



**HAL**  
open science

# Proliferating parasites in dividing cells: Kimmel's branching model revisited.

Vincent Bansaye

► **To cite this version:**

Vincent Bansaye. Proliferating parasites in dividing cells: Kimmel's branching model revisited.. 2007.  
hal-00128224v1

**HAL Id: hal-00128224**

**<https://hal.science/hal-00128224v1>**

Preprint submitted on 31 Jan 2007 (v1), last revised 27 Jun 2008 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Proliferating parasites in dividing cells : Kimmel's branching model revisited.

Vincent Bansaye \*

January 31, 2007

## Abstract

We consider a branching model introduced by M. Kimmel for cell division with parasite infection. Cells contain proliferating parasites which are shared randomly between the two daughter cells when they divide. We determine the probability that the organism recovers, meaning that the asymptotic proportion of contaminated cells vanishes. We study the tree of contaminated cells, give the asymptotic number of contaminated cells and the asymptotic proportions of contaminated cells with a given number of parasites. This depends on domains inherited from the behavior of branching processes in random environment (BPRE) and given by the bivariate value of the means of parasite offsprings. In one of these domains, the convergence of proportions holds in probability, the limit is deterministic and given by the Yaglom quasistationary distribution. Moreover we get an interpretation of the limit of the Q-process as the size-biased quasistationary distribution.

**Key words.** Bienaymé Galton Watson process (BGW). Branching processes in random environment (BPRE). Markov chain indexed by a tree. Quasistationary distribution. Empirical measures.

**A.M.S. Classification.** 60J80, 60J85, 60K37, 92C37, 92D25, 92D30.

## 1 Introduction

We consider the following model for cell division with parasite infection. Unless otherwise specified, we start with a single cell infected with a single parasite. At each generation, each parasite multiplies independently, each cell divides into two daughter cells and the offspring of each parasite is shared independently into the two daughter cells. It is convenient to distinguish a first daughter cell called 0 and a second one called

---

\*Laboratoire de Probabilités et Modèles Aléatoires. Université Pierre et Marie Curie et C.N.R.S. UMR 7599. 175, rue du Chevaleret, 75 013 Paris, France.  
*e-mail* : bansaye@ccr.jussieu.fr

1 and to write  $Z^{(0)} + Z^{(1)}$  the number of offspring of a parasite,  $Z^{(0)}$  of which go into the first daughter cell and  $Z^{(1)}$  of which into the second one. The symmetric sharing is the case when  $(Z^{(0)}, Z^{(1)}) \stackrel{d}{=} (Z^{(1)}, Z^{(0)})$ . Even in that case, the sharing of parasites can be unequal (for example when  $\mathbb{P}(Z^{(0)}Z^{(1)} = 0) = 1$ ).

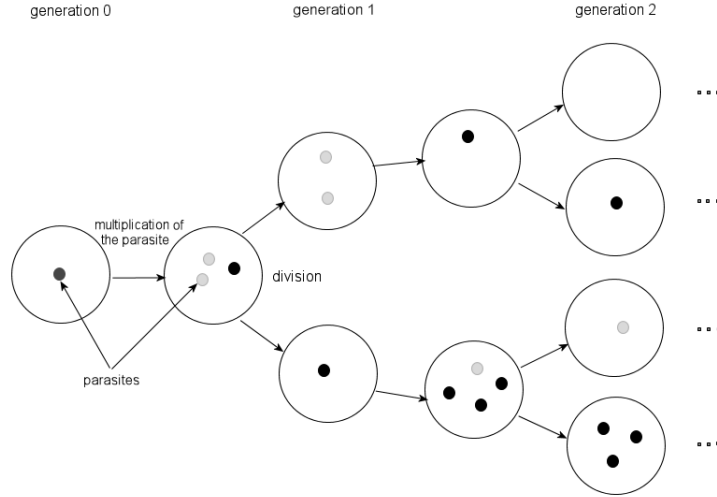
We denote by  $\mathbb{T}$  the binary genealogical tree of the cell population, by  $\mathbb{G}_n$  (resp.  $\mathbb{G}_n^*$ ) the set of cells at generation  $n$  (resp. the set of contaminated cells at generation  $n$ ) and by  $Z_{\mathbf{i}}$  the number of parasites of cell  $\mathbf{i} \in \mathbb{T}$ . That is,

$$\mathbb{G}_n := \{0, 1\}^n, \quad \mathbb{G}_n^* := \{\mathbf{i} \in \mathbb{G}_n : Z_{\mathbf{i}} > 0\}, \quad \mathbb{T} := \bigcup_{n \in \mathbb{N}} \mathbb{G}_n.$$

For every cell  $\mathbf{i} \in \mathbb{T}$ , conditionally on  $Z_{\mathbf{i}} = x$ , the numbers of parasites  $(Z_{\mathbf{i}0}, Z_{\mathbf{i}1})$  of its two daughter cells is distributed as

$$\sum_{k=1}^x (Z_k^{(0)}(\mathbf{i}), Z_k^{(1)}(\mathbf{i})),$$

where  $(Z_k^{(0)}(\mathbf{i}), Z_k^{(1)}(\mathbf{i}))_{\mathbf{i} \in \mathbb{T}, k \geq 1}$  is an iid sequence with common distribution  $(Z^{(0)}, Z^{(1)})$ .



This is a discrete version of the model introduced by M. Kimmel in [14]. Note that this model contains every model where the reproduction of parasites and the sharings of their offspring are independent and identically distributed. For example, it contains the following model with binomial repartition of parasites. Let  $Z$  be a random variable in  $\mathbb{N}$  and  $p \in [0, 1]$ . At each generation, every parasite multiplies independently with the same reproduction law  $Z$  and then every parasite of this cell chooses independently the first daughter cell with probability  $p$  (and the second one with probability  $1 - p$ ). It contains also the case when every parasite gives birth to a random cluster of parasites of size  $Z$  which goes to the first cell with probability  $p$  (and to the second one with probability  $1 - p$ ).

We introduce for  $a \in \{0, 1\}$

$$m_a := \mathbb{E}(Z^{(a)}), \quad \forall s \geq 0, f_a(s) := \mathbb{E}(s^{Z^{(a)}}), \quad p_a := \inf\{s \in [0, 1] : f_a(s) = s\}. \quad (1)$$

We assume  $0 < m_0 < \infty$ ,  $0 < m_1 < \infty$  and to avoid trivial cases, we require that

$$\mathbb{P}((Z^{(0)}, Z^{(1)}) = (1, 1)) < 1, \quad \mathbb{P}((Z^{(0)}, Z^{(1)}) \in \{(1, 0), (0, 1)\}) < 1. \quad (2)$$

This model is a Markov chain indexed by a tree. This subject has been studied in the literature (see e.g. [5, 6, 7]) in the symmetric independent case. That is,  $\forall(\mathbf{i}, k) \in \mathbb{T} \times \mathbb{N}$ ,

$$\mathbb{P}((Z_{\mathbf{i}0}, Z_{\mathbf{i}1}) = (k_0, k_1) \mid Z_{\mathbf{i}} = k) = \mathbb{P}(Z_{\mathbf{i}0} = k_0 \mid Z_{\mathbf{i}} = k)\mathbb{P}(Z_{\mathbf{i}1} = k_1 \mid Z_{\mathbf{i}} = k)$$

which require that  $Z^{(0)}$  and  $Z^{(1)}$  are iid in this model. Guyon [13] studies a Markov chain indexed by a binary tree where asymmetry and dependence are allowed and limit theorems are proved. But the case where his results apply is degenerate (this is the case  $m_0 m_1 \leq 1$  and the limit of the number of parasites in a random cell line is zero). Moreover adapting his arguments for the theorems stated here reveals cumbersome (see the remark in Section 5.2 for details). In the same vein, we refer to [9, 19] (cellular aging).

The total population of parasites at generation  $n$ , which we denote by  $\mathcal{Z}_n$ , is a Bienaymé Galton Watson process (BGW) with reproduction law  $Z^{(0)} + Z^{(1)}$ . We call Ext (resp Ext<sup>c</sup>) the event Extinction of the parasites (resp Non extinction of the parasites),

$$\mathcal{Z}_n := \sum_{\mathbf{i} \in \mathbb{G}_n} Z_{\mathbf{i}}, \quad \text{Ext} := \{\exists n \in \mathbb{N} : \mathcal{Z}_n = 0\}, \quad \text{Ext}^c := \{\forall n \in \mathbb{N} : \mathcal{Z}_n > 0\}. \quad (3)$$

Another process that appears naturally is the number of parasites in a random cell line. More precisely, let  $(a_i)_{i \in \mathbb{N}}$  be an iid sequence independent of  $(Z_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$  such that

$$\mathbb{P}(a_1 = 0) = \mathbb{P}(a_1 = 1) = 1/2. \quad (4)$$

Then  $(Z_n)_{n \in \mathbb{N}} = (Z_{(a_1, a_2, \dots, a_n)})_{n \in \mathbb{N}}$  is a Branching Process in Random Environment (BPRE). Under suitable conditions, this process conditioned to be non zero converges to the Yaglom quasistationary distribution  $\Upsilon$  [4, 11].

The first question arose from observations made by M. de Paepe, G. Paul and F. Taddei at TaMaRa's Laboratory (Hôpital Necker, Paris). They have infected the bacteria *E. Coli* with a parasite (lysogen bacteriophage M13). A fluorescent marker allows them to see the level of contamination of cells. They observed that a very contaminated cell often gives birth to a very contaminated cell which dies fast and to a much less contaminated cell whose descendance may survive. So cells tend to share unequally their parasites when they divide so that there are lots of healthy cells. It is a little surprising since one could think that cells share equally all their biological content (including parasites). In Section 3, we determine the probability of recovery of the organism (meaning that the number of infected cells becomes negligible compared to the number of cells when  $n \rightarrow \infty$ ) in function of  $(m_0, m_1)$ .

**Theorem 3.1.** *There exists a random variable  $L$  taking values in  $[0, 1]$  such that*

$$\#\mathbb{G}_n^*/2^n \xrightarrow{n \rightarrow \infty} L.$$

*If  $m_0 m_1 \leq 1$  then  $\mathbb{P}(L = 0) = 1$ . Otherwise  $\mathbb{P}(L = 0) < 1$  and  $\{L = 0\} = \text{Ext}$ .*

In Section 4, we consider the tree of contaminated cells. We denote by  $\partial\mathbb{T}$  the boundary of the cell tree  $\mathbb{T}$  and by  $\partial\mathbb{T}^*$  the infinite lines of contaminated cells, that is,

$$\partial\mathbb{T} = \{0, 1\}^{\mathbb{N}}, \quad \partial\mathbb{T}^* = \{\mathbf{i} \in \partial\mathbb{T} : \forall n \in \mathbb{N}, Z_{\mathbf{i}|n} \neq 0\}.$$

We shall prove that the contaminated cells are not concentrated in a cell line. Note that if  $m_0 + m_1 > 1$ , conditionally on  $\text{Ext}^c$ ,  $\partial\mathbb{T}^* \neq \emptyset$  since at each generation, one can choose a daughter cell whose parasite descendance does not become extinct.

The rest of the work is devoted to the convergence of the number of contaminated cells in generation  $n$  and the convergence of proportions of contaminated cells with a given number of parasites (Section 5). These asymptotics depend on  $(m_0, m_1)$  and we distinguish five different cases given by the following domains, which come from the behavior of the BGW process  $\mathcal{Z}_n$  and the BPRE  $Z_n$ .

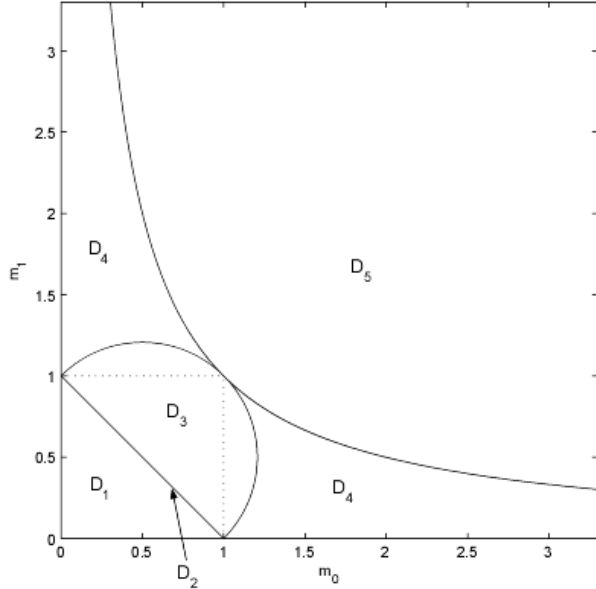
$$D_1 = \{(m_0, m_1) : m_0 + m_1 < 1\}$$

$$D_2 = \{(m_0, m_1) : m_0 + m_1 = 1\}$$

$$D_3 = \{(m_0, m_1) : m_0 + m_1 > 1, \\ m_0 + m_1 > m_0^2 + m_1^2\}$$

$$D_4 = \{(m_0, m_1) : m_0 m_1 \leq 1, \\ m_0 + m_1 \leq m_0^2 + m_1^2\}$$

$$D_5 = \{(m_0, m_1) : m_0 m_1 > 1\}$$



If  $(m_0, m_1) \in D_5$ , the contaminated cells become largely infected (Theorem 5.1). The main two results correspond to cases  $(m_0, m_1) \in D_3$  and  $(m_0, m_1) \in D_1$  and are given by the following two theorems.

**Theorem 5.2.** *If  $(m_0, m_1) \in D_3$ , conditionally on  $\text{Ext}^c$ , the following convergence holds in probability for every  $k \in \mathbb{N}$ ,*

$$\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\} / \#\mathbb{G}_n^* \xrightarrow{n \rightarrow \infty} \mathbb{P}(\Upsilon = k).$$

Note that the limit is deterministic and given by the Yaglom quasistationary distribution  $\Upsilon$  of the BPRE  $(Z_n)_{n \in \mathbb{N}}$ , which depends solely on the marginal laws of  $(Z^{(0)}, Z^{(1)})$  (see Proposition 2.3). This gives then a way to compute  $\Upsilon$  as a deterministic limit, although it is defined by conditioning on a vanishing event. Kimmel [14] considers the symmetric case  $((Z^{(0)}, Z^{(1)}) \stackrel{d}{=} (Z^{(1)}, Z^{(0)}))$  with  $m_0 = m_1 < 1 < m_0 + m_1$  in a continuous analogue of this model (cells divide after an exponential time). The counterpart of his result in the discrete case is easy to prove (see (21)) and makes a first link with  $\Upsilon$  :

$$\lim_{n \rightarrow \infty} \mathbb{E}(\#\{\mathbf{i} \in \mathbb{G}_n : Z_{\mathbf{i}} = k\}) / \mathbb{E}(\#\mathbb{G}_n^*) = \mathbb{P}(\Upsilon = k).$$

Moreover the proportions of contaminated cells on the boundary of the tree whose ancestors at generation  $n$  have a given number of parasites converge to the size-biased distribution of  $\Upsilon$  letting  $n \rightarrow \infty$  (Corollary 5.4). In the case  $Z^{(0)} \stackrel{d}{=} Z^{(1)}$ , this gives a pathwise interpretation that the limit of the Q-process associated to  $Z^{(0)}$  is the size-biased quasistationary distribution.

**Theorem 5.7.** *If  $(m_0, m_1) \in D_1$ ,  $(\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\})_{k \in \mathbb{N}}$  conditioned on  $Z_n > 0$  converges in distribution as  $n \rightarrow \infty$  to a finite random sequence  $(N_k)_{k \in \mathbb{N}}$ .*

We obtain a similar result in the case  $(m_0, m_1) \in D_2$  (Theorem 5.5). Recall that conditionally on  $\text{Ext}^c$ ,  $\#\mathbb{G}_n^*$  is asymptotically proportional to  $2^n$  if  $m_0 m_1 > 1$  (by Theorem 3.1). The previous theorems enable us to complete this result with

- ★ If  $(m_0, m_1) \in D_3$ , then conditionally on  $\text{Ext}^c$ ,  $\#\mathbb{G}_n^*/(m_0 + m_1)^n$  converges in probability to a finite positive random variable.
- ★ If  $(m_0, m_1) \in D_1$  (resp  $D_2$ ), then  $\#\mathbb{G}_n^*$  (resp  $\#\mathbb{G}_n^*/n$ ) conditioned by  $\#\mathbb{G}_n^* > 0$  converges in distribution to a finite positive random variable.

In the case  $(m_0, m_1) \in D_4$ , we get only some estimates of the asymptotic of  $\#\mathbb{G}_n^*$ . Our conjecture is that, as in the other cases,  $\#\mathbb{G}_n^*$  has a deterministic asymptotic, which depends on five subdomains (two disjoint sets and their boundaries given in Lemma 3.2). As a perspective, we are also interested in determining which types of convergences hold in  $D_4$  for the proportions of contaminated cells with a given number of parasites (see Section 5.5).

Moreover we wonder if the convergences stated above hold a.s. and if they extend to the continuous case and complement the results of Kimmel. Finally, in a work in progress with Julien Beresticky and Amaury Lambert, we aim at determining the localizations of contaminated cells and the presence of cells filled-in by parasites on the boundary of the tree (branching measure and multifractal analysis).

## 2 Preliminaries

In this section, we give some useful results about the two processes introduced above. First define :

$$m := \frac{1}{2}(m_0 + m_1), \quad \bar{m} := \frac{1}{2}(m_0^2 + m_1^2). \quad (5)$$

And recall Ulam-Harris-Neveu notation, where for every  $\mathbf{i} = (\alpha_1, \dots, \alpha_n) \in \mathbb{G}_n$ ,

$$|\mathbf{i}| = n, \quad \mathbf{i}|k = (\alpha_1, \dots, \alpha_k) \text{ for every } k \leq n, \quad \mathbf{j} < \mathbf{i} \text{ if } \exists k < n : \mathbf{i}|k = \mathbf{j}.$$

## 2.1 Results on the BGW $(Z_n)_{n \in \mathbb{N}}$

The results stated hereafter are well known and can be found in [4]. First the probability of extinction of the parasites satisfies

$$\mathbb{P}(\text{Ext}) = \inf\{s \in [0, 1] : \mathbb{E}(s^{Z^{(0)}+Z^{(1)}}) = s\} \quad ; \quad \mathbb{P}(\text{Ext}) = 1 \quad \text{iff} \quad m_0 + m_1 \leq 1/2$$

By now, we assume

$$\tilde{m} := \mathbb{E}((Z^{(0)} + Z^{(1)})\log^+(Z^{(0)} + Z^{(1)})) < \infty.$$

Then there exists a random variable  $W$  such that

$$\frac{Z_n}{(m_0 + m_1)^n} \xrightarrow{n \rightarrow \infty} W, \quad \mathbb{P}(W = 0) = \mathbb{P}(\text{Ext}), \quad \mathbb{E}(W) = 1. \quad (6)$$

In the case  $m_0 + m_1 < 1$ , there exists  $b > 0$  such that  $\mathbb{P}(Z_n > 0) \stackrel{n \rightarrow \infty}{\sim} b(m_0 + m_1)^n$ . Then, there exists a  $U > 0$  such that

$$\mathbb{P}(Z_n > 0) \geq U(m_0 + m_1)^n. \quad (7)$$

Moreover  $(Z_n)_{n \in \mathbb{N}}$  conditioned to be non zero converges to a variable called the Yaglom quasistationary distribution and we set

$$\mathcal{B}(s) := \lim_{n \rightarrow \infty} \mathbb{E}(s^{Z_n} \mid Z_n > 0). \quad (8)$$

We consider then  $\mathcal{B}_{n,k}(s) := \mathbb{E}(s^{Z_n} \mid Z_{n+k} > 0)$  which satisfies

$$\lim_{n \rightarrow \infty} \mathcal{B}_{n,k}(s) = \frac{\mathcal{B}(s) - \mathcal{B}(sf_k(0))}{1 - \mathcal{B}(f_k(0))}. \quad (9)$$

Moreover  $\mathcal{B}$  is differentiable at 1 (Lemma 1 on page 44 in [4]) and we get

$$\lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \mathcal{B}_{n,k}(s) = \frac{s\mathcal{B}'(s)}{\mathcal{B}'(1)}. \quad (10)$$

This is the probability generating function of the size-biased Yaglom quasistationary distribution, which is also the stationary distribution of the Q-process.

Finally if  $\hat{m} := \mathbb{E}((Z^{(0)} + Z^{(1)})(Z^{(0)} + Z^{(1)} - 1)) < \infty$ , we have

$$\text{If } 2m \neq 1, \mathbb{E}(Z_n(Z_n - 1)) = \hat{m}(2m)^n \frac{(2m)^n - 1}{(2m)^2 - 2m} \quad \text{If } 2m = 1, \mathbb{E}(Z_n(Z_n - 1)) = \hat{m}n \quad (11)$$

## 2.2 Properties of the BPRE $(Z_n)_{n \in \mathbb{N}}$

Recall that  $(Z_n)_{n \in \mathbb{N}}$  is the population of parasites in a uniform random cell line. Then  $(Z_n)_{n \in \mathbb{N}}$  is the amount of parasites with the following reproduction law. For each  $n \in \mathbb{N}$ , conditionally on  $a_n = a$  with  $a \in \{0, 1\}$  (see (4)), all parasites behave independently of one another and each of them gives birth to  $Z^{(a)}$  children. The size of

the population at generation 0 is denoted by  $k$  and we note  $\mathbb{P}_k$  the associated probability. Unless otherwise mentioned, the initial state is equal to 1. Thus  $(Z_n)_{n \in \mathbb{N}}$  is a BPRE with two equiprobable environments. For the general theory, see e.g. [8, 11, 12, 18]. In the case  $Z^{(0)} \stackrel{d}{=} Z^{(1)}$ ,  $(Z_n)_{n \in \mathbb{N}}$  is a BGW with reproduction law  $Z^{(0)}$ .

For  $\mathbf{i} = (\alpha_1, \dots, \alpha_n) \in \mathbb{G}_n$ , we define

$$f_{\mathbf{i}} := f_{\alpha_1} \circ \dots \circ f_{\alpha_n}, \quad m_{\mathbf{i}} = \prod_{i=1}^n m_{\alpha_i},$$

and for all  $(n, k) \in \mathbb{N} \times \mathbb{N}^*$  and  $\mathbf{i} \in \mathbb{G}_n$ ,

$$\mathbb{E}_k(s^{Z_n} \mid (a_1, \dots, a_n) = \mathbf{i}) = f_{\mathbf{i}}(s)^k.$$

Then for all  $(n, k) \in \mathbb{N} \times \mathbb{N}^*$  and  $s \in [0, 1]$ ,

$$\mathbb{E}_k(s^{Z_n}) = 2^{-n} \sum_{\mathbf{i} \in \mathbb{G}_n} f_{\mathbf{i}}(s)^k. \quad (12)$$

**Lemma 2.1.** For every  $n \in \mathbb{N}$ ,  $\mathbb{E}(Z_{n+1} \mid Z_n) = mZ_n$ ,  $\mathbb{E}(Z_n) = m^n$ .

Assume  $\tilde{m} := \frac{1}{2}(\mathbb{E}(Z^{(0)}(Z^{(0)} - 1) + Z^{(1)}(Z^{(1)} - 1))) < \infty$ , then

$$\text{If } m \neq \tilde{m}, \mathbb{E}(Z_n(Z_n - 1)) = \tilde{m} \frac{m^n - \tilde{m}^n}{m - \tilde{m}}. \quad \text{If } m = \tilde{m}, \mathbb{E}(Z_n(Z_n - 1)) = n\tilde{m}m^{n-1}.$$

*Proof.* We write  $u_n = \mathbb{E}(Z_n(Z_n - 1))$ , then :

$$\begin{aligned} u_{n+1} &= \sum_{\mathbf{i} \in \mathbb{G}_{n+1}} 2^{-(n+1)} f_{\mathbf{i}}''(1) = \sum_{\mathbf{i} \in \mathbb{G}_n, a \in \{0,1\}} 2^{-(n+1)} (f_a \circ f_{\mathbf{i}})''(1) \\ &= \tilde{m} \sum_{\mathbf{i} \in \mathbb{G}_n} 2^{-n} f_{\mathbf{i}}'(1)^2 + mu_n = \tilde{m}\bar{m}^n + mu_n \end{aligned}$$

since expanding  $(m_0^2 + m_1^2)^n$ , we get

$$\bar{m}^n = \sum_{\mathbf{i} \in \mathbb{G}_n} 2^{-n} m_{\mathbf{i}}^2 = \sum_{\mathbf{i} \in \mathbb{G}_n} 2^{-n} f_{\mathbf{i}}'(1)^2. \quad (13)$$

This gives the result by induction.  $\square$

As  $(\mathbb{P}(Z_n = 0))_{n \in \mathbb{N}}$  is an increasing sequence, it converges to the probability of extinction  $p$  of the process. Recalling (1) and using [18] or [2], we get the following result.

**Proposition 2.2.** If  $m_0 m_1 \leq 1$ ,  $p = 1$ .

Otherwise  $p < 1$  and :  $p \in ]p_0, p_1[$  if  $p_0 < p_1$ ,  $p = p_0$  if  $p_0 = p_1$ .



*Proof.* Use  $\mathbb{P}(Z_n = 0) = 2^{-n} \sum_{\mathbf{i} \in \mathbb{G}_n} f_{\mathbf{i}}(0)$  and

$$\min\{f_{\mathbf{i}}(0) : \mathbf{i} \in \{0, 1\}^n\} \uparrow_{n \rightarrow \infty} \min(p_0, p_1), \quad \sup\{f_{\mathbf{i}}(0) : \mathbf{i} \in \{0, 1\}^n\} \downarrow_{n \rightarrow \infty} \max(p_0, p_1).$$

to compare  $p$ ,  $p_0$  and  $p_1$ .  $\square$

We can also estimate  $(\mathbb{P}(Z_n > 0))_{n \in \mathbb{N}}$  when  $n$  is large. In that purpose, we require the following condition (which entails that  $m_0 m_1 \leq 1$ )

$$m_0 \log(m_0) + m_1 \log(m_1) < 0 \quad ; \quad \mathbb{E}(Z_a \log^+(Z_a)) < \infty. \quad (14)$$

Then there exists  $c > 0$  such that as  $n$  tends to  $\infty$  (Theorem 1.1 in [11]),

$$\mathbb{P}(Z_n > 0) \sim cm^n. \quad (15)$$

The last displayed relation ensures the weak convergence of the process  $Z_n$  conditioned to be non zero which is denoted by  $Z_n^*$  (Theorem 1.1 in [11]). By analogy with BGW, we call its limit distribution the Yaglom quasistationary distribution and denote it by  $\Upsilon$ . That is,

$$\forall s \in [0, 1], \quad \mathbb{E}(s^{Z_n} | Z_n > 0) \xrightarrow{n \rightarrow \infty} \mathbb{E}(s^\Upsilon) = G(s).$$

Moreover the limit  $\Upsilon$  is characterized by

**Proposition 2.3.**  *$G$  is the unique probability generating function solution of*

$$G(0) = 0, \quad G'(1) < \infty, \quad \frac{G(f_0(s)) + G(f_1(s))}{2} = mG(s) + (1 - m). \quad (16)$$

To prove the uniqueness, we need and prove below the following result.

**Lemma 2.4.** *If  $H : [0, 1] \mapsto \mathbb{R}$  is continuous,  $H(1) = 0$  and*

$$H = \frac{H \circ f_0 \cdot f_0' + H \circ f_1 \cdot f_1'}{2m}, \quad (17)$$

*then  $H = 0$ .*

*Proof of Proposition 2.3.* The finiteness of  $G'(1) = \mathbb{E}(\Upsilon)$  is the second part of Theorem 1.1 in [11].

We characterize now the probability generating function  $G$  of the limit distribution :

$$\begin{aligned} & 1 - \mathbb{E}(s^{Z_{n+1}} | Z_{n+1} > 0) \\ = & \frac{1 - \mathbb{E}(s^{Z_{n+1}})}{\mathbb{P}(Z_{n+1} > 0)} = \frac{1}{\mathbb{P}(Z_{n+1} > 0)} \sum_{i=1}^{\infty} \mathbb{P}(Z_n = i) (1 - \mathbb{E}_i(s^{Z_1})^i) \\ = & \frac{\mathbb{P}(Z_n > 0)}{\mathbb{P}(Z_{n+1} > 0)} \frac{1}{\mathbb{P}(Z_n > 0)} \sum_{i=1}^{\infty} \mathbb{P}(Z_n = i) \left(1 - \frac{f_0(s)^i + f_1(s)^i}{2}\right) \\ = & \frac{\mathbb{P}(Z_n > 0)}{\mathbb{P}(Z_{n+1} > 0)} (1 - \mathbb{E}(f_0(s)^{Z_n} | Z_n > 0) + 1 - \mathbb{E}(f_1(s)^{Z_n} | Z_n > 0))/2 \end{aligned}$$

And (15) ensures that  $\mathbb{P}(Z_n > 0)/\mathbb{P}(Z_{n+1} > 0) \xrightarrow{n \rightarrow \infty} m^{-1}$ , so that

$$1 - G(s) = \frac{1 - G(f_0(s)) + 1 - G(f_1(s))}{2m}.$$

Finally we prove the uniqueness of solutions of this equation. Let  $G$  and  $F$  be two probability generating functions which are solutions of (16). Choose  $\alpha > 0$  such that  $G'(1) = \alpha F'(1)$ . Putting  $H := G - \alpha F$ ,  $H'$  satisfies equation (17) and  $H'(1) = 0$ . Thus Lemma 2.4 gives  $H' = 0$ . As  $H(0) = 0$ ,  $H = 0$ . Moreover  $F(1) = G(1) = 1$ , so  $\alpha = 1$  and  $F = G$ .  $\square$

*Proof of Lemma 2.4.* If  $H \neq 0$  then there exists  $\alpha \in [0, 1[$  such that

$$\sup\{| H(s) | : s \in [0, \alpha]\} \neq 0$$

Let  $\alpha_n \in [0, 1[$  such that  $\alpha_n \xrightarrow{n \rightarrow \infty} 1$  and  $\alpha \leq \min(f_0(\alpha_n), f_1(\alpha_n)) \leq 1$ . Then for every  $n \in \mathbb{N}$ , there exists  $\beta_n \in [0, \alpha_n]$  such that :

$$\begin{aligned} \sup\{| H(s) | : s \in [0, \alpha_n]\} &= H(\beta_n) \\ &\leq \frac{H(f_0(\beta_n))f'_0(\beta_n) + H(f_1(\beta_n))f'_1(\beta_n)}{2m} \quad \text{using (17)} \\ &\leq \sup\{| H(s) | : s \in [0, \max(f_0(\alpha_n), f_1(\alpha_n))]\} \frac{f'_0(\beta_n) + f'_1(\beta_n)}{2m} \\ &< \sup\{| H(s) | : s \in [0, \max(f_0(\alpha_n), f_1(\alpha_n))]\} \end{aligned}$$

since  $\sup\{| H(s) | : s \in [0, \max(f_0(\alpha_n), f_1(\alpha_n))]\} \neq 0$  and  $(2m)^{-1}(f'_0(\beta_n) + f'_1(\beta_n)) < 1$ . As  $\sup(I) < \sup(I \sqcup J) \Rightarrow \sup(I) < \sup(J)$  for  $I, J$  disjoint subsets of  $\mathbb{R}$ , we get

$$\sup\{| H(s) | : s \in [0, \alpha_n]\} < \sup\{| H(s) | : s \in ]\alpha_n, \max(f_0(\alpha_n), f_1(\alpha_n))]\}.$$

And  $H(s) \xrightarrow{s \rightarrow 1} 0$  leads to a contradiction letting  $n \rightarrow \infty$ . So  $H = 0$ .  $\square$

Finally we have also the following expected result in the supercritical case [3].

**Proposition 2.5.** *If  $m_0 m_1 > 1$ ,  $\mathbb{P}(Z_n \xrightarrow{n \rightarrow \infty} \infty \mid \forall n \in \mathbb{N} : Z_n > 0) = 1$ .*

### 3 Probability of recovery

We say that the organism recovers if the number of contaminated cells becomes negligible compared to the number of cells when  $n \rightarrow \infty$ . We determine here the probability of this event. Actually if this probability is not equal to 1, then the parasites must die out for the organism to recover.

**Theorem 3.1.** *There exists a random variable  $L$  taking values in  $[0, 1]$  such that*

$$\#\mathbb{G}_n^*/2^n \xrightarrow{n \rightarrow \infty} L$$

*If  $m_0 m_1 \leq 1$  then  $\mathbb{P}(L = 0) = 1$ . Otherwise  $\mathbb{P}(L = 0) < 1$  and  $\{L = 0\} = \text{Ext}$ .*

*Remark.* This theorem states how unequal the sharing of parasites must be for the organism to recover. More precisely, let  $M = m_0 + m_1 > 0$  be the parasite growth rate and  $m_0 = \alpha M$ ,  $m_1 = (1 - \alpha)M$ . Then the organism recovers with probability 1 iff and

$$M \leq 2 \quad \text{or} \quad \alpha \notin \left] (1 - \sqrt{1 - 4/M^2})/2, (1 + \sqrt{1 - 4/M^2})/2 \right[ \quad (M > 2).$$

Note that in the case of a binomial repartition of parasites with parameter  $p$  (see Introduction),  $\alpha = p$ .

In the case  $m > 1/2$  and  $m_0 m_1 \leq 1$ , the population of parasites may explode although the organism recovers. We will focus more closely on this interesting case in Section 5.

Note that for all  $n \in \mathbb{N}$ ,

$$\mathbb{E}\left(\frac{\#\mathbb{G}_n^*}{2^n}\right) = \frac{\mathbb{E}(\sum_{i \in \mathbb{G}_n} \mathbb{1}_{Z_i > 0})}{2^n} = \mathbb{P}(Z_n > 0)$$

and recall that  $p$  is the probability of extinction of  $(Z_n)_{n \in \mathbb{N}}$  given by Proposition 2.2.

**Lemma 3.2.** *For all  $n \in \mathbb{N}$ ,*

$$\mathbb{E}\left(\frac{\#\mathbb{G}_n^*}{2^n}\right) = \mathbb{P}(Z_n > 0) \xrightarrow{n \rightarrow \infty} 1 - p$$

The last equality gives also the asymptotic of  $\mathbb{E}(\#\mathbb{G}_n^*)/2^n$  as  $n \rightarrow \infty$  in the case  $m_0 m_1 < 1$  (it depends on the sign of  $m_0 \log(m_0) + m_1 \log(m_1)$ , use [11]) and in the case  $m_0 m_1 = 1$  (see [1, 15]).

*Proof of Theorem 3.1.* As  $\#\mathbb{G}_n^*/2^n$  decreases as  $n$  increases, it converges as  $n \rightarrow \infty$ .

Monotone convergence of  $\#\mathbb{G}_n^*/2^n$  to  $L$  as  $n \rightarrow \infty$  and the previous lemma ensure  $\mathbb{E}(L) = 1 - p$ . This proves  $\mathbb{P}(L = 0) = 1$  iff  $m_0 m_1 \leq 1$ .

Obviously  $\{L = 0\} \supset \text{Ext}$ . Denote by  $\mathcal{P}(n)$  the set of parasites at generation  $n$  and for every  $\mathbf{p} \in \mathcal{P}(n)$ , denote by  $N_k(\mathbf{p})$  the number of cells at generation  $n + k$  which contain at least a parasite whose ancestor is  $\mathbf{p}$ . Then, for every  $n \in \mathbb{N}$ ,

$$\{L = 0\} = \bigcap_{\mathbf{p} \in \mathcal{P}(n)} \left\{ \frac{N_k(\mathbf{p})}{2^k} \xrightarrow{k \rightarrow \infty} 0 \right\}.$$

As  $T_n := \inf\{k \geq 0 : \mathcal{Z}_k \geq n\}$  is a stopping time associated to the natural filtration of  $(Z_i)_{i \leq n}$ , strong Markov property gives

$$\mathbb{P}(L = 0) \leq \mathbb{P}(T_n < \infty) \mathbb{P}(L = 0)^n + \mathbb{P}(T_n = \infty)$$

If  $\mathbb{P}(L = 0) < 1$ , letting  $n \rightarrow \infty$  gives

$$\mathbb{P}(L = 0) \leq \lim_{n \rightarrow \infty} \mathbb{P}(T_n = \infty) = \mathbb{P}(\mathcal{Z}_n \text{ is bounded}) = \mathbb{P}(\text{Ext})$$

since  $\mathcal{Z}_n$  is a BGW. This completes the proof. One can also use a coupling argument : the number of contaminated cells starting with one single cell with  $n$  parasites is less than the number of contaminated cells starting from  $n$  cells with one single parasite.  $\square$

## 4 Tree of contaminated cells

Here, we prove that contaminated cells are not concentrated in a cell line. If  $m_0 + m_1 \leq 1$ , contaminated cells die out but conditionally on the survival of parasites at generation  $n$ , the number of leaves of the tree of contaminated cells tends to  $\infty$  as  $n \rightarrow \infty$ . The proof of this result will also ensure that, if  $m_0 + m_1 > 1$ , the number of contaminated cells tends to  $\infty$  provided that they do not die out.

**Theorem 4.1.** *If  $m_0 + m_1 \leq 1$ ,  $\#\{\mathbf{i} \in \mathbb{T} : Z_{\mathbf{i}} \neq 0, Z_{\mathbf{i}0} = 0, Z_{\mathbf{i}1} = 0\}$  conditioned by  $\#\mathbb{G}_n^* > 0$  converges in probability as  $n \rightarrow \infty$  to  $\infty$ . If  $m_0 + m_1 > 1$ , conditionally on  $\text{Ext}^c$ ,  $\#\mathbb{G}_n^* \xrightarrow{n \rightarrow \infty} \infty$  a.s.*

*Remark.* In the conditions of the theorem,  $\#\mathbb{G}_n^*$  (resp. the number of leaves) grows at least linearly with respect to  $n$ . In Section 5, we determine the asymptotic of  $\#\mathbb{G}_n^*$  (resp.  $\#\mathbb{G}_n^*$  conditioned by  $\#\mathbb{G}_n^* > 0$ ) when  $m_0 + m_1 > 1$  (resp.  $m_0 + m_1 \leq 1$ ). In the case  $m_0 + m_1 \leq 1$ , conditionally on  $\#\mathbb{G}_n^* > 0$ , the tree of contaminated cells is a spine with finite subtrees, as for BGW conditioned to survive (see [10, 16]).

We need two lemmas for the proof. First we prove that the ancestor of a contaminated cell has given birth to two contaminated cells with a probability bounded from below. We have to distinguish the case where  $\mathbb{P}(Z^{(0)}Z^{(1)} = 0) = 1$ , since in that case a cell must contain at least two parasites so that it can give birth to two contaminated cells.

**Lemma 4.2.** *There exists  $\alpha > 0$  such that for all  $N \in \mathbb{N}$ ,  $\mathbf{i} \in \mathbb{G}_N$ ,  $n < N$  and  $k \geq 2$ ,*

$$\mathbb{P}(Z_{\mathbf{j}0} \neq 0, Z_{\mathbf{j}1} \neq 0 \mid Z_{\mathbf{j}} = k, Z_{\mathbf{i}} > 0) \geq \alpha$$

denoting  $\mathbf{j} = \mathbf{i} \mid n$ . If  $\mathbb{P}(Z^{(0)}Z^{(1)} = 0) \neq 1$ , this result also holds for  $k = 1$ .

*Proof.* We consider first the case  $\mathbb{P}(Z^{(0)}Z^{(1)} = 0) \neq 1$  and we choose  $(k_0, k_1) \in \mathbb{N}^{*2}$  such that  $\mathbb{P}((Z^{(0)}, Z^{(1)}) = (k_0, k_1)) > 0$ . For every  $k \in \mathbb{N}^*$ , we have

$$\mathbb{P}(Z_{\mathbf{j}0} \neq 0, Z_{\mathbf{j}1} \neq 0 \mid Z_{\mathbf{j}} = k, Z_{\mathbf{i}} > 0) \geq \mathbb{P}(Z_{\mathbf{j}0} \neq 0, Z_{\mathbf{j}1} \neq 0 \mid Z_{\mathbf{j}} = 1, Z_{\mathbf{i}} > 0)$$

Moreover, as the function  $\mathbb{R}_+^* \ni u \mapsto \frac{1-e^{-u}}{u}$  decreases, we have for all  $y, x > 0$  and  $p \in [0, 1[$ ,

$$\frac{1-p^x}{1-p^y} \geq \frac{x}{\max\{y, x\}}. \quad (18)$$

Let  $a \in \{0, 1\}$  and  $\mathbf{k}$  such that  $\mathbf{i} = \mathbf{j}a\mathbf{k}$ . Then for all  $(k'_0, k'_1) \in \mathbb{N}^2 - (0, 0)$ , we have

$$\begin{aligned} & \frac{\mathbb{P}(Z_{\mathbf{j}0} = k_0, Z_{\mathbf{j}1} = k_1 \mid Z_{\mathbf{j}} = 1, Z_{\mathbf{i}} > 0)}{\mathbb{P}(Z_{\mathbf{j}0} = k'_0, Z_{\mathbf{j}1} = k'_1 \mid Z_{\mathbf{j}} = 1, Z_{\mathbf{i}} > 0)} \\ &= \frac{\mathbb{P}(Z^{(0)} = k_0, Z^{(1)} = k_1 \mid Z_{a\mathbf{k}} > 0)}{\mathbb{P}(Z^{(0)} = k'_0, Z^{(1)} = k'_1 \mid Z_{a\mathbf{k}} > 0)} \\ &= \frac{\mathbb{P}(Z_{a\mathbf{k}} > 0 \mid Z^{(0)} = k_0, Z^{(1)} = k_1) \mathbb{P}(Z^{(0)} = k_0, Z^{(1)} = k_1)}{\mathbb{P}(Z_{a\mathbf{k}} > 0 \mid Z^{(0)} = k'_0, Z^{(1)} = k'_1) \mathbb{P}(Z^{(0)} = k'_0, Z^{(1)} = k'_1)} \\ &= \frac{1 - \mathbb{P}(Z_{\mathbf{k}} = 0)^{k_a} \mathbb{P}((Z^{(0)}, Z^{(1)}) = (k_0, k_1))}{1 - \mathbb{P}(Z_{\mathbf{k}} = 0)^{k'_a} \mathbb{P}((Z^{(0)}, Z^{(1)}) = (k'_0, k'_1))} \\ &\geq \frac{\min\{k_0, k_1\}}{k_0 + k_1 + k'_0 + k'_1} \frac{\mathbb{P}((Z^{(0)}, Z^{(1)}) = (k_0, k_1))}{\mathbb{P}((Z^{(0)}, Z^{(1)}) = (k'_0, k'_1))} \quad \text{using (18)} \end{aligned}$$

Cross product and sum over  $(k'_0, k'_1)$  give

$$\begin{aligned} & [\mathbb{E}(Z^{(0)} + Z^{(1)}) + k_0 + k_1] \mathbb{P}(Z_{j0} = k_0, Z_{j1} = k_1 \mid Z_j = 1, Z_i > 0) \\ & \geq \min\{k_0, k_1\} \mathbb{P}((Z^{(0)}, Z^{(1)}) = (k_0, k_1)). \end{aligned}$$

This gives the result since  $\mathbb{P}(Z_{j0} = k_0, Z_{j1} = k_1 \mid Z_j = 1, Z_i > 0) \geq \alpha$  with

$$\alpha = \frac{\min\{k_0, k_1\} \mathbb{P}((Z^{(0)}, Z^{(1)}) = (k_0, k_1))}{\mathbb{E}(Z^{(0)} + Z^{(1)}) + k_0 + k_1} > 0.$$

In the case  $\mathbb{P}(Z^{(0)}Z^{(1)} = 0) = 1$ , we choose  $(k_0, k_1) \in \mathbb{N}^{*2}$  such that  $\mathbb{P}_2((Z_0, Z_1) = (k_0, k_1)) > 0$  (using (2)). We make then the same proof as above with  $Z_j = 2$  and

$$\alpha = \frac{\min\{k_0, k_1\} \mathbb{P}_2((Z_0, Z_1) = (k_0, k_1))}{\mathbb{E}_2(Z_0 + Z_1) + k_0 + k_1},$$

so that the result follows as previously.  $\square$

Thus if  $\mathbb{P}(Z^{(0)}Z^{(1)} = 0) = 1$ , we need to prove that there are many cells with more than two parasites in a contaminated cell line.

**Lemma 4.3.** *If  $\beta := \mathbb{P}(Z^{(0)} \geq 2 \text{ or } Z^{(1)} \geq 2) > 0$  then*

$$\inf_{\mathbf{i} \in \mathbb{G}_n} \mathbb{P}(\#\{\mathbf{j} < \mathbf{i} : Z_{j0} \geq 2 \text{ or } Z_{j1} \geq 2\} \geq \beta n/2 \mid Z_i > 0) \xrightarrow{n \rightarrow \infty} 1.$$

*Proof.* For all  $\mathbf{i} \in \mathbb{G}_n$  and  $\mathbf{j} < \mathbf{i}$ , let  $\mathbf{k}$  such that  $\mathbf{i} = \mathbf{j}\mathbf{k}$ , then for all  $\alpha > 0$ ,

$$\mathbb{P}(Z_{j0} \geq 2 \text{ or } Z_{j1} \geq 2 \mid Z_j = \alpha, Z_i > 0) \geq \mathbb{P}(Z_0 \geq 2 \text{ or } Z_1 \geq 2 \mid Z_{\mathbf{k}} > 0) \geq \beta$$

Then conditionally on  $Z_i > 0$ ,  $\#\{\mathbf{j} < \mathbf{i} : Z_{j0} \geq 2 \text{ or } Z_{j1} \geq 2\} \geq \sum_{k=0}^n \beta_k$ , where  $(\beta_k)_{1 \leq k \leq n}$  are iid and distributed as a Bernoulli( $\beta$ ). Conclude with the law of large numbers.  $\square$

*Proof of Theorem 4.1.* We consider first the case when  $m_0 + m_1 > 1$ , work conditionally on  $\text{Ext}^c$  and choose  $\mathbf{i} \in \delta\mathbb{T}^*$ .

If  $\mathbb{P}(Z^{(0)}Z^{(1)} = 0) \neq 1$ , Lemma 4.2 (with  $k = 1$ ) entails that a.s. under  $\mathbb{P}(\cdot \mid Z_i > 0)$ ,

$$\#\{\mathbf{j} < \mathbf{i} : Z_{j0} > 0, Z_{j1} > 0\} = \infty.$$

Using the branching property and the fact that the probability of non-extinction of parasites is positive ensures that  $\#\mathbb{G}_n^* \xrightarrow{n \rightarrow \infty} \infty$ .

If  $\mathbb{P}(Z^{(0)}Z^{(1)} = 0) = 1$  then  $\mathbb{P}(Z^{(0)} \geq 2 \text{ or } Z^{(1)} \geq 2) > 0$  and by Lemma 4.3, we have a.s. on  $\mathbb{P}(\cdot \mid Z_i > 0)$ ,

$$\#\{\mathbf{j} < \mathbf{i} : Z_{j0} \geq 2 \text{ or } Z_{j1} \geq 2\} = \infty.$$

Using as above Lemma 4.2 (with  $k = 2$ ) and the fact that the probability of non-extinction of parasites is positive ensures that  $\#\mathbb{G}_n^* \xrightarrow{n \rightarrow \infty} \infty$  a.s on  $\text{Ext}^c$ .

We consider now the case when  $m_0 + m_1 \leq 1$  and work conditionally on  $\mathbf{i} = (\alpha_1, \dots, \alpha_n) \in \mathbb{G}_n^*$ . We denote  $\mathbf{i}_j := (\alpha_1, \dots, \alpha_{j-1}, 1 - \alpha_j)$  for  $1 \leq j \leq n$ .

If  $\mathbb{P}(Z^{(0)}Z^{(1)} = 0) \neq 1$ , Lemma 4.2 entails that

$$\forall 1 \leq j \leq n, k \geq 1, \quad \mathbb{P}(Z_{\mathbf{i}_j} > 0 \mid Z_{\mathbf{i}_{|j-1}} = k, Z_{\mathbf{i}} > 0) \geq \alpha. \quad (19)$$

Moreover if  $Z_{\mathbf{i}_j} > 0$ , then the tree of contaminated cells rooted in  $\mathbf{i}_j$  dies out and so has at least one leaf. So by the branching property, the number of leaves converges in probability to infinity as  $n$  tends to infinity.

If  $\mathbb{P}(Z^{(0)}Z^{(1)} = 0) = 1$ , (19) holds for  $k \geq 2$  and Lemma 4.3 allows to conclude similarly in this case.  $\square$

## 5 Proportion of contaminated cells with a given number of parasites

We determine here the asymptotic behavior of the proportion  $F_k$  of cells with  $k$  parasites among the contaminated cells, defined as

$$F_k(n) := \frac{\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\}}{\#\mathbb{G}_n^*}.$$

In that view, we introduce the Banach space  $l^1(\mathbb{N})$  and the subset of frequencies  $\mathbb{S}^1(\mathbb{N})$  which we endow with the norm  $\|\cdot\|_1$  defined by :

$$l^1(\mathbb{N}) := \{(x_i)_{i \in \mathbb{N}} : \sum_{i=0}^{\infty} |x_i| < \infty\}, \quad \|(x_i)_{i \in \mathbb{N}}\|_1 = \sum_{i=0}^{\infty} |x_i|,$$

$$\mathbb{S}^1(\mathbb{N}) := \{(f_i)_{i \in \mathbb{N}} : \forall i \in \mathbb{N}, f_i \in \mathbb{R}^+, \sum_{i=0}^{\infty} f_i = 1\}.$$

We shall work conditionally on  $\text{Ext}^c$  or  $\mathcal{Z}_n > 0$  and introduce the probabilities

$$\mathbb{P}^* := \mathbb{P}(\cdot \mid \text{Ext}^c), \quad \mathbb{P}^n := \mathbb{P}(\cdot \mid \mathcal{Z}_n > 0). \quad (20)$$

The asymptotics of the proportions depend naturally on the distribution of  $(Z^{(0)}, Z^{(1)})$  and we determine five different behaviors according to the value of  $(m_0, m_1)$ .

The proofs of the convergences use the asymptotic distribution of the number of parasites of a typical contaminated cell at generation  $n$ , which is equal to  $\mathbb{P}^n(Z_{\mathbf{U}_n} \in \cdot)$ ,

where  $\mathbf{U}_n$  is a uniform random variable in  $\mathbb{G}_n^*$  independent of  $(Z_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}^*}$ . This distribution is different from the distribution of  $Z_n^*$ , that is the number of parasites of a random cell line conditioned to be contaminated at generation  $n$ . The following example even proves that  $\mathbb{P}^n(Z_{\mathbf{U}_n} \in \cdot)$  and  $\mathbb{P}(Z_n^* \in \cdot)$  could be *a priori* very different.

*Example.* Suppose that generation  $n$  (fixed) contains 100 cells with 1 parasite (and no other contaminated cells) with probability 1/2 and it contains 1 cell with 100 parasites with probability 1/2 (and no other contaminated cells). Compare then

$$\begin{aligned}\mathbb{P}^n(Z_{\mathbf{U}_n} = 1) &= 1/2, & \mathbb{P}^n(Z_{\mathbf{U}_n} = 100) &= 1/2; \\ \mathbb{P}(Z_n^* = 1) &= 100/101, & \mathbb{P}(Z_n^* = 100) &= 1/101.\end{aligned}$$

Actually the convergence of  $(Z_n^*)_{n \in \mathbb{N}}$  leads to the result obtained by Kimmel [14] in the continuous analogue of this model

$$\frac{\mathbb{P}(Z_n = k)}{\mathbb{P}(Z_n > 0)} = \frac{2^{-n} \sum_{\mathbf{i} \in \mathbb{G}_n} \mathbb{E}(\mathbb{1}_{Z_{\mathbf{i}}=k})}{2^{-n} \sum_{\mathbf{i} \in \mathbb{G}_n} \mathbb{E}(\mathbb{1}_{Z_{\mathbf{i}}>0})} = \frac{\mathbb{E}(\#\{\mathbf{i} \in \mathbb{G}_n : Z_{\mathbf{i}} = k\})}{\mathbb{E}(\#\mathbb{G}_n^*)} \xrightarrow{n \rightarrow \infty} \mathbb{P}(\Upsilon = k), \quad (21)$$

whereas we are here interested in the expectation of  $F_k(n)$ .

A sufficient condition to get the equality of the two distributions is that  $\#\mathbb{G}_n^*$  is deterministic, which does not hold here. But in the case when  $(m_0, m_1) \in D_3$ , we shall prove that  $\#\mathbb{G}_n^*$  is asymptotically proportional to  $(m_0 + m_1)^n$  as  $n \rightarrow \infty$  (forthcoming Proposition 6.4). This enables us to control  $\mathbb{P}^n(Z_{\mathbf{U}_n} \in \cdot)$  by the distribution of  $\mathbb{P}(Z_n^* \in \cdot)$ . More precisely, it is sufficient to prove the separation of descendances of parasites (Proposition 6.5) and the control of filled-in cells (Lemma 6.6) using the results of Section 2.2 about the BPFE  $Z_n^*$ . These two results are the keys for Theorems 5.2, 5.5 and 5.7. Similarly, when  $(m_0, m_1) \in D_5$ , we already know that  $\#\mathbb{G}_n^*$  is approximatively equivalent to  $2^n$ . Then the fact that  $Z_n^*$  explodes as  $n \rightarrow \infty$  will ensure that the proportion of filled-in cells among contaminated cells tends to one (Theorem 5.1 below).

### 5.1 Case $(m_0, m_1) \in D_5$ ( $m > 1$ )

In that case, recall that conditionally on  $\text{Ext}^c$ ,  $\#\mathbb{G}_n^*$  is asymptotically proportional to  $2^n$  (by Theorem 3.1). Moreover the contaminated cells become largely infected, as stated below.

**Theorem 5.1.** *Conditionally on  $\text{Ext}^c$ , for every  $k \in \mathbb{N}$ ,  $F_k(n)$  converges in probability to 0 as  $n \rightarrow \infty$ . i.e.*

$$\forall K, \epsilon > 0, \quad \mathbb{P}^*\left(\frac{\#\{\mathbf{i} \in \mathbb{G}_n : Z_{\mathbf{i}} \geq K\}}{\#\mathbb{G}_n^*} \geq 1 - \epsilon\right) \xrightarrow{n \rightarrow \infty} 1$$

If  $m_0 = m_1$ , the number of parasites in a contaminated cell is of the same order as  $m^n$ . More precisely, for every  $\epsilon > 0$ ,

$$\sup_{n \in \mathbb{N}} \left\{ \mathbb{P}^*\left(\frac{\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} \leq \alpha m^n\}}{\#\mathbb{G}_n^*} \geq \epsilon\right) \right\} \xrightarrow{\alpha \rightarrow 0} 0$$

*Proof.* In that case, use Theorem 3.1 and (20) to get that there exists a non negative random variable  $\tilde{L}$  such that

$$\#\mathbb{G}_n^* \geq 2^n \tilde{L}, \quad \mathbb{P}^*(\tilde{L} = 0) = 0. \quad (22)$$

Let  $K, \eta$  and  $\epsilon > 0$  and put  $B_n(K, \eta) := \left\{ \frac{\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} \leq K\}}{\#\mathbb{G}_n^*} \geq \eta \right\} \cap \text{Ext}^c$ , then

$$\sum_{\mathbf{i} \in \mathbb{G}_n^*} \mathbb{1}_{\{Z_{\mathbf{i}} \leq K\}} \geq \eta 2^n \tilde{L} \mathbb{1}_{B_n(K, \eta)}$$

which gives, taking expectations,

$$\mathbb{E}(\tilde{L} \mathbb{1}_{B_n(K, \eta)}) \leq \frac{\mathbb{E}(\sum_{\mathbf{i} \in \mathbb{G}_n^*} 2^{-n} \mathbb{1}_{\{Z_{\mathbf{i}} \leq K\}})}{\eta} = \frac{\mathbb{P}(0 < Z_n \leq K)}{\eta}.$$

Use then Proposition 2.5 and (22) to choose  $n$  large enough so that

$$\mathbb{P}(B_n(K, \eta)) \leq \epsilon,$$

which completes the proof of the theorem. In the case  $m_0 = m_1 = m$ , follow the proof above and use that  $Z_n/m^n$  converges to a positive limit on  $\text{Ext}^c$  (see [3]) to get the finer result given after the theorem.  $\square$

## 5.2 Case $(m_0, m_1) \in D_3$ ( $m \leq 1$ )

We assume here  $\mathbb{E}(Z^{(a)})^2 < \infty$  and prove that  $(F_k(n))_{k \in \mathbb{N}}$  converges to a deterministic limit. We prove the convergence thanks to the Cauchy criterion (using completeness of  $l^1(\mathbb{N})$ ). The fact that the limit is deterministic is a consequence of the separation of the descendances of parasites and the law of large numbers. Once we know this limit is deterministic, we identify it with the Yaglom limit  $\Upsilon$  (see Section 6.1 for proofs).

**Theorem 5.2.** *Conditionally on  $\text{Ext}^c$ , as  $n \rightarrow \infty$ ,  $(F_k(n))_{k \in \mathbb{N}}$  converges in probability in  $\mathbb{S}^1(\mathbb{N})$  to  $(\mathbb{P}(\Upsilon = k))_{k \in \mathbb{N}}$ .*

*Remark.* We get here a realization of the Yaglom distribution  $\Upsilon$ . Note also that

- ★ The limit just depends on the one-dimensional distributions of  $(Z^{(0)}, Z^{(1)})$ . More precisely, recall that the probability generating function  $G$  of  $\Upsilon$  is characterized by (16).
- ★ This theorem still holds starting from  $k$  parasites. We get also easily a similar result in the case when a cell gives birth to  $N$  cells ( $N \in \mathbb{N}$ ).



- ★ As an application, we can obtain numerically the Yaglom quasistationary distribution of any BGW. Let  $Z$  be the reproduction law of a BGW with mean  $m < 1$  and choose  $N$  such that  $Nm > 1$ . Consider Kimmel's model where each cell divides into  $N$  daughter cells and  $Z^{(0)} \stackrel{d}{=} Z^{(1)} \stackrel{d}{=} \dots \stackrel{d}{=} Z^{(N)} \stackrel{d}{=} Z$ . Computing then the asymptotic of the proportions of contaminated cells with  $k$  parasites gives the Yaglom quasistationary distribution associated to  $Z$ . If  $\mathbb{P}(\text{Ext}) \neq 0$ , one can start from many parasites 'to avoid' extinction. More generally, we can obtain similarly the Yaglom quasistationary distribution of any BPRE with finite number  $k$  of environments such that  $\sum_1^k m_i^2 < \sum_1^k m_i$ .
- ★ This theorem is in the same vein as Theorem 11 in [13]. But we can not follow the same approach as Guyon for the proof. Indeed we have to consider here the proportions among the contaminated cells in generation  $n$  whereas Guyon considers proportions among all cells in generation  $n$ . Unfortunately, the subtree of contaminated cells is itself random and induces long-range dependences between cells lines, so that Guyon's arguments do not hold here. Moreover Theorem 11 in [13] relies on an ergodicity hypothesis which cannot be circumvented.

*Example.* We give two examples when the limit can be calculated.

- ★ Trivial case :  $\mathbb{P}(Z^{(0)} \in \{0, 1\}, Z^{(1)} \in \{0, 1\}) = 1$  leads to  $\mathbb{P}(\Upsilon = 1) = 1$  ( $G = id$ ).
- ★ Symmetric linear fractional case :  $p \in ]0, 1[, b \in ]0, (1 - p)^2[$  and

$$\mathbb{P}(Z^{(0)} = k) = \mathbb{P}(Z^{(1)} = k) = bp^{k-1} \text{ if } k \geq 1$$

and  $\mathbb{P}(Z^{(0)} = 0) = \mathbb{P}(Z^{(1)} = 0) = (1 - b - p)/(1 - p)$ .

Then  $m_0 = m_1 = b/(1 - p)^2 < 1$  and letting  $s_0$  be the root of  $f_0(s) = s$  larger than 1 yields

$$\forall k \geq 1, \quad \mathbb{P}(\Upsilon = k) = (s_0 - 1)/s_0^k.$$

As asymptotically we know the number of parasites and the proportion of cells with  $k$  parasites, we get the number of contaminated cells (recall that  $W$  is given by (6)).

**Corollary 5.3.** *Conditionally on  $\text{Ext}^c$ , the following convergences hold in probability*

$$\frac{\#\mathbb{G}_n^*}{Z_n} \xrightarrow{n \rightarrow \infty} \frac{1}{\mathbb{E}(\Upsilon)}, \quad \frac{\#\mathbb{G}_n^*}{(m_0 + m_1)^n} \xrightarrow{n \rightarrow \infty} \frac{W}{\mathbb{E}(\Upsilon)}.$$

We can also consider the ancestors at generation  $n$  of the cells of  $\partial\mathbb{T}^*$ , which amounts to considering

$$F_k(n, p) = \frac{\#\{\mathbf{i} \in \mathbb{G}_{n+p}^* : Z_{\mathbf{i}|n} = k\}}{\#\mathbb{G}_{n+p}^*}$$

and let  $p \rightarrow \infty$ . Letting then  $n \rightarrow \infty$  yields the biased Yaglom quasistationary distribution, thanks to the separation of descandances of parasites.

**Corollary 5.4.** *Conditionally on  $\text{Ext}^c$ , for every  $k \in \mathbb{N}$ ,  $F_k(n, p)$  converges in probability in  $\mathbb{S}^1(\mathbb{N})$  as  $p$  tends to infinity. This limit converges in probability in  $\mathbb{S}^1(\mathbb{N})$  as  $n \rightarrow \infty$ .*

$$\forall k \in \mathbb{N}, \quad \lim_{n \rightarrow \infty} \lim_{p \rightarrow \infty} F_k(n, p) \stackrel{\mathbb{P}}{=} \frac{k\mathbb{P}(\Upsilon = k)}{\mathbb{E}(\Upsilon)}.$$

If  $Z^{(0)} \stackrel{d}{=} Z^{(1)}$ , we get here an interpretation of the fact that the stationary distribution of the Q-process associated to  $Z^{(0)}$  is the size-biased Yaglom limit.

### 5.3 Case $(m_0, m_1) \in D_2$

In that case, the parasites die out. So we condition by  $\mathcal{Z}_n > 0$ , we still assume  $\mathbb{E}(Z^{(a)})^2 < \infty$  and we get a similar result.

**Theorem 5.5.** *As  $n \rightarrow \infty$ ,  $(F_k(n))_{k \in \mathbb{N}}$  conditioned by  $\mathcal{Z}_n > 0$  converges in distribution on  $\mathbb{S}^1(\mathbb{N})$  to  $(\mathbb{P}(\Upsilon = k))_{k \in \mathbb{N}}$ .*

The proof follows that of the previous theorem. Indeed we still have  $m < \bar{m}$ , so (14) is satisfied (by Lemma 6.1) and we can use the results on the BPRE  $(Z_n)_{n \in \mathbb{N}}$ . There are only two differences. First, we work under  $\mathbb{P}^n$  instead of  $\mathbb{P}^*$ . Moreover  $\mathcal{Z}_n$  satisfies now  $\mathbb{P}(\mathcal{Z}_n > 0) \stackrel{n \rightarrow \infty}{\sim} 2/(\text{Var}(Z^{(0)} + Z^{(1)})n)$  and  $\mathcal{Z}_n/n$  conditioned to be non zero converges in distribution as  $n \rightarrow \infty$  to an exponential variable  $\mathcal{E}$  of parameter  $2/(\hat{m} + 1)$  (see section 2.1). As above, we can derive the following result.

**Corollary 5.6.** *As  $n \rightarrow \infty$ ,  $\#\mathbb{G}_n^*/n$  conditioned by  $\#\mathbb{G}_n^* > 0$  converges in distribution to  $\mathcal{E}/\mathbb{E}(\Upsilon)$ .*

### 5.4 Case $(m_0, m_1) \in D_1$

In this case, the number of contaminated cells does not explode and the number of cells of type  $k$  at generation  $n$  conditioned by the survival of parasites in this generation converges weakly to a non deterministic limit (see Section 7 for proofs).

**Theorem 5.7.** *As  $n \rightarrow \infty$ ,  $(\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\})_{k \in \mathbb{N}}$  conditioned on  $\mathcal{Z}_n > 0$  converges in distribution on  $l^1(\mathbb{N})$  to a random sequence  $(N_k)_{k \in \mathbb{N}}$  which satisfies  $\mathbb{E}(\sum_{k \in \mathbb{N}} kN_k) < \infty$ .*

As above, we get

**Corollary 5.8.**  *$\#\mathbb{G}_n^*$  conditioned by  $\#\mathbb{G}_n^* > 0$  converges in distribution to a positive finite random variable.*

Picking a cell uniformly on  $\partial\mathbb{T}^*$  leads again to the size-biased distribution :

**Corollary 5.9.** *For every  $n \in \mathbb{N}$ ,  $(\#\{\mathbf{i} \in \mathbb{G}_{n+p}^* : Z_{\mathbf{i}|n} = k\})_{k \in \mathbb{N}}$  conditioned on  $\mathcal{Z}_{n+p} > 0$  converges weakly in  $l^1(\mathbb{N})$  to a random sequence as  $p$  tends to infinity. This limit converges weakly as  $n \rightarrow \infty$ .*

$$\forall k \in \mathbb{N}, \quad \lim_{n \rightarrow \infty} \lim_{p \rightarrow \infty} \#\{\mathbf{i} \in \mathbb{G}_{n+p}^* : Z_{\mathbf{i}|n} = k\} | \mathcal{Z}_{n+p} > 0 = \frac{kN_k}{\sum_{k' \in \mathbb{N}} k'N_{k'}}.$$

## 5.5 Remaining domain : $(m_0, m_1) \in D_4$

If  $(m_0, m_1) \in D_4$  and  $m_0 < 1 < m_1$ , one can prove using Lemma 3.2 and a coupling argument with Corollary 5.3 that

$$\sup_{n \in \mathbb{N}} \left\{ \mathbb{P} \left( \frac{\#\mathbb{G}_n^*}{2^n \mathbb{P}(Z_n > 0)} \geq A, \frac{\#\mathbb{G}_n^*}{(m_0 + \widetilde{m}_0)^n} \leq 1/A \right) \right\} \xrightarrow{A \rightarrow 0} 0$$

where  $\widetilde{m}_0 = (1 + \sqrt{1 + 4(m_0 - m_0^2)})/2 > 1$ . Thus  $\#\mathbb{G}_n^*$  grows geometrically and one can naturally conjecture that  $\#\mathbb{G}_n^*$  is asymptotically proportional to  $\mathbb{E}(\#\mathbb{G}_n^*) = 2^n \mathbb{P}(Z_n > 0)$ . Recall that the asymptotics of  $2^n \mathbb{P}(Z_n > 0)$  depends on  $m_0 m_1 = 1$  or  $m_0 m_1 < 1$  and  $m_0 \log(m_0) + m_1 \log(m_1)$  is positive, zero or negative (see [1, 11, 15]).

But in  $D_4$ ,  $\sup_{n \in \mathbb{N}} \{\mathbb{E}(Z_n \mathbb{1}_{Z_n > K})\} \xrightarrow{K \rightarrow \infty} 0$  does not hold *a priori* and this result was needed to control filled-in cells and get Theorem 5.2. Moreover, if  $m_0 \log(m_0) + m_1 \log(m_1) \geq 0$  (weakly subcritical case), separation of descendances of parasites and Corollary 5.4 do not hold. Finally recall that if  $m_0 m_1 = 1$ , then the number of contaminated cells becomes negligible compared to  $2^n$ . But as  $(Z_n^*)_{n \in \mathbb{N}}$  explodes in this case (see [1]), the asymptotic proportion of contaminated cells which are arbitrarily largely contaminated should be equal to 1 as in Theorem 5.1.

In this sense, the domains  $D_i$  ( $i \in \{1, 2, 3, 5\}$ ) studied above are maximal and the remaining domain  $D_4$ , which includes the frontiers of  $D_3$  and  $D_5$ , needs a different approach. The limiting behavior is thought to be different as that of other domains and to depend on several other subdomains. Nevertheless, Theorem 5.2 and Corollaries 5.3 and 5.4 may hold if  $m_0 \log(m_0) + m_1 \log(m_1) < 0$  but  $m_0^2 + m_1^2 \geq m_0 + m_1$ .

## 6 Proofs in the case $(m_0, m_1) \in D_3$

We assume in this section that  $\mathbb{E}(Z^{(a)2}) < \infty$  (i.e.  $\widetilde{m} < \infty$ ) and we start with giving some technical results.

### 6.1 Preliminaries

First, note that for all  $u, v \in l^1(\mathbb{N}^*)$ , we have,

$$\left\| \frac{u}{\|u\|_1} - \frac{v}{\|v\|_1} \right\|_1 = \left\| \frac{u-v}{\|u\|_1} + \frac{v}{\|v\|_1} \frac{\|v\|_1 - \|u\|_1}{\|u\|_1} \right\|_1 \leq 2 \frac{\|u-v\|_1}{\|u\|_1}. \quad (23)$$

Moreover by (6), there exist two random variables  $C$  and  $D$  a.s finite such that

$$\forall n \in \mathbb{N}, C \leq \frac{Z_n}{(2m)^n} \leq D \quad \text{a.s.}, \quad \mathbb{P}^*(C=0) = \mathbb{P}^*(D=0) = 0 \quad (24)$$

and as  $\cap_{n \in \mathbb{N}} \{Z_n > 0\} = \{\forall n \in \mathbb{N} : Z_n > 0\}$ , we have,

$$\sup_A \left\{ \left| \mathbb{P}^n(A) - \mathbb{P}^*(A) \right| \right\} \xrightarrow{n \rightarrow \infty} 0. \quad (25)$$

We focus now on the BPRE  $(Z_n)_{n \in \mathbb{N}}$ . By induction and convexity of  $f_a$ , we have for all  $\mathbf{i} \in \mathbb{G}_n$ , (see Section 2.2 for the Notation)

$$\mathbb{P}(Z_{\mathbf{i}} > 0) = 1 - f_{\mathbf{i}}(0) \leq m_{\mathbf{i}}. \quad (26)$$

Moreover, in this case, we have also

**Lemma 6.1.** *If  $(m_0, m_1) \in D_3$ , then  $(Z^{(0)}, Z^{(1)})$  satisfies condition (14).*

*Proof.* First  $\mathbb{E}(Z^{(a)2}) < \infty$  entails that  $\mathbb{E}(Z_a \log^+(Z_a)) < \infty$ . Moreover  $m_0^2 + m_1^2 \leq m_0 + m_1$  ensures  $m_0 \log(m_0) + m_1 \log(m_1) < 0$ .

To see that : Consider the non trivial case  $m_1 > 1$  or  $m_0 > 1$ , assume  $m_0 \log(m_0) + m_1 \log(m_1) \geq 0$  and introduce  $g : \alpha \in [0, \infty] \mapsto m_0^\alpha + m_1^\alpha$ . As  $g'' > 0$  then  $g'$  is strictly increasing. Moreover  $g'(1) = m_0 \log(m_0) + m_1 \log(m_1) \geq 0$  then  $g(2) > g(1)$  which gives  $m_0^2 + m_1^2 > m_0 + m_1$ .  $\square$

Then identities (26) and (15) entail that there exists  $M > 0$  such that

$$M \leq \frac{\mathbb{P}(Z_n > 0)}{m^n} \leq 1. \quad (27)$$

## 6.2 Estimation of $\#\mathbb{G}_n^*$

The number of parasites which belong to filled-in cells is negligible compared to the total number of parasites (see also Lemma 6.6 for a result of the same kind). To prove this result, we use its counterpart for BPRE  $(Z_n)_{n \in \mathbb{N}}$  conditioned to be non zero.

**Lemma 6.2.** *For every  $\eta > 0$ ,*

$$\sup_{n \in \mathbb{N}} \left\{ \mathbb{P}^* \left( \frac{\sum_{\mathbf{i} \in \mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} > K\}}}{Z_n} \geq \eta \right) \right\} \xrightarrow{K \rightarrow \infty} 0.$$

*Proof.* Let  $\eta > 0$  and write  $A_n(K, \eta) := \left\{ \frac{\sum_{\mathbf{i} \in \mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} > K\}}}{Z_n} \geq \eta \right\} \cap \text{Ext}^c$ . Then

$$\mathbb{1}_{A_n(K, \eta)} \sum_{\mathbf{i} \in \mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} > K\}} \geq \mathbb{1}_{A_n(K, \eta)} Z_n \eta.$$

Then using (24), we have,

$$\mathbb{1}_{A_n(K, \eta)} (2m)^{-n} \sum_{\mathbf{i} \in \mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} > K\}} \geq \eta \mathbb{1}_{A_n(K, \eta)} C$$

so that taking expectations,

$$m^{-n} \mathbb{E} \left( 2^{-n} \sum_{\mathbf{i} \in \mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} > K\}} \right) \geq \mathbb{E}(\mathbb{1}_{A_n(K, \eta)} C) \eta$$

Next, notice that  $\mathbb{E}(2^{-n} \sum_{\mathbf{i} \in \mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} > K\}}) = \mathbb{E}(Z_n \mathbb{1}_{\{Z_n > K\}}) \leq K^{-1} \mathbb{E}(Z_n(Z_n - 1))$ . Then using Lemma 2.1 and  $\bar{m} < m$ , we have

$$\mathbb{E}(\mathbb{1}_{A_n(K, \eta)} C) \leq \frac{\tilde{m} \frac{m^n - \bar{m}^n}{m - \bar{m}}}{\eta K m^n} \leq \frac{\tilde{m}}{(m - \bar{m}) \eta K}$$

Then observe that  $\forall \alpha > 0$ ,  $\inf_{\mathbb{P}^*(A) \geq \alpha} \{\mathbb{E}(C \mathbb{1}_A)\} > 0$ . So  $\exists K_0 \geq 0$  such that  $\forall n \in \mathbb{N}, \forall K \geq K_0$ ,

$$\mathbb{P}^*(A_n(K, \eta)) < \alpha,$$

which completes the proof.  $\square$

First, for any  $\epsilon > 0$ , choose  $K$  using the previous lemma such that

$$\mathbb{P}^*\left(\frac{\sum_{\mathbf{i} \in \mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} \leq K\}}}{Z_n} \geq 1/2\right) = 1 - \mathbb{P}^*\left(\frac{\sum_{\mathbf{i} \in \mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} > K\}}}{Z_n} < 1/2\right) \geq 1 - \epsilon/2.$$

Adding that conditionally on  $\text{Ext}^c$ ,  $Z_n \xrightarrow{n \rightarrow \infty} \infty$  a.s, gives the following result.

**Proposition 6.3.** *Let  $\epsilon > 0$ , there exists  $K \in \mathbb{N}$  such that  $\forall N \in \mathbb{N}$ ,  $\exists n_0 \in \mathbb{N}$  such that  $\forall n \geq n_0$ ,*

$$\mathbb{P}^*\left(\sum_{\mathbf{i} \in \mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} \leq K\}} \geq N\right) \geq 1 - \epsilon.$$

Second, we derive an estimation of  $\#\mathbb{G}_n^*$ . By Lemma 6.2, the cells are not very contaminated so the number of contaminated cells is asymptotically proportional to the number of parasites, which is a Bienaymé Galton Watson process.

**Proposition 6.4.** *For every  $\epsilon > 0$ , there exist  $A, B > 0$  such that for every  $n \in \mathbb{N}$ ,*

$$\mathbb{P}^*\left(\frac{\#\mathbb{G}_n^*}{(2m)^n} \in [A, B]\right) \geq 1 - \epsilon.$$

*Proof.* First use (24) to get

$$\frac{\#\mathbb{G}_n^*}{(2m)^n} \leq \frac{Z_n}{(2m)^n} \leq D.$$

Moreover using again (24), we have

$$\frac{\#\mathbb{G}_n^*}{(2m)^n} \geq \frac{\sum_{\mathbf{i} \in \mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} \leq K\}}}{K(2m)^n} \geq \frac{C}{K} \frac{\sum_{\mathbf{i} \in \mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} \leq K\}}}{Z_n}$$

and Lemma 6.2 gives the result.  $\square$

### 6.3 Separation of the descendances of parasites

Start with two parasites and consider the BPRE  $(Z_n)_{n \in \mathbb{N}}$ . Even when conditioning on the survival of their descendance, the descendance of one of them dies out. This ensures that two distinct parasites in generation  $n$  do not have descendants which belong to the same cell in generation  $n + q$  if  $q$  is large enough. More precisely, we define  $N_n(\mathbf{i})$  as the number of parasites of cell  $\mathbf{i}|n$  whose descendance is still alive in cell  $\mathbf{i}$  and we prove the following result.

**Proposition 6.5.**  $\forall K \in \mathbb{N}, \forall \epsilon, \eta > 0, \exists q \in \mathbb{N}$  such that  $\forall n \in \mathbb{N}$ , we have

$$\mathbb{P}^* \left( \frac{\#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}|n} \leq K, N_n(\mathbf{i}) \geq 2\}}{\#\mathbb{G}_{n+q}^*} \geq \eta \right) \leq \epsilon.$$

*Proof.* We consider for  $A > 0$ ,

$$E_n^q(\eta) = \left\{ \frac{\#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}|n} \leq K, N_n(\mathbf{i}) \geq 2\}}{\#\mathbb{G}_{n+q}^*} \geq \eta \right\} \cap \left\{ \frac{\#\mathbb{G}_{n+q}^*}{(2m)^{n+q}} \geq A \right\}.$$

Then

$$\mathbb{1}_{E_n^q(\eta)} \#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}|n} \leq K, N_n(\mathbf{i}) \geq 2\} \geq \mathbb{1}_{E_n^q(\eta)} \eta A (2m)^{n+q}$$

so that taking expectations,

$$\begin{aligned} \mathbb{P}(E_n^q(\eta)) &\leq \frac{2^{-(n+q)} \mathbb{E}(\sum_{\mathbf{i} \in \mathbb{G}_{n+q}} \mathbb{1}_{\{Z_{\mathbf{i}|n} \leq K, N_n(\mathbf{i}) \geq 2\}})}{\eta A m^{n+q}} \\ &\leq \frac{2^{-n} \sum_{\mathbf{j} \in \mathbb{G}_n} \mathbb{P}(0 < Z_{\mathbf{j}} \leq K) 2^{-q} \sum_{\mathbf{i} \in \mathbb{G}_q} \mathbb{P}_K(N_0(\mathbf{i}) \geq 2)}{\eta A m^{n+q}} \\ &\leq \frac{\mathbb{P}(Z_n > 0) 2^{-q} \sum_{\mathbf{i} \in \mathbb{G}_q} \mathbb{P}_K(N_0(\mathbf{i}) \geq 2)}{\eta A m^{n+q}} \end{aligned}$$

As we have  $\binom{K}{2}$  ways to choose two parasites among  $K$  and they both survive along  $\mathbf{i}$  with probability  $\mathbb{P}(Z_{\mathbf{i}} > 0)^2$ , we have  $\mathbb{P}_K(N_0(\mathbf{i}) \geq 2) \leq \binom{K}{2} \mathbb{P}(Z_{\mathbf{i}} > 0)^2$ . Using also (27) gives

$$\begin{aligned} \mathbb{P}(E_n^q(\eta)) &\leq \frac{\binom{K}{2} 2^{-q} \sum_{\mathbf{i} \in \mathbb{G}_q} \mathbb{P}(Z_{\mathbf{i}} > 0)^2}{\eta A m^q} \\ &\leq \frac{\binom{K}{2} 2^{-q} \sum_{\mathbf{i} \in \mathbb{G}_q} m_{\mathbf{i}}^2}{\eta A m^q} = \frac{\binom{K}{2} \bar{m}^q}{\eta A m^q} \quad \text{using (26) and (13).} \end{aligned}$$

Then  $\bar{m} < m$  and Proposition 6.4 give the result.  $\square$

## 6.4 Control of filled-in cells

Here we prove that filled-in cells have asymptotically no impact on the proportions of cells with a given number of parasites.

**Lemma 6.6.**  $\forall \epsilon, \eta > 0, \exists K \in \mathbb{N}$  such that  $\forall n, q \in \mathbb{N}$ , we have

$$\mathbb{P}^* \left( \frac{\#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}|n} > K\}}{\#\mathbb{G}_{n+q}^*} \geq \eta \right) \leq \epsilon.$$

*Proof.* Let  $A > 0$  and consider

$$F_n^q(\eta) = \left\{ \frac{\#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}|n} > K\}}{\#\mathbb{G}_{n+q}^*} \geq \eta \right\} \cap \left\{ \frac{\#\mathbb{G}_{n+q}^*}{(2m)^{n+q}} \geq A \right\}$$

then

$$\mathbb{1}_{F_n^q(\eta)} \#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}|n} > K\} \geq \mathbb{1}_{F_n^q(\eta)} \eta A (2m)^{n+q}$$

Taking expectations leads to

$$\begin{aligned} \mathbb{P}(F_n^q(\eta)) &\leq \frac{2^{-(n+q)} \mathbb{E}(\sum_{\mathbf{i} \in \mathbb{G}_{n+q}} \mathbb{1}_{\{Z_{\mathbf{i}|n} > K, Z_{\mathbf{i}} > 0\}})}{\eta A m^{n+q}} \\ &\leq \frac{2^{-(n+q)} \sum_{\mathbf{i} \in \mathbb{G}_{n+q}} \mathbb{P}(Z_{\mathbf{i}|n} > K, Z_{\mathbf{i}} > 0)}{\eta A m^{n+q}} \\ &\leq \frac{\sum_{k > K} 2^{-n} \sum_{\mathbf{j} \in \mathbb{G}_n} \mathbb{P}(Z_{\mathbf{j}} = k) 2^{-q} \sum_{\mathbf{i} \in \mathbb{G}_q} \mathbb{P}_k(Z_{\mathbf{i}} > 0)}{\eta A m^{n+q}} \end{aligned}$$

Moreover  $\mathbb{P}_k(Z_{\mathbf{i}} > 0) = 1 - (1 - \mathbb{P}(Z_{\mathbf{i}} > 0))^k \leq k \mathbb{P}(Z_{\mathbf{i}} > 0)$  and we have

$$\begin{aligned} \mathbb{P}(F_n^q(\eta)) &\leq \frac{\sum_{k > K} 2^{-n} \sum_{\mathbf{j} \in \mathbb{G}_n} k \mathbb{P}(Z_{\mathbf{j}} = k) \mathbb{P}(Z_{\mathbf{i}} > 0)}{\eta A m^{n+q}} \\ &\leq \frac{\mathbb{E}(Z_n \mathbb{1}_{\{Z_n > K\}})}{\eta A m^n} && \text{using (27)} \\ &\leq \frac{\mathbb{E}(Z_n(Z_n - 1))}{\eta A K m^n} \\ &\leq \frac{\tilde{m}(m^n - \bar{m}^n)}{\eta A K (m - \bar{m}) m^n} && \text{using Lemma 2.1} \\ &\leq \frac{\tilde{m}}{\eta A K (m - \bar{m})} && \text{since } \bar{m} < m \end{aligned}$$

which completes the proof choosing  $A$  in agreement with Proposition 6.4.  $\square$

## 6.5 Proof of Theorem 5.2

Consider the contaminated cells in generation  $n + q$ . Their ancestors in generation  $n$  are cells which are not very contaminated (by Lemma 6.6). Then taking  $q$  large, the parasites of a contaminated cell in generation  $n + q$  come from a same parasite in generation  $n$  (separation of the descendances of parasites, Proposition 6.5). Then at generation  $n + q$ , everything occurs as if all parasites from generation  $n$  belonged to different cells. As the number of parasites at generation  $n$  tends to infinity ( $m_0 + m_1 > 1$ ), we have a law of large numbers phenomenon and get a deterministic limit.

STEP 1 : We prove that for all  $\epsilon, \eta > 0$ , there exist  $n_0 \in \mathbb{N}$  and  $\vec{f} \in \mathbb{S}^1(\mathbb{N})$  such that for every  $n \geq n_0$ ,

$$\mathbb{P}^*(\|(F_k(n))_{k \in \mathbb{N}} - \vec{f}\|_1 \geq \eta) \leq \epsilon$$

- Denote by  $\mathcal{P}_{>K}(n)$  (resp  $\mathcal{P}_{\leq K}(n)$ ) the set of parasites in generation  $n$  which belong to a cell containing more than (resp. at most) than  $K$  parasites. Consider a parasite  $\mathbf{p}$  in generation  $n$ . For every  $k \in \mathbb{N}^*$ , we denote by  $Y_k^q(\mathbf{p})$  the number of cells in generation  $n + q$  which contain at least  $k$  parasites, exactly  $k$  of which have  $\mathbf{p}$  as an ancestor. By

convention,  $Y_0^q(\mathbf{p}) = 0$ . That is, writing for  $\mathbf{p}$  parasite,  $\mathbf{p} \leftrightarrow \mathbf{i}$  when  $\mathbf{p}$  belongs to the cell  $\mathbf{i}$  and  $\mathbf{p}|n$  its ancestor (parasite) in generation  $n$ ,

$$Y_k^q(\mathbf{p}) = \sum_{\mathbf{i} \in \mathbb{G}_{n+q}} \mathbb{1}_{\#\{\mathbf{r} : \mathbf{r} \leftrightarrow \mathbf{i}, \mathbf{r}|n=\mathbf{p}\}=k}, \quad k \in \mathbb{N}^*.$$

By the branching property,  $(Y_k^q(\mathbf{p}))_{k \in \mathbb{N}} (\mathbf{p} \in \mathcal{P}(n))$  are iid and we denote by  $(Y_k^q)_{k \in \mathbb{N}}$  a random variable with this distribution.

$$\begin{aligned} & \sum_{k \in \mathbb{N}^*} \left| \#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}} = k\} - \sum_{\mathbf{p} \in \mathcal{P}_{\leq K}(n)} Y_k^q(\mathbf{p}) \right| \\ \leq & \sum_{k \in \mathbb{N}^*} \left| \#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}} = k, Z_{\mathbf{i}|n} \leq K\} - \sum_{\mathbf{p} \in \mathcal{P}_{\leq K}(n)} Y_k^q(\mathbf{p}) \right| + \#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}|n} > K\} \\ \leq & (K+1)\#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}|n} \leq K, N_n(\mathbf{i}) \geq 2\} + \#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}|n} > K\} \end{aligned} \quad (28)$$

Indeed, recalling that  $N_n(\mathbf{i})$  is the number of parasites of cell  $\mathbf{i}|n$  whose descendance is still alive in cell  $\mathbf{i}$ , we get the following three equalities

$$\begin{aligned} \sum_{\mathbf{p} \in \mathcal{P}_{\leq K}(n)} Y_k^q(\mathbf{p}) &= \sum_{\mathbf{i} \in \mathbb{G}_{n+q}} \sum_{\mathbf{p} \in \mathcal{P}_{\leq K}(n)} \mathbb{1}_{\#\{\mathbf{r} : \mathbf{r} \leftrightarrow \mathbf{i}, \mathbf{r}|n=\mathbf{p}\}=k} \\ \#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}} = k, Z_{\mathbf{i}|n} \leq K\} &= \sum_{\mathbf{i} \in \mathbb{G}_{n+q}} \mathbb{1}_{Z_{\mathbf{i}}=k, Z_{\mathbf{i}|n} \leq K} \\ \mathbb{1}_{Z_{\mathbf{i}}=k, Z_{\mathbf{i}|n} \leq K, N_n(\mathbf{i})=1} &= \mathbb{1}_{N_n(\mathbf{i})=1} \sum_{\mathbf{p} \in \mathcal{P}_{\leq K}(n)} \mathbb{1}_{\#\{\mathbf{r} : \mathbf{r} \leftrightarrow \mathbf{i}, \mathbf{r}|n=\mathbf{p}\}=k} \end{aligned}$$

which ensure

$$\begin{aligned} & \sum_{k \in \mathbb{N}^*} \left| \#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}} = k, Z_{\mathbf{i}|n} \leq K\} - \sum_{\mathbf{p} \in \mathcal{P}_{\leq K}(n)} Y_k^q(\mathbf{p}) \right| \\ \leq & \sum_{k \in \mathbb{N}^*} \sum_{\substack{\mathbf{i} \in \mathbb{G}_{n+q} \\ N_n(\mathbf{i}) \geq 2}} \left| \mathbb{1}_{Z_{\mathbf{i}}=k, Z_{\mathbf{i}|n} \leq K} - \sum_{\mathbf{p} \in \mathcal{P}_{\leq K}(n)} \mathbb{1}_{\#\{\mathbf{r} : \mathbf{r} \leftrightarrow \mathbf{i}, \mathbf{r}|n=\mathbf{p}\}=k} \right| \\ \leq & \#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}|n} \leq K, N_n(\mathbf{i}) \geq 2\} + \sum_{\substack{\mathbf{i} \in \mathbb{G}_{n+q} \\ N_n(\mathbf{i}) \geq 2}} \sum_{\mathbf{p} \in \mathcal{P}_{\leq K}(n)} \mathbb{1}_{\#\{\mathbf{r} : \mathbf{r} \leftrightarrow \mathbf{i}, \mathbf{r}|n=\mathbf{p}\} > 0} \\ \leq & \#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}|n} \leq K, N_n(\mathbf{i}) \geq 2\} + \sum_{\substack{\mathbf{i} \in \mathbb{G}_{n+q} \\ N_n(\mathbf{i}) \geq 2}} K \mathbb{1}_{Z_{\mathbf{i}|n} \leq K} \\ = & (K+1)\#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}|n} \leq K, N_n(\mathbf{i}) \geq 2\} \end{aligned}$$

We shall now prove that the quantities on the right hand side of (28) are small when  $n$  and  $q$  are large enough and that  $\sum_{\mathbf{p} \in \mathcal{P}_{\leq K}(n)} Y_k^q(\mathbf{p})$  follow a law of large number. To that purpose, let  $\epsilon, \eta > 0$  and for all  $K, k, n, q \geq 0$  define  $G_k^K(n, q)$  as

$$G_k^K(n, q) := \frac{\sum_{\mathbf{p} \in \mathcal{P}_{\leq K}(n)} Y_k^q(\mathbf{p})}{\sum_{k \in \mathbb{N}} \sum_{\mathbf{p} \in \mathcal{P}_{\leq K}(n)} Y_k^q(\mathbf{p})}.$$



• First, by Proposition 6.3 and (25),  $\exists K_1 \in \mathbb{N}$  such that  $\forall N \in \mathbb{N}$ ,  $\exists n_1 \in \mathbb{N}$  such that  $\forall K \geq K_1$ ,  $\forall n \geq n_1$ ,

$$\mathbb{P}^n(|\mathcal{P}_{\leq K}(n)| \geq N) \geq 1 - \epsilon. \quad (29)$$

Moreover by Lemma 6.6,  $\exists K_2 \geq K_1$  such that  $\forall n, q \in \mathbb{N}$ ,

$$\mathbb{P}^*\left(\frac{\#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}|n} > K_2\}}{\#\mathbb{G}_{n+q}^*} \geq \eta\right) \leq \epsilon. \quad (30)$$

And by Proposition 6.5,  $\exists q_0 \in \mathbb{N}$  such that  $\forall n \in \mathbb{N}$

$$\mathbb{P}^*\left(\frac{\#\{\mathbf{i} \in \mathbb{G}_{n+q_0}^* : Z_{\mathbf{i}|n} \leq K_2, N_n(\mathbf{i}) \geq 2\}}{\#\mathbb{G}_{n+q_0}^*} \geq \eta/(K_2 + 1)\right) \leq \epsilon. \quad (31)$$

Use then (28), (30) and (31) to get

$$\mathbb{P}^*\left(\frac{\left|\sum_{k \in \mathbb{N}^*} \#\{\mathbf{i} \in \mathbb{G}_{n+q_0} : Z_{\mathbf{i}} = k\} - \sum_{\mathbf{p} \in \mathcal{P}_{\leq K_2}(n)} Y_k^{q_0}(\mathbf{p})\right|}{\#\mathbb{G}_{n+q_0}^*} \geq 2\eta\right) \leq 2\epsilon.$$

Then by (23), for every  $n \in \mathbb{N}$ , we have

$$\mathbb{P}^*(\|(F_k(n + q_0))_{k \in \mathbb{N}} - (G_k^{K_2}(n, q_0))_{k \in \mathbb{N}}\|_1 \geq 4\eta) \leq 2\epsilon. \quad (32)$$

• Second, conditionally on  $\mathcal{Z}_n > 0$ ,  $Y_k^{q_0}(\mathbf{p})$  ( $\mathbf{p} \in \mathcal{P}_{\leq K_2}(n)$ ) are iid. Then the law of large numbers (LLN) ensures that  $\forall k \in \mathbb{N}$ , as  $n$  and so  $\mathcal{P}_{\leq K_2}(n)$  becomes large :

$$G_k^{K_2}(n, q_0) \longrightarrow f_k(q_0) \quad \text{where} \quad f_k(q_0) := \frac{\mathbb{E}(Y_k^{q_0})}{\sum_{k' \in \mathbb{N}} \mathbb{E}(Y_{k'}^{q_0})}.$$

To see that, divide the numerator and denominator of  $G_k^{K_2}(n, q_0)$  by  $\#\mathcal{P}_{\leq K_2}(n)$ . More precisely, by the LLN, there exists  $N > 0$  such that for all  $n \in \mathbb{N}$ ,

$$\mathbb{P}^n(\|(G_k^{K_2}(n, q_0))_{k \in \mathbb{N}^*} - \vec{f}(q_0)\|_1 \geq \eta, \mathcal{P}_{\leq K_2}(n) \geq N) \leq \epsilon.$$

So using (29), there exists  $n_1 \in \mathbb{N}$  such that for every  $\forall n \geq n_1$ ,

$$\mathbb{P}^n(\|(G_k^{K_2}(n, q_0))_{k \in \mathbb{N}^*} - \vec{f}(q_0)\|_1 \geq \eta) \leq 2\epsilon.$$

Finally by (25), there exists  $n_2 \geq n_1$  such that for every  $n \geq n_2$ ,

$$\mathbb{P}^*(\|(G_k^{K_2}(n, q_0))_{k \in \mathbb{N}^*} - \vec{f}(q_0)\|_1 \geq \eta) \leq 3\epsilon. \quad (33)$$

As a conclusion, using (32) and (33), we have proved that for all  $\epsilon, \eta > 0$ , and for every  $n \geq n_2 + q_0$ ,

$$\mathbb{P}^*(\|(F_k(n))_{k \in \mathbb{N}^*} - \vec{f}(q_0)\|_1 \geq 5\eta) \leq 3\epsilon.$$

STEP 2 : Existence of the limit.

For every  $l \in \mathbb{N}$ , there exist  $n_0(l) \in \mathbb{N}$  and  $\vec{f}(l) \in \mathbb{S}^1(\mathbb{N})$  such that for every  $n \geq n_0(l)$

$$\mathbb{P}(\|F(n) - \vec{f}(l)\|_1 \geq 1/2^{l+1}) \leq 1/2^l.$$

Then for all  $l, l'$  such that  $2 \leq l \leq l' : \|\vec{f}(l') - \vec{f}(l)\|_1 \leq 1/2^l$  and completeness of  $l^1(\mathbb{N})$  ensures that  $(\vec{f}(l))_{l \in \mathbb{N}}$  converges in  $\mathbb{S}^1(\mathbb{N})$  to a limit  $\vec{f}$ . Moreover  $\|\vec{f}(l) - \vec{f}\|_1 \leq 1/2^l$  so for every  $n \geq n_0(l)$ ,

$$\mathbb{P}(\|F(n) - \vec{f}\|_1 \geq 1/2^l) \leq 1/2^l$$

which ensures the convergence in probability of  $(F_k(n))_{n \in \mathbb{N}}$  to  $\vec{f}$  as  $n \rightarrow \infty$ .

STEP 3 : Characterization of the limit as  $f_k = \mathbb{P}(\Upsilon = k)$ .

By Proposition 2.3, we have

$$\forall k \in \mathbb{N} \quad \mathbb{P}(Z_n = k \mid Z_n \neq 0) \xrightarrow{n \rightarrow \infty} \mathbb{P}(\Upsilon = k) \quad (34)$$

Moreover for every  $k \in \mathbb{N}^*$ , using (21),

$$\mathbb{P}(Z_n = k \mid Z_n \neq 0) = \frac{\mathbb{E}(\#\{\mathbf{i} \in \mathbb{G}_n : Z_{\mathbf{i}} = k \mid\})}{\mathbb{E}(\#\mathbb{G}_n^*)} = \frac{\mathbb{E}(F_k(n)\#\mathbb{G}_n^*)}{\mathbb{E}(\#\mathbb{G}_n^*)}$$

As  $F_k(n)$  converges in probability to a deterministic limit  $f_k$ , we get

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(Z_n = k \mid Z_n \neq 0) \xrightarrow{n \rightarrow \infty} f_k. \quad (35)$$

Indeed, by Proposition 6.4, there exists  $A > 0$  such that

$$\frac{\mathbb{E}(\#\mathbb{G}_n^*)}{(2m)^n} \geq A.$$

Then for every  $\eta > 0$ , using  $|F_k(n) - f_k| \leq 1$ , we have

$$\begin{aligned} \left| \frac{\mathbb{E}(F_k(n)\#\mathbb{G}_n^*)}{\mathbb{E}(\#\mathbb{G}_n^*)} - f_k \right| &\leq \frac{\mathbb{E}(\#\mathbb{G}_n^* | F_k(n) - f_k | \mathbb{1}_{\{|F_k(n) - f_k| < \eta\}})}{\mathbb{E}(\#\mathbb{G}_n^*)} + \frac{\mathbb{E}(\#\mathbb{G}_n^* \mathbb{1}_{\{|F_k(n) - f_k| \geq \eta\}})}{\mathbb{E}(\#\mathbb{G}_n^*)} \\ &\leq \eta + \frac{\mathbb{E}(Z_n \mathbb{1}_{\{|F_k(n) - f_k| \geq \eta\}})}{A(2m)^n} \end{aligned}$$

By (11),  $Z_n/(2m)^n$  is bounded in  $L^2$  and it is uniformly integrable. Then, thanks to the previous steps, the second term in the last displayed equation vanishes as  $n$  grows and we get (35). Putting (34) and (35) together proves that  $f_k = \mathbb{P}(\Upsilon = k)$ .

## 6.6 Proof of corollaries

*Proof of Corollary 5.3.* Recall that  $\mathbb{E}(\Upsilon) < \infty$  (Proposition 2.3) and note also that for every  $K \in \mathbb{N}^*$ ,

$$\#\mathbb{G}_n^* = \frac{\sum_{\mathbf{i} \in \mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} \leq K\}}}{\sum_{k=1}^K k F_k(n)}.$$

Then using  $\sum_{\mathbf{i} \in \mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} \leq K\}} \leq \mathcal{Z}_n$  gives

$$\begin{aligned} \left| \frac{\#\mathbb{G}_n^*}{\mathcal{Z}_n} - \frac{1}{\mathbb{E}(\Upsilon)} \right| &= \left| \frac{1}{\sum_{k=1}^K k F_k(n)} \frac{\sum_{\mathbf{i} \in \mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} \leq K\}}}{\mathcal{Z}_n} - \frac{1}{\mathbb{E}(\Upsilon)} \right| \\ &\leq \left| \frac{1}{\sum_{k=1}^K k F_k(n)} - \frac{1}{\mathbb{E}(\Upsilon)} \right| + \frac{1}{\mathbb{E}(\Upsilon)} \left| \frac{\sum_{\mathbf{i} \in \mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} \leq K\}}}{\mathcal{Z}_n} - 1 \right| \end{aligned}$$

Let  $\eta, \epsilon > 0$ . We use Lemma 6.2 to choose  $K \in \mathbb{N}^*$  such that

$$\forall n \in \mathbb{N}, \mathbb{P}^* \left( \frac{\sum_{\mathbf{i} \in \mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} \leq K\}}}{\mathcal{Z}_n} \geq 1 - \eta \right) \geq 1 - \epsilon \quad ; \quad \left| \frac{1}{\mathbb{E}(\Upsilon \mathbb{1}_{\Upsilon \leq K})} - \frac{1}{\mathbb{E}(\Upsilon)} \right| \leq \eta$$

Choose  $n_0 \in \mathbb{N}$  using Theorem 5.2 so that for every  $n \geq n_0$ ,

$$\mathbb{P}^* \left( \left| \frac{1}{\sum_{k=1}^K k F_k(n)} - \frac{1}{\mathbb{E}(\Upsilon \mathbb{1}_{\Upsilon \leq K})} \right| \leq \eta \right) \geq 1 - \epsilon.$$

Then for every  $n \geq n_0$ ,

$$\mathbb{P}^* \left( \left| \frac{\#\mathbb{G}_n^*}{\mathcal{Z}_n} - \frac{1}{\mathbb{E}(\Upsilon)} \right| \geq 2\eta + \frac{1}{\mathbb{E}(\Upsilon)} \eta \right) \leq 2\epsilon,$$

which proves the convergence in probability of  $\#\mathbb{G}_n^*/\mathcal{Z}_n$  to  $1/\mathbb{E}(\Upsilon)$ . The second convergence follows from (6).  $\square$

*Proof of Corollary 5.4.* We write for  $n, p, k \in \mathbb{N}$ ,

$$\frac{\#\{\mathbf{i} \in \mathbb{G}_{n+p}^* : Z_{\mathbf{i}|n} = k\}}{\#\mathbb{G}_{n+p}^*} = \frac{(2m)^p}{\#\mathbb{G}_{n+p}^*} \sum_{\mathbf{j} \in \mathbb{G}_n^* : Z_{\mathbf{j}} = k} \frac{\#\{\mathbf{i} \in \mathbb{G}_{n+p}^* : \mathbf{i}|n = \mathbf{j}\}}{(2m)^p}$$

Conditionally on  $Z_{\mathbf{j}} = k$ , by Corollary 5.3 and separation of descendances of parasites, we have the following convergence in probability

$$\frac{\#\{\mathbf{i} \in \mathbb{G}_{n+p}^* : \mathbf{i}|n = \mathbf{j}\}}{(2m)^p} \xrightarrow{p \rightarrow \infty} W_k(\mathbf{j})$$

where  $W_k(\mathbf{j})$  is the sum of  $k$  iid variables distributed as  $W/\mathbb{E}(\Upsilon)$ . Then, using also (6),

$$\mathbb{E}(W_k(\mathbf{j})) = \frac{k\mathbb{E}(W)}{\mathbb{E}(\Upsilon)} = \frac{k}{\mathbb{E}(\Upsilon)}. \quad (36)$$

Using again Corollary 5.3, we get the first limit of the Corollary

$$\lim_{p \rightarrow \infty} \frac{\#\{\mathbf{i} \in \mathbb{G}_{n+p}^* : Z_{\mathbf{i}|n} = k\}}{\#\mathbb{G}_{n+p}^*} \stackrel{\mathbb{P}}{=} \frac{\mathbb{E}(\Upsilon)}{W} \frac{\sum_{\mathbf{j} \in \mathbb{G}_n^* : Z_{\mathbf{j}} = k} W_k(\mathbf{j})}{(2m)^n}.$$

Moreover, Theorem 5.2 ensures that

$$\frac{\#\{\mathbf{j} \in \mathbb{G}_n^* : Z_{\mathbf{j}} = k\}}{(2m)^n} = F_k(n) \frac{\mathcal{Z}_n}{(2m)^n} \xrightarrow{n \rightarrow \infty} \frac{W}{\mathbb{E}(\Upsilon)} f_k.$$

And conditionally on  $\#\mathbb{G}_n^* > 0$ ,  $W_k(\mathbf{j})$  ( $\mathbf{j} \in \mathbb{G}_n^*$ ) is iid by the branching property and  $\#\mathbb{G}_n^*$  tends to infinity. So the law of large numbers and (36) ensure that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\mathbb{E}(\Upsilon) \sum_{\mathbf{j} \in \mathbb{G}_n^* : Z_{\mathbf{j}}=k} W_k(\mathbf{j})}{W (2m)^n} &= \lim_{n \rightarrow \infty} \frac{\mathbb{E}(\Upsilon) \#\{\mathbf{j} \in \mathbb{G}_n^* : Z_{\mathbf{j}}=k\}}{W (2m)^n} \frac{\sum_{\mathbf{j} \in \mathbb{G}_n^* : Z_{\mathbf{j}}=k} W_k(\mathbf{j})}{\#\{\mathbf{j} \in \mathbb{G}_n^* : Z_{\mathbf{j}}=k\}} \\ &\stackrel{\mathbb{P}^*}{=} \frac{k f_k}{\mathbb{E}(\Upsilon)} \end{aligned}$$

which ends the proof.  $\square$

## 7 Proofs in the case $(m_0, m_1) \in D_1$

We still assume  $\mathbb{E}(Z^{(a)2}) < \infty$ , the proof is in the same vein as the proof in the previous section and use the separation of the descendances of the parasites. The main difference is that  $\mathcal{Z}_n$  does not explode so the limit is not deterministic and the convergence holds in distribution.

**Lemma 7.1.** *For every  $K > 0$ , there exists  $q_0 \in \mathbb{N}$  such that for all  $q \geq q_0$  and  $n \in \mathbb{N}$*

$$\mathbb{P}^{n+q}(\{\mathbf{i} \in \mathbb{G}_{n+q}^* : N_n(\mathbf{i}) \geq 2\} \neq \emptyset, \mathcal{Z}_n \leq K) \leq \epsilon.$$

*Proof.* Denote by  $E_n^q$  the event  $\{\{\mathbf{i} \in \mathbb{G}_{n+q}^* : N_n(\mathbf{i}) \geq 2\} \neq \emptyset, \mathcal{Z}_n \leq K\}$ , so that

$$\mathbb{1}_{E_n^q} \leq \sum_{\mathbf{i} \in \mathbb{G}_{n+q}} \mathbb{1}_{\{N_n(\mathbf{i}) \geq 2, \mathcal{Z}_n \leq K\}}.$$

Thus we can follow the proof of Lemma 6.5.

$$\begin{aligned} \mathbb{P}^{n+q}(E_n^q) &\leq \sum_{\mathbf{i} \in \mathbb{G}_{n+q}} \frac{\mathbb{P}(N_n(\mathbf{i}) \geq 2, \mathcal{Z}_n \leq K)}{\mathbb{P}(\mathcal{Z}_{n+q} > 0)} \\ &\leq \frac{\sum_{\mathbf{i} \in \mathbb{G}_{n+q}} \mathbb{P}(N_n(\mathbf{i}) \geq 2, Z_{\mathbf{i}|n} \leq K)}{U(2m)^{n+q}} \quad \text{using (7)} \\ &\leq \frac{\mathbb{P}(0 < \mathcal{Z}_n \leq K) 2^{-q} \sum_{\mathbf{i} \in \mathbb{G}_q} \mathbb{P}_K(N_0(\mathbf{i}) \geq 2)}{Um^{n+q}} \\ &\leq \frac{\binom{K}{2} 2^{-q} \sum_{\mathbf{i} \in \mathbb{G}_q} \mathbb{P}(Z_{\mathbf{i}} > 0)^2}{Um^q} \quad \text{using (27)} \\ &\leq \frac{\binom{K}{2} \bar{m}^q}{Um^q} \quad \text{using (26) and (13)} \end{aligned}$$

Adding that  $\bar{m} < m$  (since  $m < 1/2 \Rightarrow m_0 < 1, m_1 < 1$ ) completes the proof.  $\square$

*Proof of Theorem 5.7. STEP 1 :* We recall that  $\mathcal{P}_n$  is the set of parasites in generation  $n$ , follow STEP 1 in the proof of Theorem 5.2 and use its Notation. Thus, we begin with proving that for every  $\epsilon > 0$ , there exists  $n_0 \in \mathbb{N}$  such that for every  $n \geq n_0$ ,

$$\mathbb{P}^{n+q}(\|(\#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}} = k\})_{k \in \mathbb{N}} - (N_k(n, q))_{k \in \mathbb{N}}\|_1 \neq 0) \leq \epsilon$$

where for all  $n, q, k \geq 0$ ,  $N_k(n, q) := \sum_{\mathbf{p} \in \mathcal{P}(n)} Y_k^q(\mathbf{p})$ .

First, by (10), there exist  $K, q_0 \in \mathbb{N}$  such that for every  $q \geq q_0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}^{n+q}(\mathcal{Z}_n > K) \leq \epsilon. \quad (37)$$

By Lemma 7.1, there exists  $q_1 \geq q_0$  such that for every  $n \in \mathbb{N}$ , we have

$$\mathbb{P}^{n+q_1}(\{\mathbf{i} \in \mathbb{G}_{n+q_1}^* : N_n(\mathbf{i}) \geq 2\} \neq \emptyset, \mathcal{Z}_n \leq K) \leq \epsilon. \quad (38)$$

And by (37), there exists  $n_0 \geq 0$  such that for every  $n \geq n_0$ ,

$$\mathbb{P}^{n+q_1}(\mathcal{Z}_n \geq K) \leq 2\epsilon.$$

Then

$$\mathbb{P}^{n+q_1}(\#\{\mathbf{i} \in \mathbb{G}_{n+q_1}^* : N_n(\mathbf{i}) \geq 2\} \neq 0) \leq 3\epsilon.$$

Moreover

$$\#\{\mathbf{i} \in \mathbb{G}_{n+q_1}^* : N_n(\mathbf{i}) \geq 2\} = 0 \Rightarrow (\#\{\mathbf{i} \in \mathbb{G}_{n+q_1}^* : Z_{\mathbf{i}} = k\})_{k \in \mathbb{N}} = (N_k(n, q_1))_{k \in \mathbb{N}}$$

Then for every  $n \geq n_0$ ,

$$\mathbb{P}^{n+q_1}(\|(\#\{\mathbf{i} \in \mathbb{G}_{n+q_1}^* : Z_{\mathbf{i}} = k\})_{k \in \mathbb{N}} - (N_k(n, q_1))_{k \in \mathbb{N}}\|_1 \neq 0) \leq 3\epsilon.$$

STEP 2 : As  $l^1(\mathbb{N})$  is separable, we can consider the distance  $d$  associated with the weak convergence of probabilities on  $l^1(\mathbb{N})$ . It is defined for any  $\mathbb{P}_1$  and  $\mathbb{P}_2$  probabilities by (see Theorem 6.2 chapter II in [17])

$$d(\mathbb{P}_1, \mathbb{P}_2) = \sup\left\{ \left| \int f(w) \mathbb{P}_1(dw) - \int f(w) \mathbb{P}_2(dw) \right| : \|f\|_\infty \leq 1, \|f\|_{Lips} \leq 1 \right\}$$

where

$$\|f\|_{Lips} = \sup\left\{ \frac{\|f(x) - f(y)\|_1}{\|x - y\|_1} : x, y \in \mathbb{S}^1(\mathbb{N}), x \neq y \right\}.$$

We prove now that for every  $l \geq 1$ , there exist  $n_0(l) \in \mathbb{N}$  and a measure  $\mu(l)$  on  $\mathbb{N}^*$  such that for every  $n \geq n_0(l)$ ,

$$d(\mathbb{P}^n((\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\})_{k \in \mathbb{N}} \in \cdot), \mu(l)) \leq 1/2^l. \quad (39)$$

For that purpose, let  $l \in \mathbb{N}$ . By STEP 1, choose  $q, n_0 \in \mathbb{N}$  such that for all  $n \geq n_0$ ,

$$d(\mathbb{P}^{n+q}((\#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}} = k\})_{k \in \mathbb{N}} \in \cdot), \mathbb{P}^{n+q}((N_k(n, q))_{k \in \mathbb{N}} \in \cdot)) \leq 1/2^{l+1}. \quad (40)$$

Recall that  $(Y_k^q(\mathbf{p}))_{k \in \mathbb{N}}(\mathbf{p} \in \mathcal{P}(n))$  is an iid sequence distributed as  $(Y_k^q)_{k \in \mathbb{N}}$  and  $\#\mathcal{P}(n) = \mathcal{Z}_n$ . Thus, under  $\mathbb{P}^{n+q}$ ,  $N_k(n, q)$  is the sum of  $\mathcal{Z}_n$  variables which are iid, distributed as  $Y_k^q$  and independent of  $\mathcal{Z}_n$ , conditionally on  $\sum_{k \in \mathbb{N}} \sum_{\mathbf{p} \in \mathcal{P}(n)} Y_k^q(\mathbf{p}) > 0$ .

Moreover  $\mathbb{P}^{n+q}(\mathcal{Z}_n \in \cdot)$  converges weakly as  $n \rightarrow \infty$  to a probability  $\nu$  (see (9)) and

we denote by  $\mathcal{N}$  a random variable with distribution  $\nu$  and by  $(Y_k^q(i))_{k \in \mathbb{N}} (i \in \mathbb{N})$  an iid sequence independent of  $\mathcal{N}$  and distributed as  $(Y_k^q)_{k \in \mathbb{N}}$ . Then we have for  $n$  large enough,

$$d(\mathbb{P}^{n+q}((N_k(n, q))_{k \in \mathbb{N}} \in \cdot), \mu(l)) \leq 1/2^l \quad (41)$$

where  $\mu(l)$  is the distribution of  $(\sum_{1 \leq i \leq \mathcal{N}} Y_k^q(i))_{k \in \mathbb{N}}$  conditionally on  $\sum_{k \in \mathbb{N}} \sum_{1 \leq i \leq \mathcal{N}} Y_k^q(i) > 0$ .

Putting (40) and (41) together gives (39).

**CONCLUSION :** As  $l^1(\mathbb{N})$  is complete, the space of probabilities on  $l^1(\mathbb{N})$  endowed with  $d$  is complete (see Theorem 6.5 chapter II in [17]),  $(\mu(l))_{l \in \mathbb{N}}$  converges and we get the convergence of Theorem 5.7.

We now prove that  $\mathbb{E}(\sum_{k \in \mathbb{N}^*} k N_k) < \infty$ . For every  $\epsilon > 0$ , there exists  $K \geq 0$  such that

$$\mathbb{E}(\sum_{k \geq K} k \#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\} \mid \mathcal{Z}_n > 0) \leq \mathbb{E}(\mathcal{Z}_n \mathbb{1}_{\{\mathcal{Z}_n \geq K\}} \mid \mathcal{Z}_n > 0) \leq \frac{\mathbb{E}(\mathcal{Z}_n^2)}{\mathbb{P}(\mathcal{Z}_n > 0)K} \leq \epsilon$$

and using Theorem 5.7 and  $k \#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\} \leq \mathcal{Z}_n$ , there exists  $n_0 \geq 0$  such that for every  $n \geq n_0$ ,

$$|\mathbb{E}(\sum_{1 \leq k \leq K} k \#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\} \mid \mathcal{Z}_n > 0) - \mathbb{E}(\sum_{1 \leq k \leq K} k N_k)| \leq \epsilon.$$

This ensures

$$\mathbb{E}(\sum_{k \in \mathbb{N}} k \#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\} \mid \mathcal{Z}_n > 0) \xrightarrow{n \rightarrow \infty} \mathbb{E}(\sum_{k \in \mathbb{N}^*} k N_k)$$

and recalling Section 2.1, we have also

$$\mathbb{E}(\sum_{k \in \mathbb{N}^*} k \#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\} \mid \mathcal{Z}_n > 0) = \mathbb{E}(\mathcal{Z}_n \mid \mathcal{Z}_n > 0) \xrightarrow{n \rightarrow \infty} \mathcal{B}'(1) < \infty.$$

which completes the proof. □

The proofs of the corollaries follow those of the previous section.

**Acknowledgments :** I am very grateful to Amaury Lambert who introduced me to this topic. This work has largely benefited from his pedagogical qualities and various suggestions.

## References

- [1] V. I. Afanasyev, J. Geiger, G. Kersting, V.A. Vatutin (2005). Criticality for branching processes in random environment. *Ann. Probab.* 33. 645-673.
- [2] K. B. Athreya, S. Karlin (1971). On branching processes with random environments, I : Extinction probability. *Ann. Math. Stat.* 42. 1499-1520.

- [3] K. B. Athreya, S. Karlin (1971). On branching processes with random environments, II : limit theorems. *Ann. Math. Stat.* 42. 1843-1858.
- [4] K. B. Athreya, P. E. Ney (2004). *Branching processes*. Dover Publications, Inc., Mineola, NY.
- [5] K. Athreya, H.J. Kang (1998). Some limit theorems for positive recurrent branching Markov chains I. *Adv. Appl. Prob.* 30 (3). 693-710.
- [6] K. Athreya, H.J. Kang (1998). Some limit theorems for positive recurrent branching Markov chains II. *Adv. Appl. Prob.* 30 (3). 711-722.
- [7] I. Benjamini, Y. Peres (1994). Markov chains indexed by trees. *Ann. Probab.* 22 (1). 219-243.
- [8] F.M. Dekking (1988). On the survival probability of a branching process in a finite state iid environment. *Stochastic Process. Appl.* 27 p151-157.
- [9] S. N. Evans, D. Steinsaltz (2006). Damage segregation at fissioning may increase growth rates : a superprocess model. Available via [http://arxiv.org/PS\\_cache/q-bio/pdf/0608/0608008.pdf](http://arxiv.org/PS_cache/q-bio/pdf/0608/0608008.pdf).
- [10] J. Geiger (1999). Elementary new proofs of classical limit theorems for Galton-Watson processes. *J. Appl. Prob.* 36, 301-309.
- [11] J. Geiger, G. Kersting, V A. Vatutin (2003). Limit Theorems for subcritical branching processes in random environment. *Ann. I. H. Poincaré* 39, no. 4. 593-620.
- [12] Y. Guivarc'h, Q. Liu (2001). Asymptotic properties of branching processes in random environment. *C.R. Acad. Sci. Paris*, t.332, Série I. 339-344.
- [13] J. Guyon (2005). Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Preprint*.
- [14] M. Kimmel (1997). Quasistationarity in a branching model of division-within-division. *Classical and modern branching processes (Minneapolis, MN, 1994)*. 157-164, IMA Vol. Math. Appl., 84, Springer, New York.
- [15] M. V. Kozlov (1976). The asymptotic behavior of the probability of non-extinction of critical branching processes in a random environment. *Theor. Probability Appl.* 21, no. 4, 813-825.
- [16] R. Lyons, R. Pemantle, Y. Peres (1995). Conceptual proofs of  $L \log L$  criteria for mean behavior of branching processes. *Ann. Probab.* 23 (3), 1125-1138.
- [17] K.R. Parthasarathy (1967). *Probability measures on metric spaces*. Probability and Mathematical Statistics, No. 3 Academic Press, Inc., New York-London.
- [18] W.L. Smith, W. Wilkinson (1969). On branching processes in random environments. *Ann. Math. Stat.* Vol40, No 3, p814-827.
- [19] E. J. Stewart, R. Madden, G. Paul, F. Taddei (2005). Aging and Death in a Organism that reproduces by Morphologically Symmetric Division. *PLoS Biol*, 3 (2) : e45.