  $\psi_{10}(1) > 0$ , there exists a  $k$  such that  $p_1(k)a_{10}(k) > 0$ . Thus, in view of condition (4.3) we can conclude that

$$\limsup_{\mu \rightarrow 0} \frac{\mu}{Q_{1m}^{(\mu)}} < \infty. \tag{5.2}$$

By Lemma 5.1,

$$f_1^{(\mu)}(0, Q_{1u}^{(\mu)}) - \phi_1(Q_{1u}^{(\mu)}) + \mu Q_{1u}^{(\mu)} \psi_{10}(Q_{1u}^{(\mu)}) = o(\mu) + O\left(\mu^2 \frac{\eta(Q_{1u}^{(\mu)})}{Q_{1m}^{(\mu)}}\right),$$

which combined with (2.2) and (5.2) yields

$$Q_{1u}^{(\mu)} - \phi_1(Q_{1u}^{(\mu)}) + \mu Q_{1u}^{(\mu)} \psi_{10}(Q_{1u}^{(\mu)}) = o(\mu) + O(\mu\eta(Q_{1u}^{(\mu)})).$$

It follows immediately that  $Q_{1u}^{(\mu)} \rightarrow 1$  and, therefore,

$$\frac{\phi_1(Q_{1u}^{(\mu)}) - Q_{1u}^{(\mu)}}{\mu} \rightarrow \psi_{10}(1).$$

This implies (4.5), since  $\phi_1(s) - s \sim (1 - m_1)(1 - s)$  as  $s \rightarrow 1$ .

Applying Lemma 5.1 once again we obtain

$$\begin{aligned} f_1^{(\mu)}(q_0^{(\mu)}, 1 - Q_{1e}^{(\mu)}) - \phi_1(1 - Q_{1e}^{(\mu)}) \\ = \mu(q_0^{(\mu)} - 1 + Q_{1e}^{(\mu)})\psi_{10}(1 - Q_{1e}^{(\mu)}) + o(\mu) + O\left(\frac{\mu^2\eta(1 - Q_{1e}^{(\mu)})}{Q_{1e}^{(\mu)}}\right), \end{aligned}$$

where  $q_0^{(\mu)} \rightarrow q$  with  $q \in [0, 1)$  given  $m_0 > 1$ . Using (3.1), we can derive

$$\begin{aligned} 1 - Q_{1e}^{(\mu)} - \phi_1(1 - Q_{1e}^{(\mu)}) \\ = \mu(q - 1 + Q_{1e}^{(\mu)})\psi_{10}(1 - Q_{1e}^{(\mu)}) + o(\mu) + O(\mu\eta(1 - Q_{1e}^{(\mu)})), \end{aligned}$$

since  $Q_{1e}^{(\mu)}/\mu$  is bounded away from 0. Dividing by  $\mu$  gives

$$\frac{\phi_1(1 - Q_{1e}^{(\mu)}) - (1 - Q_{1e}^{(\mu)})}{\mu} \rightarrow (1 - q)\psi_{10}(1),$$

and this implies (4.6), since  $\phi_1(s) - s \sim (1 - m_1)(1 - s)$  as  $s \rightarrow 1$ .

*Proof of (4.7)–(4.9).* In view of Proposition 2.1, relation (4.7) is obvious. The other two relations have similar proofs—here we give a proof of (4.9) based on the next observation. If  $0 \leq s_i \leq s_i + \delta_i \leq 1$  for  $i = 0, 1$  and  $s_1 < 1$  then, according to (5.1),

$$\begin{aligned} 0 \leq f_1^{(\mu)}(s_0 + \delta_0, s_1 + \delta_1) - f_1^{(\mu)}(s_0, s_1) - R^{(\mu)}(s_0, s_1, \delta_0, \delta_1) \\ \leq \sum_{k=1}^{\infty} kp_1(k)(\delta_1 + \mu(\delta_0 - \delta_1)a_{10}^{(\mu)}(k))^2 \frac{1 - (s_1 + \mu(s_0 - s_1)a_{10}^{(\mu)}(k))^{k-1}}{1 - s_1 - \mu(s_0 - s_1)a_{10}^{(\mu)}(k)}, \end{aligned} \tag{5.3}$$

where

$$R^{(\mu)}(s_0, s_1, \delta_0, \delta_1) = \sum_{k=1}^{\infty} kp_1(k)(s_1 + \mu(s_0 - s_1)a_{10}^{(\mu)}(k))^{k-1}(\delta_1 + \mu(\delta_0 - \delta_1)a_{10}^{(\mu)}(k)).$$

Relations (3.4) and (5.3) yield

$$\begin{aligned} 0 &\leq Q_{1_e}^{(\mu)} E_{1_e}^{(\mu)}(s^{Z_{1_e}(1)}v^{Z_{1_d}(1)}t^{Z_{1_u}(1)}) - R^{(\mu)}(q_0^{(\mu)}, s^{(\mu)}, 1 - q_0^{(\mu)}, sQ_{1_e}^{(\mu)}) \\ &\leq \sum_{k=1}^{\infty} kp_1(k)(sQ_{1_e}^{(\mu)} + \mu(1 - q_0^{(\mu)} - sQ_{1_e}^{(\mu)})a_{10}^{(\mu)}(k))^2 \\ &\quad \times \frac{1 - (s^{(\mu)} + \mu(q_0^{(\mu)} - s^{(\mu)})a_{10}^{(\mu)}(k))^{k-1}}{1 - s^{(\mu)} - \mu(q_0^{(\mu)} - s^{(\mu)})a_{10}^{(\mu)}(k)} \end{aligned}$$

with  $s^{(\mu)} = tQ_{1_u}^{(\mu)} + vQ_{1_d}^{(\mu)} \rightarrow t$ . It remains to observe that the right-hand side is  $O(\mu^2)$  and

$$\begin{aligned} &\frac{R^{(\mu)}(q_0^{(\mu)}, s^{(\mu)}, 1 - q_0^{(\mu)}, sQ_{1_e}^{(\mu)})}{Q_{1_e}^{(\mu)}} \\ &= \sum_{k=1}^{\infty} kp_1(k)(s^{(\mu)} + \mu(q_0^{(\mu)} - s^{(\mu)})a_{10}^{(\mu)}(k)t)^{k-1} \left( s + \mu \left( \frac{1 - q_0^{(\mu)}}{Q_{1_e}^{(\mu)}} - s \right) a_{10}^{(\mu)}(k) \right) \\ &\rightarrow m_1s \frac{\phi'_1(t)}{\phi'_1(1)} + (1 - m_1) \frac{\psi_{10}(t)}{\psi_{10}(1)}. \end{aligned}$$

### 6. The sequential mutation model

Suppose that we can distinguish between  $L + 1$  types of particle, labeled  $0, \dots, L$ . Type- $i$  particles can only produce particles of the types  $0, \dots, i$ , for all  $i \in [0, L]$ . Note that this sequential mutation model only partially extends the previous two-type model. We prohibit the reverse mutations for the sake of simplicity. As the asymptotic analysis of the two-type case shows, the more general sequential model with reversed mutations should lead to the same asymptotic behavior.

Let  $Z_j(n)$  be the number of type- $j$  particles existing at time  $n$  given that the branching process stems from a single particle whose type is specified by the index of the probability measure  $P_i$ . Adjusting the notation of the two-type case, set

$$f_i(s_0, s_1, \dots, s_i) = E_i(s_0^{Z_0(1)}s_1^{Z_1(1)} \dots s_i^{Z_i(1)}), \quad (s_0, s_1, \dots, s_i) \in [0, 1]^{i+1},$$

and

$$Q_{i_u} = P_i(Z_0(n) = 0 \text{ for all } n \geq 0), \quad Q_{i_m} = 1 - Q_{i_u}.$$

Then

$$Q_{i_u} = f_i(0, Q_{1_u}, Q_{2_u}, \dots, Q_{i_u}), \tag{6.1}$$

since, to avoid descendants of type 0, the progenitor itself should not have daughters of type 0 and the progenitor’s daughters should not have descendants of type 0. We split each type into two subtypes in a way similar to the decomposition of Section 2. Consider the future of a

particle of type  $i$ : with probability  $Q_{i_m}$ , it will eventually manage to produce a particle of type 0, in which case the particle is labeled  $i_m$ , otherwise, with probability  $Q_{i_u}$ , it is labeled  $i_u$ . Arguing as in Lemma 2.1, we see that the process  $\{Z_{L_m}(n), Z_{L_u}(n), \dots, Z_{1_m}(n), Z_{1_u}(n)\}_{n \geq 0}$  is a decomposable  $2L$ -type BGW process.

To describe the reproduction law in the  $2L$ -type BGW process, observe that

$$\begin{aligned} & E_i(s_i^{Z_{i_m}(1)} t_i^{Z_{i_u}(1)} \dots s_1^{Z_{1_m}(1)} t_1^{Z_{1_u}(1)}) \\ &= Q_{i_m} E_{i_m}(s_i^{Z_{i_m}(1)} t_i^{Z_{i_u}(1)} \dots s_1^{Z_{1_m}(1)} t_1^{Z_{1_u}(1)}) \\ &+ Q_{i_u} E_{i_u}(s_i^{Z_{i_m}(1)} t_i^{Z_{i_u}(1)} \dots s_1^{Z_{1_m}(1)} t_1^{Z_{1_u}(1)}), \end{aligned}$$

where the left-hand side is equal to

$$f_i(1, s_1 Q_{1_m} + t_1 Q_{1_u}, \dots, s_i Q_{i_m} + t_i Q_{i_u})$$

and

$$Q_{i_u} E_{i_u}(s_i^{Z_{i_m}(1)} t_i^{Z_{i_u}(1)} \dots s_1^{Z_{1_m}(1)} t_1^{Z_{1_u}(1)}) = f_i(0, t_1 Q_{1_u}, t_2 Q_{2_u}, \dots, t_i Q_{i_u}).$$

It follows that

$$\begin{aligned} E_{i_u}(s_i^{Z_{i_m}(1)} t_i^{Z_{i_u}(1)} \dots s_1^{Z_{1_m}(1)} t_1^{Z_{1_u}(1)}) &= \frac{f_i(0, t_1 Q_{1_u}, \dots, t_i Q_{i_u})}{Q_{i_u}}, \\ E_{i_m}(s_i^{Z_{i_m}(1)} t_i^{Z_{i_u}(1)} \dots s_1^{Z_{1_m}(1)} t_1^{Z_{1_u}(1)}) &= \frac{f_i(1, s_1 Q_{1_m} + t_1 Q_{1_u}, \dots, s_i Q_{i_m} + t_i Q_{i_u}) - f_i(0, t_1 Q_{1_u}, \dots, t_i Q_{i_u})}{Q_{i_m}}. \end{aligned} \tag{6.2}$$

Our asymptotic analysis below is an extension of the two-type case, (4.1):

$$\begin{aligned} f_i^{(\mu)}(s_0, s_1, \dots, s_i) &= E_i^{(\mu)}(s_0^{Z_0(1)} s_1^{Z_1(1)} \dots s_i^{Z_i(1)}) \\ &= \sum_{k=0}^{\infty} p_i(k) \left( s_i \left( 1 - \sum_{j=0}^{i-1} \mu^{i-j} a_{ij}^{(\mu)}(k) \right) + \sum_{j=0}^{i-1} \mu^{i-j} a_{ij}^{(\mu)}(k) s_j \right)^k \\ &= \sum_{k=0}^{\infty} p_i(k) \left( s_i + \sum_{j=0}^{i-1} \mu^{i-j} a_{ij}^{(\mu)}(k) (s_j - s_i) \right)^k, \end{aligned} \tag{6.3}$$

which says that each out of the  $k$  offspring of an  $i$ -particle independently chooses its type: it mutates to a type  $j \in [0, i - 1]$  with probability  $\mu^{i-j} a_{ij}^{(\mu)}(k)$  or retains the maternal type  $i$  with probability  $1 - \sum_{j=0}^{i-1} \mu^{i-j} a_{ij}^{(\mu)}(k)$ . Observe that, similar to the two-type case, the quantities  $a_{ij}^{(\mu)}(k)$  allow us to have mutation probabilities depending on the family size,  $k$ , on the mother's type,  $i$ , and on the daughter's type,  $j$ . Here again parameter  $\mu$  controls mutation rates so that, as  $\mu \rightarrow 0$ , mutations become rare and

$$f_i^{(\mu)}(s_0, s_1, \dots, s_i) \rightarrow \phi_i(s_i), \quad \phi_i(s) = \sum_{k=0}^{\infty} p_i(k) s^k. \tag{6.4}$$

Set  $m_i = \sum_{k=1}^{\infty} k p_i(k)$ . We will assume that all types, possibly except 0, are asymptotically subcritical, i.e.

$$0 < m_i < 1, \quad i = 1, \dots, L, \quad 0 < m_0 < \infty. \tag{6.5}$$

As in Section 2, we will assume uniform convergence, i.e.

$$\sup_{0 \leq j < i \leq L} \sup_{k \geq 0} |a_{ij}^{(\mu)}(k) - a_{ij}(k)| \rightarrow 0, \quad \mu \rightarrow 0, \tag{6.6}$$

where the limit sequences are uniformly bounded, i.e.

$$\sup_{0 \leq j < i \leq L} \sup_{k \geq 0} a_{ij}(k) < \infty. \tag{6.7}$$

Set

$$\psi_{ij}(s) = \sum_{k=1}^{\infty} k p_i(k) a_{ij}(k) s^{k-1}, \quad 0 \leq j < i \leq L, \tag{6.8}$$

and define a matrix  $A = [A_{ij}]_{i,j=0}^L$  by

$$A_{ij} = \begin{cases} 1, & i = j = 0, \\ \frac{\psi_{ij}(1)}{1 - m_i}, & 0 \leq j \leq i - 1, \\ 0, & \text{otherwise.} \end{cases}$$

Define a vector  $(\chi_0, \dots, \chi_L)$  recursively:

$$\chi_i = \sum_{j=0}^{i-1} A_{ij} \chi_j, \quad \chi_0 = 1. \tag{6.9}$$

If all  $\psi_{ij}(1) > 0$  then all components of this vector are strictly positive.

In terms of the matrix powers  $A^n = [A_{ij}^{(n)}]_{i,j=0}^L$ , we can write

$$\begin{aligned} \chi_i &= A_{i0} + \sum_{j=1}^{i-1} A_{ij} \chi_j \\ &= A_{i0} + \sum_{j=1}^{i-1} A_{ij} \left( A_{j0} + \sum_{k=1}^{j-1} A_{jk} \chi_k \right) \\ &= A_{i0}^{(2)} + \sum_{j=1}^{i-2} A_{ij}^{(2)} \chi_j \\ &= \dots \\ &= A_{i0}^{(i)} \\ &= \sum_{k=0}^{i-1} \sum_{0=j_0 < j_1 < j_2 < \dots < j_k < i} A_{ij_k} \cdots A_{j_1 0}. \end{aligned}$$

It follows from (6.9) that the  $i$ th row of the matrix  $B = [B_{ij}]_{i,j=0}^L$ , where

$$B_{ij} = \frac{\chi_j}{\chi_i} A_{ij}, \tag{6.10}$$

defines a probability distribution on the set  $\{0, \dots, i - 1\}$ . Note that the matrix powers  $A^n$  and  $B^n$  are connected by  $B_{ij}^{(n)} = (\chi_j / \chi_i) A_{ij}^{(n)}$ .

### 7. Limit theorem for the sequential mutation model

The following result partially extends the two-type Theorem 4.1. It is clear what a full extension would look like.

**Theorem 7.1.** *Consider the  $\mu$ -labeled  $2L$ -type BGW process described in Section 6 that starts from a type- $L$  particle and satisfies conditions (6.3)–(6.7). Let all  $\psi_{ij}(1) > 0$ . The probability  $Q_{L_m}$  that the process produces at least one particle of type 0 has the asymptotic behavior*

$$\lim_{\mu \rightarrow 0} Q_{L_m} \mu^{-L} = \chi_L. \tag{7.1}$$

Conditioned on the event that a 0-particle is produced, the process

$$\{Z_{L_m}(n), Z_{L_u}(n), \dots, Z_{1_m}(n), Z_{1_u}(n)\}_{n \geq 0}$$

converges in distribution to a limit process  $\{X_{L_m^*}(n), X_{L_u^*}(n), \dots, X_{1_m^*}(n), X_{1_u^*}(n)\}_{n \geq 0}$ , which is a decomposable  $2L$ -type BGW process described below.

The limit process  $\{X_{L_m^*}(n), X_{L_u^*}(n), \dots, X_{1_m^*}(n), X_{1_u^*}(n)\}_{n \geq 0}$  starts with a single particle of type  $L_m^*$ . This particle lives a geometric number  $T_L$  of generations with

$$P(T_L = n) = m_L^{n-1} (1 - m_L).$$

At times  $1, 2, \dots, T_L - 1$ , the stem  $L_m^*$ -particle gives birth to particles of type  $L_u^*$  according to a size-biased distribution with generating function  $\phi'_L(s)/m_L$ . Each particle of type  $L_u^*$  initiates an independent single-type subcritical BGW process with the offspring generating function  $\phi_L(s)$ . At time  $T_L$ , the stem particle of type  $L_m^*$  is replaced by a stem particle of type  $i_m^*$ , where the index  $i$  is chosen from the set  $\{0, 1, \dots, L - 1\}$  according to the distribution  $\{B_{L0}, \dots, B_{L,L-1}\}$ ; see (6.10). The number of  $L_u^*$ -particles born at time  $T_L$  has a different distribution: its probability generating function is given by  $\psi_{Li}(s)/\psi_{Li}(1)$ .

If  $i = 0$  then the stem lineage stops at time  $T_L$ , but otherwise, after time  $T_L$ , the scheme above is repeated with  $L$  being replaced by  $i$ . The particle of type  $i_m^*$  lives a geometric time  $T_i$  with mean  $(1 - m_i)^{-1}$ , in that

$$X_{i_m^*}(n) = \begin{cases} 1, & n \in \{T_L, \dots, T_L + T_i - 1\}, \\ 0, & \text{otherwise.} \end{cases}$$

At times  $T_L + 1, \dots, T_L + T_i - 1$ , particles of type  $i_u^*$  appear from the stem particle according to the size-biased distribution  $\phi'_i(s)/m_i$  and each one of them initiates an independent single-type subcritical BGW process with offspring generating function  $\phi_i(s)$ . At time  $T_L + T_i$ , the stem particle changes its type to  $j_m^*$ , where  $j$  is chosen from  $\{0, 1, \dots, i - 1\}$  according to the probability measure  $\{B_{i0}, \dots, B_{i,i-1}\}$ . The distribution of the number of type- $i_u^*$  particles produced at time  $T_L + T_i$  has generating function  $\psi_{ij}(s)/\psi_{ij}(1)$ . The process proceeds in the same manner until index 0 is generated by the stem lineage algorithm.

*Proof of Theorem 7.1.* The proof is similar to the proof of Theorem 4.1; therefore, here we will outline only the major changes. Lemma 5.1 can be extended to: uniformly over



$$(s_0, s_1, \dots, s_{i-1}, s_i) \in [0, 1] \times [0, 1] \times \dots \times [0, 1] \times [0, 1),$$

$$\begin{aligned} & \left| f_i^{(\mu)}(s_0, s_1, \dots, s_i) - \phi_i(s_i) - \sum_{j=0}^{i-1} \mu^{i-j} (s_j - s_i) \psi_{ij}(s_i) \right| \\ &= o\left(\sum_{j=0}^{i-1} \mu^{i-j} |s_j - s_i|\right) + O\left(\mu^{2i} \frac{\eta_i(s_i)}{1 - s_i} \left(\sum_{j=0}^{i-1} \frac{|s_j - s_i|}{\mu^j}\right)^2\right), \end{aligned} \tag{7.2}$$

where the  $\psi_{ij}$  are the functions defined in (6.8) and  $\eta_i(s) = m_i - \phi'_i(s)$ .

Convergence (7.1) is proven by induction over  $L$ . The case in which  $L = 1$  is covered by Theorem 4.1. Now assume that, for any  $j \in \{1, \dots, i - 1\}$ , it is known that  $Q_{j0}\mu^{-j} \rightarrow \chi_j$ . We prove that  $Q_{i0}\mu^{-i} \rightarrow \chi_i$  using (6.1). First observe that  $Q_{i0}\mu^{-i}$  is bounded away from 0, since  $\psi_{i,i-1}(1) > 0$ , there exists a  $k \geq 1$  such that  $p_i(k)a_{i,i-1}(k) > 0$ , and the inequality

$$Q_{im} \geq p_i(k)k\mu a_{i,i-1}^{(\mu)}(k)Q_{(i-1)m}(1 - \mu a_{i,i-1}^{(\mu)}(k))^{k-1}$$

implies that  $\overline{\lim}_{\mu \rightarrow 0} \mu^i / Q_{im} < \infty$  due to the induction assumption. Therefore, (7.2) gives

$$\begin{aligned} & \left| f_i^{(\mu)}(0, Q_{1u}, \dots, Q_{iu}) - \phi_i(Q_{iu}) - \sum_{j=0}^{i-1} \mu^{i-j} (Q_{ju} - Q_{iu}) \psi_{ij}(Q_{iu}) \right| \\ &= o\left(\sum_{j=0}^{i-1} \mu^{i-j} |Q_{ju} - Q_{iu}|\right) + O\left(\mu^i \eta_i(Q_{iu}) \left(\sum_{j=0}^{i-1} \frac{|Q_{ju} - Q_{iu}|}{\mu^j}\right)^2\right), \end{aligned}$$

which combined with (6.1) yields  $Q_{iu} \rightarrow 1$  and

$$\begin{aligned} & 1 - Q_{im} - \phi_i(1 - Q_{im}) - \sum_{j=0}^{i-1} \mu^{i-j} (Q_{im} - Q_{jm}) \psi_{ij}(1) \\ &= o\left(\sum_{j=0}^{i-1} \mu^{i-j} |Q_{im} - Q_{jm}|\right) + o\left(\mu^i \left(\sum_{j=0}^{i-1} \frac{|Q_{im} - Q_{jm}|}{\mu^j}\right)^2\right). \end{aligned}$$

It follows that

$$\phi_i(1 - Q_{im}) - 1 + Q_{im} = \mu^i \sum_{j=0}^{i-1} \chi_j \psi_{ij}(1) + O(\mu Q_{im}),$$

and (7.1) for  $L = i$  is derived from (6.9).

In order to obtain the generating functions of the reproduction law of the limit process, we need an extension of (5.3): if  $0 \leq s_j \leq s_j + \delta_j \leq 1$  for  $j \in \{0, 1, \dots, i\}$  and  $s_i < 1$ , then

$$\begin{aligned} & 0 \leq f_i^{(\mu)}(s_0 + \delta_0, \dots, s_i + \delta_i) - f_i^{(\mu)}(s_0, \dots, s_i) - R_i^{(\mu)}(s_0, \dots, s_i, \delta_0, \dots, \delta_i) \\ & \leq \sum_{k=0}^{\infty} kp_i(k) d_{k,\mu}^2(\delta_0, \dots, \delta_i) \frac{1 - d_{k,\mu}^{k-1}(s_0, \dots, s_i)}{1 - d_{k,\mu}(s_0, \dots, s_i)}, \end{aligned} \tag{7.3}$$

where  $d_{k,\mu}(s_0, \dots, s_i) = s_i + \sum_{j=0}^{i-1} \mu^{i-j}(s_j - s_i)a_{ij}^{(\mu)}(k)$  and

$$R_i^{(\mu)}(s_0, \dots, s_i, \delta_0, \dots, \delta_i) = \sum_{k=1}^{\infty} kp_i(k)d_{k,\mu}^{k-1}(s_0, \dots, s_i)d_{k,\mu}(\delta_0, \dots, \delta_i).$$

From (6.2) and (7.3), it follows that

$$\begin{aligned} & Q_{i_m} E_{i_m}(s_i^{Z_{i_m}(1)} t_i^{Z_{i_u}(1)} \dots s_1^{Z_{1_m}(1)} t_1^{Z_{1_u}(1)}) \\ &= R_i^{(\mu)}(0, t_1 Q_{1_u}, \dots, t_i Q_{i_u}, 1, s_1 Q_{1_m}, \dots, s_i Q_{i_m}) + O(\mu^{2i}), \end{aligned}$$

and it remains to check that

$$\begin{aligned} & Q_{i_m}^{-1} R_i^{(\mu)}(0, t_1 Q_{1_u}, \dots, t_i Q_{i_u}, 1, s_1 Q_{1_m}, \dots, s_i Q_{i_m}) \\ & \rightarrow s_i \phi'_i(t_i) + \sum_{j=0}^{i-1} s_j \frac{\chi_j}{\chi_i} \psi_{ij}(t_i) \\ &= m_i s_i \frac{\phi'_i(t_i)}{m_i} + (1 - m_i) \sum_{j=0}^{i-1} B_{ij} s_j \frac{\psi_{ij}(t_i)}{\psi_{ij}(1)}. \end{aligned}$$

**8. The total time to escape**

For application purposes, it is important to study the waiting time  $W_L$  to produce the escape type along an asymptotically viable path of mutations. For the sequential mutation model studied in Sections 6 and 7,  $W_L$  is a sum of a random number of independent geometric random variables. In terms of a Markov chain  $\{Y(n)\}_{n \geq 0}$ , with the transition matrix

$$D = [D_{ij}]_{i,j=0}^L, \quad D_{ij} = (1 - m_i)B_{ij} + m_i \mathbf{1}_{\{i=j\}},$$

where the  $B_{ij}$  are given by (6.10), this is the waiting time until absorption at state 0:

$$P(W_L \leq n) = P(Y(n) = 0 \mid Y(0) = L).$$

The last probability is the element  $D_{L0}^{(n)}$  of the  $n$ th step transition matrix  $D^n$ , which can be computed from the Chapman–Kolmogorov equation:

$$\begin{aligned} D_{L0}^{(n)} &= D_{L0} D_{00}^{(n-1)} + \dots + D_{LL} D_{L0}^{(n-1)} \\ &= (1 - m_L)(B_{L0} + B_{L1} D_{10}^{(n-1)} + \dots + B_{L,L-1} D_{L-1,0}^{(n-1)}) + m_L D_{L0}^{(n-1)}. \end{aligned}$$

Subtracting a similar formula for  $D_{L0}^{(n-1)}$  we obtain a recursion for the probability  $P_L(n) = P(W_L = n)$ :

$$P_L(n) = m_L P_L(n - 1) + (1 - m_L) \sum_{j=1}^{L-1} B_{Lj} P_j(n - 1).$$

Turning to the expected waiting time,

$$M_L = E(W_L) = \sum_{n=1}^{\infty} n P_L(n),$$

we derive

$$\begin{aligned}
 M_L &= \frac{1}{1 - m_L} + \sum_{j=1}^{L-1} B_{Lj} M_j \\
 &= \frac{1}{1 - m_L} + \sum_{j=1}^{L-1} \frac{B_{Lj}}{1 - m_j} + \sum_{j=1}^{L-2} B_{Lj}^{(2)} M_j \\
 &= \frac{1}{1 - m_L} + \sum_{j=1}^{L-1} \frac{B_{Lj} + B_{Lj}^{(2)} + \dots + B_{Lj}^{(L-j)}}{1 - m_j} \\
 &= \frac{1}{1 - m_L} + \sum_{j=1}^{L-1} \frac{\chi_j (A_{Lj} + \dots + A_{Lj}^{(L-j)})}{\chi_L (1 - m_j)}.
 \end{aligned}$$

Observe that the last formula is a weighted sum of the individual waiting times  $E(T_j) = (1 - m_j)^{-1}$ . The corresponding weight

$$\begin{aligned}
 \frac{\chi_j}{\chi_L} (A_{Lj} + \dots + A_{Lj}^{(L-j)}) &= \frac{A_{Lj} A_{j0}^{(j)} + \dots + A_{Lj}^{(L-j)} A_{j0}^{(j)}}{A_{L0}^{(L)}} \\
 &= P(Y(n) = j \text{ for some } n)
 \end{aligned} \tag{8.1}$$

gives the probability that the chain  $Y(n)$  visits the state  $j$  before it is absorbed at 0. Note that in the case of ‘neutral mutation’ with  $m_j = m$ ,  $j = 1, \dots, L$ , we obtain

$$\begin{aligned}
 M_L &= \frac{1}{1 - m} + \frac{1}{1 - m} \sum_{j=1}^{L-1} \frac{\chi_j (A_{Lj} + \dots + A_{Lj}^{(L-j)})}{\chi_L} \\
 &= \frac{1}{1 - m} \left( 1 + \frac{\chi_L - A_{L0} + \dots + \chi_L - A_{L0}^{(L-1)}}{\chi_L} \right) \\
 &= \frac{1}{1 - m} \left( L - \frac{A_{L0} + \dots + A_{L0}^{(L-1)}}{\chi_L} \right).
 \end{aligned}$$

Finally, we describe a case where there is a simple formula for the coefficients  $\chi_i$ . Suppose that  $a_{ij}(k) \equiv a_i(k)$  is the same for all daughter types  $j$  given the mother type  $i$ . Then, with the simplified notation  $\psi_{ij}(1) = c_i$ , we obtain

$$\begin{aligned}
 \chi_i &= \frac{c_i}{1 - m_i} \left( 1 + \sum_{k=1}^{i-1} \sum_{0 < j_1 < j_2 < \dots < j_k < i} \frac{c_{j_k}}{1 - m_{j_k}} \dots \frac{c_{j_1}}{1 - m_{j_1}} \right) \\
 &= \frac{c_i}{1 - m_i} \left( 1 + \frac{c_{i-1}}{1 - m_{i-1}} \right) \dots \left( 1 + \frac{c_1}{1 - m_1} \right).
 \end{aligned}$$

In this case we can also compute the asymptotic probability (8.1) that the random path from

type  $L$  towards type 0 visits type  $j$ :

$$\begin{aligned}
 &P(Y(n) = j \text{ for some } n) \\
 &= \frac{\chi_j}{\chi_L} \frac{c_L}{1 - m_L} \left( 1 + \sum_{k=1}^{L-j} \sum_{j < j_1 < j_2 < \dots < j_k < L} \frac{c_{j_k}}{1 - m_{j_k}} \dots \frac{c_{j_1}}{1 - m_{j_1}} \right) \\
 &= \frac{c_j}{1 - m_j} \left( 1 + \frac{c_{L-1}}{1 - m_{L-1}} \right)^{-1} \dots \left( 1 + \frac{c_j}{1 - m_j} \right)^{-1} \\
 &\quad \times \left( 1 + \frac{c_{L-1}}{1 - m_{L-1}} \right) \dots \left( 1 + \frac{c_{j+1}}{1 - m_{j+1}} \right) \\
 &= \frac{c_j}{1 + c_j - m_j}.
 \end{aligned}$$

Thus, the expected total time to escape becomes

$$M_L = \frac{1}{1 - m_L} + \sum_{j=1}^{L-1} \frac{c_j}{(1 + c_j - m_j)(1 - m_j)}.$$

In particular, if  $a_{ij}(k) \equiv 1$  then  $c_j = m_j$  and

$$M_L = \frac{1}{1 - m_L} + \sum_{j=1}^{L-1} \frac{m_j}{1 - m_j}.$$

If, furthermore,  $m_j \equiv m$  then  $\chi_j = m(1 - m)^{-j}$  and  $P(Y(n) = j \text{ for some } n) = m$ . In this special case the number of intermediate types has a binomial distribution  $\text{Bin}(L - 1, m)$  and

$$M_L = \frac{1 + (L - 1)m}{1 - m}.$$

### 9. The network mutation model

We now return to the network model described in the introduction. This model was introduced in [5] and [6]. Here particles are coded with binary sequences  $\mathbf{u} = (u_1, \dots, u_L)$  of length  $L$ . A mutation occurs if one of the  $L$  sites changes from 1 to 0 or 0 to 1. Therefore, there are  $2^L$  possible sequences which we will group into  $L + 1$  types, each containing sequences  $\mathbf{u}$  with the same number of 1s. Observe that, given a sequence  $\mathbf{u} = (u_1, \dots, u_L)$ , the number of 1s is equal to  $|\mathbf{u}| = u_1 + \dots + u_L$ . Assuming that all sequences within a type  $i$  have the same offspring number distribution described by the generating function  $\phi_i(s)$ , we arrive at an important example of the sequential mutation model allowing for backward mutations.

Let  $m_i = \phi'_i(1) \in (0, 1)$  be the mean offspring number for the virus of type  $i$  whose sequence contains  $i \in [1, L]$  1s and  $(L - i)$  0s. The sequence with all 0s,  $\mathbf{0} = (0, \dots, 0)$ , will be assigned a supercritical reproduction number  $m_0 \in (1, \infty)$ . Given the mutation rate  $\mu$  per site per generation, the mutation probability between two sequences which differ in  $j$  sites becomes  $\mu^j (1 - \mu)^{L-j}$ . Clearly, for  $j < i$ , the mutation probability between types  $i$  and  $j$  is asymptotically equivalent to  $\binom{i}{j} \mu^{i-j}$  and does not depend on the family size. Thus,

$$\psi_{ij}(s) = \binom{i}{j} \phi'_i(s), \quad \psi_{ij}(1) = \binom{i}{j} m_i,$$

implying that

$$A_{ij} = \frac{m_i}{1 - m_i} \binom{i}{j} \mathbf{1}_{\{0 \leq j \leq i-1\}} + \mathbf{1}_{\{i=j=0\}}, \quad B_{ij} = \frac{\chi_j}{\chi_i} A_{ij},$$

where  $(\chi_0, \dots, \chi_L)$  is defined recursively by (6.9).

A proper extension of Theorem 7.1 allowing for backward mutations provides an asymptotic picture of the network mutation model conditional on escape. On the sequence level the limit process starts with the sequence  $\mathbf{1} = (1, \dots, 1)$  initiating a ‘stem lineage’. Each next generation, the stem sequence either remains  $\mathbf{1}$  with probability  $m_L$  or switches to a 0-1 sequence  $\mathbf{u}_1 = (u_{11}, \dots, u_{1L})$  with probability

$$P(\mathbf{1} \rightarrow \mathbf{u}_1) = (1 - m_L) B_{L, i_1} / \binom{L}{i_1}, \quad i_1 = |\mathbf{u}_1|.$$

For a geometric number of generations,  $T_L \sim \text{Geom}(1 - m_L)$ , including the time it switches to  $\mathbf{u}_1$ , the stem sequence  $\mathbf{1}$  produces random numbers of side lineages of mutation-free  $\mathbf{1}$ -viruses. The number of such lineages per generation has generating function  $\phi'_L(s)/m_L$ . Each mutation-free  $\mathbf{1}$ -lineage is a single-type BGW process with the offspring generating function  $\phi_L(s)$ .

The mutant stem sequence  $\mathbf{u}_1$  follows the same pattern but with  $L$  replaced by  $i_1$ , which is strictly less than  $L$ , unless  $i_1 = 0$ , in which case the system stops after hitting the escape form  $\mathbf{0}$  of the virus. Thus, the sequence dynamics from  $\mathbf{1}$  towards  $\mathbf{0}$  is described by a random path  $\mathbf{1} \rightarrow \mathbf{u}_1 \rightarrow \dots \rightarrow \mathbf{u}_k = \mathbf{0}$  of a random length  $k \in [1, L + 1]$  through intermediate sequences with strictly decreasing numbers of 1s  $L > i_1 > \dots > i_k = 0$ ,  $i_j = |\mathbf{u}_j|$ . The random path forms a Markov chain with transition probabilities

$$P(\mathbf{u}_j \rightarrow \mathbf{u}_i) = (1 - m_{i_j}) B_{i_j, i_i} / \binom{i_j}{i_i}.$$

The stem lineage spends at the type  $\mathbf{u}_j$  a geometric number of generations with mean  $\lambda(i_j)$ , where  $\lambda(i) = (1 - m_i)^{-1}$ . During this time, it generates mutation-free  $\mathbf{u}_j$ -lineages. The number of such lineages per generation has generating function  $\phi'_{i_j}(s)/m_{i_j}$ . Each mutation-free  $\mathbf{u}_j$ -lineage is a single-type BGW process with the offspring generating function  $\phi_{i_j}(s)$ .

### Acknowledgements

This work has been supported by the Bank of Sweden Tercentenary Foundation. M. C. Serra thanks the Centre of Mathematics at the University of Minho and the Portuguese Foundation for Science and Technology for supporting her stay at the Chalmers University of Technology. Valuable comments of an anonymous referee have prompted a careful revision of the initial version of the paper.

### References

- [1] ATHREYA, K. B. AND NEY, P. E. (1972). *Branching Processes*. Springer, New York.
- [2] ETHIER, S. N. AND KURTZ, T. G. (1986). *Markov Processes*. John Wiley, New York.
- [3] GEIGER, J. (1999). Elementary new proofs of classical limit theorems for Galton–Watson processes. *J. Appl. Prob.* **36**, 301–309.
- [4] HACCOU, P., JAGERS, P. AND VATUTIN, V. A. (2007). *Branching Processes: Variation, Growth and Extinction of Populations*. Cambridge University Press.
- [5] IWASA, Y., MICHOR, F. AND NOWAK, M. A. (2003). Evolutionary dynamics of escape from biomedical intervention. *Proc. R. Soc. London B* **270**, 2573–2578.

- [6] IWASA, Y., MICHOR, F. AND NOWAK, M. A. (2004). Evolutionary dynamics of invasion and escape. *J. Theoret. Biol.* **226**, 205–214.
- [7] LYONS, R., PEMANTLE, R. AND PERES, Y. (1995). Conceptual proofs of  $l\log l$  criteria for mean behavior of branching processes. *Ann. Prob.* **23**, 1125–1138.
- [8] SERRA, M. C. AND HACCOU, P. (2007). Dynamics of escape mutants. *Theoret. Pop. Biol.* **72**, 167–178.
- [9] SEVAST'YANOV, B. A. (1971). *Branching Processes*. Nauka, Moscow (in Russian).
- [10] TAIB, Z. (1992). *Branching Processes and Neutral Evolution* (Lecture Notes Biomath. **93**). Springer, Berlin.