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-00128224v2

HAL Id: hal-00128224
https://hal.science/hal-00128224v2

Preprint submitted on 27 Jun 2008

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.

The Annals of Applied Probability
2008, Vol. 18, No. 3, 967–996
DOI: 10.1214/07-AAP465
© Institute of Mathematical Statistics, 2008

PROLIFERATING PARASITES IN DIVIDING CELLS: KIMMEL'S BRANCHING MODEL REVISITED

By Vincent Bansaye

Université Paris 6

We consider a branching model introduced by 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.

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 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 [e.g., when $\mathbb{P}(Z^{(0)}Z^{(1)} = 0) = 1$].

Received February 2007; revised February 2007.

AMS 2000 subject classifications. Primary 60J80, 60J85, 60K37; secondary 92C37, 92D25, 92D30.

Key words and phrases. Bienaymé Galton Watson process (BGW), branching processes in random environment (BPRE), Markov chain indexed by a tree, quasistationary distribution, empirical measures.

This is an electronic reprint of the original article published by the Institute of Mathematical Statistics in *The Annals of Applied Probability*, 2008, Vol. 18, No. 3, 967–996. This reprint differs from the original in pagination and typographic detail.

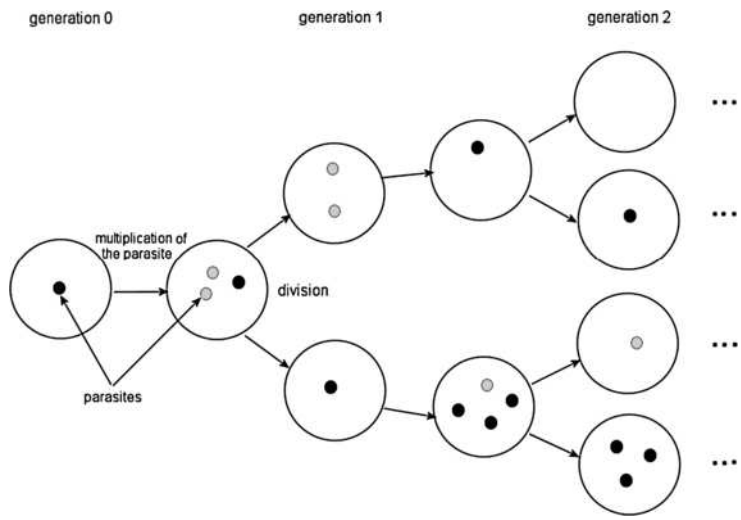

FIG. 1. *Multiplication of parasites and cell division.*

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,

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

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 given by

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

where $(Z_k^{(0)}(\mathbf{i}), Z_k^{(1)}(\mathbf{i}))_{\mathbf{i} \in \mathbb{T}, k \geq 1}$ is an i.i.d. sequence distributed as $(Z^{(0)}, Z^{(1)})$ (see Figure 1).

This is a discrete version of the model introduced by Kimmel in [15]. In particular, 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. When the cells divides, every parasite 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 $a \in \{0,1\}$

$$m_a := \mathbb{E}(Z^{(a)}) \qquad \forall s \geq 0, f_a(s) := \mathbb{E}(s^{Z^{(a)}}). \tag{1}$$

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

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

This model is a Markov chain indexed by a tree. This subject has been studied in the literature (see e.g., [6, 8]) in the symmetric independent case. In this case, for every $(\mathbf{i}, k) \in \mathbb{T} \times \mathbb{N}$, we have

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

which require that $Z^{(0)}$ and $Z^{(1)}$ are i.i.d. in this model. Guyon [14] 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 appears to be cumbersome (see the remark in Section 5.2 for details). In the same vein, we refer to [10, 20] (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^c) the event extinction of the parasites (resp. nonextinction of the parasites),

$$\begin{aligned} \mathcal{Z}_n &= \sum_{\mathbf{i} \in \mathbb{G}_n} Z_{\mathbf{i}}, \\ \mathrm{Ext} &= \{\exists n \in \mathbb{N} : \mathcal{Z}_n = 0\}, \\ \mathrm{Ext}^c &= \{\forall n \in \mathbb{N} : \mathcal{Z}_n > 0\}. \end{aligned} \tag{3}$$

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 i.i.d. sequence independent of $(Z_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ such that

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

Then $(Z_n)_{n \in \mathbb{N}} = (Z_{(a_1, a_2, \ldots, a_n)})_{n \in \mathbb{N}}$ is a Branching Process in Random Environment (BPRE).

The first question we answer here arose from observations made by de Paepe, Paul and 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 their parasites unequally when they divide so that there are lots of healthy cells. This is a little surprising since one could think that cells share equally all their biological content (including parasites). In Section 3, we prove that if

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

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

$D_3 = \{(m_0, m_1) : m_0 + m_1 > 1,$
$m_0 \log(m_0) + m_1 \log(m_1) < 0\}$

$D_4 = \{(m_0, m_1) : m_0 m_1 \leq 1,$
$m_0 \log(m_0) + m_1 \log(m_1) \geq 0\}$

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

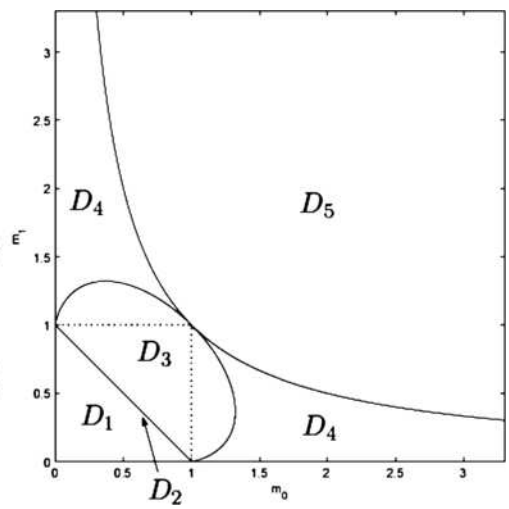

FIG. 2.

$m_0 m_1 \leq 1$, the organism recovers a.s. (meaning that the number of infected cells becomes negligible compared to the number of cells when $n \to \infty$). Otherwise the organism recovers iff parasites die out (and the probability is less than 1).

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

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

We shall prove that the contaminated cells are not concentrated in a cell line. Note that if $m_0 + m_1 > 1$, conditionally on Ext^c, $\partial\mathbb{T}^* \neq \varnothing$ 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 which come from the behavior of the BGW process $\mathcal{Z}_n$ and the BPRE Z_n (Section 2), shown in Figure 2.

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

$$\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\} / \#\mathbb{G}_n^* \overset{n\to\infty}{\longrightarrow} \mathbb{P}(\Upsilon = k),$$

where Υ is the Yaglom quasistationary distribution of the BPRE $(Z_n)_{n\in\mathbb{N}}$ (see [5, 12]). Note that the limit is deterministic and depends solely on the marginal laws of $(Z^{(0)}, Z^{(1)})$ (see Proposition 2.2). This gives then a way to compute Υ as a deterministic limit, although it is defined by conditioning on a vanishing event. Kimmel [15] 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 Υ.

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

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 Υ letting $n \to \infty$ (Corollary 5.4). This gives a pathwise interpretation that the limit of the Q-process associated to Z_n (see [1, 5]) is the size-biased quasistationary distribution.

THEOREM 1.2. *If $(m_0, m_1) \in D_1$, $(\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\})_{k\in\mathbb{N}}$ conditioned on $\mathcal{Z}_n > 0$ converges in distribution as $n \to \infty$ 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) and we get the following asymptotics (Theorem 3.1 and Corollaries 5.3, 5.6, 5.8).

If $(m_0, m_1) \in D_3$ (resp. D_5), then conditionally on Ext^c, $\#\mathbb{G}_n^*/(m_0+m_1)^n$ (resp. $\#\mathbb{G}_n^*/2^n$) converges in probability to a finite positive r.v.

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 r.v.

In the case $(m_0, m_1) \in D_4$, we get only some estimates of the asymptotic of $\#\mathbb{G}_n^*$ which are different from those which hold in the other domains. Our conjecture is that $\#\mathbb{G}_n^*$ has also a deterministic asymptotic, which depends on three subdomains (the interior of D_4 and its boundaries). 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 Beresticky and 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:

$$m := \tfrac{1}{2}(m_0 + m_1). \tag{5}$$

We use the classical notation, where for every $\mathbf{i} = (\alpha_1, \dots, \alpha_n) \in \mathbb{G}_n$,

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

2.1. *Results on the BGW process* $(\mathcal{Z}_n)_{n\in\mathbb{N}}$. The results stated hereafter are well known and can be found in [5]. First, the probability of extinction of the parasites satisfies

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

From now, we assume

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

Then there exists a random variable W such that

$$\frac{\mathcal{Z}_n}{(m_0+m_1)^n} \overset{n\to\infty}{\longrightarrow} W, \qquad \mathbb{P}(W = 0) = \mathbb{P}(\text{Ext}), \qquad \mathbb{E}(W) = 1. \tag{6}$$

In the case $m_0 + m_1 < 1$, there exists $b > 0$ such that $\mathbb{P}(\mathcal{Z}_n > 0) \overset{n\to\infty}{\sim} b(m_0+m_1)^n$. Then, there exists $U > 0$ such that

$$\mathbb{P}(\mathcal{Z}_n > 0) \ge U(m_0 + m_1)^n. \tag{7}$$

Moreover $(\mathcal{Z}_n)_{n\in\mathbb{N}}$ conditioned to be nonzero converges to a variable called the Yaglom quasistationary distribution and we set

$$\mathcal{B}(s) := \lim_{n\to\infty} \mathbb{E}(s^{\mathcal{Z}_n} \mid \mathcal{Z}_n > 0). \tag{8}$$

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

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

Moreover $\mathcal{B}$ is differentiable at 1 (Lemma 1 on page 44 in [5]) and we get

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

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 $\hat{m} := \mathbb{E}((Z^{(0)} + Z^{(1)})((Z^{(0)} + Z^{(1)}) - 1)) < \infty$ and $2m \ne 1$, then

$$\mathbb{E}(\mathcal{Z}_n(\mathcal{Z}_n - 1)) = \hat{m}(2m)^n \frac{(2m)^n - 1}{(2m)^2 - 2m}. \tag{11}$$

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 a BPRE with two equiprobable environments. More precisely, 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. For the general theory, see, for example, [9, 12, 13, 19]. 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

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

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 $s\in[0,1]$,

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

First, for every $n\in\mathbb{N}$, $\mathbb{E}(Z_{n+1}\mid Z_n) = mZ_n$ and $\mathbb{E}(Z_n) = m^n$.

Moreover, 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), we have the following result (see [19] or [3]).

PROPOSITION 2.1. *If $m_0m_1\le 1$, then $p = 1$. Otherwise $p < 1$.*

In the subcritical case ($m_0m_1 < 1$), the process Z_n conditioned to be nonzero which is denoted by Z_n^* converges weakly (Theorem 1.1 in [12]). By analogy with BGW, we call its limit distribution the Yaglom quasistationary distribution and denote it by Υ. That is,

$$\forall s\in[0,1] \qquad \mathbb{E}(s^{Z_n}\mid Z_n > 0) \stackrel{n\to\infty}{\longrightarrow} \mathbb{E}(s^{\Upsilon}) = G(s).$$

In the subcritical case, the asymptotics of $(\mathbb{P}(Z_n > 0))_{n\in\mathbb{N}}$ when n is large depends on the sign of $m_0\log(m_0) + m_1\log(m_1)$ (see [12]). Now, we require also that

$$m_0\log(m_0) + m_1\log(m_1) < 0; \qquad \mathbb{E}(Z_a\log^+(Z_a)) < \infty. \tag{13}$$

Then we say that Z_n is strongly subcritical and there exists $c > 0$ such that as n tends to ∞ (Theorem 1.1 in [12]),

$$\mathbb{P}(Z_n > 0) \sim cm^n. \tag{14}$$

Moreover, in that case, Υ is characterized by

PROPOSITION 2.2. *G is the unique probability generating function which satisfies*

$$G(0)=0, \qquad G'(1)<\infty, \tag{15}$$
$$\frac{G(f_0(s))+G(f_1(s))}{2}=mG(s)+(1-m).$$

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

LEMMA 2.3. *If $H:[0,1]\mapsto\mathbb{R}$ is continuous, $H(1)=0$ and*

$$H=\frac{H\circ f_0\cdot f_0'+H\circ f_1\cdot f_1'}{2m}, \tag{16}$$

then $H=0$.

PROOF OF PROPOSITION 2.2. The finiteness of $G'(1)=\mathbb{E}(\Upsilon)$ is the second part of Theorem 1.1 in [12].

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

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

And (14) ensures that $\mathbb{P}(Z_n>0)/\mathbb{P}(Z_{n+1}>0)\overset{n\to\infty}{\longrightarrow}m^{-1}$, 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 (15). Choose $\alpha>0$ such that $G'(1)=\alpha F'(1)$. Putting $H:=G-\alpha F$, H' satisfies equation (16) and $H'(1)=0$. Thus Lemma 2.3 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 2.3. If $H\neq 0$ then there exists $\alpha\in[0,1[$ such that

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

Let $\alpha_n \in [0,1[$ such that $\alpha_n \overset{n\to\infty}{\longrightarrow} 1$ and $\alpha \le \alpha_n \le 1$. Then, for every $n \in \mathbb{N}$, there exists $\beta_n \in [0,\alpha_n]$ such that:

$$\begin{aligned}\sup\{|H(s)| : s \in [0,\alpha_n]\} &= |H(\beta_n)| \\ &\le \frac{|H(f_0(\beta_n))|f_0'(\beta_n) + |H(f_1(\beta_n))|f_1'(\beta_n)}{2m} \\ &< \sup\{|H(s)| : s \in [0,1]\},\end{aligned}$$

since $\sup\{|H(s)| : s \in [0,1]\} \neq 0$ and $(2m)^{-1}(f_0'(\beta_n) + f_1'(\beta_n)) < 1$. As $I \cap J = \varnothing$, $\sup I < \sup(I \cup J) \Rightarrow \sup I < \sup J$, we get

$$\beta \le \sup\{|H(s)| : s \in [0,\alpha_n]\} < \sup\{|H(s)| : s \in]\alpha_n, 1]\}.$$

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

In the subcritical case ($m_0 m_1 \le 1$), if $m_0 \log(m_0) + m_1 \log(m_1) > 0$ [resp. $m_0 \log(m_0) + m_1 \log(m_1) = 0$], we say that Z_n is weakly subcritical (resp. intermediate subcritical) and we have $\mathbb{P}(Z_n > 0) \sim c' n^{-3/2} \gamma^n$ [resp. $\mathbb{P}(Z_n > 0) \sim c'' n^{-1/2} m^n$] for some $\gamma < m, c' > 0, c'' > 0$ (see [12] for details).

Finally we have the following expected result in the supercritical case [4].

PROPOSITION 2.4. *If $m_0 m_1 > 1$, $\mathbb{P}(Z_n \overset{n\to\infty}{\longrightarrow} \infty \mid \forall n \in \mathbb{N} : Z_n > 0) = 1$.*

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 $n \to \infty$. 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 $L \in [0,1]$ such that*

$$\#\mathbb{G}_n^*/2^n \overset{n\to\infty}{\longrightarrow} L.$$

If $m_0 m_1 \le 1$ then $\mathbb{P}(L = 0) = 1$.

Otherwise $\mathbb{P}(L = 0) < 1$ and $\{L = 0\} = \mathrm{Ext}$.

REMARK. In the case $m_0 + m_1 > 1$ and $m_0 m_1 \le 1$, the population of parasites may explode although the organism recovers.

This theorem states how unequal the sharing of parasites must be for the organism to recover. More precisely, let $m_0 = \alpha M$, $m_1 = (1-\alpha)M$ where $M > 0$ is the parasite growth rate. Then the organism recovers a.s. iff

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

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

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

Recalling that p is the probability of extinction of $(Z_n)_{n\in\mathbb{N}}$,

$$\forall n \in \mathbb{N} \qquad \mathbb{E}\left(\frac{\#\mathbb{G}_n^*}{2^n}\right) = \mathbb{P}(Z_n > 0) \overset{n\to\infty}{\longrightarrow} 1 - p. \tag{17}$$

The last equality gives also the asymptotic of $\mathbb{E}(\#\mathbb{G}_n^*)$ as $n \to \infty$ in the case $m_0 m_1 < 1$ [see Section 2.2 for the asymptotic of $\mathbb{P}(Z_n > 0)$, which depends on the sign of $m_0 \log(m_0) + m_1 \log(m_1)$] and in the case $m_0 m_1 = 1$ (see [2, 16]).

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

Monotone convergence of $\#\mathbb{G}_n^*/2^n$ to L as $n \to \infty$ and (17) ensure that $\mathbb{E}(L) = 1 - p$. Using Proposition 2.1, we get $\mathbb{P}(L = 0) = 1$ iff $m_0 m_1 \le 1$.

Obviously $\{L = 0\} \supset \mathrm{Ext}$. 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 $n \in \mathbb{N}$,

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

As $T_n := \inf\{k \ge 0 : \mathcal{Z}_k \ge n\}$ is a stopping time with respect to the natural filtration of $(Z_{\mathbf{i}})_{|\mathbf{i}|\le n}$, strong Markov property gives

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

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

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

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 $m_0 + m_1 \le 1$, 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 ∞ as $n \to \infty$. The proof of this result will also ensure that, if $m_0 + m_1 > 1$, the number of contaminated cells tends to ∞ provided that they do not die out.

THEOREM 4.1. *If $m_0+m_1\le 1$, $\#\{\mathbf{i}\in\mathbb{T}: Z_{\mathbf{i}}\neq 0, Z_{\mathbf{i}0}=0, Z_{\mathbf{i}1}=0\}$ conditioned by $\#\mathbb{G}_n^*>0$ converges in probability as $n\to\infty$ to ∞.*

If $m_0+m_1>1$, conditionally on Ext^c, $\#\mathbb{G}_n^* \overset{n\to\infty}{\longrightarrow} \infty$ 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 (see Section 5 for further results). In the case $m_0+m_1\le 1$, conditionally on $\#\mathbb{G}_n^*>0$, the tree of contaminated cells is a spine with finite subtrees, as for BGW conditioned to survive (see [11, 17]).

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 $k\ge 2$,*

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

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

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

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

$$\frac{1-p^x}{1-p^y}\ge\frac{x}{\max\{y,x\}}. \tag{18}$$

Let $a\in\{0,1\}$ and $\mathbf{k}$ such that $\mathbf{i}=\mathbf{j}a\mathbf{k}$. Then for all $(k_0',k_1')\in\mathbb{N}^2-(0,0)$,

$$\begin{aligned}
&\frac{\mathbb{P}(Z_{\mathbf{j}0}=k_0, Z_{\mathbf{j}1}=k_1 \mid Z_{\mathbf{j}}=1, Z_{\mathbf{i}}>0)}{\mathbb{P}(Z_{\mathbf{j}0}=k_0', Z_{\mathbf{j}1}=k_1' \mid Z_{\mathbf{j}}=1, Z_{\mathbf{i}}>0)}\\
&\qquad=\frac{\mathbb{P}(Z^{(0)}=k_0, Z^{(1)}=k_1 \mid Z_{a\mathbf{k}}>0)}{\mathbb{P}(Z^{(0)}=k_0', Z^{(1)}=k_1' \mid Z_{a\mathbf{k}}>0)}\\
&\qquad=\frac{\mathbb{P}(Z_{a\mathbf{k}}>0 \mid Z^{(0)}=k_0, Z^{(1)}=k_1)}{\mathbb{P}(Z_{a\mathbf{k}}>0 \mid Z^{(0)}=k_0', Z^{(1)}=k_1')}\frac{\mathbb{P}(Z^{(0)}=k_0, Z^{(1)}=k_1)}{\mathbb{P}(Z^{(0)}=k_0', Z^{(1)}=k_1')}
\end{aligned}$$

$$= \frac{1-\mathbb{P}(Z_{\mathbf{k}}=0)^{k_a}}{1-\mathbb{P}(Z_{\mathbf{k}}=0)^{k'_a}} \frac{\mathbb{P}((Z^{(0)},Z^{(1)})=(k_0,k_1))}{\mathbb{P}((Z^{(0)},Z^{(1)})=(k'_0,k'_1))}$$

$$\geq \frac{\min\{k_0,k_1\}}{k_0+k_1+k'_0+k'_1} \frac{\mathbb{P}((Z^{(0)},Z^{(1)})=(k_0,k_1))}{\mathbb{P}((Z^{(0)},Z^{(1)})=(k'_0,k'_1))} \qquad \text{using (18).}$$

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

$$[\mathbb{E}(Z^{(0)}+Z^{(1)})+k_0+k_1]\mathbb{P}(Z_{\mathbf{j}0}=k_0,Z_{\mathbf{j}1}=k_1 \mid Z_{\mathbf{j}}=1,Z_{\mathbf{i}}>0)$$
$$\geq \min\{k_0,k_1\}\mathbb{P}((Z^{(0)},Z^{(1)})=(k_0,k_1)).$$

This gives the result since $\mathbb{P}(Z_{\mathbf{j}0}=k_0,Z_{\mathbf{j}1}=k_1 \mid Z_{\mathbf{j}}=1,Z_{\mathbf{i}}>0)\geq\alpha$ 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 $Z_{\mathbf{j}}=2$ 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. □

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

$$\inf_{\mathbf{i}\in\mathbb{G}_n} \mathbb{P}(\#\{\mathbf{j}<\mathbf{i}: Z_{\mathbf{j}0}\geq 2 \text{or } Z_{\mathbf{j}1}\geq 2\}\geq \beta n/2 | Z_{\mathbf{i}}>0) \stackrel{n\to\infty}{\longrightarrow} 1.$$

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

$$\mathbb{P}(Z_{\mathbf{j}0}\geq 2 \text{ or } Z_{\mathbf{j}1}\geq 2 \mid Z_{\mathbf{j}}=\alpha, Z_{\mathbf{i}}>0)\geq \mathbb{P}(Z_0\geq 2 \text{ or } Z_1\geq 2 \mid Z_{\mathbf{k}}>0)\geq\beta.$$

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

Proof of Theorem 4.1. We consider first the case when $m_0+m_1>1$, work conditionally on 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 $\mathbb{P}(\cdot|Z_{\mathbf{i}}>0)$,

$$\#\{\mathbf{j}<\mathbf{i}: Z_{\mathbf{j}0}>0, Z_{\mathbf{j}1}>0\}=\infty.$$

Using the branching property and the fact that the probability of nonextinction of parasites is positive ensures that $\#\mathbb{G}_n^* \overset{n\to\infty}{\longrightarrow} \infty$ a.s.

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 $\mathbb{P}(\cdot|Z_{\mathbf{i}}>0)$,

$$\#\{\mathbf{j}<\mathbf{i}: Z_{\mathbf{j}0}\geq 2 \text{ or } Z_{\mathbf{j}1}\geq 2\}=\infty.$$

Using as above Lemma 4.2 (with $k=2$) and the fact that the probability of nonextinction of parasites is positive ensures that $\#\mathbb{G}_n^* \overset{n\to\infty}{\longrightarrow} \infty$ a.s.

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

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

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. □

5. Proportion of contaminated cells with a given number of parasites. We determine here the asymptotics of the number of contaminated cells and the proportion F_k of cells with k parasites, defined as

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

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:

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

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

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

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

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

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

Actually the convergence of $(Z_n^*)_{n\in\mathbb{N}}$ leads to the result obtained by Kimmel [15] in the continuous analogue of this model. That is,

$$\frac{\mathbb{P}(Z_n = k)}{\mathbb{P}(Z_n > 0)} = \frac{\sum_{\mathbf{i}\in\mathbb{G}_n} \mathbb{E}(\mathbb{1}_{Z_{\mathbf{i}}=k})}{\sum_{\mathbf{i}\in\mathbb{G}_n} \mathbb{E}(\mathbb{1}_{Z_{\mathbf{i}}>0})} = \frac{\mathbb{E}(\#\{\mathbf{i}\in\mathbb{G}_n : Z_{\mathbf{i}} = k\})}{\mathbb{E}(\#\mathbb{G}_n^*)} \tag{21}$$

tends to $\mathbb{P}(\Upsilon = k)$ 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 $n \to \infty$ (forthcoming Proposition 6.3). 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 6.4) and the control of filled-in cells (Lemma 6.5) using the results about the BPRE 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 $n \to \infty$ (by Proposition 2.4) 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 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* Ext^c, *for every* $k \in \mathbb{N}$, $F_k(n)$ *converges in probability to* 0 *as* $n \to \infty$, *that is,*

$$\forall K, \varepsilon > 0 \qquad \mathbb{P}^*\left(\frac{\#\{\mathbf{i}\in\mathbb{G}_n : Z_{\mathbf{i}} \geq K\}}{\#\mathbb{G}_n^*} \geq 1 - \varepsilon\right) \overset{n\to\infty}{\longrightarrow} 1.$$

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

$$\sup_{n\in\mathbb{N}}\left\{\mathbb{P}^*\left(\frac{\#\{\mathbf{i}\in\mathbb{G}_n^* : Z_{\mathbf{i}} \le \alpha m_0^n\}}{\#\mathbb{G}_n^*} \ge \varepsilon\right)\right\} \xrightarrow{\alpha\to 0} 0.$$

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

$$\#\mathbb{G}_n^* \ge 2^n \widetilde{L}, \qquad \mathbb{P}^*(\widetilde{L} = 0) = 0. \tag{22}$$

Let K, η and $\varepsilon > 0$ and put $B_n(K,\eta) := \{\frac{\#\{\mathbf{i}\in\mathbb{G}_n^* : Z_{\mathbf{i}}\le K\}}{\#\mathbb{G}_n^*} \ge \eta\} \cap \mathrm{Ext}^c$, then

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

which gives, taking expectations,

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

Use then Proposition 2.4 and (22) to choose n large enough so that

$$\mathbb{P}(B_n(K,\eta)) \le \varepsilon,$$

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 Ext^c (see [4]) to get the finer result given after the theorem. □

5.2. *Case* $(m_0, m_1) \in D_3$ $(m \le 1)$. 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 Υ (see Section 6.1 for proofs).

Theorem 5.2. *Conditionally on* Ext^c, *as* $n \to \infty$, $(F_k(n))_{k\in\mathbb{N}}$ *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 Υ.

The limit just depends on the one-dimensional distributions of $(Z^{(0)}, Z^{(1)})$. More precisely, recall that the probability generating function G of Υ is characterized by (15).

This theorem still holds starting from k parasites. We also easily get 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 $\mathbb{P}(\mathrm{Ext}) \neq 0$, 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 [14]. 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 [14] 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 $\mathbb{P}(\Upsilon = 1) = 1$.

Symmetric linear fractional case: $p \in]0,1[$, $b \in]0,(1-p)^2[$ and

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

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,

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

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* Ext^c, *the following convergences hold in probability*

$$\frac{\#\mathbb{G}_n^*}{\mathcal{Z}_n} \stackrel{n\to\infty}{\longrightarrow} \frac{1}{\mathbb{E}(\Upsilon)}, \qquad \frac{\#\mathbb{G}_n^*}{(m_0+m_1)^n} \stackrel{n\to\infty}{\longrightarrow} \frac{W}{\mathbb{E}(\Upsilon)}.$$

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

COROLLARY 5.4. *Conditionally on* 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* $n \to \infty$:

$$\forall k \in \mathbb{N} \qquad \lim_{n\to\infty} \lim_{p\to\infty} F_k(n,p) \overset{\mathbb{P}}{=} \frac{k\mathbb{P}(\Upsilon = k)}{\mathbb{E}(\Upsilon)}.$$

We get here an interpretation of the fact that the stationary distribution of the Q-process associated to the BPRE $(Z_n)_{n\in\mathbb{N}}$ is the size-biased Yaglom limit (see [1]).

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* $n \to \infty$, $(F_k(n))_{k\in\mathbb{N}}$ *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 (13) is still satisfied and we can use the same 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 $\mathbb{P}(\mathcal{Z}_n > 0) \overset{n\to\infty}{\sim} 2/(\mathrm{Var}(Z^{(0)} + Z^{(1)})n)$ and $\mathcal{Z}_n/n$ conditioned to be nonzero converges in distribution as $n \to \infty$ 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* $n \to \infty$, $\#\mathbb{G}_n^*/n$ *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 nondeterministic limit (see Section 7 for proofs).

THEOREM 5.7. *As* $n \to \infty$, $(\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\})_{k\in\mathbb{N}}$ *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 $n \to \infty$.*

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

5.5. *Remaining domain*: $(m_0, m_1) \in D_4$. In this domain, the asymptotic of the mean of the number of contaminated cells, that is $\mathbb{E}(\#\mathbb{G}^*_n) = 2^n \mathbb{P}(Z_n > 0)$, is different from the previous ones.

Recalling Section 2.2, this asymptotic depends on three subdomains, the interior of D_4 and the two connex components of its boundary. More precisely, it 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 or zero.

If $(m_0, m_1) \in D_4$ and $m_0 < 1 < m_1$, using (17) and a coupling argument with Corollary 5.3, one can prove that

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

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 $\mathbb{E}(\#\mathbb{G}^*_n) = 2^n \mathbb{P}(Z_n > 0)$.

Moreover separation of descendances of parasites, control of filled-in cells and Corollary 5.4 do not hold in this case. Thus determining the limit behaviors here requires a different approach.

Finally, note that in the subdomain $m_0 m_1 = 1$ (boundary of D_5), $(Z^*_n)_{n\in\mathbb{N}}$ explodes (see [2]) so the asymptotic proportion of contaminated cells which are arbitrarily largely contaminated should be equal to 1 as in Theorem 5.1.

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

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

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

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

and as $\bigcap_{n\in\mathbb{N}}\{\mathcal{Z}_n>0\}=\{\forall n\in\mathbb{N}:\mathcal{Z}_n>0\}$, we have

$$\sup_A\{|\mathbb{P}^n(A)-\mathbb{P}^*(A)|\}\overset{n\to\infty}{\longrightarrow}0. \tag{25}$$

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

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

Then identities (26) and (14) entail that there exists $M>0$ such that

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

Moreover, by Corolary 2.3 in [1], we have

$$\lim_{K\to\infty}\sup_{n\in\mathbb{N}}\{\mathbb{E}(Z_n\mathbb{1}_{Z_n\ge K}\mid Z_n>0)\}=0. \tag{28}$$

Finally, following the proof of Theorem 1.2 in [13] (see [7], Section 2.1 for details) ensures that, if $(Z_n^{(1)})_{n\in\mathbb{N}}$ and $(Z_n^{(2)})_{n\in\mathbb{N}}$ are two independent BPRE distributed as $(Z_n)_{n\in\mathbb{N}}$, we have

$$\mathbb{P}(Z_n^{(1)}>0,Z_n^{(2)}>0)=o(\mathbb{P}(Z_n>0))=o(m^n)\qquad(n\to\infty).$$

Then, we have

$$2^{-n}\sum_{\mathbf{i}\in\mathbb{G}_n}\mathbb{P}(Z_{\mathbf{i}}>0)^2=o(m^n)\qquad(n\to\infty). \tag{29}$$

6.2. *Estimation of* $\#\mathbb{G}_n^*$. We prove here that the number of parasites which belong to filled-in cells is negligible compared to the total number of parasites (see also Lemma 6.5 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 nonzero.

LEMMA 6.1. *For every* $\eta>0$,

$$\sup_{n\in\mathbb{N}}\left\{\mathbb{P}^*\left(\frac{\sum_{\mathbf{i}\in\mathbb{G}_n^*}Z_{\mathbf{i}}\mathbb{1}_{\{Z_{\mathbf{i}}>K\}}}{\mathcal{Z}_n}\ge\eta\right)\right\}\overset{K\to\infty}{\longrightarrow}0.$$

PROOF. Let $\eta>0$ and write

$$A_n(K,\eta):=\left\{\frac{\sum_{\mathbf{i}\in\mathbb{G}_n^*}Z_{\mathbf{i}}\mathbb{1}_{\{Z_{\mathbf{i}}>K\}}}{\mathcal{Z}_n}\ge\eta\right\}\cap\mathrm{Ext}^c.$$

Then

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

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,

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

$$m^{-n}\mathbb{E}(Z_n\mathbb{1}_{\{Z_n>K\}})/\eta \geq \mathbb{E}(\mathbb{1}_{A_n(K,\eta)}C).$$

Then, by (28), we have

$$\lim_{K\to\infty}\sup_{n\in\mathbb{N}}\{\mathbb{E}(\mathbb{1}_{A_n(K,\eta)}C)\} = 0.$$

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 $\forall K\geq K_0$, $\forall n\in\mathbb{N}$,

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

which completes the proof. □

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

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

Adding that conditionally on Ext^c, $\mathcal{Z}_n \overset{n\to\infty}{\longrightarrow} \infty$ a.s, gives the following result.

Proposition 6.2. *Let $\varepsilon>0$, there exists $K\in\mathbb{N}$ such that $\forall N\in\mathbb{N}$, $\exists n_0\in\mathbb{N}$ such that $\forall n\geq n_0$,*

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

Second, we derive an estimation of $\#\mathbb{G}_n^*$. By Lemma 6.1, 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 6.3. *For every $\varepsilon>0$, there exist $A,B>0$ such that for every $n\in\mathbb{N}$,*

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

PROOF. First use (24) to get

$$\frac{\#\mathbb{G}_n^*}{(2m)^n} \le \frac{\mathcal{Z}_n}{(2m)^n} \le D.$$

Moreover using again (24), we have

$$\frac{\#\mathbb{G}_n^*}{(2m)^n} \ge \frac{\sum_{\mathbf{i}\in\mathbb{G}_n^*} Z_{\mathbf{i}}\mathbb{1}_{\{Z_{\mathbf{i}}\le K\}}}{K(2m)^n} \ge \frac{C}{K}\frac{\sum_{\mathbf{i}\in\mathbb{G}_n^*} Z_{\mathbf{i}}\mathbb{1}_{\{Z_{\mathbf{i}}\le K\}}}{\mathcal{Z}_n}$$

and Lemma 6.1 gives the result. □

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

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

PROOF. Let $K\in\mathbb{N}$, $\eta>0$ and consider for $A>0$,

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

Then

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

so that taking expectations,

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

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

$$\mathbb{P}_K(N_0(\mathbf{i})\ge 2)\le\binom{K}{2}\mathbb{P}(Z_{\mathbf{i}}>0)^2.$$

Then

$$\mathbb{P}(E_n^q(\eta)) \leq \frac{\binom{K}{2} 2^{-q} \sum_{\mathbf{i}\in\mathbb{G}_q} \mathbb{P}(Z_{\mathbf{i}} > 0)^2}{\eta A m^q}.$$

Conclude choosing A in agreement with Proposition 6.3 and q with (29). □

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

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

PROOF. Let $\eta > 0$, $A > 0$ 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

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

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

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

By (29), we get

$$\lim_{K\to\infty} \sup_{n\in\mathbb{N}} \{\mathbb{P}(F_n^q(\eta))\} = 0.$$

Complete the proof choosing A in agreement with Proposition 6.3. □

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 6.5). 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 6.4). Thus 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 ($n \to \infty$, $m_0 + m_1 > 1$), we have a law of large numbers phenomenon and get a deterministic limit.

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

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

For every $k \in \mathbb{N}^*$ and every parasite $\mathbf{p}$ in generation 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, $\mathbf{p} \hookrightarrow \mathbf{i}$ when $\mathbf{p}$ belongs to the cell $\mathbf{i}$ and $\mathbf{p}|n$ its ancestor (parasite) in generation n,

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

By the branching property, $(Y_k^q(\mathbf{p}))_{k\in\mathbb{N}}$ $[\mathbf{p} \in \mathcal{P}(n)]$ are i.i.d. and we denote by $(Y_k^q)_{k\in\mathbb{N}}$ a random variable with this common distribution. Denoting by $\mathcal{P}_K(n)$ the set of parasites in generation n which belong to a cell containing at most K parasites, we have

$$\begin{aligned}\sum_{k\in\mathbb{N}^*} &\left|\#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}} = k\} - \sum_{\mathbf{p}\in\mathcal{P}_K(n)} Y_k^q(\mathbf{p})\right| \\ &\leq (K+1)\#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}|n} \leq K, N_n(\mathbf{i}) \geq 2\} + \#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}|n} > K\}.\end{aligned} \tag{30}$$

Indeed, the left-hand side of (30) is less than

$$\sum_{k\in\mathbb{N}^*} \left|\#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}} = k, Z_{\mathbf{i}|n} \leq K\} - \sum_{\mathbf{p}\in\mathcal{P}_K(n)} Y_k^q(\mathbf{p})\right| + \#\{\mathbf{i} \in \mathbb{G}_{n+q}^* : Z_{\mathbf{i}|n} > K\}.$$

And 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 equalities:

$$\sum_{\mathbf{p}\in\mathcal{P}_K(n)} Y_k^q(\mathbf{p}) = \sum_{\mathbf{i}\in\mathbb{G}_{n+q}} \sum_{\mathbf{p}\in\mathcal{P}_K(n)} \mathbb{1}_{\#\{\mathbf{r}\,:\,\mathbf{r}\hookrightarrow\mathbf{i},\mathbf{r}|n=\mathbf{p}\}=k}$$

and

$$\mathbb{1}_{Z_{\mathbf{i}}=k, Z_{\mathbf{i}|n}\leq K, N_n(\mathbf{i})=1} = \mathbb{1}_{N_n(\mathbf{i})=1} \sum_{\mathbf{p}\in\mathcal{P}_K(n)} \mathbb{1}_{\#\{\mathbf{r}\,:\,\mathbf{r}\hookrightarrow\mathbf{i},\mathbf{r}|n=\mathbf{p}\}=k}$$

which ensure

$$\begin{aligned}
&\sum_{k\in\mathbb{N}^*}\left|\#\{\mathbf{i}\in\mathbb{G}^*_{n+q}: Z_{\mathbf{i}}=k, Z_{\mathbf{i}|n}\le K\}-\sum_{\mathbf{p}\in\mathcal{P}_K(n)}Y^q_k(\mathbf{p})\right|\\
&\quad\le\sum_{k\in\mathbb{N}^*}\sum_{\mathbf{i}\in\mathbb{G}_{n+q},N_n(\mathbf{i})\ge 2}\left|\mathbb{1}_{Z_{\mathbf{i}}=k,Z_{\mathbf{i}|n}\le K}-\sum_{\mathbf{p}\in\mathcal{P}_K(n)}\mathbb{1}_{\#\{\mathbf{r}:\mathbf{r}\hookrightarrow\mathbf{i},\mathbf{r}|n=\mathbf{p}\}=k}\right|\\
&\quad\le\#\{\mathbf{i}\in\mathbb{G}^*_{n+q}: Z_{\mathbf{i}|n}\le K,\ N_n(\mathbf{i})\ge 2\}+\sum_{\substack{\mathbf{i}\in\mathbb{G}_{n+q},N_n(\mathbf{i})\ge 2\\ \mathbf{p}\in\mathcal{P}_K(n)}}\mathbb{1}_{\#\{\mathbf{r}:\mathbf{r}\hookrightarrow\mathbf{i},\mathbf{r}|n=\mathbf{p}\}>0}\\
&\quad\le\#\{\mathbf{i}\in\mathbb{G}^*_{n+q}: Z_{\mathbf{i}|n}\le K,\ N_n(\mathbf{i})\ge 2\}+\sum_{\mathbf{i}\in\mathbb{G}_{n+q},N_n(\mathbf{i})\ge 2}K\mathbb{1}_{Z_{\mathbf{i}|n}\le K}\\
&\quad=(K+1)\#\{\mathbf{i}\in\mathbb{G}^*_{n+q}: Z_{\mathbf{i}|n}\le K, N_n(\mathbf{i})\ge 2\}.
\end{aligned}$$

We shall now prove that the quantities on the right-hand side of (30) are small when n and q are large enough and that $\sum_{\mathbf{p}\in\mathcal{P}_K(n)}Y^q_k(\mathbf{p})$ follow a law of large number. To that purpose, let $\varepsilon,\eta>0$ and for all $K,k,n,q\ge 0$ define

$$G^K_k(n,q):=\frac{\sum_{\mathbf{p}\in\mathcal{P}_K(n)}Y^q_k(\mathbf{p})}{\sum_{k\in\mathbb{N}}\sum_{\mathbf{p}\in\mathcal{P}_K(n)}Y^q_k(\mathbf{p})}.$$

First, by Proposition 6.2 and (25), $\exists K_1\in\mathbb{N}$ such that $\forall N\in\mathbb{N}$, $\exists n_1\in\mathbb{N}$ such that $\forall K\ge K_1$, $\forall n\ge n_1$,

$$\mathbb{P}^n(|\mathcal{P}_K(n)|\ge N)\ge 1-\varepsilon. \tag{31}$$

Moreover by Lemma 6.5, $\exists K_2\ge K_1$ such that $\forall n,q\in\mathbb{N}$,

$$\mathbb{P}^*\left(\frac{\#\{\mathbf{i}\in\mathbb{G}^*_{n+q}: Z_{\mathbf{i}|n}>K_2\}}{\#\mathbb{G}^*_{n+q}}\ge\eta\right)\le\varepsilon. \tag{32}$$

And by Proposition 6.4, $\exists q_0\in\mathbb{N}$ such that $\forall n\in\mathbb{N}$,

$$\mathbb{P}^*\left(\frac{\#\{\mathbf{i}\in\mathbb{G}^*_{n+q_0}: Z_{\mathbf{i}|n}\le K_2, N_n(\mathbf{i})\ge 2\}}{\#\mathbb{G}^*_{n+q_0}}\ge\eta/(K_2+1)\right)\le\varepsilon. \tag{33}$$

Use then (30), (32) and (33) to get

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

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

$$\mathbb{P}^*(\|(F_k(n+q_0))_{k\in\mathbb{N}}-(G^{K_2}_k(n,q_0))_{k\in\mathbb{N}}\|_1\ge 4\eta)\le 2\varepsilon. \tag{34}$$

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

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

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

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

So using (31), there exists $n_1 \in \mathbb{N}$ such that for every $\forall n \geq n_1$,

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

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

$$\mathbb{P}^*(\|(G_k^{K_2}(n, q_0))_{k \in \mathbb{N}^*} - \vec{f}(q_0)\|_1 \geq \eta) \leq 3\varepsilon. \tag{35}$$

As a conclusion, using (34) and (35), we have proved that for all $\varepsilon, \eta > 0$, and for every $n \geq n_2 + q_0$,

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

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 $n \to \infty$.

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

By Proposition 2.2, we have

$$\forall k \in \mathbb{N} \qquad \mathbb{P}(Z_n = k \mid Z_n \neq 0) \overset{n \to \infty}{\longrightarrow} \mathbb{P}(\Upsilon = k). \tag{36}$$

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

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

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

$$\forall k \in \mathbb{N} \qquad \mathbb{P}(Z_n = k \mid Z_n \neq 0) \overset{n \to \infty}{\longrightarrow} f_k. \tag{37}$$

Indeed, by Proposition 6.3, 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

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

By (11), $\mathcal{Z}_n/(2m)^n$ 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 (37). Putting (36) and (37) 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 2.2) and note also that for every $K\in\mathbb{N}^*$,

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

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

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

Let $\eta,\varepsilon>0$. We use Lemma 6.1 to choose $K\in\mathbb{N}^*$ such that

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

$$\left|\frac{1}{\mathbb{E}(\Upsilon\mathbb{1}_{\Upsilon\leq K})}-\frac{1}{\mathbb{E}(\Upsilon)}\right|\leq\eta.$$

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

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

Then for every $n \geq n_0$,

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

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

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

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

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

$$\mathbb{E}(W_k(\mathbf{j})) = \frac{k\mathbb{E}(W)}{\mathbb{E}(\Upsilon)} = \frac{k}{\mathbb{E}(\Upsilon)}. \tag{38}$$

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

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

Moreover, Theorem 5.2 ensures that

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

And conditionally on $\#\mathbb{G}_n^* > 0$, $W_k(\mathbf{j})$ $(\mathbf{j} \in \mathbb{G}_n^*)$ is i.i.d. by the branching property and $\#\mathbb{G}_n^*$ tends to infinity. So the law of large numbers and (38) ensure that

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

which ends the proof. □

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 $n \in \mathbb{N}$,*

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

PROOF. Denoting by E^q_n the event

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

we have

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

Thus we can follow the proof of Lemma 6.4.

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

Conclude with (29). □

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 $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$,

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

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

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

$$\lim_{n \to \infty} \mathbb{P}^{n+q}(\mathcal{Z}_n > K) \leq \varepsilon. \tag{39}$$

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

$$\mathbb{P}^{n+q_1}(\{\mathbf{i} \in \mathbb{G}^*_{n+q_1} : N_n(\mathbf{i}) \geq 2\} \neq \varnothing, \mathcal{Z}_n \leq K) \leq \varepsilon. \tag{40}$$

And by (39), there exists $n_0 \geq 0$ such that for every $n \geq n_0$,

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

Then

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

Moreover,

$$\begin{aligned} &\#\{\mathbf{i} \in \mathbb{G}^*_{n+q_1} : N_n(\mathbf{i}) \geq 2\} = 0 \\ &\quad \Longrightarrow \quad (\#\{\mathbf{i} \in \mathbb{G}^*_{n+q_1} : Z_{\mathbf{i}} = k\})_{k\in\mathbb{N}} = (N_k(n, q_1))_{k\in\mathbb{N}}. \end{aligned}$$

Then for every $n \geq n_0$,

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

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 [18])

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

where

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

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 $n \geq n_0(l)$,

$$d(\mathbb{P}^n((\#\{\mathbf{i} \in \mathbb{G}^*_n : Z_{\mathbf{i}} = k\})_{k\in\mathbb{N}} \in \cdot), \mu(l)) \leq 1/2^l. \tag{41}$$

For that purpose, let $l \in \mathbb{N}$. By Step 1, choose $q, n_0 \in \mathbb{N}$ such that

$$\begin{aligned} \forall n \geq n_0 \qquad & d(\mathbb{P}^{n+q}((\#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_{\mathbf{i}} = k\})_{k\in\mathbb{N}} \in \cdot), \\ & \qquad \mathbb{P}^{n+q}((N_k(n, q))_{k\in\mathbb{N}} \in \cdot)) \leq 1/2^{l+1}. \end{aligned} \tag{42}$$

Recall that $(Y^q_k(\mathbf{p}))_{k\in\mathbb{N}}(\mathbf{p} \in \mathcal{P}(n))$ is an i.i.d. sequence distributed as $(Y^q_k)_{k\in\mathbb{N}}$ 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 i.i.d., distributed as Y^q_k and independent of $\mathcal{Z}_n$, conditionally on $\sum_{k\in\mathbb{N}} \sum_{\mathbf{p}\in\mathcal{P}(n)} Y^q_k(\mathbf{p}) > 0$.

Moreover $\mathbb{P}^{n+q}(\mathcal{Z}_n \in \cdot)$ converges weakly as $n \to \infty$ to a probability ν [see (9)] and we denote by $\mathcal{N}$ a random variable with distribution ν and

by $(Y_k^q(i))_{k\in\mathbb{N}}(i\in\mathbb{N})$ an i.i.d. sequence independent of $\mathcal{N}$ and distributed as $(Y_k^q)_{k\in\mathbb{N}}$. Then we have for n large enough,

$$d(\mathbb{P}^{n+q}((N_k(n,q))_{k\in\ \mathbb{N}}\in\cdot),\mu(l))\le 1/2^l, \tag{43}$$

where $\mu(l)$ is the distribution of $(\sum_{1\le i\le\mathcal{N}} Y_k^q(i))_{k\in\mathbb{N}}$ conditionally on $\sum_{k\in\mathbb{N}}\sum_{1\le i\le\mathcal{N}} Y_k^q(i)>0$. Combining (42) and (43) gives (41).

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 [18]), $(\mu(l))_{l\in\mathbb{N}}$ converges and we get the convergence of Theorem 5.7.

We now prove that $\mathbb{E}(\sum_{k\in\mathbb{N}^*} kN_k)<\infty$. For all $n,K>0$, we have

$$\mathbb{E}\left(\sum_{k\ge K} k\#\{\mathbf{i}\in\mathbb{G}_n^*: Z_{\mathbf{i}}=k\}|\mathcal{Z}_n>0\right)\le \mathbb{E}(\mathcal{Z}_n\mathbb{1}_{\{\mathcal{Z}_n\ge K\}}|\mathcal{Z}_n>0)\le\frac{\mathbb{E}(\mathcal{Z}_n^2)}{\mathbb{P}(\mathcal{Z}_n>0)K}$$

which converges uniformly to 0 as $K\to\infty$ using (11). Moreover, Theorem 5.7 and $k\#\{\mathbf{i}\in\mathbb{G}_n^*: Z_{\mathbf{i}}=k\}\le\mathcal{Z}_n$ ensure that

$$\lim_{n\to\infty}\mathbb{E}\left(\sum_{1\le k\le K} k\#\{\mathbf{i}\in\mathbb{G}_n^*: Z_{\mathbf{i}}=k\}\mid\mathcal{Z}_n>0\right)=\mathbb{E}\left(\sum_{1\le k\le K} kN_k\right).$$

Thus we get the expected limit

$$\mathbb{E}\left(\sum_{k\in\mathbb{N}} k\#\{\mathbf{i}\in\mathbb{G}_n^*: Z_{\mathbf{i}}=k\}\mid\mathcal{Z}_n>0\right)\overset{n\to\infty}{\longrightarrow}\mathbb{E}\left(\sum_{k\in\mathbb{N}^*} kN_k\right)$$

and recalling Section 2.1, we have also

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

This completes the proof. □

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

Acknowledgment. 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] Afanasyev, V. I., Geiger, J., Kersting, G. and Vatutin, V. A. (2005). Functional limit theorems for strongly subcritical branching processes in random environment. *Stochastic Process. Appl.* **115** 1658–1676. MR2165338

[2] Afanasyev, V. I., Geiger, J., Kersting, G. and Vatutin, V. A. (2005). Criticality for branching processes in random environment. *Ann. Probab.* **33** 645–673. MR2123206

[3] Athreya, K. B. and Karlin, S. (1971). On branching processes with random environments. I. Extinction probability *Ann. Math. Statist.* **42** 1499–1520. MR0298780

[4] Athreya, K. B. and Karlin, S. (1971). On branching processes with random environments. II. Limit theorems. *Ann. Math. Statist.* **42** 1843–1858. MR0298781

[5] Athreya, K. B. and Ney, P. E. (2004). *Branching Processes.* Dover Publications, Inc., Mineola, NY. MR2047480

[6] Athreya, K. B. and Kang, H. J. (1998). Some limit theorems for positive recurrent branching Markov chains. I, II. *Adv. in Appl. Probab.* **30** 693–710. MR1663545

[7] Bansaye, V. (2007). Subcritical branching processes in random environment starting from many particules. To appear.

[8] Benjamini, I. and Peres, Y. (1994). Markov chains indexed by trees. *Ann. Probab.* **22** 219–243. MR1258875

[9] Dekking, F. M. (1988). On the survival probability of a branching process in a finite state iid environment. *Stochastic Process. Appl.* **27** 151–157. MR0934535

[10] Evans, S. N. and Steinsaltz, D. (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`.

[11] Geiger, J. (1999). Elementary new proofs of classical limit theorems for Galton–Watson processes. *J. Appl. Probab.* **36** 301–309. MR1724856

[12] Geiger, J., Kersting, G. and Vatutin, V. A. (2003). Limit theorems for subcritical branching processes in random environment. *Ann. Inst. H. Poincaré Probab. Statist.* **39** 593–620. MR1983172

[13] Guivarc'h, Y. and Liu, Q. (2001). Asymptotic properties of branching processes in random environment. *C. R. Acad. Sci. Paris Sér. I* **332** 339–344. MR1821473

[14] Guyon, J. (2007). Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.* **17** 1538–1569. MR2358633

[15] Kimmel, M. (1997). Quasistationarity in a branching model of division-within-division. *Classical and Modern Branching Processes* (K. B. Athreya and P. Jagers, eds.) 157–164. Springer, New York. MR1601725

[16] Kozlov, M. V. (1976). The asymptotic behavior of the probability of non-extinction of critical branching processes in a random environment. *Theor. Probab. Appl.* **21** 813–825. MR0428492

[17] Lyons, R., Pemantle, R. and Peres, Y. (1995). Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.* **23** 1125–1138. MR1349164

[18] Parthasarathy, K. R. (1967). *Probability Measures on Metric Spaces.* Academic Press, New York. MR0226684

[19] Smith, W. L. and Wilkinson, W. (1969). On branching processes in random environments. *Ann. Math. Statist.* **40** 814–827. MR0246380

[20] Stewart, E. J., Madden, R., Paul, G. and Taddei, F. (2005). Aging and death in a organism that reproduces by morphologically symmetric division. *PLoS Biol* **3** e45.

UPMC et C.N.R.S. UMR 7599
175 rue du Chevaleret
75 013 Paris
France
E-mail: bansaye@ccr.jussieu.fr