



# The distribution of the quasispecies for a Moran model on the sharp peak landscape

Raphaël Cerf, Joseba Dalmau

## ► To cite this version:

Raphaël Cerf, Joseba Dalmau. The distribution of the quasispecies for a Moran model on the sharp peak landscape. Stochastic Processes and their Applications, 2016, 126 (6), pp.1681-1709. 10.1016/j.spa.2015.12.002 . hal-03852263

**HAL Id: hal-03852263**

**<https://hal.science/hal-03852263>**

Submitted on 14 Nov 2022

**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 distribution of the quasispecies for a Moran model on the sharp peak landscape

Raphaël Cerf and Joseba Dalmau  
Université Paris Sud and ENS Paris

May 5, 2016

## Abstract

We consider the Moran model on the sharp peak landscape, in the asymptotic regime studied in [3], where a quasispecies is formed. We find explicitly the distribution of this quasispecies.

## 1 Introduction

In his paper [10], Eigen introduced the model of quasispecies to describe the evolution of a population of macromolecules which is subject to two main forces: mutation and selection. The model was developed further in a series of papers by Eigen and Schuster [12, 13, 14], and analysed in great detail by Eigen, McCaskill and Schuster in [11]. A major conclusion is that this kind of evolutionary process, rather than selecting a single dominant species, is more likely to select a master sequence (the macromolecule with the highest fitness) along with a cloud of mutants that closely resemble the master sequence. Hence the name quasispecies. One other major discovery that Eigen made was the existence of an error threshold allowing a quasispecies to form: if the mutation rate exceeds the error threshold, then the population evolves towards a totally random state, whereas if the mutation rate is below the error threshold, a quasispecies can be formed.

Even if Eigen's original goal was to explain the behaviour of a population of macromolecules, the theory of quasispecies rapidly extended to other areas of biology. In particular, experimental studies support the validity of the model in virology [9]. Some RNA viruses are known to have very high mutation rates, like the HIV virus, and this is a factor of resistance against conventional drugs. A promising strategy to combat this kind of viruses consists in developing mutagenic drugs that would increase the mutation rate

beyond the error threshold, in order to induce an error catastrophe [2, 24]. This strategy has successfully been applied to several types of RNA viruses [5]. Moreover, several similarities have been observed between the evolution of cancer cell populations and RNA viruses, in particular, the possibility of inducing an error catastrophe [23].

Two important features of Eigen’s model are its deterministic nature (the model is based on a system of differential equations derived from certain chemical and physical laws) and the fact that the population is considered to be infinite. When dealing with simple macromolecules, these assumptions are quite natural. Nevertheless, they become unrealistic if we want to apply this model to population genetics, and they are two of the major drawbacks when applying it to virus populations, as pointed out by Wilke [25]. On one hand, we have to take into account the stochastic nature of the evolution of a finite population. The higher the complexity of the individuals, the harder it is to explain the replication and mutation schemes via chemical reactions. This fact, together with the widely recognised role of randomness in evolutionary processes strongly suggest a stochastic approach to the matter. On the other hand, when dealing with populations of complex individuals, the amount of possible genotypes largely exceeds the size of the population. Therefore, if we want to use Eigen’s model in population genetics, a finite and stochastic version of the model is called for.

The interest of a finite stochastic counterpart to Eigen’s model is not new. Eigen, McCaskill and Schuster already emphasise the importance of developing such a model [11], so does Wilke in the more recent paper [25]. Several researchers have pursued this task. Demetrius, Schuster and Sigmund [7] introduce stochasticity into Eigen’s model using branching processes. McCaskill [17] also develops a stochastic version of Eigen’s model. Nowak and Schuster [20] use birth and death Markov processes to give a finite stochastic version of Eigen’s model on the sharp peak landscape. Alves and Fontanari [1] study the dependence of the error threshold on the population size for the sharp peak replication landscape. Saakian, Deem and Hu [22] compute the variance of the mean fitness in a finite population model in order to control how it approximates the infinite population model. Deem, Muñoz and Park [21] use a field theoretic representation in order to derive analytical results. Other recent papers introduce finite stochastic models that approach Eigen’s model asymptotically when the population size goes to  $\infty$ , like Musso [19] or Dixit, Srivastava, Vishnoi [8].

In [3], Cerf studies a population of size  $m$  of chromosomes of length  $\ell$  over an

alphabet  $\mathcal{A}$  of cardinality  $\kappa$ , which evolves according to a Moran model [18]. The mutation probability per locus is  $q$ . Only the sharp peak landscape is considered: the master sequence, which we denote by  $w^*$ , replicates with rate  $\sigma > 1$ , while all the other sequences replicate with rate 1. In the asymptotic regime where

$$\begin{aligned} \ell \rightarrow +\infty, \quad m \rightarrow +\infty, \quad q \rightarrow 0, \\ \ell q \rightarrow a, \quad \frac{m}{\ell} \rightarrow \alpha, \end{aligned}$$

a critical curve is obtained in the parameter space  $(a, \alpha)$ , which is given by  $\alpha\phi(a) = \ln \kappa$ . If  $\alpha\phi(a) < \ln \kappa$ , then the population is totally random, i.e., the fraction of the master sequence in a population at equilibrium converges to 0. On the contrary, if  $\alpha\phi(a) > \ln \kappa$ , then a quasispecies is formed, i.e., at equilibrium, the population contains a positive fraction of the master sequence, which in the asymptotic regime presented above converges to  $(\sigma e^{-a} - 1)/(\sigma - 1)$ .

The aim of our article is to obtain the whole distribution of the quasispecies. As it is customary with this kind of models, we introduce Hamming classes with respect to the master sequence in the space  $\mathcal{A}^\ell$  of sequences of length  $\ell$ . We say that a chromosome  $u \in \mathcal{A}^\ell$  belongs to the class  $k \in \{0, \dots, \ell\}$  if it differs from the master sequence in exactly  $k$  characters. We study then the fraction of each of these classes in a population at equilibrium. For  $k \geq 0$  fixed, in the above asymptotic regime, we recover the critical curve  $\alpha\phi(a) = \ln \kappa$ . If  $\alpha\phi(a) < \ln \kappa$ , then the fraction of the class  $k$  converges to 0, whereas if  $\alpha\phi(a) > \ln \kappa$ , then the fraction of the class  $k$  in a population at equilibrium converges to

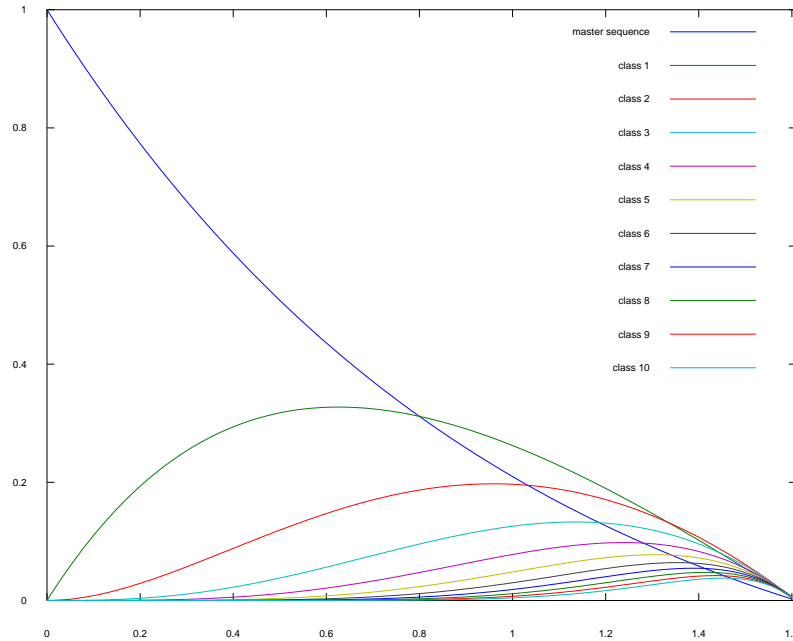
$$\rho_k^* = (\sigma e^{-a} - 1) \frac{a^k}{k!} \sum_{i \geq 1} \frac{i^k}{\sigma^i}.$$

We denote by  $\mathcal{Q}(\sigma, a)$  the probability distribution which assigns mass  $\rho_k^*$  to each non-negative integer  $k$ , and we call it the distribution of the quasispecies with parameters  $\sigma, a$ .

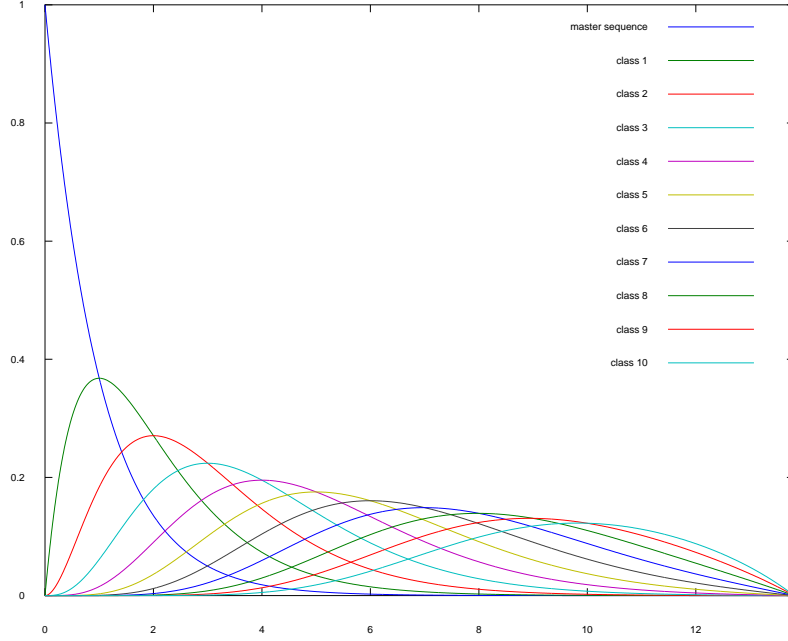
Similar results have been obtained in [4, 6] for the Wright–Fisher model. In [4] a population of size  $m$  of chromosomes of length  $\ell$  over an alphabet of cardinality  $\kappa$  is considered to evolve according to the classical Wright–Fisher model. In the asymptotic regime considered above a different critical curve is obtained separating the regime where the population at equilibrium is totally random from the regime where the population at equilibrium forms a

quasispecies, the fraction of the master sequence in the quasispecies is again given by  $(\sigma e^{-a} - 1)/(\sigma - 1)$ . The model is further studied in [6], where the distribution of the quasispecies is shown to converge to  $\mathcal{Q}(\sigma, a)$ . The results in both articles are similar and also the broad idea of the proofs, namely the construction of certain "simplified" Markov chains that can be compared to a suitably lumped version of the population process, is analogous. However, the quantitative form of the criterion for the asymptotic selection of a master sequence is different in the two models, and also the available proof techniques are quite different. In [6], large deviation estimates for multinomial random variables were used to show that for large  $m$  the comparison processes closely track a dynamical system. Here, in the case of the Moran model, a comparison with suitably constructed birth and death chains, whose equilibria can be computed explicitly, is feasible.

The article is organised as follows. First, we present our main result, along with a sketch of the proof and some background material from [3]. The remaining sections are devoted to the proof. We use coupling and monotonicity arguments to derive simpler Markov chains from the Moran model. We deal then with these simpler processes by obtaining large deviation estimates for several stopping times.



Frequency of the MS and the first 10 classes as a function of  $a$  for  $\sigma = 5$



Frequency of the MS and the first 10 classes as a function of  $a$  for  $\sigma = 10^6$

## 2 Main Result

Let  $\mathcal{A}$  be a finite alphabet of cardinality  $\kappa$  and  $\ell \geq 1$  an integer. We consider the space  $\mathcal{A}^\ell$  of sequences of length  $\ell$  over the alphabet  $\mathcal{A}$ . Elements of the space  $\mathcal{A}^\ell$  represent the chromosome of an haploid individual. We consider a population of size  $m$  of individuals from  $\mathcal{A}^\ell$ . The size of the population  $m$  is kept constant throughout the evolution.

When a reproduction occurs, the chromosome is subject to mutations. We suppose that mutations occur independently at random at each locus, with probability  $q \in ]0, 1[$ . If a mutation occurs, we replace the letter with a new one, chosen uniformly at random between the remaining  $\kappa - 1$  letters of the alphabet  $\mathcal{A}$ . The mutation mechanism is encoded in a mutation matrix  $(M(u, v), u, v \in \mathcal{A}^\ell)$ , where  $M(u, v)$  is the probability that the chromosome  $u$  is transformed into  $v$  by mutation. We have the following analytical expression for  $M(u, v)$ :

$$M(u, v) = \prod_{l=1}^{\ell} \left( (1 - q) 1_{u(l)=v(l)} + \frac{q}{\kappa - 1} 1_{u(l) \neq v(l)} \right).$$

The only allowed transformations in a population consist of replacing a chromosome of the population with a new one. For a population  $x \in (\mathcal{A}^\ell)^m$ ,  $j \in \{1, \dots, m\}$ ,  $u \in \mathcal{A}^\ell$ , we denote by  $x(j \leftarrow u)$  the population  $x$  where the  $j$ -th chromosome  $x(j)$  has been replaced by  $u$ :

$$x(j \leftarrow u) = \begin{pmatrix} x(1) \\ \vdots \\ x(j-1) \\ u \\ x(j+1) \\ \vdots \\ x(m) \end{pmatrix}.$$

The replication mechanism is encoded in a fitness function:

$$A : \mathcal{A}^\ell \longrightarrow [0, +\infty[.$$

The continuous time Moran model is the Markov process  $(X_t)_{t \geq 0}$  having the following infinitesimal generator: for  $\psi$  a function from  $(\mathcal{A}^\ell)^m$  to  $\mathbb{R}$  and for any  $x \in (\mathcal{A}^\ell)^m$ ,

$$\lim_{t \rightarrow 0} \frac{1}{t} \left( E(\psi(X_t) | X_0 = x) - \psi(x) \right) = \sum_{1 \leq i, j \leq m} \sum_{u \in \mathcal{A}^\ell} A(x(i)) M(x(i), u) \left( \psi(x(j \leftarrow u)) - \psi(x) \right).$$

We will only consider the sharp peak landscape. Let  $\sigma > 1$  be a real number. There exists a particular sequence, called the master sequence or the wild type, denoted by  $w^*$ , for which the replication rate is  $\sigma$ . The replication rate for all other sequences is 1. The fitness function is then given by

$$\forall u \in \mathcal{A}^\ell \quad A(u) = \begin{cases} 1 & \text{if } u \neq w^*, \\ \sigma & \text{if } u = w^*. \end{cases}$$

We denote by  $d_H$  the Hamming distance between two chromosomes:

$$\forall u, v \in \mathcal{A}^\ell \quad d_H(u, v) = \text{card} \{ l : 1 \leq l \leq \ell, u(l) \neq v(l) \}.$$

Let  $x$  be a population in  $(\mathcal{A}^\ell)^m$ . We fix an integer  $K \geq 0$  and we look at the number  $N^K(x)$  of chromosomes in  $x$  which are at distance  $K$  or less from the master sequence:

$$N^K(x) = \text{card} \{ i : 1 \leq i \leq m, d_H(x(i), w^*) \leq K \}.$$

Let  $\phi : \mathbb{R}^+ \longrightarrow \mathbb{R}^+ \cup \{+\infty\}$  be the map given by

$$\forall a < \ln \sigma \quad \phi(a) = \frac{\sigma(1 - e^{-a}) \ln \frac{\sigma(1 - e^{-a})}{\sigma - 1} + \ln(\sigma e^{-a})}{1 - \sigma(1 - e^{-a})},$$

and  $\phi(a) = 0$  for  $a \geq \ln \sigma$ . Let  $(\rho_k^*)_{k \geq 0}$  be the sequence given by

$$\forall k \geq 0 \quad \rho_k^* = (\sigma e^{-a} - 1) \frac{a^k}{k!} \sum_{i \geq 1} \frac{i^k}{\sigma^i}.$$

We have the following result:

**Theorem 2.1.** *Suppose that*

$$\begin{aligned} \ell &\rightarrow +\infty, & m &\rightarrow +\infty, & q &\rightarrow 0, \\ \ell q &\rightarrow a \in ]0, +\infty[, & \frac{m}{\ell} &\rightarrow \alpha \in [0, +\infty]. \end{aligned}$$

*We have the following dichotomy:*

- *If  $\alpha \phi(a) < \ln \kappa$ , then*

$$\forall K \geq 0 \quad \lim_{\substack{\ell, m \rightarrow \infty, q \rightarrow 0 \\ \ell q \rightarrow a, \frac{m}{\ell} \rightarrow \alpha}} \lim_{t \rightarrow \infty} E\left(\frac{1}{m} N^K(X_t)\right) = 0.$$

- *If  $\alpha \phi(a) > \ln \kappa$ , then*

$$\forall K \geq 0 \quad \lim_{\substack{\ell, m \rightarrow \infty, q \rightarrow 0 \\ \ell q \rightarrow a, \frac{m}{\ell} \rightarrow \alpha}} \lim_{t \rightarrow \infty} E\left(\frac{1}{m} N^K(X_t)\right) = \rho_0^* + \dots + \rho_K^*.$$

*Furthermore, in both cases*

$$\forall K \geq 0 \quad \lim_{\substack{\ell, m \rightarrow \infty, q \rightarrow 0 \\ \ell q \rightarrow a, \frac{m}{\ell} \rightarrow \alpha}} \lim_{t \rightarrow \infty} \text{Var}\left(\frac{1}{m} N^K(X_t)\right) = 0.$$

We stated the result for a continuous time Moran model. For the proof, we will work with the discrete time counterpart of the Moran model  $(X_n)_{n \geq 0}$ , suitably renormalised. Its transition matrix is given by

$$\begin{aligned} \forall x \in (\mathcal{A}^\ell)^m \quad \forall j \in \{1, \dots, \ell\} \quad \forall u \in \mathcal{A}^\ell \setminus \{x(j)\} \\ p(x, x(j \leftarrow u)) = \frac{1}{m} \sum_{1 \leq i \leq m} \frac{A(x(i))M(x(i), u)}{A(x(1)) + \dots + A(x(m))}. \end{aligned}$$



The other non diagonal coefficients of the transition matrix are zero. The diagonal coefficients are arranged so that the matrix is stochastic, i.e., the sum over each row equals 1.

As it is shown in section 5 of [3], the invariant measure of the continuous Moran model is the same as the invariant measure of the discrete normalised Moran model.

The statement of theorem 2.1 for the case  $K = 0$  is just the main result in [3]. The proof of the first assertion of theorem 2.1 is essentially the same for  $K \geq 1$  and for  $K = 0$ , with minor changes in some of the calculations. These changes do not really give a better understanding of the model, thus, we will omit the proof of the first statement of theorem 2.1 and focus on the more interesting case  $\alpha\phi(a) > \ln \kappa$ .

## 2.1 The occupancy process

Let  $(X_n)_{n \geq 0}$  be the normalised Moran model as described in the end of the previous section. The state space of the Markov chain  $(X_n)_{n \geq 0}$  has cardinality  $\kappa^{\ell m}$ , which is too big to work with. Thus, we work with a simpler process  $(O_n)_{n \geq 0}$ , called the occupancy process, whose state space is much smaller. The occupancy process  $(O_n)_{n \geq 0}$  keeps track of the number of chromosomes in each of the  $\ell + 1$  Hamming classes. It is obtained from the original Moran process  $(X_n)_{n \geq 0}$  via lumping, as shown in section 6.2 of [3]. This process will be the main subject of our study.

Let  $\mathcal{P}_{\ell+1}^m$  be the set of ordered partitions of the integer  $m$  in at most  $\ell + 1$  parts:

$$\mathcal{P}_{\ell+1}^m = \{ (o(0), \dots, o(\ell)) \in \mathbb{N}^{\ell+1} : o(0) + \dots + o(\ell) = m \}.$$

A partition  $(o(0), \dots, o(\ell))$  is interpreted as an occupancy distribution, which corresponds to a population with  $o(l)$  individuals in the Hamming class  $l$ , for  $0 \leq l \leq \ell$ . The set  $\mathcal{P}_{\ell+1}^m$  is the state space of the occupancy process  $(O_n)_{n \geq 0}$ . Since we are working with a Moran model, only one chromosome can change classes at a time, i.e., the only possible transitions for the occupancy process  $(O_n)_{n \geq 0}$  are of the form

$$o \longrightarrow o(k \rightarrow l), \quad 0 \leq k, l \leq \ell,$$

where  $o(k \rightarrow l)$  is the occupancy distribution obtained by transferring a

chromosome from the Hamming class  $k$  to the class  $l$ , i.e.,

$$\forall h \in \{0, \dots, \ell\} \quad o(k \rightarrow l)(h) = \begin{cases} o(h) & \text{if } h \neq k, l, \\ o(k) - 1 & \text{if } h = k, \\ o(l) + 1 & \text{if } h = l. \end{cases}$$

We will work with the discrete time occupancy process  $(O_n)_{n \geq 0}$ , whose transition matrix is given by

$$\forall o \in \mathcal{P}_{\ell+1}^m \quad \forall k, l \in \{0, \dots, \ell\}, \quad k \neq l, \\ p_O(o, o(k \rightarrow l)) = \frac{o(k) \left( \sigma o(0) M_H(0, l) + \sum_{h=1}^{\ell} o(h) M_H(h, l) \right)}{m((\sigma - 1)o(0) + m)},$$

where  $M_H$  is the lumped mutation matrix: for  $b, c \in \{0, \dots, \ell\}$  the coefficient  $M_H(b, c)$  is given by

$$\sum_{\substack{0 \leq k \leq \ell - b \\ 0 \leq l \leq b \\ k - l = c - b}} \binom{\ell - b}{k} \binom{b}{l} q^k (1 - q)^{\ell - b - k} \left( \frac{q}{\kappa - 1} \right)^l \left( 1 - \frac{q}{\kappa - 1} \right)^{b - l}.$$

In order to interpret this formula, note that a passage from the class  $b$  to the class  $c$  is done by mutating "towards the master sequence" in  $k$  sites, which can be done in

$$\binom{\ell - b}{k} q^k (1 - q)^{\ell - b - k}$$

ways, and mutating "away from the master sequence" in  $l$  sites, which can be done in

$$\binom{b}{l} \left( \frac{q}{\kappa - 1} \right)^l \left( 1 - \frac{q}{\kappa - 1} \right)^{b - l}$$

ways. Of course,  $k$  and  $l$  must satisfy  $b + k - l = c$ . In other words, the above formula is the probability that

$$\text{Binomial}(\ell - b, q) - \text{Binomial}(b, q/(\kappa - 1)) = c - b.$$

One of the main advantages of working with the occupancy process is that we can conveniently endow the state space  $\mathcal{P}_{\ell+1}^m$  with the following partial order: for all  $o, o' \in \mathcal{P}_{\ell+1}^m$

$$o \preceq o' \Leftrightarrow \forall l \in \{0, \dots, \ell\} \quad o(0) + \dots + o(l) \leq o'(0) + \dots + o'(l).$$

Let  $\mu_O$  be the invariant probability measure of the occupancy process, and let  $f : [0, 1] \rightarrow \mathbb{R}$  be a continuous mapping. The aim of the following sections will be to show that

$$\lim_{\substack{\ell, m \rightarrow \infty, q \rightarrow 0 \\ \ell q \rightarrow a, m/\ell \rightarrow \alpha}} \int_{\mathcal{P}_{\ell+1}^m} f\left(\frac{o(0) + \dots + o(K)}{m}\right) d\mu_O(o) = f(\rho_0^* + \dots + \rho_K^*).$$

Throughout the text, when we say that some property holds asymptotically, we mean that it holds for  $\ell, m$  large enough,  $q$  small enough,  $\ell q$  close enough to  $a$  and  $m/\ell$  close enough to  $\alpha$ .

## 2.2 Sketch of proof

In order to demonstrate theorem 2.1, we will compare the time that the process  $(O_n)_{n \geq 0}$  spends having at least a sequence in one of the Hamming classes  $0, \dots, K$  (which we call the persistence time), with the time the process  $(O_n)_{n \geq 0}$  spends having no sequences in any of the Hamming classes  $0, \dots, K$  (which we call the discovery time). Asymptotically, when  $\alpha\phi(a) < \ln \kappa$ , the persistence time becomes negligible with respect to the discovery time, whereas when  $\alpha\phi(a) > \ln \kappa$ , it is the discovery time that becomes negligible with respect to the persistence time. This already settles the first assertion in theorem 2.1. These facts are rigorously proven in [3] for the case  $K = 0$ . When  $K \geq 1$  the proofs are essentially the same, and since the minor changes required for them do not improve the understanding of the model, we will omit the proof of the first statement of theorem 2.1.

The second statement of the theorem requires much more work. We suppose that  $\alpha\phi(a) > \ln \kappa$ . From the previous paragraph (taking  $K = 0$ ) we know that asymptotically, the time the master sequence is absent in the population becomes negligible with respect to the time it is present. Therefore, we focus on the dynamics of  $(O_n)_{n \geq 0}$  when the master sequence is present. In order to understand the heuristics of the proof, let  $o \in \mathcal{P}_{\ell+1}^m$ , and let us compute, for  $k \in \{0, \dots, \ell\}$ , the following expectation:

$$\begin{aligned} E(O_{n+1}(k) - O_n(k) | O_n = o) &= \sum_{i: i \neq k} p(o, o(i \rightarrow k)) - \sum_{i: i \neq k} p(o, o(k \rightarrow i)) \\ &= \frac{1}{(\sigma - 1)o(0) + m} \left( \frac{m - o(k)}{m} \left( \sigma o(0) M_H(0, k) + \sum_{l=1}^{\ell} o(l) M_H(l, k) \right) \right) \end{aligned}$$

$$\begin{aligned}
& - \frac{o(k)}{m} \left( \sigma o(0)(1 - M_H(0, k)) + \sum_{l=1}^{\ell} o(l)(1 - M_H(l, k)) \right) \Bigg) \\
& = \left( (\sigma - 1) \frac{o(0)}{m} + 1 \right)^{-1} \left( \sigma \frac{o(0)}{m} M_H(0, k) + \sum_{l=1}^{\ell} \frac{o(l)}{m} M_H(l, k) \right) - \frac{o(k)}{m}.
\end{aligned}$$

Thus, when  $m$  is large, we can interpret the Markov chain  $(O_n/m)_{n \geq 0}$  as a random perturbation of the system of differential equations

$$\dot{x}_k = ((\sigma - 1)x_0 + 1)^{-1} \left( \sigma x_0 M_H(0, k) + \sum_{l=1}^{\ell} x_l M_H(l, k) \right) - x_k, \quad 0 \leq k \leq \ell.$$

In other words,  $(O_n)_{n \geq 0}$  is a density dependent process converging to the above system of differential equations in the infinite population limit. Taking the asymptotic regime

$$\ell \rightarrow \infty, \quad q \rightarrow 0, \quad \ell q \rightarrow a,$$

and recalling that each row in the mutation kernel  $M_H$  is obtained as the difference of two binomial variables, we see that, asymptotically, mutations towards the master sequence are negligible, while mutations away from the master sequence follow a Poisson distribution of parameter  $a$ , i.e.

$$\lim_{\substack{\ell \rightarrow \infty, q \rightarrow 0 \\ \ell q \rightarrow a}} M_H(l, k) = \begin{cases} \frac{a^{k-l}}{(k-l)!} e^{-a} & \text{if } l \leq k, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, it is natural to expect that asymptotically, the stationary distribution of the process  $(O_n/m)_{n \geq 0}$  will be concentrated around the stable stationary solutions of the limiting system

$$\dot{x}_k = ((\sigma - 1)x_0 + 1)^{-1} \left( \sigma x_0 e^{-a} \frac{a^k}{k!} + \sum_{l=1}^k x_l e^{-a} \frac{a^{k-l}}{(k-l)!} \right) - x_k, \quad 0 \leq k \leq \ell.$$

We are assuming that  $\alpha\phi(a) > \ln \kappa$ , which in particular means that  $\sigma e^{-a} > 1$ , and when  $\sigma e^{-a} > 1$ , the only stable stationary solution of the limiting system is given by the distribution of the quasispecies  $(\rho_k^*)_{k \geq 0}$ .

The goal is to show that the invariant probability measure of the occupancy process converges to the distribution of the quasispecies  $(\rho_k^*)_{k \geq 0}$ . In order to do so we construct several stochastic processes: in section 3, we build a pair of processes  $(O_n^\ell)_{n \geq 0}, (O_n^1)_{n \geq 0}$ , which bound the original occupancy

process, and have the property of exiting always at the same point from the set of distributions where the master sequence is present. This allows us, via a renewal result, to express the invariant probability measures of the bounding processes in terms of their dynamics when the master sequence is present in the population, i.e., we can forget the neutral phase. We finish section 3 by constructing two Markov chains  $(Z_n^\ell)_{n \geq 0}, (Z_n^1)_{n \geq 0}$  that replicate the dynamics of the bounding occupancy processes, but which do not have a neutral phase, i.e., as soon as the master sequence disappears from the population, it is created again. The proof of the theorem will be achieved if we manage to show that the invariant measures of  $(Z_n^\ell)_{n \geq 0}, (Z_n^1)_{n \geq 0}$  both converge to  $\mathcal{Q}(\sigma, a)$ .

In section 4, we fix  $K \geq 0$  and we estimate the typical time that the processes  $(Z_n^\ell)_{n \geq 0}, (Z_n^1)_{n \geq 0}$  spend inside and outside a neighbourhood of  $(\rho_0^*, \dots, \rho_K^*)$ . The time they spend inside such a neighbourhood is typically of exponential order in  $m$ , while the time they spend outside it is typically of polynomial order in  $m$ . These large deviations estimates are enough to prove theorem 2.1. The estimates are obtained by induction on  $k$ . Both at the base case and at the induction step, we can bound stochastically the dynamics of each coordinate  $k$  of  $(Z_n^\ell)_{n \geq 0}$  and  $(Z_n^1)_{n \geq 0}$  with a pair of birth and death Markov chains  $(\underline{Z}_n^k)_{n \geq 0}$  and  $(\overline{Z}_n^k)_{n \geq 0}$ . These birth and death chains have the following properties: for any point in a certain neighbourhood of  $(\rho_0^*, \dots, \rho_{k-1}^*)$ , the probability of losing an individual in the class  $k$  for  $(Z_n^\ell)_{n \geq 0}, (Z_n^1)_{n \geq 0}$ , is bounded between the probabilities of a death happening in  $(\underline{Z}_n^k)_{n \geq 0}$  and  $(\overline{Z}_n^k)_{n \geq 0}$ , while the probability of producing an individual in the class  $k$  for  $(Z_n^\ell)_{n \geq 0}, (Z_n^1)_{n \geq 0}$ , is bounded between the probabilities of a birth happening in  $(\underline{Z}_n^k)_{n \geq 0}$  and  $(\overline{Z}_n^k)_{n \geq 0}$ . For these chains, we use classical explicit formulas to compute their hitting times. Finally, we use the above estimates to prove the desired convergence.

### 3 Stochastic bounds

In this section we modify the occupancy process in order to simplify its study. We can regard the dynamics of the occupancy process as having two different phases: the neutral phase, where no master sequence is present, and the phase where the master sequence remains present in the population. We are only interested in the latter, and the presence of the neutral phase is inconvenient for our purposes. Thus, we get rid of it by bounding the invariant probability measure of  $(O_n)_{n \geq 0}$  with the invariant measures of a

pair of new Markov chains. These Markov chains have exactly the same dynamics as the occupancy process whenever the master sequence is present in the population, but every time they reach the neutral phase they jump back out of it immediately. We formalise these ideas in the rest of the section. We denote by  $\mathcal{W}^*$  the set of the occupancy distributions containing the master sequence, i.e.,

$$\mathcal{W}^* = \{o \in \mathcal{P}_{\ell+1}^m : o(0) \geq 1\},$$

and by  $\mathcal{N}$  the set of the occupancy distributions which do not contain the master sequence, i.e.,

$$\mathcal{N} = \{o \in \mathcal{P}_{\ell+1}^m : o(0) = 0\}.$$

We define the following occupancy distributions:

$$\begin{aligned} o_{\text{enter}}^\ell &= (1, 0, \dots, 0, m-1), & o_{\text{exit}}^\ell &= (0, \dots, 0, m), \\ o_{\text{enter}}^1 &= (1, m-1, \dots, 0), & o_{\text{exit}}^1 &= (0, m, 0, \dots, 0). \end{aligned}$$

Note that  $o_{\text{exit}}^\ell$  and  $o_{\text{exit}}^1$  are the extreme points of  $\mathcal{N}$ , while  $o_{\text{enter}}^\ell$  and  $o_{\text{enter}}^1$  are the extreme points of  $\mathcal{W}^*$  that the occupancy process can reach when jumping from  $\mathcal{N}$  to  $\mathcal{W}^*$ , i.e.,

$$\begin{aligned} \forall o \in \mathcal{N} & & o_{\text{exit}}^\ell &\preceq o \preceq o_{\text{exit}}^1, \\ \forall o \in \{o \in \mathcal{P}_{\ell+1}^m : o(0) = 1\} & & o_{\text{enter}}^\ell &\preceq o \preceq o_{\text{enter}}^1. \end{aligned}$$

Let  $\mathcal{R}$  and  $\Phi_O$  be the set and the coupling map defined in section 7.1 of [3]. The set  $\mathcal{R}$  and the map  $\Phi_O$  are such that, given a sequence of i.i.d. random variables  $(R_n)_{n \geq 1}$  on  $\mathcal{R}$ , we can build the process  $(O_n)_{n \geq 0}$  by setting  $O_0 = o \in \mathcal{P}_{\ell+1}^m$  and

$$\forall n \geq 1 \quad O_n = \Phi_O(O_{n-1}, R_n).$$

Moreover (lemma 7.5 of [3]), the map  $\Phi_O$  is non-decreasing with respect to the occupancy distribution, i.e.,

$$\forall o, o' \in \mathcal{P}_{\ell+1}^m \quad \forall r \in \mathcal{R} \quad o \preceq o' \Rightarrow \Phi_O(o, r) \preceq \Phi_O(o', r).$$

We define a lower map and an upper map  $\Phi_O^\ell, \Phi_O^1 : \mathcal{P}_{\ell+1}^m \times \mathcal{R} \longrightarrow \mathcal{P}_{\ell+1}^m$  by setting, for  $o \in \mathcal{P}_{\ell+1}^m$ ,  $r \in \mathcal{R}$  and  $\theta = 1$  or  $\ell$ ,

$$\Phi_O^\theta(o, r) = \begin{cases} o_{\text{exit}}^\theta & \text{if } o \in \mathcal{W}^* \text{ and } \Phi_O(o, r) \in \mathcal{N} \\ o_{\text{enter}}^\theta & \text{if } o \in \mathcal{N} \text{ and } \Phi_O(o, r) \in \mathcal{W}^* \\ \Phi_O(o, r) & \text{otherwise} \end{cases}$$

We use these two maps, and the i.i.d. sequence  $(R_n)_{n \geq 1}$ , to build a lower process  $(O_n^\ell)_{n \geq 0}$  and an upper process  $(O_n^1)_{n \geq 0}$ . We set  $O_0^\ell = O_0^1 = o \in \mathcal{P}_{\ell+1}^m$  and

$$\forall n \geq 1 \quad O_n^\ell = \Phi_O^\ell(O_{n-1}^\ell, R_n), \quad O_n^1 = \Phi_O^1(O_{n-1}^1, R_n).$$

If all the processes  $(O_n)_{n \geq 0}$ ,  $(O_n^\ell)_{n \geq 0}$ ,  $(O_n^1)_{n \geq 0}$  start from the same occupancy distribution  $o$ , then

$$\forall n \geq 0 \quad O_n^\ell \preceq O_n \preceq O_n^1.$$

The process  $(O_n^\ell)_{n \geq 0}$  behaves pretty much like  $(O_n)_{n \geq 0}$ . The only difference being that whenever it jumps from the set  $\mathcal{N}$  to  $\mathcal{W}^*$  it jumps to the lowest possible occupancy distribution of  $\mathcal{W}^*$  that can be reached from  $\mathcal{N}$ , and whenever it jumps from the set  $\mathcal{W}^*$  to  $\mathcal{N}$  it jumps to the lowest possible occupancy distribution of  $\mathcal{N}$ . The process  $(O_n^1)_{n \geq 0}$  is built in a similar way. For  $k \in \{0, \dots, \ell\}$ , let  $\pi_k : \mathcal{P}_{\ell+1}^m \rightarrow \mathbb{N}$  be the function given by

$$\forall o \in \mathcal{P}_{\ell+1}^m \quad \pi_k(o) = o(0) + \dots + o(k).$$

Let us denote by  $\mu_O, \mu_O^\ell, \mu_O^1$  the invariant probability measures of the processes  $(O_n)_{n \geq 0}, (O_n^\ell)_{n \geq 0}, (O_n^1)_{n \geq 0}$ . Let us also fix a non-decreasing function  $f : [0, 1] \rightarrow \mathbb{R}$ . By stochastic domination, we have

$$\forall n \geq 0 \quad f\left(\frac{\pi_K(O_n^\ell)}{m}\right) \leq f\left(\frac{\pi_K(O_n)}{m}\right) \leq f\left(\frac{\pi_K(O_n^1)}{m}\right).$$

By the ergodic theorem for Markov chains, taking the expectation and sending  $n$  to  $\infty$ , we get

$$\int_{\mathcal{P}_{\ell+1}^m} f\left(\frac{\pi_K(o)}{m}\right) d\mu_O^\ell(o) \leq \int_{\mathcal{P}_{\ell+1}^m} f\left(\frac{\pi_K(o)}{m}\right) d\mu_O(o) \leq \int_{\mathcal{P}_{\ell+1}^m} f\left(\frac{\pi_K(o)}{m}\right) d\mu_O^1(o).$$

Our next goal is to find estimates on the above integrals. The strategy is the same for the lower and upper integrals. Let  $\theta$  be either  $\ell$  or  $1$  and let us study the invariant probability measure  $\mu_O^\theta$ . We will rely on a renewal result, which we state next.

Let  $(Y_n)_{n \geq 0}$  be a discrete time Markov chain taking values in a finite space  $\mathcal{E}$ . We suppose that  $(Y_n)_{n \geq 0}$  is irreducible and we call  $\mu$  its invariant probability measure.

**Proposition 3.1.** *Let  $\mathcal{W}^*$  be a subset of  $\mathcal{E}$  and let  $e$  be a point in  $\mathcal{E} \setminus \mathcal{W}^*$ . Let  $f$  be a function from  $\mathcal{E}$  to  $\mathbb{R}$ . We define*

$$\tau^* = \inf\{n \geq 0 : Y_n \in \mathcal{W}^*\}, \quad \tau = \inf\{n \geq \tau^* : Y_n = e\}.$$

We have

$$\int_{\mathcal{E}} f(y) d\mu(y) = \frac{1}{E(\tau | Y_0 = e)} E\left(\sum_{n=0}^{\tau-1} f(Y_n) \middle| Y_0 = e\right).$$

The proof is standard and similar to the proof of proposition 8.2 in [3], so we omit it. We apply this renewal result to the process  $(O_n^\theta)_{n \geq 0}$ , the set  $\mathcal{W}^*$ , the occupancy distribution  $o_{\text{exit}}^\theta$  and the function  $o \mapsto f(\pi_K(o)/m)$ . Set

$$\tau^* = \inf\{n \geq 0 : O_n^\theta \in \mathcal{W}^*\}, \quad \tau = \inf\{n \geq \tau^* : O_n^\theta = o_{\text{exit}}^\theta\}.$$

We have then

$$\begin{aligned} \int_{\mathcal{P}_{\ell+1}^m} f\left(\frac{\pi_K(o)}{m}\right) d\mu_O^\theta(o) &= \frac{E\left(\sum_{n=0}^{\tau-1} f\left(\frac{\pi_K(O_n^\theta)}{m}\right) \middle| O_0^\theta = o_{\text{exit}}^\theta\right)}{E(\tau | O_0^\theta = o_{\text{exit}}^\theta)} = \\ &= \frac{E\left(\sum_{n=0}^{\tau^*-1} f\left(\frac{\pi_K(O_n^\theta)}{m}\right) \middle| O_0^\theta = o_{\text{exit}}^\theta\right)}{E(\tau | O_0^\theta = o_{\text{exit}}^\theta)} + \frac{E\left(\sum_{n=\tau^*}^{\tau-1} f\left(\frac{\pi_K(O_n^\theta)}{m}\right) \middle| O_0^\theta = o_{\text{exit}}^\theta\right)}{E(\tau | O_0^\theta = o_{\text{exit}}^\theta)}. \end{aligned}$$

Our goal is to prove that, when  $\alpha\phi(a) > \ln \kappa$ ,

$$\lim_{\substack{\ell, m \rightarrow \infty, q \rightarrow 0 \\ \ell q \rightarrow a, m/\ell \rightarrow \alpha}} \int_{\mathcal{P}_{\ell+1}^m} f\left(\frac{\pi_K(o)}{m}\right) d\mu_O^\theta = f(\rho_0^* + \dots + \rho_K^*).$$

The first term in the above sum can be bounded by

$$f(1) \frac{E(\tau^* | O_0^\theta = o_{\text{exit}}^\theta)}{E(\tau^* | O_0^\theta = o_{\text{exit}}^\theta) + E(\tau - \tau^* | O_0^\theta = o_{\text{exit}}^\theta)}.$$

We have the following large deviations estimates from [3] (corollary 9.2 and propositions 10.13, 10.16):

$$\begin{aligned} \lim_{\substack{\ell, m \rightarrow \infty, q \rightarrow 0 \\ \ell q \rightarrow a, m/\ell \rightarrow \alpha}} \frac{1}{m} \ln E(\tau - \tau^* | O_0^\theta = o_{\text{exit}}^\theta) &= \phi(a), \\ \lim_{\substack{\ell, m \rightarrow \infty, q \rightarrow 0 \\ \ell q \rightarrow a, m/\ell \rightarrow \alpha}} \frac{1}{\ell} \ln E(\tau^* | O_0^\theta = o_{\text{exit}}^\theta) &= \ln \kappa. \end{aligned}$$



Thus, when  $\alpha\phi(a) > \ln \kappa$ , the first term goes to 0, and we only need to take care of the second term, which we can rewrite as follows:

$$\left( \frac{E(\tau^* | O_0^\theta = o_{\text{exit}}^\theta)}{E(\tau - \tau^* | O_0^\theta = o_{\text{exit}}^\theta)} + 1 \right)^{-1} \times \frac{E\left( \sum_{n=\tau^*}^{\tau-1} f\left( \frac{\pi_K(O_n^\theta)}{m} \right) \middle| O_0^\theta = o_{\text{exit}}^\theta \right)}{E(\tau - \tau^* | O_0^\theta = o_{\text{exit}}^\theta)}.$$

The previous observations imply that the first term of the product goes to 1, so it suffices to prove that

$$\frac{1}{E(\tau - \tau^* | O_0^\theta = o_{\text{exit}}^\theta)} E\left( \sum_{n=\tau^*}^{\tau-1} f\left( \frac{\pi_K(O_n^\theta)}{m} \right) \middle| O_0^\theta = o_{\text{exit}}^\theta \right) \rightarrow f(\rho_0^* + \dots + \rho_K^*).$$

We have reduced the problem to show the convergence of an expression which only depends on the dynamics of  $(O_n^\theta)_{n \geq 0}$  inside  $\mathcal{W}^*$ . We build now a Markov chain  $(Z_n^\theta)_{n \geq 0}$ , which will replicate the dynamics of  $(O_n^\theta)_{n \geq 0}$  in  $\mathcal{W}^*$ , but with no neutral phase. Let  $(Z_n^\theta)_{n \geq 0}$  be a Markov chain with state space  $\mathcal{P}_{\ell+1}^m$  and transition matrix  $p^\theta$  given by:

$$\begin{aligned} \forall o \in \mathcal{W}^* \quad \forall o' \in \mathcal{P}_{\ell+1}^m \quad p^\theta(o, o') &= P(O_1^\theta = o' | O_0^\theta = o), \\ \forall o \in \mathcal{N} \quad p^\theta(o, o_{\text{enter}}^\theta) &= 1. \end{aligned}$$

The other non-diagonal elements of the matrix are null. The diagonal coefficients are arranged so that the matrix is stochastic, i.e., the sum over each row equals 1. We define

$$\tau_0 = \inf\{n \geq 0 : Z_n^\theta(0) = 0\}.$$

We then have

$$\begin{aligned} E\left( \sum_{n=\tau^*}^{\tau-1} f\left( \frac{\pi_K(O_n^\theta)}{m} \right) \middle| O_0^\theta = o_{\text{exit}}^\theta \right) &= E\left( \sum_{n=0}^{\tau_0-1} f\left( \frac{\pi_K(Z_n^\theta)}{m} \right) \middle| Z_0^\theta = o_{\text{enter}}^\theta \right), \\ E(\tau - \tau^* | O_0^\theta = o_{\text{exit}}^\theta) &= E(\tau_0 | Z_0^\theta = o_{\text{enter}}^\theta). \end{aligned}$$

Let  $\nu^\theta$  be the invariant probability measure of  $(Z_n^\theta)_{n \geq 0}$ . We apply the renewal result of proposition 3.1 to the process  $(Z_n^\theta)_{n \geq 0}$ , the set  $\mathcal{N}$ , the occupancy distribution  $o_{\text{enter}}^\theta$  and the function  $o \mapsto f(\pi_K(o)/m)$ , which, with the help of the Markov property, gives

$$\int_{\mathcal{P}_{\ell+1}^m} f\left( \frac{\pi_K(o)}{m} \right) d\nu^\theta = \frac{E\left( \sum_{n=0}^{\tau_0-1} f\left( \frac{\pi_K(Z_n^\theta)}{m} \right) \middle| Z_0^\theta = o_{\text{enter}}^\theta \right)}{1 + E(\tau_0 | Z_0^\theta = o_{\text{enter}}^\theta)}.$$

In view of the above remarks, our problem now boils down to show the convergence

$$\lim_{\substack{\ell, m \rightarrow \infty, q \rightarrow 0 \\ \ell q \rightarrow a, m/\ell \rightarrow \alpha}} \int_{\mathcal{P}_{\ell+1}^m} f\left(\frac{\pi_K(o)}{m}\right) d\nu^\theta(o) = f(\rho_0^* + \cdots + \rho_K^*).$$

In passing from the process  $(O_n^\theta)_{n \geq 0}$  to the process  $(Z_n^\theta)_{n \geq 0}$ , we have eliminated the neutral phase, the newly built process  $(Z_n^\theta)_{n \geq 0}$  has the same dynamics as the original occupancy process inside  $\mathcal{W}^*$ , but whenever it enters the neutral phase, it automatically jumps back to  $\mathcal{W}^*$ . The upcoming sections will be concerned with the study of the Markov chain  $(Z_n^\theta)_{n \geq 0}$ .

## 4 Induction and estimates

We estimate next some hitting times associated to the Markov chain  $(Z_n^\theta)_{n \geq 0}$ . From now on, we will denote by  $P_o, E_o$  the probabilities and expectations for the process  $(Z_n^\theta)_{n \geq 0}$  starting from the occupancy distribution  $o$ . Let us define for  $0 \leq k \leq K$  and for  $\varepsilon > 0$ ,

$$U_k(\varepsilon) = \left\{ o \in \mathcal{P}_{\ell+1}^m : \left| \frac{o(i)}{m} - \rho_i^* \right| < \varepsilon, 0 \leq i \leq k \right\}.$$

We also define, for any subset  $A \subset \mathcal{P}_{\ell+1}^m$  the hitting time of  $A$ :

$$\tau(A) = \inf\{n \geq 0 : Z_n^\theta \in A\}.$$

**Theorem 4.1.** *Let  $\varepsilon > 0$ . For all  $0 \leq k \leq K$ , there exist real numbers  $\alpha_k, \alpha'_k, \lambda_k, \lambda'_k > 0$  (depending on  $\varepsilon$ ), such that asymptotically:*

$$\begin{aligned} \forall o \in \mathcal{P}_{\ell+1}^m \quad P_o(\tau(U_k(\varepsilon)) \geq m^{\alpha_k}) &\leq \exp(-\alpha'_k m). \\ \forall o \in U_k(\varepsilon) \quad P_o(\tau(U_k(2\varepsilon)^c) \leq \exp(\lambda_k m)) &\leq \exp(-\lambda'_k m). \end{aligned}$$

We will prove this theorem by induction on  $k$ . The strategy is as follows. From the definition of  $U_k(\varepsilon)$  we see that

$$U_0(\varepsilon) \supset U_1(\varepsilon) \supset \cdots \supset U_K(\varepsilon),$$

and that in order to know whether the process  $(Z_n^\theta)_{n \geq 0}$  is in  $U_k(\varepsilon)$ , it is enough to check the coordinates 0 to  $k$ . To deal with the base case,  $k = 0$ ,

we bound stochastically the dynamics of  $(Z_n^\theta(0))_{n \geq 0}$  with a pair of birth and death Markov chains, which are very similar to the ones studied in section 9 of [3]. If the estimates hold at rank  $k - 1$ , then we know that the process  $(Z_n^\theta)_{n \geq 0}$  spends almost all of its time inside the set  $U_{k-1}(\varepsilon)$ . As long as the process is in  $U_{k-1}(\varepsilon)$ , we can again bound stochastically the dynamics of  $(Z_n^\theta(k))_{n \geq 0}$  with a pair of birth and death Markov chains; these Markov chains can also be studied with the same techniques as the base case.

## 4.1 Bounding coordinate $k$

In this section we show how to use birth and death Markov chains in order to bound stochastically the dynamics of  $(Z_n^\theta(k))_{n \geq 0}$  when the process is in  $U_{k-1}(\varepsilon)$ , for  $\varepsilon > 0$  and  $1 \leq k \leq K$ , as well as the dynamics of  $(Z_n^\theta(0))_{n \geq 0}$ . This is accomplished by building, for  $0 \leq k \leq K$ , birth and death Markov chains  $(\underline{Z}_n^k)_{n \geq 0}$ ,  $(\overline{Z}_n^k)_{n \geq 0}$  with the following properties: for  $k = 0$  the probability of  $Z_n^\theta(0)$  losing an individual is bounded between the death probabilities of  $\underline{Z}_n^0$ ,  $\overline{Z}_n^0$ . Likewise, the probability of  $Z_n^\theta(0)$  earning an individual is bounded between the birth probabilities of  $\underline{Z}_n^0$ ,  $\overline{Z}_n^0$ . For  $k \geq 1$ , provided that  $Z_n^\theta$  is in  $U_{k-1}(\varepsilon)$ , the probability of  $Z_n^\theta(k)$  losing an individual is bounded between the death probabilities of  $(\underline{Z}_n^k)_{n \geq 0}$ ,  $(\overline{Z}_n^k)_{n \geq 0}$ . Likewise, provided that  $Z_n^\theta$  is in  $U_{k-1}(\varepsilon)$ , the probability of  $Z_n^\theta(k)$  earning an individual is bounded between the birth probabilities of  $(\underline{Z}_n^k)_{n \geq 0}$ ,  $(\overline{Z}_n^k)_{n \geq 0}$ . Let us define the bounding birth and death Markov chains first. We set  $\underline{\beta}_0^0 = \overline{\beta}_0^0 = 1$  and

$$\begin{aligned} \forall i \in \{1, \dots, m\} \quad \underline{\beta}_i^0 &= \frac{m-i}{m} \frac{\sigma i M_H(0,0)}{(\sigma-1)i+m}, \\ \overline{\beta}_i^0 &= \frac{m-i}{m} \frac{\sigma i M_H(0,0) + (m-i) M_H(1,0)}{(\sigma-1)i+m}, \\ \forall i \in \{0, \dots, m\} \quad \underline{\delta}_i^0 &= \frac{i}{m} \frac{\sigma i (1 - M_H(0,0)) + m - i}{(\sigma-1)i+m}, \\ \overline{\delta}_i^0 &= \frac{i}{m} \frac{\sigma i (1 - M_H(0,0)) + (m-i)(1 - M_H(1,0))}{(\sigma-1)i+m}. \end{aligned}$$

For  $\rho \in [0, 1]^k$ , let  $\|\rho\|_1$  be the 1-norm of  $\rho$ , i.e.,

$$\|\rho\|_1 = \rho_0 + \dots + \rho_{k-1}.$$

For  $1 \leq k \leq K$ ,  $0 \leq i \leq m$ , and  $\rho \in [0, 1]^k$  we set

$$\begin{aligned}\underline{b}_i^k(\rho) &= \frac{1 - i/m}{(\sigma - 1)\rho_0 + 1} \left( \sigma \rho_0 M_H(0, k) + \sum_{l=1}^{k-1} \rho_l M_H(l, k) + \frac{i}{m} M_H(k, k) \right), \\ \underline{d}_i^k(\rho) &= \frac{i/m}{(\sigma - 1)\rho_0 + 1} \left( \sigma \rho_0 (1 - M_H(0, k)) + \sum_{l=1}^{k-1} \rho_l (1 - M_H(l, k)) \right. \\ &\quad \left. + \frac{i}{m} (1 - M_H(k, k)) + 1 - \|\rho\|_1 - \frac{i}{m} \right), \\ \bar{b}_i^k(\rho) &= \frac{1 - i/m}{(\sigma - 1)\rho_0 + 1} \left( \sigma \rho_0 M_H(0, k) + \sum_{l=1}^{k-1} \rho_l M_H(l, k) + \frac{i}{m} M_H(k, k) \right. \\ &\quad \left. + (1 - \|\rho\|_1 - \frac{i}{m}) M_H(k + 1, k) \right), \\ \bar{d}_i^k(\rho) &= \frac{i/m}{(\sigma - 1)\rho_0 + 1} \left( \sigma \rho_0 (1 - M_H(0, k)) + \sum_{l=1}^{k-1} \rho_l (1 - M_H(l, k)) \right. \\ &\quad \left. + \frac{i}{m} (1 - M_H(k, k)) + (1 - \|\rho\|_1 - \frac{i}{m}) (1 - M_H(k + 1, k)) \right).\end{aligned}$$

Let  $\varepsilon > 0$ . We also define, for  $1 \leq k \leq K$ ,  $0 \leq i \leq m$ , and  $\rho \in [0, 1]^k$ ,

$$\begin{aligned}\underline{\beta}_i^k &= \min\{ \underline{b}_i^k(\rho) : |\rho_l - \rho_l^*| < \varepsilon, 0 \leq l \leq k-1 \}, \\ \bar{\beta}_i^k &= \max\{ \bar{b}_i^k(\rho) : |\rho_l - \rho_l^*| < \varepsilon, 0 \leq l \leq k-1 \}, \\ \underline{\delta}_i^k &= \max\{ \underline{d}_i^k(\rho) : |\rho_l - \rho_l^*| < \varepsilon, 0 \leq l \leq k-1 \}, \\ \bar{\delta}_i^k &= \min\{ \bar{d}_i^k(\rho) : |\rho_l - \rho_l^*| < \varepsilon, 0 \leq l \leq k-1 \}.\end{aligned}$$

Asymptotically, we have  $\bar{\beta}_i^k + \bar{\delta}_i^k \leq 1$  and  $\underline{\beta}_i^k + \underline{\delta}_i^k \leq 1$ . For  $k \in \{0, \dots, K\}$  and  $\varepsilon > 0$ ,  $(\underline{Z}_n^k)_{n \geq 0}$  will be a Markov chain with state space  $\{0, \dots, m\}$  and transition probabilities given by: for all  $n \geq 0$  and  $i \in \{0, \dots, m\}$

$$P(\underline{Z}_{n+1}^k = i + 1 \mid \underline{Z}_n^k = i) = \underline{\beta}_i^k, \quad P(\underline{Z}_{n+1}^k = i - 1 \mid \underline{Z}_n^k = i) = \underline{\delta}_i^k.$$

The remaining off-diagonal coefficients of the transition matrix are null, the diagonal coefficients are arranged so that the matrix is stochastic. We define a Markov chain  $(\bar{Z}_n^k)_{n \geq 0}$  in an analogous way.

**Theorem 4.2.** *i) For all  $o \in \mathcal{P}_{\ell+1}^m$ , if  $Z_0^\theta = o$  and  $\underline{Z}_0^0 = \bar{Z}_0^0 = o(0)$ , then*

$$\forall n \geq 0 \quad \underline{Z}_n^0 \leq Z_n^\theta(0) \leq \bar{Z}_n^0.$$

ii) For all  $o \in U_{k-1}(\varepsilon)$ , if  $Z_0^\theta = o$  and  $\underline{Z}_0^k = \overline{Z}_0^k = o(k)$ , then

$$\forall n \in \{0, \dots, \tau(U_{k-1}(\varepsilon)^c)\} \quad \underline{Z}_n^k \leq Z_n^\theta(k) \leq \overline{Z}_n^k.$$

The remaining of the section is devoted to the proof of the theorem. We will only prove the second statement; the first statement is much simpler and can almost be seen as a particular case of the second one. We wish to couple the processes  $(Z_n^\theta)_{n \geq 0}$ ,  $(\underline{Z}_n^k)_{n \geq 0}$ ,  $(\overline{Z}_n^k)_{n \geq 0}$  together. We start by defining a coupling map for the process  $(Z_n^\theta)_{n \geq 0}$ , which will be well suited to compare it with the other two processes. For each  $o \in \mathcal{P}_{\ell+1}^m$ , let  $o_1(o), \dots, o_{N(o)}(o)$  be the set of the states  $o' \in \mathcal{P}_{\ell+1}^m$  such that  $p^\theta(o, o') > 0$ , numbered according to the following rule: there exist  $n_1(o) < n_2(o)$  in  $\{1, \dots, N(o)\}$  such that

$$\begin{aligned} \forall n \in \{1, \dots, n_1(o)\} & \quad o_n(o)(k) = o(k) - 1, \\ \forall n \in \{n_1(o) + 1, \dots, n_2(o)\} & \quad o_n(o)(k) = o(k), \\ \forall n \in \{n_2(o) + 1, \dots, N(o)\} & \quad o_n(o)(k) = o(k) + 1. \end{aligned}$$

In other words, all the states with one less individual in the class  $k$  are piled up at the beginning, and all those with one more individual in the class  $k$  are piled up in the end. We define a coupling map  $C : \mathcal{P}_{\ell+1}^m \times [0, 1] \rightarrow \mathcal{P}_{\ell+1}^m$  as follows:

$$\forall o \in \mathcal{P}_{\ell+1}^m \quad \forall u \in [0, 1] \quad C(o, u) = o_{N(o, u)}(o),$$

where  $N(o, u)$  is the only index  $n \in \{1, \dots, N(o)\}$  such that

$$p^\theta(o, o_1(o)) + \dots + p^\theta(o, o_{n-1}(o)) \leq u < p^\theta(o, o_1(o)) + \dots + p^\theta(o, o_n(o)).$$

The coupling map  $C$  is defined so that if  $U$  is a uniform random variable on  $[0, 1]$ , then

$$\forall o, o' \in \mathcal{P}_{\ell+1}^m \quad P(C(o, U) = o') = p^\theta(o, o').$$

Moreover, for all  $n \geq 0$  and  $o \in \mathcal{P}_{\ell+1}^m$

$$\begin{aligned} C(o, u)(k) = o(k) - 1 & \iff u < P_o(Z_1^\theta(k) = o(k) - 1), \\ C(o, u)(k) = o(k) + 1 & \iff u \geq 1 - P_o(Z_1^\theta(k) = o(k) + 1). \end{aligned}$$

Let  $(U_n)_{n \geq 1}$  be an i.i.d. sequence of uniform random variables on  $[0, 1]$ , and let  $o \in \mathcal{P}_{\ell+1}^m$  be a starting point. We build the process  $(Z_n^\theta)_{n \geq 0}$  as follows: we set  $Z_0^\theta = o$  and

$$\forall n \geq 1 \quad Z_n^\theta = C(Z_{n-1}^\theta, U_n).$$

Next we define the coupling maps for the processes  $(\underline{Z}_n^k)_{n \geq 0}$  and  $(\overline{Z}_n^k)_{n \geq 0}$ . Define  $\underline{C}, \overline{C} : \{0, \dots, m\} \times [0, 1] \longrightarrow \{0, \dots, m\}$  by setting, for  $i \in \{0, \dots, m\}$  and  $u \in [0, 1]$

$$\underline{C}(i, u) = i - 1_{u < \underline{\delta}_i^k} + 1_{u \geq 1 - \underline{\beta}_i^k}, \quad \overline{C}(i, u) = i - 1_{u < \overline{\delta}_i^k} + 1_{u \geq 1 - \overline{\beta}_i^k}.$$

We define the processes  $(\underline{Z}_n^k)_{n \geq 0}, (\overline{Z}_n^k)_{n \geq 0}$  with the help of the same sequence  $(U_n)_{n \geq 1}$  as for  $(Z_n^\theta)_{n \geq 0}$ . Let  $i \in \{0, \dots, m\}$  be the starting point of the processes. We set  $\underline{Z}_0^k = \overline{Z}_0^k = i$  and

$$\forall n \geq 1 \quad \underline{Z}_n^k = \underline{C}(\underline{Z}_{n-1}^k, U_n), \quad \overline{Z}_n^k = \overline{C}(\overline{Z}_{n-1}^k, U_n).$$

**Lemma 4.3.** *For all  $o \in U_{k-1}(\varepsilon)$  and  $0 \leq i \leq o(k) \leq j \leq m$ , we have, asymptotically:*

$$\forall u \in [0, 1] \quad \underline{C}(i, u) \leq C(o, u)(k) \leq \overline{C}(j, u).$$

*Proof.* The proof is the same for both inequalities, so we take care of the first one only. The only non-trivial cases are  $i = o(k)$  and  $i = o(k) - 1$ . Let  $o \in U_{k-1}(\varepsilon)$ . We set

$$\delta(o) = P_o(Z_1^\theta(k) = o(k) - 1), \quad \beta(o) = P_o(Z_1^\theta(k) = o(k) + 1).$$

If  $i = o(k)$ , we have, by the above remarks,

$$C(o, u)(k) - \underline{C}(i, u) = -1_{u < \delta(o)} + 1_{u < \underline{\delta}_i^k} + 1_{u > 1 - \beta(o)} - 1_{u > 1 - \underline{\beta}_i^k}.$$

For all  $k \in \{0, \dots, K\}$ , asymptotically

$$\forall l > k + 1 \quad 0 < M_H(l, k) < M_H(k + 1, k).$$

Thus, from the definition of  $\underline{\delta}_i^k$  and  $\underline{\beta}_i^k$  and since  $o \in U_{k-1}(\varepsilon)$ , the above expression is necessarily non-negative. Now let  $i = o(k) - 1$ . We have

$$C(o, u)(k) - \underline{C}(i, u) = 1 - (1_{u < \delta(o)} - 1_{u < \underline{\delta}_i^k}) + (1_{u > 1 - \beta(o)} - 1_{u > 1 - \underline{\beta}_i^k}).$$

If this quantity were to be negative, then we would have  $\delta(o) + \underline{\beta}_i^k > 1$ . We will show that, asymptotically,  $\delta(o) + \underline{\beta}_i^k$  is strictly lower than 1. Since

$o \in U_{k-1}(\varepsilon)$ , taking  $\rho = o/m$ , the expression  $\delta(o) + \underline{\beta}_i^k$  is bounded by

$$\begin{aligned}
& \frac{\rho_k}{(\sigma-1)\rho_0+1} \left( \sigma\rho_0(1-M_H(0,k)) + \sum_{l=1}^{\ell} \rho_l(1-M_H(l,k)) \right) \\
& + \frac{1-\rho_k+1/m}{(\sigma-1)\rho_0+1} \left( \sigma\rho_0 M_H(0,k) + \sum_{l=1}^k \rho_l M_H(l,k) - \frac{1}{m} M_H(k,k) \right) \\
& \leq \frac{1}{(\sigma-1)\rho_0+1} \left( \sigma\rho_0 \left( \rho_k(1-M_H(0,k)) + (1-\rho_k)M_H(0,k) \right) \right. \\
& \quad \left. + \sum_{l=1}^k \rho_l \left( \rho_k(1-M_H(l,k)) + (1-\rho_k)M_H(l,k) \right) + \sum_{l=k+1}^{\ell} \rho_l \right) \\
& \quad + \frac{1}{m} \left( \sigma\rho_0 M_H(0,k) + \sum_{l=1}^k M_H(l,k) \right).
\end{aligned}$$

Yet, for  $u, v \in [0, 1]$  we have  $u(1-v) + (1-u)v \leq \max\{v, 1-v\}$ . Setting

$$M^* = \max\{M_H(l,k), 1-M_H(l,k) : 0 \leq l \leq k\},$$

the above inequality is bounded by

$$\begin{aligned}
& \frac{1}{(\sigma-1)\rho_0+1} \left( \left( \sigma\rho_0 + \sum_{l=1}^k \rho_l \right) M^* + \sum_{l=k+1}^{\ell} \rho_l \right) \\
& + \frac{1}{m} \left( \sigma\rho_0 M_H(0,k) + \sum_{l=1}^k M_H(l,k) \right),
\end{aligned}$$

which is asymptotically strictly lower than 1.  $\square$

*Proof of theorem 4.2.* We can now finish the proof by induction. The case  $n = 0$  is a hypothesis of the theorem, and the induction step is a direct conclusion of the induction hypothesis and the lemma we just proved.  $\square$

## 4.2 Birth and death Markov chains

Let  $k \in \{0, \dots, K\}$  and  $\varepsilon > 0$ . The aim of this section is to study the asymptotic behaviour of the birth and death Markov chains  $(\underline{Z}_n^k)_{n \geq 0}, (\overline{Z}_n^k)_{n \geq 0}$

defined in the previous section (recall that they depend on  $\varepsilon$ ). Let  $\varepsilon' > 0$  and define

$$V(\varepsilon') = \{i \in \{0, \dots, m\} : |i/m - \rho_k^*| < \varepsilon'\}.$$

We define the hitting time of a subset  $A \subset \{0, \dots, m\}$  for the processes  $(\underline{Z}_n^k)_{n \geq 0}, (\overline{Z}_n^k)_{n \geq 0}$  as follows:

$$\tau(A) = \inf\{n \geq 0 : \underline{Z}_n^k \in A\}, \quad \bar{\tau}(A) = \inf\{n \geq 0 : \overline{Z}_n^k \in A\}.$$

**Theorem 4.4.** *Let  $\varepsilon' > 0$ . For  $\varepsilon$  small enough, there exist real numbers  $\alpha, \alpha', \lambda, \lambda' > 0$  (depending on  $\varepsilon, \varepsilon'$ ) such that, asymptotically:*

- For all  $i \in \{0, \dots, m\}$

$$P_i(\tau(V(\varepsilon')) \geq m^\alpha) \vee P_i(\bar{\tau}(V(\varepsilon')) \geq m^\alpha) \leq \exp(-\alpha'm).$$

- For all  $i \in V(\varepsilon')$

$$P_i(\tau(V(2\varepsilon')^c) \leq \exp(\lambda m)) \vee P_i(\bar{\tau}(V(2\varepsilon')^c) \leq \exp(\lambda m)) \leq \exp(-\lambda'm).$$

The rest of the section is devoted to the proof of the theorem. We will show the result only for the chain  $(\underline{Z}_n^k)_{n \geq 0}$  when  $k \in \{1, \dots, K\}$ . The case of  $(\overline{Z}_n^k)_{n \geq 0}$  is dealt with in an analogous way. The case  $k = 0$  is slightly different, but the techniques to study it are the same as for  $k \in \{1, \dots, K\}$ , while the calculations become simpler in the case  $k = 0$ , since there is no parameter  $\varepsilon$ , indeed, for  $k \geq 1$  we need the chain  $(Z_n^\theta/m)_{n \geq 0}$  to be in an  $\varepsilon$ -neighbourhood of  $(\rho_0^*, \dots, \rho_{k-1}^*)$ , while for  $k = 0$  there is no assumption of this kind. In order to ease the notation, let us fix  $k \in \{1, \dots, K\}$  and  $\varepsilon > 0$ , and let us rename the process  $(\underline{Z}_n^k)_{n \geq 0}$  with parameters  $\varepsilon$  and  $k$  as simply  $(Z_n)_{n \geq 0}$ , we also omit the underline and the superscript  $k$  in the rest of the notation, so for example  $\underline{b}_i^k(\rho), \underline{d}_i^k(\rho), \underline{\beta}_i^k, \underline{\delta}_i^k$  become  $b_i(\rho), d_i(\rho), \beta_i, \delta_i$ . First of all, we look for the points  $\rho$  that minimise and maximise the maps  $b_i(\rho)$  and  $d_i(\rho)$ . The function  $b_i(\rho)$  is non-decreasing with respect to the variables  $\rho_1, \dots, \rho_{k-1}$ . Likewise, the function  $d_i(\rho)$  is non-increasing with respect to the variables  $\rho_1, \dots, \rho_{k-1}$ . Therefore, for all  $i \in \{0, \dots, m\}$

$$\begin{aligned} \beta_i &= \min\{b_i(\rho_0, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon) : |\rho_0 - \rho_0^*| < \varepsilon\}, \\ \delta_i &= \max\{d_i(\rho_0, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon) : |\rho_0 - \rho_0^*| < \varepsilon\}. \end{aligned}$$



Let us take the partial derivatives with respect to  $\rho_0$ :

$$\begin{aligned}\frac{\partial b_i(\rho)}{\partial \rho_0} &= \frac{1 - i/m}{((\sigma - 1)\rho_0 + 1)^2} \times \\ &\quad \left( \sigma M_H(0, k) - (\sigma - 1) \left( \sum_{l=1}^{k-1} \rho_l M_H(l, k) + \frac{i}{m} M_H(k, k) \right) \right), \\ \frac{\partial d_i(\rho)}{\partial \rho_0} &= \frac{i/m}{((\sigma - 1)\rho_0 + 1)^2} \times \\ &\quad \left( -\sigma M_H(0, k) + (\sigma - 1) \left( \sum_{l=1}^{k-1} \rho_l M_H(l, k) + \frac{i}{m} M_H(k, k) \right) \right).\end{aligned}$$

The sign of these partial derivatives does not depend on  $\rho_0$ . In particular, for fixed  $\rho_1, \dots, \rho_{k-1}$ , the functions  $b_i(\rho), d_i(\rho)$  are monotonous with respect to  $\rho_0$ . Furthermore, the partial derivatives above have opposite signs, and their value is 0 if and only if

$$\sigma M_H(0, k) = (\sigma - 1) \left( \sum_{l=1}^{k-1} \rho_l M_H(l, k) + \frac{i}{m} M_H(k, k) \right).$$

Since this equation is linear with respect to  $i$ , we conclude that there exists an  $\eta^* \in [0, 1]$  (depending on  $m, \rho_1^*, \dots, \rho_{k-1}^*, \varepsilon$ ) such that:

- If  $0 \leq i < \eta^* m$ , the function  $\rho_0 \mapsto b_i(\rho_0, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon)$  is non-decreasing, the function  $\rho_0 \mapsto d_i(\rho_0, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon)$  is non-increasing, and

$$\begin{aligned}\beta_i &= b_i(\rho_0^* - \varepsilon, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon), \\ \delta_i &= d_i(\rho_0^* - \varepsilon, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon).\end{aligned}$$

- If  $\eta^* m < i \leq m$ , the function  $\rho_0 \mapsto b_i(\rho_0, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon)$  is non-increasing, the function  $\rho_0 \mapsto d_i(\rho_0, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon)$  is non-decreasing, and

$$\begin{aligned}\beta_i &= b_i(\rho_0^* + \varepsilon, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon), \\ \varepsilon_i &= d_i(\rho_0^* + \varepsilon, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon).\end{aligned}$$

For  $i \in \{0, \dots, m-1\}$  we define the products  $\pi(i)$  by setting  $\pi(0) = 1$  and

$$\forall i \in \{1, \dots, m-1\} \quad \pi(i) = \frac{\beta_1 \cdots \beta_i}{\delta_1 \cdots \delta_i}.$$

These products are important in the study of the Markov chain  $(Z_n)_{n \geq 0}$  (chapter 4 of [15]). In order to understand their behaviour, we study first the ratio  $b_i(\rho)/d_i(\rho)$ . For  $1 \leq i \leq m-1$  we have

$$\frac{b_i(\rho)}{d_i(\rho)} = g(M_H(0, k), \dots, M_H(k, k), \rho, i/m),$$

where the function  $g : ]0, 1]^{k+1} \times ]0, 1[ \times ]0, 1[ \rightarrow ]0, \infty[$  is given by

$$g(\gamma, \rho, \eta) = \frac{(1 - \eta) \left( \sigma \rho_0 \gamma_0 + \sum_{l=1}^{k-1} \rho_l \gamma_l + \eta \gamma_k \right)}{\eta \left( \sigma \rho_0 (1 - \gamma_0) + \sum_{l=1}^{k-1} \rho_l (1 - \gamma_l) + \eta (1 - \gamma_k) + 1 - \|\rho\|_1 - \eta \right)}.$$

The behaviour of the products  $\pi(i)$  depends on whether the value of  $g$  is larger or smaller than 1. The equation  $g(\gamma, \rho, \eta) = 1$  is linear with respect to  $\eta$ , its only root being

$$r(\gamma, \rho) = \frac{1}{(\sigma - 1)\rho_0 + 1 - \gamma_k} \left( \sigma \rho_0 \gamma_0 + \sum_{l=1}^{k-1} \rho_l \gamma_l \right).$$

Therefore,

$$\begin{aligned} g(\gamma, \rho, \eta) &> 1 & \text{if } \eta < r(\gamma, \rho), \\ g(\gamma, \rho, \eta) &< 1 & \text{if } \eta > r(\gamma, \rho). \end{aligned}$$

Moreover, the function  $g(\gamma, \rho, \eta)$  is continuous and non-decreasing with respect to each component of the variable  $\gamma$ . Take  $\psi : ]0, 1]^{k+1} \times ]0, 1[ \rightarrow ]0, +\infty[$  to be the function defined by:

$$\begin{aligned} \forall \gamma \in ]0, 1]^{k+1} \quad \forall \eta \in ]0, 1[ \\ \psi(\gamma, \eta) = \begin{cases} g(\gamma, \rho_0^* - \varepsilon, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon, \eta) & \text{if } \eta \leq \eta^* \\ g(\gamma, \rho_0^* + \varepsilon, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon, \eta) & \text{if } \eta > \eta^* \end{cases} \end{aligned}$$

Recall (section 2.2) that we have the following limits for the mutation probabilities:

$$\lim_{\substack{\ell \rightarrow \infty, q \rightarrow 0 \\ \ell q \rightarrow a}} M_H(l, k) = \begin{cases} \frac{a^{k-l}}{(k-l)!} e^{-a} & \text{if } l \leq k \\ 0 & \text{otherwise} \end{cases}$$

We have the following large deviations estimates for the products  $\pi(i)$ :

**Proposition 4.5.** *Let  $a \in ]0, +\infty[$ . For  $\eta \in [0, 1]$ , we have*

$$\lim_{\substack{\ell, m \rightarrow \infty \\ q \rightarrow 0, \ell q \rightarrow a}} \frac{1}{m} \ln \pi(\lfloor \eta m \rfloor) = \int_0^\eta \ln \psi\left(e^{-a} \frac{a^k}{k!}, \dots, e^{-a}, s\right) ds.$$

By the above development, the limit on the left-hand side can be rewritten as

$$\lim_{\substack{\ell, m \rightarrow \infty \\ q \rightarrow 0, \ell q \rightarrow a}} \frac{1}{m} \sum_{i=0}^{\lfloor \eta m \rfloor} \ln \psi(M_H(0, k), \dots, M_H(k, k), i/m),$$

which can be interpreted as a Riemann approximation of the integral on the right-hand side. The rigorous proof of the proposition is very similar to that of proposition 9.1 of [3], so we omit it. Let us define

$$\begin{aligned} \rho^- &= \min \left\{ r\left(e^{-a} \frac{a^k}{k!}, \dots, e^{-a}, \rho_0^* - \varepsilon, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon\right), \right. \\ &\quad \left. r\left(e^{-a} \frac{a^k}{k!}, \dots, e^{-a}, \rho_0^* + \varepsilon, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon\right) \right\}, \\ \rho^+ &= \max \left\{ r\left(e^{-a} \frac{a^k}{k!}, \dots, e^{-a}, \rho_0^* - \varepsilon, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon\right), \right. \\ &\quad \left. r\left(e^{-a} \frac{a^k}{k!}, \dots, e^{-a}, \rho_0^* + \varepsilon, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon\right) \right\}. \end{aligned}$$

From the definitions, we see that

$$\begin{aligned} \psi\left(e^{-a} \frac{a^k}{k!}, \dots, e^{-a}, \eta\right) &> 1 && \text{for } \eta < \rho^-, \\ \psi\left(e^{-a} \frac{a^k}{k!}, \dots, e^{-a}, \eta\right) &< 1 && \text{for } \eta > \rho^+. \end{aligned}$$

In particular, the function

$$\eta \mapsto \int_0^\eta \ln \psi\left(e^{-a} \frac{a^k}{k!}, \dots, e^{-a}, s\right) ds$$

is non-decreasing on  $]0, \rho^-[$  and non-increasing on  $]\rho^+, 1[$ . Furthermore, when  $\varepsilon$  goes to 0, both the points  $\rho^-$  and  $\rho^+$  converge to the point

$$r\left(e^{-a} \frac{a^k}{k!}, \dots, e^{-a}, \rho_0^*, \dots, \rho_{k-1}^*\right).$$

Calling this point  $\rho_k$ , and using the definition of the function  $r$ , we see that  $\rho_k$  is precisely the solution to the recurrence relation

$$\rho_k = \frac{1}{(\sigma - 1)\rho_0 + 1 - e^{-a}} \left( \sigma \rho_0 e^{-a} \frac{a^k}{k!} + \sum_{l=1}^{k-1} \rho_l e^{-a} \frac{a^{k-l}}{(k-l)!} \right),$$

with initial condition  $\rho_0 = \rho_0^* = (\sigma e^{-a} - 1)/(\sigma - 1)$ . Solving this recurrence relation, for example by the method of generating functions, we see that

$$\rho_k = \rho_k^* = (\sigma e^{-a} - 1) \frac{a^k}{k!} \sum_{i \geq 1} \frac{i^k}{\sigma^i}.$$

We also define

$$\begin{aligned} r^- &= \min \left\{ r(M_H(0, k), \dots, M_H(k, k), \rho_0^* - \varepsilon, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon), \right. \\ &\quad \left. r(M_H(0, k), \dots, M_H(k, k), \rho_0^* + \varepsilon, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon) \right\}, \\ r^+ &= \max \left\{ r(M_H(0, k), \dots, M_H(k, k), \rho_0^* + \varepsilon, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon), \right. \\ &\quad \left. r(M_H(0, k), \dots, M_H(k, k), \rho_0^* - \varepsilon, \rho_1^* - \varepsilon, \dots, \rho_{k-1}^* - \varepsilon) \right\}. \end{aligned}$$

We then have

$$\begin{aligned} 1 \leq i \leq j \leq r^- m &\implies \pi(i) \leq \pi(j), \\ r^+ m \leq i \leq j \leq m &\implies \pi(i) \geq \pi(j), \end{aligned}$$

and the situation between  $r^- m$  and  $r^+ m$  is somewhat more delicate. Anyhow, we have

$$\lim_{\substack{\ell, m \rightarrow \infty \\ q \rightarrow 0, \ell q \rightarrow a}} r^- = \rho^-, \quad \lim_{\substack{\ell, m \rightarrow \infty \\ q \rightarrow 0, \ell q \rightarrow a}} r^+ = \rho^+.$$

We are now ready to prove theorem 4.4.

*Proof of Theorem 4.4.* We begin by showing the first statement in the theorem. Let  $\varepsilon' > 0$ . The cases  $i \leq \rho_k^* m$  and  $i \geq \rho_k^* m$  are dealt with in a similar way, thus, we will only show the result for  $i \leq \rho_k^* m$ . Let us call  $b$  the minimum of the discrete interval  $V(\varepsilon')$ , i.e.,

$$b = \lfloor (\rho_k^* - \delta') m \rfloor + 1.$$

Starting from  $i < b$ , the expectation of the hitting time of  $b$  is given by the formula

$$E_i(\tau(\{b\})) = \sum_{j=i}^{b-1} \sum_{k=j}^{b-1} \frac{1}{\beta_j} \frac{\pi(j)}{\pi(k)}.$$

The formula is classical (see for example formula 5.2.4 in [16]). From the definition of  $\beta_j$ , we see that, for  $\varepsilon$  small enough, we have asymptotically

$$\forall j \in \{0, \dots, m-1\} \quad \beta_j \geq \frac{C}{m^2},$$

where the constant  $C$  might depend on  $\varepsilon$  but not on  $m$ ; and that  $r^- \in V(\varepsilon')$ , so that  $b \leq r^-$  and therefore,

$$1 \leq j \leq k \leq b-1 \implies \frac{\pi(j)}{\pi(k)} \leq 1.$$

It follows that

$$E_i(\tau(V(\varepsilon'))) = E_i(\tau(\{b\})) \leq \frac{m^4}{C}.$$

Let  $\kappa > 4$ . By the Markov inequality

$$P_i(\tau(V(\varepsilon')) \geq m^\kappa) \leq \frac{m^{-(\kappa-4)}}{C}.$$

We estimate next, for  $n \geq 1$ ,

$$P_i(\tau(V(\varepsilon')) \geq nm^\kappa).$$

We decompose the interval  $\{0, \dots, nm^\kappa\}$  into subintervals of length  $m^\kappa$  and with the help of the Markov property we apply repeatedly the previous inequality, in order to get

$$P_i(\tau(V(\varepsilon')) \geq nm^\kappa) \leq \exp\left(-n((\kappa - k - 2) \ln m + \ln C)\right).$$

Thus, setting  $n = m$ , we obtain the desired result with  $\alpha = \kappa + 1$  and  $\alpha' = (\kappa - 4) \ln m_0 + \ln C$  for  $m_0$  large enough so that  $\alpha' > 0$ .

We show next the second statement of theorem 4.4. Let  $\varepsilon' > 0$ . Let  $n > 0$ ,  $i \in V(\varepsilon')$ , and let us first estimate the value of

$$P_i(\tau(V(2\varepsilon')^c) \leq n).$$

Let  $\theta$  be the last time the process  $(Z_n)_{n \geq 0}$  visits the set  $V(\varepsilon')$  before time  $\tau(V(2\varepsilon')^c)$ , i.e.,

$$\theta = \max \{s < \tau(V(2\varepsilon')^c) : Z_s \in V(\varepsilon')\}.$$

We denote by  $b$  and  $c$  the extreme points of the discrete interval  $V(\varepsilon')$ ,

$$b = \lfloor (\rho_k^* - \varepsilon')m \rfloor + 1, \quad c = \lfloor (\rho_k^* + \varepsilon')m \rfloor.$$

Likewise, we denote by  $b'$  and  $c'$  the extreme points of the discrete interval  $V(2\varepsilon')$ ,

$$b' = \lfloor (\rho_k^* - 2\varepsilon')m \rfloor + 1, \quad c' = \lfloor (\rho_k^* + 2\varepsilon')m \rfloor.$$

We have then

$$\begin{aligned} P_i(\tau(V(2\varepsilon')^c) \leq n) &= \sum_{s < n} P_i(\theta = s, \tau(V(2\varepsilon')^c) \leq n) = \\ &\sum_{s < n} \left( P_i(\theta = s, Z_s = b, \tau(V(2\varepsilon')^c) \leq n) + P_i(\theta = s, Z_s = c, \tau(V(2\varepsilon')^c) \leq n) \right) \end{aligned}$$

Let us consider the first term within the parenthesis. By the Markov property,

$$\begin{aligned} P_i(\theta = s, Z_s = b, \tau(V(2\varepsilon')^c) \leq n) &= \\ P_i(Z_s = b, Z_{s+1} = b-1, \tau(V(2\varepsilon')^c) \leq n, Z_r \notin V(\varepsilon') \text{ for } s < r \leq \tau(V(2\varepsilon')^c)) &= \\ \leq P_{b-1}(Z_r \notin V(\varepsilon') \text{ for } r \leq \tau(V(2\varepsilon')^c), \tau(V(2\varepsilon')^c) \leq n-s-1) &= \\ \leq P_{b-1}(Z_{\tau(V(2\varepsilon')^c) \cup \{b\}} \in V(2\varepsilon')^c) = P_{b-1}(Z_{\tau(\{b'-1, b\})} = b'-1). \end{aligned}$$

Once again, an explicit formula for this last probability exists:

$$P_{b-1}(Z_{\tau(\{b'-1, b\})} = b'-1) = \left( \sum_{i=b'-1}^{b-1} \frac{1}{\pi(i)} \right)^{-1} \frac{1}{\pi(b-1)} \leq \frac{\pi(b'-1)}{\pi(b-1)}.$$

Let  $\epsilon > 0$ . By proposition 4.5, asymptotically,

$$\begin{aligned} \left| \frac{1}{m} \ln \pi(b-1) - \int_0^{\rho_k^* - \varepsilon'} \ln \psi(e^{-a} \frac{a^k}{k!}, \dots, e^{-a}, s) ds \right| &< \frac{\epsilon}{2}, \\ \left| \frac{1}{m} \ln \pi(b'-1) - \int_0^{\rho_k^* - 2\varepsilon'} \ln \psi(e^{-a} \frac{a^k}{k!}, \dots, e^{-a}, s) ds \right| &< \frac{\epsilon}{2}. \end{aligned}$$

Thus,

$$\begin{aligned} \frac{\pi(b'-1)}{\pi(b-1)} &= \exp \left( m \left( \frac{1}{m} \ln \pi(b'-1) - \frac{1}{m} \ln \pi(b-1) \right) \right) \\ &\leq \exp \left( -m \left( \int_{\rho_k^* - 2\varepsilon'}^{\rho_k^* - \varepsilon'} \ln \psi(e^{-a} \frac{a^k}{k!}, \dots, e^{-a}, s) ds - \epsilon \right) \right). \end{aligned}$$

We can choose  $\epsilon$  small enough so that

$$\lambda_1 = \int_{\rho_k^* - 2\epsilon'}^{\rho_k^* - \epsilon'} \ln \psi(e^{-a} \frac{a^k}{k!}, \dots, e^{-a}, s) ds - \epsilon$$

is positive. We have then

$$P_i(\theta = s, Z_{s+1} = b - 1, \tau(V(2\epsilon')^c) \leq n) \leq \exp(-\lambda_1 m).$$

The term

$$P_i(\theta = s, Z_{s+1} = c + 1, \tau(V(2\epsilon')^c) \leq n)$$

is dealt with in a similar fashion, and we obtain that there exists  $\lambda_2 > 0$  such that

$$P_i(\theta = s, Z_s = c, \tau(V(2\epsilon')^c) \leq n) \leq \exp(-\lambda_2 m).$$

It follows that

$$P_i(\tau(V(2\epsilon')^c) \leq n) \leq n \left( \exp(-\lambda_1 m) + \exp(-\lambda_2 m) \right).$$

We choose  $\lambda < \min(\lambda_1, \lambda_2)$  and for  $m_0$  large enough

$$\lambda' = \min(\lambda_1 - \lambda, \lambda_2 - \lambda) - \frac{1}{m_0} \ln 2.$$

We apply the previous inequality at time  $n = \exp(\lambda m)$  and we obtain the desired result.  $\square$

### 4.3 Proof of theorem 4.1

In this section we conclude the proof of theorem 4.1. The base case,  $k = 0$ , is a direct application of theorems 4.2 and 4.4. The induction step relies also strongly on these two theorems, but it requires some more work. Let us suppose that the result of theorem 4.1 holds at rank  $k - 1$ . Let  $\epsilon, \epsilon' > 0$  with  $2\epsilon' < \epsilon$ . Thanks to the induction hypothesis, there exist positive real numbers  $\alpha_{k-1}, \alpha'_{k-1}, \lambda_{k-1}, \lambda'_{k-1}$  (depending on  $\epsilon'$ ) such that asymptotically,

$$\begin{aligned} \forall o \in \mathcal{P}_{\ell+1}^m \quad & P_o(\tau(U_{k-1}(\epsilon')) \geq m^{\alpha_{k-1}}) \leq \exp(-\alpha'_{k-1} m), \\ \forall o \in U_{k-1}(\epsilon') \quad & P_o(\tau(U_{k-1}(2\epsilon')^c) \leq \exp(\lambda_{k-1} m)) \leq \exp(-\lambda'_{k-1} m). \end{aligned}$$

Let  $o \in \mathcal{P}_{\ell+1}^m$  and choose  $\alpha_k > \alpha_{k-1}$ . We have

$$\begin{aligned} P_o(\tau(U_k(\epsilon)) \geq m^{\alpha_k}) &= P_o(\tau(U_{k-1}(\epsilon')) \geq m^{\alpha_{k-1}}, \tau(U_k(\epsilon)) \geq m^{\alpha_k}) \\ &\quad + P_o(\tau(U_{k-1}(\epsilon')) < m^{\alpha_{k-1}}, \tau(U_k(\epsilon)) \geq m^{\alpha_k}). \end{aligned}$$

By the induction hypothesis the first term in the sum is bounded above by  $\exp(-\alpha'_{k-1}m)$ . We use the Markov property to control the second term:

$$\begin{aligned}
& P_o(\tau(U_{k-1}(\varepsilon')) < m^{\alpha_{k-1}}, \tau(U_k(\varepsilon)) \geq m^{\alpha_k}) \\
&= \sum_{\substack{n < m^{\alpha_{k-1}} \\ o' \in U_{k-1}(\varepsilon')}} P_o(\tau(U_{k-1}(\varepsilon')) = n, Z_n^\theta = o', \tau(U_k(\varepsilon)) \geq m^{\alpha_k}) \\
&= \sum_{\substack{n < m^{\alpha_{k-1}} \\ o' \in U_{k-1}(\varepsilon')}} P_{o'}(\tau(U_k(\varepsilon)) \geq m^{\alpha_k} - n) P_o(\tau(U_{k-1}(\varepsilon')) = n, Z_n^\theta = o').
\end{aligned}$$

Let  $m$  be large enough so that  $m^{\alpha_k} - m^{\alpha_{k-1}} < \exp(\lambda_{k-1}m)$ . For  $n < m^{\alpha_{k-1}}$ ,

$$\begin{aligned}
& P_{o'}(\tau(U_k(\varepsilon)) \geq m^{\alpha_k} - n) \leq P_{o'}(\tau(U_k(\varepsilon)) \geq m^{\alpha_k} - m^{\alpha_{k-1}}) \\
&= P_{o'}(\tau(U_{k-1}(2\varepsilon')^c) \leq \exp(\lambda_{k-1}m), \tau(U_k(\varepsilon)) \geq m^{\alpha_k} - m^{\alpha_{k-1}}) \\
&\quad + P_{o'}(\tau(U_{k-1}(2\varepsilon')^c) > \exp(\lambda_{k-1}m), \tau(U_k(\varepsilon)) \geq m^{\alpha_k} - m^{\alpha_{k-1}}).
\end{aligned}$$

By the induction hypothesis, the first term in the sum is bounded above by  $\exp(-\lambda'_{k-1}m)$ . For the second term we have:

$$\begin{aligned}
& P_{o'}(\tau(U_{k-1}(2\varepsilon')^c) > \exp(\lambda_{k-1}m), \tau(U_k(\varepsilon)) \geq m^{\alpha_k} - m^{\alpha_{k-1}}) \\
&\leq P_{o'}(\tau(U_k(\varepsilon)) \geq m^{\alpha_k} - m^{\alpha_{k-1}} \mid \tau(U_{k-1}(2\varepsilon')^c) > \exp(\lambda_{k-1}m)).
\end{aligned}$$

Since  $\exp(\lambda_{k-1}m) > m^{\alpha_k} - m^{\alpha_{k-1}}$  and  $2\varepsilon' < \varepsilon$ , conditionally on the event  $\tau(U_{k-1}(2\varepsilon')^c) > \exp(\lambda_{k-1}m)$ , we have, in view of theorem 4.2,

$$\forall n < m^{\alpha_k} - m^{\alpha_{k-1}} \quad \underline{Z}_n \leq Z_n^\theta(k) \leq \overline{Z}_n.$$

Therefore,

$$\begin{aligned}
& P_{o'}(\tau(U_k(\varepsilon)) \geq m^{\alpha_k} - m^{\alpha_{k-1}} \mid \tau(U_{k-1}(2\varepsilon')^c) > \exp(\lambda_{k-1}m)) \\
&\leq P_{o'}(\forall n < m^{\alpha_k} - m^{\alpha_{k-1}}, \underline{Z}_n > \rho_k^* + \varepsilon \text{ or } \overline{Z}_n < \rho_k^* - \varepsilon) \\
&\leq P_{o'(k)}(\tau(V(\varepsilon)) \geq m^{\alpha_k} - m^{\alpha_{k-1}}) + P_{o'(k)}(\overline{\tau}(V(\varepsilon)) \geq m^{\alpha_k} - m^{\alpha_{k-1}}).
\end{aligned}$$

Let  $\alpha, \alpha' > 0$  be given by theorem 4.4. Choosing  $\alpha_k$  large enough so that  $m^{\alpha_k} - m^{\alpha_{k-1}} > m^\alpha$ , this last expression is bounded by  $2\exp(-\alpha'm)$ , and this yields the desired bound for the hitting time of  $U_k(\varepsilon)$ .

In order to show the bound on the exit time of  $U_k(2\varepsilon)$ , we argue in a similar way. Let  $o \in U_k(\varepsilon)$ . Let  $\lambda_{k-1}$  be given by the induction hypothesis and let



$\lambda_k > 0$ . We have

$$\begin{aligned} P_o(\tau(U_k(2\varepsilon)^c) \leq \exp(\lambda_k m)) &= \\ &P_o(\tau(U_{k-1}(2\varepsilon)^c) \leq \exp(\lambda_{k-1} m), \tau(U_k(2\varepsilon)^c) \leq \exp(\lambda_k m)) \\ &+ P_o(\tau(U_{k-1}(2\varepsilon)^c) > \exp(\lambda_{k-1} m), \tau(U_k(2\varepsilon)^c) \leq \exp(\lambda_k m)). \end{aligned}$$

By the induction hypothesis, the first term in the sum is bounded above by  $\exp(-\lambda'_{k-1} m)$ . For the second term we have:

$$\begin{aligned} P_o(\tau(U_{k-1}(2\varepsilon)^c) > \exp(\lambda_{k-1} m), \tau(U_k(2\varepsilon)^c) \leq \exp(\lambda_k m)) \\ \leq P_o(\tau(U_k(2\varepsilon)^c) \leq \exp(\lambda_k m) \mid \tau(U_{k-1}(2\varepsilon)^c) > \exp(\lambda_{k-1} m)). \end{aligned}$$

Let  $\lambda, \lambda'$  be given by theorem 4.4, and  $\lambda_k > 0$  such that  $\lambda_k < \lambda_{k-1} \wedge \lambda$ . Conditionally on  $\tau(U_{k-1}(2\varepsilon)^c) > \exp(\lambda_{k-1} m)$ , we have by theorem 4.2,

$$\forall n \in \{0, \dots, \exp(\lambda_{k-1} m)\} \quad \underline{Z}_n \leq Z_n^\theta(k) \leq \bar{Z}_n.$$

Therefore

$$\begin{aligned} P_o(\tau(U_k(2\varepsilon)^c) \leq \exp(\lambda_k m) \mid \tau(U_{k-1}(2\varepsilon)^c) > \exp(\lambda_{k-1} m)) \\ \leq P_{o(k)}(\underline{\tau}(V(2\varepsilon)^c) \leq \exp(\lambda_k m)) + P_{o(k)}(\bar{\tau}(V(2\varepsilon)^c) \leq \exp(\lambda_k m)) \\ \leq 2 \exp(-\lambda' m). \end{aligned}$$

This completes the induction step.

## 4.4 Convergence

In this section we will prove that when  $\sigma e^{-a} > 1$ , the invariant probability measure  $\nu^\theta$  of  $(Z_n^\theta)_{n \geq 0}$  converges to the Dirac mass at  $\rho^*$ . Let  $a > 0$  be such that  $\sigma e^{-a} > 1$ .

**Theorem 4.6.** *For every continuous and increasing function  $f : [0, 1] \rightarrow \mathbb{R}$  such that  $f(0) = 0$ , we have*

$$\lim_{\substack{\ell, m \rightarrow \infty, q \rightarrow 0 \\ \ell q \rightarrow a, \frac{m}{\ell} \rightarrow \alpha}} \int_{\mathcal{P}_{\ell+1}^m} f\left(\frac{\pi_K(o)}{m}\right) d\nu^\theta(o) = f(\rho_0^* + \dots + \rho_K^*).$$

*Proof.* Let  $\varepsilon > 0$ . We define two sequences of stopping times  $(T_n)_{n \geq 0}$  and  $(T_n^*)_{n \geq 1}$  as follows. Let  $T_0 = 0$  and set

$$\begin{aligned} T_1^* &= \inf \{ n \geq 0 : Z_n^\theta \in U_K(\varepsilon) \}, & T_1 &= \inf \{ n \geq T_1^* : Z_n^\theta \notin U_K(2\varepsilon) \}, \\ &\vdots & &\vdots \\ T_k^* &= \inf \{ n \geq T_{k-1} : Z_n^\theta \in U_K(\varepsilon) \}, & T_k &= \inf \{ n \geq T_k^* : Z_n^\theta \notin U_K(2\varepsilon) \}, \\ &\vdots & &\vdots \end{aligned}$$

The ergodic theorem for Markov chains implies that

$$\int_{\mathcal{P}_{\ell+1}^m} f\left(\frac{\pi_K(o)}{m}\right) d\nu^\theta(o) = \lim_{N \rightarrow \infty} \frac{1}{N} E\left(\sum_{n=0}^{N-1} f\left(\frac{\pi_K(Z_n^\theta)}{m}\right)\right).$$

Let  $N \geq 0$ . We decompose this last sum as follows:

$$\sum_{n=0}^{N-1} f\left(\frac{\pi_K(Z_n^\theta)}{m}\right) = \sum_{k \geq 1} \sum_{n=T_{k-1} \wedge N}^{T_k^* \wedge N-1} f\left(\frac{\pi_K(Z_n^\theta)}{m}\right) + \sum_{k \geq 1} \sum_{n=T_k^* \wedge N}^{T_k \wedge N-1} f\left(\frac{\pi_K(Z_n^\theta)}{m}\right).$$

The function  $f$  is continuous; let  $\epsilon > 0$  and let us choose  $\varepsilon$  small enough so that

$$\forall o \in U_K(2\varepsilon) \quad \left| f\left(\frac{\pi_K(o)}{m}\right) - f(\rho_0^* + \dots + \rho_K^*) \right| < \frac{\epsilon}{2}.$$

We have then

$$\begin{aligned} \left| E\left(\sum_{n=0}^{N-1} f\left(\frac{\pi_K(Z_n^\theta)}{m}\right)\right) - Nf(\rho_0^* + \dots + \rho_K^*) \right| \leq \\ 2f(1) \sum_{k \geq 1} E(T_k^* \wedge N - T_{k-1} \wedge N) + \frac{N\epsilon}{2}. \end{aligned}$$

Let us define

$$K(n) = \max \{ k \geq 0 : T_k \leq n \}.$$

We can bound the last sum as follows:

$$\sum_{k \geq 1} (T_k^* \wedge N - T_{k-1} \wedge N) \leq \sum_{k=1}^{K(N)} (T_k^* - T_{k-1}) + (N - T_{K(N)}).$$

We study now the random variable  $K(n)$ . Let  $k \in \mathbb{N}$ ,  $\lambda > 0$  and  $o \in \mathcal{P}_{\ell+1}^m$ . Let us seek estimates on the following probability:

$$P_o(K(k \exp(\lambda m)/2) \geq k).$$

From the definition of  $K(n)$ , it follows that  $K(n) \geq k$  if and only if  $T_k \leq n$ . Thus

$$P_o(K(k \exp(\lambda m)/2) \geq k) = P_o(T_k \leq k \exp(\lambda m)/2).$$

Let us define for  $i \geq 1$ ,

$$Y_i = T_i - T_{i-1}, \quad Y_i^* = T_i - T_i^*.$$

By theorem 4.1, there exist positive real numbers  $\lambda$  and  $\lambda'$  such that, for all  $i \geq 1$ ,

$$\begin{aligned} P_o(Y_i \leq \exp(\lambda m)) &\leq P_o(Y_i^* \leq \exp(\lambda m)) \\ &= \sum_{o' \in U_K(\varepsilon)} P_o(Y_i^* \leq \exp(\lambda m) \mid Z_{T_i^*}^\theta = o') P_o(Z_{T_i^*}^\theta = o') \\ &= \sum_{o' \in U_K(\varepsilon)} P_{o'}(T_1 \leq \exp(\lambda m)) P_o(Z_{T_i^*}^\theta = o') \leq \exp(-\lambda' m). \end{aligned}$$

Let us define the following Bernoulli random variables:

$$\forall i \geq 1, \quad \varepsilon_i = 1_{Y_i^* \leq \exp(\lambda m)}.$$

Notice that

$$T_k = Y_1 + \dots + Y_k \geq Y_1^* + \dots + Y_k^*.$$

If  $T_k \leq k \exp(\lambda m)/2$ , then there exist at least  $k/2$  indices in  $\{1, \dots, k\}$  such that  $Y_i^* \leq \exp(\lambda m)$ . Therefore,

$$T_k \leq \frac{1}{2} k \exp(\lambda m) \implies \varepsilon_1 + \dots + \varepsilon_k \geq \frac{k}{2}.$$

Thus

$$P_o\left(K\left(\frac{k \exp(\lambda m)}{2}\right) \geq k\right) \leq P\left(\varepsilon_1 + \dots + \varepsilon_k \geq \frac{k}{2}\right).$$

Let  $\beta \geq 0$ , thanks to Chebyshev's exponential inequality we have

$$P\left(\varepsilon_1 + \dots + \varepsilon_k \geq k/2\right) \leq \exp\left(-\beta/2 + \ln E\left(\exp(\beta \varepsilon_1/k) \cdots \exp(\beta \varepsilon_k/k)\right)\right).$$

Since  $\varepsilon_1, \dots, \varepsilon_{k-1}$  are measurable with respect to  $(Z_n^\theta, 0 \leq n \leq T_k^*)$ ,

$$\begin{aligned} &E\left(\exp(\beta \varepsilon_1/k) \cdots \exp(\beta \varepsilon_k/k)\right) \\ &= E\left(E\left(\exp(\beta \varepsilon_1/k) \cdots \exp(\beta \varepsilon_k/k) \mid Z_n^\theta, 0 \leq n \leq T_k^*\right)\right) \\ &= E\left(\exp(\beta \varepsilon_1/k) \cdots \exp(\beta \varepsilon_{k-1}/k) E\left(\exp(\beta \varepsilon_k/k) \mid Z_n^\theta, 0 \leq n \leq T_k^*\right)\right). \end{aligned}$$

Thanks to the strong Markov property, the above conditional expectation can be rewritten as follows:

$$E\left(\exp(\beta\varepsilon_k/k) \mid Z_n^\theta, 0 \leq n \leq T_k^*\right) = E\left(\exp(\beta\varepsilon_1/k) \mid Z_0^\theta = Z_{T_k^*}^\theta\right).$$

Yet, for all  $o' \in U_K(\delta)$ ,

$$E_{o'}\left(\exp(\beta\varepsilon_1/k)\right) \leq \exp\left(-\lambda'm + \frac{\beta}{k}\right) + 1 - \exp(-\lambda'm).$$

We iterate this procedure and we obtain

$$E\left(\exp(\beta\varepsilon_1/k) \cdots \exp(\beta\varepsilon_k/k)\right) \leq \left(\exp\left(-\lambda'm + \frac{\beta}{k}\right) + 1 - \exp(-\lambda'm)\right)^k.$$

The change of variables  $\beta \rightarrow k\beta$  yields

$$\begin{aligned} P\left(\varepsilon_1 + \cdots + \varepsilon_k \geq k/2\right) \\ \leq \exp\left(-k\left(\beta/2 - \ln\left(\exp(-\lambda'm + \beta) + 1 - \exp(-\lambda'm)\right)\right)\right). \end{aligned}$$

Let  $\Lambda^*(t)$  be the Cramér transform of the Bernoulli law with parameter  $p = \exp(-\lambda'm)$ :

$$\Lambda^*(t) = \sup_{\beta \geq 0} \left(\beta t - \ln(pe^\beta + 1 - p)\right) = t \ln \frac{t}{p} + (1-t) \ln \frac{1-t}{1-p}.$$

Optimising the previous inequality over  $\beta \geq 0$ , we obtain

$$P\left(\varepsilon_1 + \cdots + \varepsilon_k \geq k/2\right) \leq \exp\left(-k\Lambda^*(1/2)\right).$$

In our particular case, for  $m$  large enough,

$$\Lambda^*(1/2) = \frac{1}{2} \ln \frac{\exp(\lambda'm)}{2} + \frac{1}{2} \ln \frac{\exp(\lambda'm)}{2(\exp(\lambda'm) - 1)} \geq c(m),$$

where  $c(m)$  is a positive constant depending on  $m$  but not on  $k$ . It follows that for  $m$  large enough,

$$\forall n \geq 1 \quad P_o\left(K(k \exp(\lambda m)/2) \geq k\right) \leq \exp\left(-kc(m)\right).$$

Let  $N \geq 0$ . We seek next an upper bound for the expectation

$$E\left(\sum_{k=1}^{K(N)} (T_k^* - T_{k-1}) + (N - T_{K(N)})\right).$$

The sum inside the parenthesis is at most  $N$ , therefore, for  $i \geq 1$

$$E\left(\sum_{k=1}^{K(N)} (T_k^* - T_{k-1}) + (N - T_{K(N)})\right) \leq$$

$$E\left(\left(\sum_{k=1}^{K(N)} (T_k^* - T_{k-1}) + (N - T_{K(N)})\right) 1_{K(N) < i}\right) + NP(K(N) \geq i).$$

Let

$$i_N = \min \left\{ i \in \mathbb{N} : N \leq \frac{i \exp(\lambda m)}{2} \right\}.$$

On one hand, the analysis of the random variable  $K(k \exp(\lambda m)/2)$  shows that taking  $i = i_N$ , the second term is bounded by

$$\frac{i_N}{2} e^{\lambda m} P(K(i_N e^{\lambda m}/2) \geq i_N) \leq \frac{i_N}{2} \exp(\lambda m - i_N c(m)),$$

which goes to 0 when  $N$  goes to  $\infty$ . On the other hand, we can bound the first term thanks to theorem 4.1:

$$E\left(\left(\sum_{k=1}^{K(N)} (T_k^* - T_{k-1}) + (N - T_{K(N)})\right) 1_{K(N) < i_N}\right)$$

$$\leq E\left(\sum_{k=1}^{i_N+1} (T_k^* - T_{k-1})\right) = \sum_{k=1}^{i_N+1} E(T_k^* - T_{k-1}) \leq (i_N + 1) \frac{m^\alpha}{1 - \exp(-\alpha' m)},$$

where  $\alpha, \alpha' > 0$ . We combine the above inequalities, and we obtain for  $m$  large enough and for all  $N > 0$ ,

$$\frac{1}{N} E\left(\sum_{k \geq 1} (T_k^* \wedge N - T_{k-1} \wedge N)\right)$$

$$\leq \frac{2}{(i_N - 1) \exp(\lambda m)} \left( \frac{i_N}{2} \exp(\lambda m - i_N c(m)) + (i_N + 1) \frac{m^\alpha}{1 - \exp(-\alpha' m)} \right).$$

When  $N$  goes to  $\infty$ , this expression goes to  $2m^\alpha / \exp(\lambda m)(1 - \exp(-\alpha' m))$ , which in turn goes to 0 with  $m$ . We deduce that, for  $m$  large enough, there exists  $N_m > 0$  such that for all  $N \geq N_m$ ,

$$\left| \frac{1}{N} E\left(\sum_{n=0}^N f\left(\frac{\pi_K(Z_n^\theta)}{m}\right)\right) - f(\rho_0^* + \dots + \rho_K^*) \right| < \epsilon.$$

Thus,

$$\lim_{\substack{\ell, m \rightarrow \infty, q \rightarrow 0 \\ \ell q \rightarrow \alpha, \frac{m}{\ell} \rightarrow \alpha}} \int_{\mathcal{P}_{\ell+1}^m} f\left(\frac{\pi_K(o)}{m}\right) d\nu^\theta(o) = f(\rho_0^* + \dots + \rho_K^*).$$

□

## References

- [1] Domingos Alves and Jose Fernando Fontanari. Error threshold in finite populations. *Phys. Rev. E*, 57:7008–7013, 1998.
- [2] Jon P. Anderson, Richard Daifuku, and Lawrence A. Loeb. Viral error catastrophe by mutagenic nucleosides. *Annual Review of Microbiology*, 58(1):183–205, 2004.
- [3] Raphaël Cerf. Critical population and error threshold on the sharp peak landscape for a Moran model. *Mem. Amer. Math. Soc.*, 233(1096):vi+87, 2015.
- [4] Raphaël Cerf. Critical population and error threshold on the sharp peak landscape for a Wright–Fisher model. *Annals of Applied Probability*, 25(4):1936–1992, 2015.
- [5] Shane Crotty, Craig E. Cameron, and Raul Andino. RNA virus error catastrophe: Direct molecular test by using ribavirin. *Proceedings of the National Academy of Sciences*, 98(12):6895–6900, 2001.
- [6] Joseba Dalmau. The distribution of the quasispecies for the Wright–Fisher model on the sharp peak landscape. *Stochastic Process. Appl.*, 125(1):272–293, 2015.
- [7] Lloyd Demetrius, Peter Schuster, and Karl Sigmund. Polynucleotide evolution and branching processes. *Bulletin of Mathematical Biology*, 47(2):239 – 262, 1985.
- [8] Narendra M. Dixit, Piyush Srivastava, and Nisheeth K. Vishnoi. A finite population model of molecular evolution: theory and computation. *J. Comput. Biol.*, 19(10):1176–1202, 2012.
- [9] Esteban Domingo. Quasispecies theory in virology. *Journal of Virology*, 76(1):463–465, 2002.
- [10] Manfred Eigen. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523, 1971.
- [11] Manfred Eigen, John McCaskill, and Peter Schuster. The molecular quasi-species. *Advances in Chemical Physics*, 75:149–263, 1989.
- [12] Manfred Eigen and Peter Schuster. The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften*, 64(11):541 – 565, 1977.

- [13] Manfred Eigen and Peter Schuster. The hypercycle. A principle of natural self-organization. Part B: The abstract hypercycle. *Naturwissenschaften*, 65(1):7 – 41, 1978.
- [14] Manfred Eigen and Peter Schuster. The hypercycle. A principle of natural self-organization. Part C: The realistic hypercycle. *Naturwissenschaften*, 65(7):341 – 369, 1978.
- [15] Samuel Karlin and Howard M. Taylor. *A first course in stochastic processes*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, second edition, 1975.
- [16] Julian Keilson. *Markov chain models—rarity and exponentiality*, volume 28 of *Applied Mathematical Sciences*. Springer-Verlag, New York-Berlin, 1979.
- [17] John McCaskill. A stochastic theory of macromolecular evolution. *Biological Cybernetics*, 50:63–73, 1984.
- [18] Pat Moran. Random processes in genetics. *Proc. Cambridge Philos. Soc.*, 54:60–71, 1958.
- [19] Fabio Musso. A stochastic version of the Eigen model. *Bull. Math. Biol.*, 73(1):151–180, 2011.
- [20] Martin A. Nowak and Peter Schuster. Error thresholds of replication in finite populations. Mutation frequencies and the onset of Muller’s ratchet. *Journal of theoretical Biology*, 137 (4):375–395, 1989.
- [21] Jeong-Man Park, Enrique Muñoz, and Michael W. Deem. Quasispecies theory for finite populations. *Phys. Rev. E*, 81:011902, 2010.
- [22] David B. Saakian, Michael W. Deem, and Chin-Kun Hu. Finite population size effects in quasispecies models with single-peak fitness landscape. *Europhysics Letters*, 98(1):18001, 2012.
- [23] Ricard V. Solé and Thomas S. Deisboeck. An error catastrophe in cancer? *Journal of Theoretical Biology*, 228:47–54, 2004.
- [24] Kushal Tripathi, Rajesh Balagam, Nisheeth K. Vishnoi, and Narendra M. Dixit. Stochastic simulations suggest that HIV-1 survives close to its error threshold. *PLoS Comput. Biol.*, 8(9):e1002684, 14, 2012.
- [25] Claus Wilke. Quasispecies theory in the context of population genetics. *BMC Evolutionary Biology*, 5:1–8, 2005.