Critical population and error threshold on the sharp peak landscape for the Wright–Fisher model

Raphaël Cerf
Université Paris Sud and IUF

July 3, 2012

Abstract

We pursue the task of developing a finite population counterpart to Eigen's model. We consider the classical Wright–Fisher model describing the evolution of a population of size m of chromosomes of length ℓ over an alphabet of cardinality κ. The mutation probability per locus is q. The replication rate is $\sigma > 1$ for the master sequence and 1 for the other sequences. We study the equilibrium distribution of the process in the regime where

$$\ell \to +\infty\,, \qquad m \to +\infty\,, \qquad q \to 0\,,$$
$$\ell q \to a \in]0, +\infty[\,, \qquad \frac{m}{\ell} \to \alpha \in [0, +\infty]\,.$$

We obtain an equation $\alpha\,\psi(a) = \ln \kappa$ in the parameter space (a, α) separating the regime where the equilibrium population is totally random from the regime where a quasispecies is formed. We observe the existence of a critical population size necessary for a quasispecies to emerge and we recover the finite population counterpart of the error threshold. The result is the twin brother of the corresponding result for the Moran model. The proof is more complex and it relies on the Freidlin–Wentzell theory of random perturbations of dynamical systems.

1 Introduction.

The Wright–Fisher model is one of the most studied model in mathematical population genetics. In this work, we apply to a basic Wright–Fisher model the ideas presented in [3] for the Moran model, thereby pursuing the task of developing a finite population counterpart to Eigen's model. Numerous works have attacked this issue [1, 5, 10, 16, 20, 24]. Using different techniques, Deem, Hu and Saakian [23], Deem, Muñoz and Park [22], Musso [19] and Dixit, Srivastava, Vishnoi [6] considered finite population models which approximate Eigen's model when the population size goes to infinity. These models are variants or generalizations of the classical Wright–Fisher model of population genetics. The problem is to understand how the error threshold phenomenon present in Eigen's model in the infinite population limit shows up in the finite population model. We refer to the introduction of [3] for a detailed discussion of this question and the heuristics guiding our strategy. We consider here the classical Wright–Fisher model describing the evolution of a population of size m of chromosomes of length ℓ over an alphabet of cardinality κ. The mutation probability per locus is q. The replication rate is $\sigma > 1$ for the master sequence and 1 for the other sequences. We study the equilibrium distribution of the process in the regime where

$$\begin{gathered}\ell \to +\infty\,, \qquad m \to +\infty\,, \qquad q \to 0\,,\\ \ell q \to a \in]0,+\infty[\,, \qquad \frac{m}{\ell} \to \alpha \in [0,+\infty]\,.\end{gathered}$$

We obtain an equation $\alpha\,\psi(a) = \ln\kappa$ in the parameter space (a,α) separating the regime where the equilibrium population is totally random from the regime where a quasispecies is formed. We observe the existence of a critical population size necessary for a quasispecies to emerge and we recover the finite population counterpart of the error threshold. It is a classical fact that the Moran model and the Wright–Fisher model have similar dynamics. Indeed, the main result here is the twin brother of the main result of [3], the only difference being the equation of the critical curve. While we could compute exactly the critical curve for the Moran model, here the critical curve is defined through a variational problem depending on the parameter a. Apart from this point, the scaling and the associated exponents are the same in both cases. This confirms a conjecture of [3] and it sustains the hope that this kind of analysis is robust.

A potential application of the result concerns genetic algorithms. Indeed, the Wright–Fisher model is identical to the genetic algorithm without crossover. In her PhD thesis [21], Ochoa investigated the role of the error threshold phenomenon for genetic algorithms and she concluded that there exists a relationship between the optimal mutation rate and the error threshold. The result proved here provides a theoretical basis for some heuristics to control efficiently the genetic algorithms proposed in [2].

On the technical side, the Wright–Fisher model is much more difficult to handle than the Moran model. In the Moran model, the estimates of the selection drift relied on a birth and death model introduced by Nowak and Schuster [20]. In the Wright–Fisher model, the bounding processes are more complicated, they involve three dependent binomial laws. As the size of the population grows, their transition probabilities satisfy a large deviation principle, derived with the help of the classical Cramér theorem. In the set of the populations containing the master sequence, the process can be seen as the random perturbation of a discrete dynamical system. This discrete dynamical system is simply the sequence of the iterates of a rational map $F : [0,1] \to [0,1]$. Depending on the parameters, this map has either one stable fixed point or two fixed points, one stable and the other unstable. This opens the way to the application of the general scheme developed by Freidlin and Wentzell [9] to study the random perturbations of dynamical systems. Originally, Freidlin and Wentzell studied diffusion processes arising as brownian perturbations of a differential equation. These processes are continuous time Markov processes with a continuous state space. However their approach is robust and it can be applied in other contexts. Kifer [13, 14] reworked this theory in the discrete time case. Unfortunately, our bounding processes do not fit the hypothesis of Kifer's model, for the following two reasons. In Kifer's model, the large deviation rate function of the transition probabilities is not allowed to be infinite, and the large deviation principle for the transition probabilities is assumed to be uniform with respect to the starting point. Certainly the general framework considered by Kifer could be adjusted to include our case, with the help of some relaxed hypothesis. Yet in our case, we have only two attractors, one unstable and one stable, and we need only two specific estimates from the general theory, which is concerned with a finite number of attractors of any type. In fact, the kind of estimates we need have been computed in two other works handling closely related models. In an unpublished work [4] (transmitted to me by courtesy of Gregory Morrow), Darden analyzed a Wright–Fisher model with two alleles and no mutation with the help of the Freidlin–Wentzell theory. What we have to do essentially is to obtain results analogous to Darden for the model with mutations. Morrow and Sawyer [18] considered a more general model of Markov chains evolving in a convex subset of $\mathbb{R}^d$ around one stable attractor. Our bounding processes would fit this framework, were it for the uniform assumption on the variance of the transition probabilities. In our case, this condition is violated close to the unstable attractor 0. We can apply their results outside a neighborhood of 0, but this would lead to a messy construction. It appears that, in any case, if we try to apply the results of Kifer or of Morrow and

Sawyer, we have to make a specific study of our process in the vicinity of the unstable fixed point 0. In the end, it seems that the most efficient presentation consists in deriving from scratch the required estimates, following the initial ideas of Freidlin and Wentzell. The techniques involved in the proof are classical and go back to the seminal work of Freidlin and Wentzell. However there is an important simplifying feature in our case. Indeed, the bounding processes are monotone. This allows to avoid uniform large deviation estimates and to provide substantially simpler proofs.

We describe the model in the next section and we present the main result in section 3. The rest of the paper is devoted to the proofs. The global strategy is identical to the case of the Moran model. The lumping is performed in section 4. In section 5, we build a coupling and we prove the monotonicity of the occupancy process. This allows us to define simple bounding processes in section 6. Section 7 which analyzes the dynamics of the bounding processes is much more complicated than for the Moran model. Section 8 is slightly simpler than for the Moran model. Some results for the mutation dynamics from [3] are restated without proof, otherwise the estimates in the neutral case are adapted easily to the case of the Wright–Fisher model.

2 The Wright–Fisher model.

Let $\mathcal{A}$ be a finite alphabet and let $\kappa = \mathrm{card}\,\mathcal{A}$ be its cardinality. Let $\ell \geq 1$ be an integer. We consider the space $\mathcal{A}^\ell$ of sequences of length ℓ over the alphabet $\mathcal{A}$. Elements of this space represent the chromosome of an haploid individual, or equivalently its genotype. In our model, all the genes have the same set of alleles and each letter of the alphabet $\mathcal{A}$ is a possible allele. Typical examples are $\mathcal{A} = \{\, A, T, G, C \,\}$ to model standard DNA, or $\mathcal{A} = \{\, 0, 1 \,\}$ to deal with binary sequences. Generic elements of $\mathcal{A}^\ell$ will be denoted by the letters u, v, w. A population is an m–tuple of elements of $\mathcal{A}^\ell$. Generic populations will be denoted by the letters x, y, z. Thus a population x is a vector

$$x = \begin{pmatrix} x(1) \\ \vdots \\ x(m) \end{pmatrix}$$

whose components are chromosomes. For $i \in \{\, 1, \dots, m \,\}$, we denote by

$$x(i,1), \dots, x(i,\ell)$$

the letters of the sequence $x(i)$. This way a population x can be represented as an array

$$x = \begin{pmatrix} x(1,1) & \cdots & x(1,\ell) \\ \vdots & & \vdots \\ x(m,1) & \cdots & x(m,\ell) \end{pmatrix}$$

of size $m \times \ell$ of elements of $\mathcal{A}$, the i–th line being the i–th chromosome. The evolution of the population is random and it is driven by two antagonistic forces: replication and mutation.

Replication. The replication favors the development of fit chromosomes. The fitness of a chromosome is encoded in a fitness function

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

With the help of the fitness function A, we define a selection function $F : \mathcal{A}^\ell \times \big(\mathcal{A}^\ell\big)^m \to [0,1]$ by setting

$\forall u \in \mathcal{A}^\ell \quad \forall x \in \big(\mathcal{A}^\ell\big)^m$

$$F(u,x) = \frac{A(u)}{A(x(1)) + \cdots + A(x(m))} \sum_{1 \le i \le m} 1_{x(i)=u}\,.$$

The population x being fixed, the values $F(u,x)$, $u \in \mathcal{A}^\ell$, define a probability distribution over $\mathcal{A}^\ell$. The value $F(u,x)$ is the probability of choosing u when sampling from the population x.

Mutation. The mutation mechanism is the same for all the loci, and mutations occur independently. We denote by $q \in]0, 1-1/\kappa[$ the probability that a mutation occurs at one particular locus. If a mutation occurs, then the letter is replaced randomly by another letter, chosen uniformly over the $\kappa - 1$ remaining letters. We encode this mechanism in a mutation matrix

$$M(u,v)\,, \quad u, v \in \mathcal{A}^\ell\,,$$

where $M(u,v)$ is the probability that the chromosome u is transformed by mutation into the chromosome v. The analytical formula for $M(u,v)$ is

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

Transition matrix. We consider the classical Wright–Fisher model. In this model, generations do not overlap. The mechanism to build a new generation is divided in two steps. In the first step, m chromosomes are sampled with replacement from the population. The sampling law is given by the selection function. In the second step, each chromosome mutates according to the law specified by the mutation matrix. For $n \geq 0$, we denote by X_n the n–th generation. The Wright–Fisher model is the Markov chain $(X_n)_{n\in\mathbb{N}}$ on the space $\big(\mathcal{A}^\ell\big)^m$ whose transition matrix is given by

$$\forall n \in \mathbb{N} \quad \forall x, y \in \big(\mathcal{A}^\ell\big)^m$$
$$P\big(X_{n+1} = y \,|\, X_n = x\big) \,=\, \prod_{1\leq i\leq m} \Big(\sum_{u\in\mathcal{A}^\ell} F\big(u, x\big)\, M\big(u, y(i)\big)\Big)\,.$$

The other non diagonal coefficients of the transition matrix are zero. The diagonal terms are chosen so that the sum of each line is equal to one.

3 Main results.

We present the main results in this section.

Sharp peak landscape. We will consider only the sharp peak landscape defined as follows. We fix a specific sequence, denoted by w^*, called the wild type or the master sequence. Let $\sigma > 1$ be a fixed real number. The fitness function A is given by

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

Density of the master sequence. We denote by $N(x)$ the number of copies of the master sequence w^* present in the population x:

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

We are interested in the expected density of the master sequence in the steady state distribution of the process, that is,

$$\text{Master}(\sigma, \ell, m, p) \,=\, \lim_{n\to\infty} E\Big(\frac{1}{m} N(X_n)\Big)\,,$$

as well as the variance

$$\text{Variance}(\sigma, \ell, m, p) \,=\, \lim_{n\to\infty} E\bigg(\Big(\frac{1}{m} N(X_n) - \text{Master}(\sigma, \ell, m, p)\Big)^2\bigg)\,.$$

The ergodic theorem for Markov chains ensures that the above limits exist. We denote by $I(p, t)$ the rate function governing the large deviations of the binomial law of parameter $p \in [0, 1]$, given by

$$\forall t \in [0, 1] \qquad I(p, t) \,=\, t \ln \frac{t}{p} + (1 - t) \ln \frac{1 - t}{1 - p}\,.$$

We define, for $a \in]0,+\infty[$,

$$\rho^*(a) \;=\; \begin{cases} \dfrac{\sigma e^{-a}-1}{\sigma-1} & \text{if } \sigma e^{-a}>1 \\ 0 & \text{if } \sigma e^{-a}\leq 1 \end{cases},$$

$$\Psi(a) \;=\; \inf_{l\in\mathbb{N}} \inf \Big\{ \sum_{k=0}^{l-1} I\Big(\frac{\sigma\rho_k}{(\sigma-1)\rho_k+1},\gamma_k\Big) + \gamma_k\, I\Big(e^{-a},\frac{\rho_{k+1}}{\gamma_k}\Big) : \\ \rho_0=\rho^*(a),\ \rho_l=0,\ \rho_k,\gamma_k\in[0,1] \text{ for } 0\leq k<l \Big\}.$$

The function Ψ is finite positive on $]0,\ln\sigma[$ and it vanishes on $[\ln\sigma,+\infty[$.

Theorem 3.1 We suppose that

$$\ell\to+\infty\,,\qquad m\to+\infty\,,\qquad q\to 0\,,$$

in such a way that

$$\ell q\to a\in]0,+\infty[\,,\qquad \frac{m}{\ell}\to\alpha\in[0,+\infty]\,.$$

We have the following dichotomy:

- If $\alpha\,\psi(a)<\ln\kappa$ then $\text{Master}\big(\sigma,\ell,m,q\big)\;\to\;0$.
- If $\alpha\,\psi(a)>\ln\kappa$ then $\text{Master}\big(\sigma,\ell,m,q\big)\;\to\;\rho^*(a)$.

In both cases, we have $\text{Variance}\big(\sigma,\ell,m,q\big)\;\to\;0$.

The statement of the theorem holds also in the case where α is null or infinite, but a must belong to $]0,+\infty[$. This result is very similar to the result for the Moran model. Therefore all the comments done for the Moran model apply here as well. The main difference is that the function $\Psi(a)$ is more complicated. While we could obtain an explicit formula in the case of the Moran model, here the function $\Psi(a)$ is the solution of a complicated variational problem. The general structure of the proof is similar to the one for the Moran model. We use the lumping theorem to reduce the size of the state space. We couple the lumped processes with different initial conditions. The coupling for the occupancy process is monotone. We construct then a lower and an upper process. These processes behave like the original process in the neutral region and like a Wright–Fisher model with two alleles whenever the master sequence is present in the population. The dynamics of these models is analyzed with a specific implementation of the Freidlin–Wentzell theory. We compute estimates of the persistence time of the master sequence, as well as its equilibrium density. Although the results are similar to the case of the Moran model, this part is much more technical in the case of the Wright–Fisher model. Indeed, in the case of the Moran model, we needed simply to estimate some explicit formula associated to the birth and death model introduced by Nowak and Schuster [20]. The approach used here to handle the Wright–Fisher model is quite robust and it should work for other variants of the model. In the final section we analyze the discovery time of the master sequence. This part is similar to the case of the Moran model, it is even simpler.

4 Lumping

We denote by d_H the Hamming distance between two chromosomes:

$$\forall u,v\in\mathcal{A}^\ell\qquad d_H(u,v)\,=\,\mathrm{card}\,\big\{\,j:1\leq j\leq\ell,\,u(j)\neq v(j)\,\big\}\,.$$

We define a function $H:\mathcal{A}^\ell\to\{\,0,\dots,\ell\,\}$ by setting

$$\forall u\in\mathcal{A}^\ell\qquad H(u)\,=\,d_H\big(u,w^*\big)\,.$$

We define further a vector function $\mathbb{H}:\big(\mathcal{A}^\ell\big)^m\to\{\,0,\dots,\ell\,\}^m$ by setting

$$\forall x\,=\,\begin{pmatrix}x(1)\\ \vdots\\ x(m)\end{pmatrix}\in\big(\mathcal{A}^\ell\big)^m\qquad \mathbb{H}(x)\,=\,\begin{pmatrix}H\big(x(1)\big)\\ \vdots\\ H\big(x(m)\big)\end{pmatrix}\,.$$

4.1 Mutation and replication

We state some results on the mutation matrix that have been proved in [3]. The mutation matrix is lumpable with respect to the function H. Let $b,c\in\{\,0,\dots,\ell\,\}$ and let $u\in\mathcal{A}^\ell$ such that $H(u)=b$. The sum

$$\sum_{\substack{w\in\mathcal{A}^\ell\\ H(w)=c}}M(u,w)$$

does not depend on u in $H^{-1}(\{\,b\,\})$, it is a function of b and c only, which we denote by $M_H(b,c)$. The coefficient $M_H(b,c)$ is equal to

$$\sum_{\substack{0\leq k\leq\ell-b\\ 0\leq l\leq b\\ k-l=c-b}}\binom{\ell-b}{k}\binom{b}{l}\Big(p\Big(1-\frac{1}{\kappa}\Big)\Big)^k\Big(1-p\Big(1-\frac{1}{\kappa}\Big)\Big)^{\ell-b-k}\Big(\frac{p}{\kappa}\Big)^l\Big(1-\frac{p}{\kappa}\Big)^{b-l}\,.$$

The fitness function A of the sharp peak landscape can be factorized through H. If we define

$$\forall b\in\{\,0,\dots,\ell\,\}\qquad A_H(b)\,=\,\begin{cases}\sigma & \text{if } b=0\\ 1 & \text{if } b\geq 1\end{cases}$$

then we have

$$\forall u \in \mathcal{A}^\ell \qquad A(u) \,=\, A_H(H(u))\,.$$

The selection function F can also be factorized through H. We define a selection function $F_H : \{\,0,\dots,\ell\,\} \times \{\,0,\dots,\ell\,\}^m \to [0,1]$ by setting

$$\forall k \in \{\,0,\dots,\ell\,\} \quad \forall d \in \{\,0,\dots,\ell\,\}^m$$
$$F_H(k,d) \,=\, \frac{A_H(k)}{A_H(d(1)) + \cdots + A_H(d(m))} \sum_{1 \le i \le m} 1_{d(i)=k}\,.$$

We have then

$$\forall k \in \{\,0,\dots,\ell\,\} \quad \forall x \in \big(\mathcal{A}^\ell\big)^m \qquad \sum_{\substack{u \in \mathcal{A}^\ell \\ H(u)=k}} F\big(u,x\big) \,=\, F_H\big(k, \mathbb{H}(x)\big)\,.$$

4.2 Exchangeability

The symmetric group $\mathfrak{S}_m$ of the permutations of $\{\,1,\dots,m\,\}$ acts in a natural way on the populations through the following group operation:

$$\forall x \in \big(\mathcal{A}^\ell\big)^m \quad \forall \rho \in \mathfrak{S}_m \quad \forall j \in \{\,1,\dots,m\,\} \qquad (\rho \cdot x)(j) \,=\, x(\rho(j))\,.$$

A probability measure μ on $\big(\mathcal{A}^\ell\big)^m$ is exchangeable if it is invariant under the action of $\mathfrak{S}_m$:

$$\forall \rho \in \mathfrak{S}_m \quad \forall x \in \big(\mathcal{A}^\ell\big)^m \quad \mu(\rho \cdot x) \,=\, \mu(x)\,.$$

A process $(X_n)_{n \ge 0}$ with values in $\big(\mathcal{A}^\ell\big)^m$ is exchangeable if and only if, for any $n \ge 0$, the law of X_n is exchangeable.

Lemma 4.1 The transition matrix p is invariant under the action of $\mathfrak{S}_m$:

$$\forall x, y \in \big(\mathcal{A}^\ell\big)^m \quad \forall \rho \in \mathfrak{S}_m \qquad p\big(\rho \cdot x, \rho \cdot y\big) \,=\, p\big(x,y\big)\,.$$

Proof. Let $\rho \in \mathfrak{S}_m$ and let $x, y \in \big(\mathcal{A}^\ell\big)^m$. We have

$$\begin{aligned}
p\big(\rho \cdot x, \rho \cdot y\big) \,&=\, \sum_{z \in (\mathcal{A}^\ell)^m} \prod_{1 \le i \le m} \Big(F\big(z(i), \rho \cdot x\big)\, M\big(z(i), \rho \cdot y(i)\big)\Big) \\
&=\, \sum_{z \in (\mathcal{A}^\ell)^m} \prod_{1 \le i \le m} \Big(F\big(\rho \cdot z(i), \rho \cdot x\big)\, M\big(\rho \cdot z(i), \rho \cdot y(i)\big)\Big) \\
&=\, \sum_{z \in (\mathcal{A}^\ell)^m} \prod_{1 \le i \le m} \Big(F\big(z(i), x\big)\, M\big(z(i), y(i)\big)\Big) \,=\, p\big(x,y\big)\,.
\end{aligned}$$

Thus the matrix p satisfies the required invariance property. $\square$

Corollary 4.2 Let μ be an exchangeable probability distribution on the population space $\big(\mathcal{A}^\ell\big)^m$. The Wright–Fisher model $(X_n)_{n\geq 0}$ starting with μ as the initial distribution is exchangeable.

4.3 Distance process

We define the distance process $(D_n)_{n\geq 0}$ by

$$\forall n\geq 0 \qquad D_n \;=\; \mathbb{H}\big(X_n\big)\,.$$

We prove next that the Markov chain $(X_n)_{n\geq 0}$ is lumpable with respect to the partition of $\big(\mathcal{A}^\ell\big)^m$ induced by the map $\mathbb{H}$, so that the distance process $(D_n)_{n\geq 0}$ is a genuine Markov chain.

Proposition 4.3 ($\mathbb{H}$ Lumpability) Let p be the transition matrix of the Wright–Fisher model. We have

$$\begin{aligned}
&\forall e\in\{\,0,\dots,\ell\,\}^m \quad \forall x,y\in\big(\mathcal{A}^\ell\big)^m\,,\\
&\qquad\mathbb{H}(x)=\mathbb{H}(y)\quad\Longrightarrow\quad \sum_{\substack{z\in(\mathcal{A}^\ell)^m\\ \mathbb{H}(z)=e}} p(x,z)\;=\;\sum_{\substack{z\in(\mathcal{A}^\ell)^m\\ \mathbb{H}(z)=e}} p(y,z)\,.
\end{aligned}$$

Proof. Let $d,e\in\{\,0,\dots,\ell\,\}^m$ and let $x\in\big(\mathcal{A}^\ell\big)^m$ be such that $\mathbb{H}(x)=d$. We have

$$\begin{aligned}
\sum_{\substack{z\in(\mathcal{A}^\ell)^m\\ \mathbb{H}(z)=e}} p(x,z) \;&=\; \sum_{\substack{z\in(\mathcal{A}^\ell)^m\\ \mathbb{H}(z)=e}}\;\prod_{1\leq i\leq m}\Big(\sum_{u\in\mathcal{A}^\ell} F\big(u,x\big)\,M\big(u,z(i)\big)\Big)\\
&=\prod_{1\leq i\leq m}\Big(\sum_{u\in\mathcal{A}^\ell} F\big(u,x\big)\sum_{\substack{v\in\mathcal{A}^\ell\\ H(v)=e(i)}} M\big(u,v\big)\Big)\\
&=\prod_{1\leq i\leq m}\Big(\sum_{u\in\mathcal{A}^\ell} F\big(u,x\big)\,M_H\big(H(u),e(i)\big)\Big)\\
&=\prod_{1\leq i\leq m}\Big(\sum_{k\in\{\,0,\dots,\ell\,\}}\;\sum_{\substack{u\in\mathcal{A}^\ell\\ H(u)=k}} F\big(u,x\big)\,M_H\big(k,e(i)\big)\Big)\\
&=\prod_{1\leq i\leq m}\Big(\sum_{k\in\{\,0,\dots,\ell\,\}} F_H\big(k,\mathbb{H}(x)\big)\,M_H\big(k,e(i)\big)\Big)\,.
\end{aligned}$$

This quantity is a function of $\mathbb{H}(x)=d$ and e only. $\square$

We apply the classical lumping result to conclude that the distance process $(D_n)_{n\geq 0}$ is a Markov chain. From the previous computations, we see that its transition matrix p_H is given by

$$\forall d,e\in\{\,0,\dots,\ell\,\}^m$$
$$p_H\big(d,e\big)\;=\;\prod_{1\leq i\leq m}\Big(\sum_{k\in\{\,0,\dots,\ell\,\}}F_H\big(k,d\big)\,M_H\big(k,e(i)\big)\Big)\,.$$

4.4 Occupancy process

We denote by $\mathcal{P}^m_{\ell+1}$ the set of the ordered partitions of the integer m in at most $\ell+1$ parts:

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

These partitions are interpreted as occupancy distributions. The partition $(o(0),\dots,o(\ell))$ corresponds to a population in which $o(l)$ chromosomes are at Hamming distance l from the master sequence, for any $l\in\{\,0,\dots,\ell\,\}$. Let $\mathcal{O}$ be the map which associates to each population x its occupancy distribution $\mathcal{O}(x)=(o(x,0),\dots,o(x,\ell))$, defined by:

$$\forall l\in\{\,0,\dots,\ell\,\}\qquad o(x,l)\;=\;\text{card}\big\{\,i:1\leq i\leq m,\,d_H(x(i),w^*)=l\,\big\}\,.$$

The map $\mathcal{O}$ can be factorized through $\mathbb{H}$. For $d\in\{\,0,\dots,\ell\,\}^m$, we set

$$o_H(d,l)\;=\;\text{card}\big\{\,i:1\leq i\leq m,\,d(i)=l\,\big\}$$

and we define a map $\mathcal{O}_H:\{\,0,\dots,\ell\,\}^m\to\mathcal{P}^m_{\ell+1}$ by setting

$$\mathcal{O}_H(d)\;=\;(o_H(d,0),\dots,o_H(d,\ell))\,.$$

We have then

$$\forall x\in\big(\mathcal{A}^\ell\big)^m\qquad\mathcal{O}(x)\;=\;\mathcal{O}_H\big(\mathbb{H}(x)\big)\,.$$

The map $\mathcal{O}$ lumps together populations which are permutations of each other:

$$\forall x\in\big(\mathcal{A}^\ell\big)^m\quad\forall\rho\in\mathfrak{S}_m\qquad\mathcal{O}(x)\;=\;\mathcal{O}(\rho\cdot x)\,.$$

We define the occupancy process $(O_n)_{n\geq 0}$ by setting

$$\forall n\geq 0\qquad O_n\;=\;\mathcal{O}(X_n)\;=\;\mathcal{O}_H(D_n)\,.$$

Proposition 4.4 ($\mathcal{O}$ Lumpability) Let p_H be the transition matrix of the distance process. We have

$$\forall o\in\mathcal{P}^m_{\ell+1}\quad\forall d,e\in\{\,0,\dots,\ell\,\}^m\,,$$
$$\mathcal{O}_H(d)=\mathcal{O}_H(e)\quad\Longrightarrow\quad\sum_{\substack{f\in\{\,0,\dots,\ell\,\}^m\\ \mathcal{O}_H(f)=o}}p_H(d,f)\;=\;\sum_{\substack{f\in\{\,0,\dots,\ell\,\}^m\\ \mathcal{O}_H(f)=o}}p_H(e,f)\,.$$

Proof. Let $o \in \mathcal{P}_{\ell+1}^m$ and $d, e \in \{0, \ldots, \ell\}^m$ such that $\mathcal{O}_H(d) = \mathcal{O}_H(e)$. Since $\mathcal{O}_H(d) = \mathcal{O}_H(e)$, then there exists a permutation $\rho \in \mathfrak{S}_m$ such that $\rho \cdot d = e$. By lemma 4.1, the transition matrices p and p_H are invariant under the action of $\mathfrak{S}_m$, therefore

$$\sum_{\substack{f \in \{0,\ldots,\ell\}^m \\ \mathcal{O}_H(f)=o}} p_H(d,f) = \sum_{\substack{f \in \{0,\ldots,\ell\}^m \\ \mathcal{O}_H(f)=o}} p_H(\rho \cdot d, \rho \cdot f)$$
$$= \sum_{\substack{f \in \{0,\ldots,\ell\}^m \\ \mathcal{O}_H(\rho^{-1} \cdot f)=o}} p_H(e,f) = \sum_{\substack{f \in \{0,\ldots,\ell\}^m \\ \mathcal{O}_H(f)=o}} p_H(e,f)$$

as requested. □

We apply the classical lumping result to conclude that the occupancy process $(O_n)_{n \geq 0}$ is a Markov chain. Let us compute its transition probabilities. We define a selection function $F_O : \{0, \ldots, \ell\} \times \mathcal{P}_{\ell+1}^m \to [0,1]$ by setting

$$\forall k \in \{0, \ldots, \ell\} \quad \forall o \in \mathcal{P}_{\ell+1}^m \qquad F_O(k,o) = \frac{o(k) A_H(k)}{\sum_{0 \leq h \leq \ell} o(h)\, A_H(h)}\,.$$

We have then

$$\forall k \in \{0, \ldots, \ell\} \quad \forall d \in \{0, \ldots, \ell\}^m \qquad F_O\big(k, \mathcal{O}_H(d)\big) = F_H\big(k,d\big)\,.$$

Let $o \in \mathcal{P}_{\ell+1}^m$ and $d \in \{0, \ldots, \ell\}^m$ be such that $\mathcal{O}_H(d) \neq o$. We write

$$\sum_{\substack{e \in \{0,\ldots,\ell\}^m \\ \mathcal{O}_H(e)=o}} p_H(d,e) = \sum_{\substack{e \in \{0,\ldots,\ell\}^m \\ \mathcal{O}_H(e)=o}} \prod_{1 \leq i \leq m} \Big(\sum_{k \in \{0,\ldots,\ell\}} F_H\big(k,d\big)\, M_H\big(k, e(i)\big) \Big)$$
$$= \prod_{0 \leq h \leq \ell} \Big(\sum_{k \in \{0,\ldots,\ell\}} F_O\big(k, \mathcal{O}_H(d)\big)\, M_H\big(k,h\big) \Big)^{o(h)}\,.$$

The last quantity depends only on $\mathcal{O}_H(d)$ and o as requested. We conclude that the transition matrix of the occupancy process is given by

$$\forall o, o' \in \mathcal{P}_{\ell+1}^m \qquad p_O\big(o,o'\big) = \prod_{0 \leq h \leq \ell} \Big(\sum_{k \in \{0,\ldots,\ell\}} F_O\big(k,o\big)\, M_H\big(k,h\big) \Big)^{o'(h)}\,.$$

5 Monotonicity

A crucial property for comparing the Wright–Fisher model with other processes is monotonicity. We will realize a coupling of the lumped Wright–Fisher processes with different initial conditions and we will deduce the monotonicity from the coupling construction. All the processes will be built on a single large probability space. We consider a probability space $(\Omega, \mathcal{F}, P)$ containing the following collection of independent random variables, all of them following the uniform law on the interval $[0,1]$:

$$\begin{gathered} U_n^{i,j}\,, \quad n \geq 1\,, \quad 1 \leq i \leq m\,, \quad 1 \leq j \leq \ell\,, \\ S_n^i\,, \quad n \geq 1\,, \quad 1 \leq i \leq m\,. \end{gathered}$$

To build the coupling, it is more convenient to replace the mutation probability q by the parameter p given by

$$p \,=\, \frac{\kappa}{\kappa - 1}\, q\,.$$

5.1 Coupling of the lumped processes

We build here a coupling of the lumped processes. We set

$$\forall n \geq 1 \qquad R_n \,=\, \begin{pmatrix} S_n^1, U_n^{1,1}, \dots, U_n^{1,\ell} \\ \vdots \quad \vdots \quad \dots \quad \vdots \\ S_n^m, U_n^{m,1}, \dots, U_n^{m,\ell} \end{pmatrix}.$$

The matrix R_n is the random input which is used to perform the n–th step of the Markov chains. We denote by $\mathcal{R}$ the set of the matrices of size $m \times (\ell + 1)$ with coefficients in $[0,1]$. The sequence $(R_n)_{n \geq 1}$ is a sequence of independent identically distributed random matrices with values in $\mathcal{R}$.

Mutation. We define a map

$$\mathcal{M}_H : \{\, 0, \dots, \ell \,\} \times [0,1]^\ell \to \{\, 0, \dots, \ell \,\}$$

in order to couple the mutation mechanism starting with different chromosomes. Let $b \in \{\, 0, \dots, \ell \,\}$ and let $u_1, \dots, u_\ell \in [0,1]^\ell$. The map $\mathcal{M}_H$ is defined by setting

$$\mathcal{M}_H(b, u_1, \dots, u_\ell) = b - \sum_{k=1}^{b} 1_{u_k < p/\kappa} + \sum_{k=b+1}^{\ell} 1_{u_k > 1 - p(1-1/\kappa)}\,.$$

The map $\mathcal{M}_H$ is built in such a way that, if $U_1, \dots, U_\ell$ are random variables with uniform law on the interval $[0,1]$, all being independent, then for any $b \in \{\, 0, \dots, \ell \,\}$, the law of $\mathcal{M}_H(b, U_1, \dots, U_\ell)$ is given by the line of the mutation matrix M_H associated to b, i.e.,

$$\forall c \in \{\, 0, \dots, \ell \,\} \qquad P\big(\mathcal{M}_H(b, U_1, \dots, U_\ell) = c\big) \,=\, M_H(b, c)\,.$$

Selection for the distance process. We realize the replication mechanism with the help of a selection map

$$\mathcal{S}_H : \{\,0,\dots,\ell\,\}^m \times [0,1] \to \{\,1,\dots,m\,\}\,.$$

Let $d \in \{\,0,\dots,\ell\,\}^m$ and let $s \in [0,1[$. We define $\mathcal{S}_H(d,s) = i$ where i is the unique index in $\{\,1,\dots,m\,\}$ satisfying

$$\frac{A_H(d(1)) + \cdots + A_H(d(i-1))}{A_H(d(1)) + \cdots + A_H(d(m))} \leq s < \frac{A_H(d(1)) + \cdots + A_H(d(i))}{A_H(d(1)) + \cdots + A_H(d(m))}\,.$$

The map $\mathcal{S}_H$ is built in such a way that, if S is a random variable with uniform law on the interval $[0,1]$, then for any $d \in \{\,0,\dots,\ell\,\}^m$, the law of $\mathcal{S}_H(d,S)$ is given by

$$\forall i \in \{\,1,\dots,m\,\} \qquad P\big(\mathcal{S}_H(d,S) = i\big) \,=\, \frac{A_H(d(i))}{A_H(d(1)) + \cdots + A_H(d(m))}\,.$$

Coupling for the distance process. We build a deterministic map

$$\Psi_H : \{\,0,\dots,\ell\,\}^m \times \mathcal{R} \to \{\,0,\dots,\ell\,\}^m$$

in order to realize the coupling between distance processes with various initial conditions. The coupling map Ψ_H is defined by

$$\forall r \in \mathcal{R}\,, \quad \forall d \in \{\,0,\dots,\ell\,\}^m$$

$$\Psi_H(d,r) \,=\, \begin{pmatrix} \mathcal{M}_H(d(\mathcal{S}_H(d,r(1,1))), r(1,2),\dots,r(1,\ell+1))) \\ \vdots \\ \mathcal{M}_H(d(\mathcal{S}_H(d,r(m,1))), r(m,2),\dots,r(m,\ell+1))) \end{pmatrix}\,.$$

The coupling is then built in a standard way with the help of the i.i.d. sequence $(R_n)_{n\geq 1}$ and the map Ψ_H. Let $d \in \{\,0,\dots,\ell\,\}^m$ be the starting point of the process. We build the distance process $(D_n)_{n\geq 0}$ by setting $D_0 = d$ and

$$\forall n \geq 1 \qquad D_n \,=\, \Psi_H\big(D_{n-1}, R_n\big)\,.$$

A routine check shows that the process $(D_n)_{n\geq 0}$ is a Markov chain starting from d with the adequate transition matrix. This way we have coupled the distance processes with various initial conditions.

Selection for the occupancy process. We realize the replication mechanism with the help of a selection map

$$\mathcal{S}_O : \mathcal{P}^m_{\ell+1} \times [0,1] \to \{\,0,\dots,\ell\,\}\,.$$

Let $o \in \mathcal{P}_{\ell+1}^m$ and let $s \in [0,1[$. We define $\mathcal{S}_O(o,s) = l$ where l is the unique index in $\{ 0, \dots, \ell \}$ satisfying

$$\frac{o(0)A_H(0) + \cdots + o(l-1)A_H(l-1)}{o(0)A_H(0) + \cdots + o(\ell)A_H(\ell)} \leq s < \frac{o(0)A_H(0) + \cdots + o(l)A_H(l)}{o(0)A_H(0) + \cdots + o(\ell)A_H(\ell)}\,.$$

The map $\mathcal{S}_O$ is built in such a way that, if S is a random variable with uniform law on the interval $[0,1]$, then for any $o \in \mathcal{P}_{\ell+1}^m$, the law of $\mathcal{S}_O(o,S)$ is given by

$$\forall l \in \{ 0, \dots, \ell \} \qquad P\big(\mathcal{S}_O(o,S) = l\big) \,=\, \frac{o(l)\,A_H(l)}{o(0)A_H(0) + \cdots + o(\ell)A_H(\ell)}\,.$$

Coupling for the occupancy process. We build a deterministic map

$$\Psi_O : \mathcal{P}_{\ell+1}^m \times \mathcal{R} \to \mathcal{P}_{\ell+1}^m$$

in order to realize the coupling between occupancy processes with various initial conditions. The coupling map Ψ_O is defined by

$$\forall r \in \mathcal{R}\,, \quad \forall o \in \mathcal{P}_{\ell+1}^m$$
$$\Psi_O(o,r) \,=\, \mathcal{O}_H \begin{pmatrix} \mathcal{M}_H(\mathcal{S}_O(o, r(1,1)), r(1,2), \dots, r(1,\ell+1))) \\ \vdots \\ \mathcal{M}_H(\mathcal{S}_O(o, r(m,1)), r(m,2), \dots, r(m,\ell+1))) \end{pmatrix}.$$

Let $o \in \mathcal{P}_{\ell+1}^m$ be the starting point of the process. We build the occupancy process $(O_n)_{n\geq 0}$ by setting $O_0 = o$ and

$$\forall n \geq 1 \qquad O_n \,=\, \Psi_O\big(O_{n-1}, R_n\big)\,.$$

A routine check shows that the process $(O_n)_{n\geq 0}$ is a Markov chain starting from o with the adequate transition matrix. This way we have coupled the occupancy processes with various initial conditions.

5.2 Monotonicity of the model

We first recall some standard definitions concerning monotonicity and correlations for stochastic processes. A classical reference is Liggett's book [15], especially for applications to particle systems. In the next two definitions, we consider a discrete time Markov chain $(X_n)_{n\geq 0}$ with values in a space $\mathcal{E}$. We suppose that the state space $\mathcal{E}$ is finite and that it is equipped with a partial order $\leq$. A function $f : \mathcal{E} \to \mathbb{R}$ is non–decreasing if

$$\forall x, y \in \mathcal{E} \qquad x \leq y \quad \Rightarrow \quad f(x) \leq f(y)\,.$$

Definition 5.1 The Markov chain $(X_n)_{n\geq 0}$ is said to be monotone if, for any non–decreasing function f, the function

$$x \in \mathcal{E} \mapsto E\big(f(X_n) \,|\, X_0 = x\big)$$

is non–decreasing.

A natural way to prove monotonicity is to construct an adequate coupling.

Definition 5.2 A coupling for the Markov chain $(X_n)_{n\geq 0}$ is a family of processes $(X_n^x)_{n\geq 0}$ indexed by $x \in \mathcal{E}$, which are all defined on the same probability space, and such that, for $x \in \mathcal{E}$, the process $(X_n^x)_{n\geq 0}$ is the Markov chain $(X_n)_{n\geq 0}$ starting from $X_0 = x$. The coupling is said to be monotone if

$$\forall x, y \in \mathcal{E} \qquad x \leq y \quad \Rightarrow \quad \forall n \geq 1 \qquad X_n^x \leq X_n^y \,.$$

If there exists a monotone coupling, then the Markov chain is monotone. The space $\{\,0,\dots,\ell\,\}^m$ is naturally endowed with a partial order:

$$d \leq e \quad \iff \quad \forall i \in \{\,1,\dots,m\,\} \quad d(i) \leq e(i) \,.$$

The map $\mathcal{M}_H$ is non–decreasing with respect to the Hamming class, i.e.,

$$\begin{aligned} &\forall b, c \in \{\,0,\dots,\ell\,\} \quad \forall u_1,\dots,u_\ell \in [0,1] \\ &\qquad\qquad b \leq c \quad \Rightarrow \quad \mathcal{M}_H(b,u_1,\dots,u_\ell) \,\leq\, \mathcal{M}_H(c,u_1,\dots,u_\ell) \,. \end{aligned}$$

(See [3] for a detailed proof). In the neutral case $\sigma = 1$, the map $\mathcal{S}_H$ does not depend on the population, in fact,

$$\forall d \in \{\,0,\dots,\ell\,\}^m \quad \forall s \in [0,1] \qquad \mathcal{S}_H(d,s) \,=\, \lfloor ms \rfloor \,.$$

As a consequence, we have

$$\begin{aligned} &\forall d, e \in \{\,0,\dots,\ell\,\}^m \quad \forall s \in [0,1] \\ &\qquad\qquad\qquad d \leq e \quad \Rightarrow \quad d\big(\mathcal{S}_H(d,s)\big) \,\leq\, e\big(\mathcal{S}_H(e,s)\big) \,. \end{aligned}$$

Lemma 5.3 In the neutral case $\sigma = 1$, the map Ψ_H is non–decreasing with respect to the distances, i.e.,

$$\forall d, e \in \{\,0,\dots,\ell\,\}^m \quad \forall r \in \mathcal{R}\,, \qquad d \leq e \quad \Rightarrow \quad \Psi_H(d,r) \,\leq\, \Psi_H(e,r) \,.$$

Proof. Let $r \in \mathcal{R}$ and let $d, e \in \{\,0,\dots,\ell\,\}^m$, $d \leq e$. Let $i \in \{\,1,\dots,m\,\}$. Since

$$\mathcal{S}_H(d,r(i,1)) \,=\, \mathcal{S}_H(e,r(i,1)) \,=\, \lfloor mr(i,1) \rfloor \,,$$

then

$$d(\mathcal{S}_H(d, r(i,1))) \,\leq\, e(\mathcal{S}_H(e, r(i,1)))\,.$$

This inequality and the monotonicity of the map $\mathcal{M}_H$ imply that

$$\begin{aligned}\mathcal{M}_H\big(d(\mathcal{S}_H(d, r(i,1))), r(i,2), \dots, r(i,\ell+1))\big)&\\ \leq\, \mathcal{M}_H\big(e(\mathcal{S}_H(e, r(i,1))), r(i,2), \dots, r(i,\ell+1))\big)&\,.\end{aligned}$$

Therefore $\Psi_H(d,r) \,\leq\, \Psi_H(e,r)$ as requested. $\square$

Corollary 5.4 In the neutral case $\sigma = 1$, the distance process $(D_n)_{n\geq 0}$ is monotone.

Unfortunately, the map Ψ_H is not monotone for $\sigma > 1$. Indeed, suppose that

$$\begin{gathered}\kappa = 3\,, \quad \sigma = 2\,, \quad m = 3\,, \quad \ell \geq 2\,,\\ \frac{2}{3} < s_1 < \frac{3}{4}\,, \quad \frac{3}{4} < s_2 < 1\,, \quad \frac{3}{4} < s_3 < 1\,,\\ \forall i \in \{\,1,2,3\,\} \quad \forall j \in \{\,1,\dots,\ell\,\} \qquad u_{i,j} \in \Big[\frac{p}{3}, 1 - \frac{2p}{3}\Big]\,,\end{gathered}$$

then

$$\Psi_H\begin{pmatrix}0\\2\\1\end{pmatrix} \;=\; \begin{pmatrix}2\\1\\1\end{pmatrix}\,, \qquad \Psi_H\begin{pmatrix}1\\2\\1\end{pmatrix} \;=\; \begin{pmatrix}1\\1\\1\end{pmatrix}\,.$$

This creates a serious complication. To get around this problem, we lump further the distance process in order to build the occupancy process. It turns out that the occupancy process is monotone even in the non neutral case. We define an order $\preceq$ on $\mathcal{P}^m_{\ell+1}$ as follows. Let $o = (o(0), \dots, o(\ell))$ and $o' = (o'(0), \dots, o'(\ell))$ belong to $\mathcal{P}^m_{\ell+1}$. We say that o is smaller than or equal to o', which we denote by $o \preceq o'$, if

$$\forall l \leq \ell \qquad o(0) + \cdots + o(l) \,\leq\, o'(0) + \cdots + o'(l)\,.$$

As shown in [3], the map $\mathcal{S}_O$ is non–increasing with respect to the occupancy distribution, i.e.,

$$\begin{aligned}&\forall o, o' \in \mathcal{P}^m_{\ell+1} \quad \forall s \in [0,1]\\ &\qquad\qquad o \preceq o' \quad \Rightarrow \quad \mathcal{S}_O(o,s) \,\geq\, \mathcal{S}_O(o',s)\,.\end{aligned}$$

Lemma 5.5 The map Ψ_O is non–decreasing with respect to the occupancy distributions, i.e.,

$$\forall o, o' \in \mathcal{P}^m_{\ell+1} \quad \forall r \in \mathcal{R} \qquad o \preceq o' \quad \Rightarrow \quad \Psi_O(o,r) \,\preceq\, \Psi_O(o',r)\,.$$

Proof. Let $r \in \mathcal{R}$ and let $o, o' \in \mathcal{P}^m_{\ell+1}$ be such that $o \preceq o'$. Using the monotonicity of the map $\mathcal{S}_O$, we have

$$\forall i \in \{1, \dots, m\} \qquad \mathcal{S}_O(o, r(i,1)) \geq \mathcal{S}_O(o', r(i,1)) .$$

This inequality and the monotonicity of the map $\mathcal{M}_H$ imply that

$$\begin{aligned}\forall i \in \{1, \dots, m\} \qquad & \mathcal{M}_H\big(\mathcal{S}_O(o, r(i,1)), r(i,2), \dots, r(i, \ell+1))\big) \\ & \qquad \geq \mathcal{M}_H\big(\mathcal{S}_O(o', r(i,1)), r(i,2), \dots, r(i, \ell+1))\big) .\end{aligned}$$

Therefore $\Psi_O(o, r) \leq \Psi_O(o', r)$ as requested. □

Corollary 5.6 The occupancy process $(O_n)_{n \geq 0}$ is monotone.

5.3 The FKG inequality

We consider here the product space $\{0, \dots, \ell\}^m$ equipped with the natural product order:

$$d \leq e \quad \Longleftrightarrow \quad \forall i \in \{1, \dots, m\} \quad d(i) \leq e(i) .$$

Definition 5.7 A probability measure μ on $\{0, \dots, \ell\}^m$ is said to have positive correlations if for any functions $f, g : \{0, \dots, \ell\}^m \to \mathbb{R}$ which are non–decreasing, we have

$$\sum_{d \in \{0, \dots, \ell\}^m} f(d) g(d)\, \mu(d) \geq \Big(\sum_{d \in \{0, \dots, \ell\}^m} f(d)\, \mu(d) \Big) \Big(\sum_{d \in \{0, \dots, \ell\}^m} g(d)\, \mu(d) \Big) .$$

The Harris inequality, or the FKG inequality in this context, says that any product probability measure on $\{0, \dots, \ell\}^m$ has positive correlations. The FKG inequality is in fact true for any product probability measure on a product of the interval $[0, 1]$ (see section 2.2 of Grimmett's book [11]). As far as correlations are concerned, there is not much to do with the original Wright–Fisher model, because its state space is not partially ordered. So we examine the distance process.

Proposition 5.8 Suppose that we are in the neutral case $\sigma = 1$. If the law of D_0 has positive correlations, then for any $n \geq 0$, the law of D_n also has positive correlations.

Proof. The Wright–Fisher model $(X_n)_{n \geq 0}$ can be seen as a probabilistic cellular automaton. Indeed, given the population $X_n = x$ at time n, the individuals $(X_{n+1}(i), 1 \leq i \leq m)$ of the population at time $n+1$ are independent. This still holds for the distance process. By corollary 5.4, the neutral distance process $(D_n)_{n\geq 0}$ is monotone. Monotone probabilistic cellular automata preserve the FKG inequality. This is explained in detail by Mezić [17] and it was first observed by Harris [12] at the very end of his article on continuous time processes. Because the argument is very short, we reproduce it here. Suppose that the initial law μ of D_0 has positive correlations. Let $f, g : \{\, 0, \dots, \ell \,\}^m \to \mathbb{R}$ be two non–decreasing functions. For any $d \in \{\, 0, \dots, \ell \,\}^m$, the conditional law of D_1 knowing that $D_0 = d$ is a product measure on $\{\, 0, \dots, \ell \,\}^m$, thus it satisfies the FKG inequality, whence

$$\forall d \in \{\, 0, \dots, \ell \,\}^m \qquad E\big(f(D_1)g(D_1) \,|\, D_0 = d\big) \;\geq\; E\big(f(D_1) \,|\, D_0 = d\big)\, E\big(g(D_1) \,|\, D_0 = d\big)\,.$$

We integrate the inequality with respect to the initial law μ:

$$\sum_{d \in \{\, 0,\dots,\ell \,\}^m} E\big(f(D_1)g(D_1) \,|\, D_0 = d\big)\, \mu(d) \;\geq\; \sum_{d \in \{\, 0,\dots,\ell \,\}^m} E\big(f(D_1) \,|\, D_0 = d\big) E\big(g(D_1) \,|\, D_0 = d\big)\, \mu(d)\,.$$

Since $(D_n)_{n\geq 0}$ is monotone, the maps

$$\begin{aligned} &d \in \{\, 0, \dots, \ell \,\}^m \mapsto E\big(f(D_1) \,|\, D_0 = d\big)\,,\\ &d \in \{\, 0, \dots, \ell \,\}^m \mapsto E\big(g(D_1) \,|\, D_0 = d\big)\,, \end{aligned}$$

are non–decreasing. By hypothesis, the initial law μ has positive correlations, therefore

$$\sum_{d \in \{\, 0,\dots,\ell \,\}^m} E\big(f(D_1) \,|\, D_0 = d\big) E\big(g(D_1) \,|\, D_0 = d\big)\, \mu(d) \;\geq\; \Big(\sum_{d \in \{\, 0,\dots,\ell \,\}^m} E\big(f(D_1) \,|\, D_0 = d\big)\, \mu(d)\Big)\Big(\sum_{d \in \{\, 0,\dots,\ell \,\}^m} E\big(g(D_1) \,|\, D_0 = d\big)\, \mu(d)\Big)\,.$$

The two above inequalities imply that the law of D_1 satisfies the FKG inequality. We conclude by iterating the argument. $\square$

6 Stochastic bounds

In this section, we take advantage of the monotonicity of the map Ψ_O to compare the process $(O_n)_{n\geq 0}$ with simpler processes.

6.1 Lower and upper processes

We shall construct a lower process $(O^\ell_n)_{n\geq 0}$ and an upper process $(O^1_n)_{n\geq 0}$ satisfying

$$\forall n\geq 0\qquad O^\ell_n\,\preceq\,O_n\,\preceq\,O^1_n\,.$$

Loosely speaking, the lower process evolves as follows. As long as there is no master sequence present in the population, the process $(O^1_n)_{n\geq 0}$ evolves exactly as the initial process $(O_n)_{n\geq 0}$. When the first master sequence appears, all the other chromosomes are set in the Hamming class 1, i.e., the process jumps to the state $(1,m-1,0,\dots,0)$. As long as the master sequence is present, the mutations on non master sequences leading to non master sequences are suppressed, and any mutation of a master sequence leads to a chromosome in the first Hamming class. The dynamics of the upper process is similar, except that the chromosomes distinct from the master sequence are sent to the last Hamming class ℓ instead of the first one. We shall next construct precisely these dynamics. We define two maps $\pi_\ell,\pi_1:\mathcal{P}^m_{\ell+1}\to\mathcal{P}^m_{\ell+1}$ by setting

$$\begin{aligned}\forall o\in\mathcal{P}^m_{\ell+1}\qquad \pi_\ell(o)\,&=\,\big(o(0),0,\dots,0,m-o(0)\big)\,,\\ \pi_1(o)\,&=\,\big(o(0),m-o(0),0,\dots,0\big)\,.\end{aligned}$$

Obviously,

$$\forall o\in\mathcal{P}^m_{\ell+1}\qquad \pi_\ell(o)\,\preceq\,o\,\preceq\,\pi_1(o)\,.$$

We denote by $\mathcal{W}^*$ the set of the occupancy distributions containing the master sequence, i.e.,

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

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

$$\mathcal{N}\,=\,\big\{\,o\in\mathcal{P}^m_{\ell+1}:o(0)=0\,\big\}\,.$$

Let Ψ_O be the coupling map defined in section 5.1 We define a lower map Ψ^ℓ_O by setting, for $o\in\mathcal{P}^m_{\ell+1}$ and $r\in\mathcal{R}$,

$$\Psi^\ell_O(o,r)\,=\,\begin{cases}\Psi_O(o,r) & \text{if } o\in\mathcal{N}\ \text{ and }\ \Psi_O(o,r)\notin\mathcal{W}^*\\ \pi_\ell\big(\Psi_O(o,r)\big) & \text{if } o\in\mathcal{N}\ \text{ and }\ \Psi_O(o,r)\in\mathcal{W}^*\\ \pi_\ell\big(\Psi_O(\pi_\ell(o),r)\big) & \text{if } o\in\mathcal{W}^*\end{cases}$$

Similarly, we define an upper map Ψ^1_O by setting, for $o\in\mathcal{P}^m_{\ell+1}$ and $r\in\mathcal{R}$,

$$\Psi^1_O(o,r)\,=\,\begin{cases}\Psi_O(o,r) & \text{if } o\in\mathcal{N}\ \text{ and }\ \Psi_O(o,r)\notin\mathcal{W}^*\\ \pi_1\big(\Psi_O(o,r)\big) & \text{if } o\in\mathcal{N}\ \text{ and }\ \Psi_O(o,r)\in\mathcal{W}^*\\ \pi_1\big(\Psi_O(\pi_1(o),r)\big) & \text{if } o\in\mathcal{W}^*\end{cases}$$

A direct application of lemma 5.5 yields that the map Ψ_O^ℓ is below the map Ψ_O and the map Ψ_O^1 is above the map Ψ_O in the following sense:

$$\forall r\in\mathcal{R}\quad\forall o\in\mathcal{P}^m_{\ell+1}\qquad \Psi_O^\ell(o,r)\,\preceq\,\Psi_O(o,r)\,\preceq\,\Psi_O^1(o,r)\,.$$

We define a lower process $(O_n^\ell)_{n\geq 0}$ and an upper process $(O_n^1)_{n\geq 0}$ with the help of the i.i.d. sequence $(R_n)_{n\geq 1}$ and the maps Ψ_O^ℓ, Ψ_O^1 as follows. Let $o\in\mathcal{P}^m_{\ell+1}$ be the starting point of the process. We set $O_0^\ell=O_0^1=o$ and

$$\forall n\geq 1\qquad O_n^\ell\,=\,\Psi_O^\ell\big(O_{n-1}^\ell,R_n\big)\,,\qquad O_n^1\,=\,\Psi_O^1\big(O_{n-1}^1,R_n\big)\,.$$

Proposition 6.1 Suppose that the processes $(O_n^\ell)_{n\geq 0}$, $(O_n)_{n\geq 0}$, $(O_n^1)_{n\geq 0}$, start from the same occupancy distribution o. We have

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

Proof. We prove the inequality by induction over $n\in\mathbb{N}$. For $n=0$ we have $O_0=O_0^\ell=O_0^1=o$. Suppose that the inequality has been proved at time $n\in\mathbb{N}$, so that $O_n^\ell\,\preceq\,O_n\,\preceq\,O_n^1$. By construction, we have

$$O_{n+1}^\ell\,=\,\Psi_O^\ell\big(O_n^\ell,R_n\big)\,,\quad O_{n+1}\,=\,\Psi_O\big(O_n,R_n\big)\,,\quad O_{n+1}^1\,=\,\Psi_O^1\big(O_n^1,R_n\big)\,.$$

We use the induction hypothesis and we apply lemma 5.5 to get

$$\Psi_O\big(O_n^\ell,R_n\big)\,\preceq\,\Psi_O\big(O_n,R_n\big)\,\preceq\Psi_O\big(O_n^1,R_n\big)\,.$$

Yet the map Ψ_O^ℓ is below the map Ψ_O and the map Ψ_O^1 is above the map Ψ_O, thus

$$\Psi_O^\ell\big(O_n^\ell,R_n\big)\,\preceq\,\Psi_O\big(O_n^\ell,R_n\big)\,,\qquad \Psi_O\big(O_n^1,R_n\big)\,\preceq\,\Psi_O^1\big(O_n^1,R_n\big).$$

Putting together these inequalities we obtain that $O_{n+1}^\ell\,\preceq\,O_{n+1}\,\preceq\,O_{n+1}^1$ and the induction step is completed. □

6.2 Dynamics of the bounding processes

We study next the dynamics of the processes $(O_n^\ell)_{n\geq 0}$ and $(O_n^1)_{n\geq 0}$ in $\mathcal{W}^*$. The computations are the same for both processes. Throughout the section, we fix θ to be either 1 or ℓ and we denote by $(O_n^\theta)_{n\geq 0}$ the corresponding process. For the process $(O_n^\theta)_{n\geq 0}$, the states

$$\mathcal{T}^\theta\,=\,\big\{\,o\in\mathcal{P}^m_{\ell+1}:o(0)\geq 1\text{ and }o(0)+o(\theta)<m\,\big\}$$

are transient, while the populations in $\mathcal{N}\cup\big(\mathcal{W}^*\setminus\mathcal{T}^\theta\big)$ form a recurrent class. Let us look at the transition mechanism of the process restricted to $\mathcal{W}^*\setminus\mathcal{T}^\theta$. Since

$$\mathcal{W}^*\setminus\mathcal{T}^\theta\,=\,\big\{\,o\in\mathcal{P}^m_{\ell+1}:o(0)\geq 1\text{ and }o(0)+o(\theta)=m\,\big\}\,,$$

we see that a state of $\mathcal{W}^* \setminus \mathcal{T}^\theta$ is completely determined by the first occupancy number, or equivalently the number of copies of the master sequence present in the population. From the previous observations, we conclude that, whenever $(O^\theta_n)_{n\geq 0}$ starts in $\mathcal{W}^* \setminus \mathcal{T}^\theta$, the dynamics of the number of master sequences $(O^\theta_n(0))_{n\geq 0}$ is markovian until the time of exit from $\mathcal{W}^* \setminus \mathcal{T}^\theta$. We denote by $(Z^\theta_n)_{n\geq 0}$ a Markov chain on $\{\,0,\dots,m\,\}$ with the following transition probabilities: for $h\in\{\,1,\dots,m\,\}$ and $k\in\{\,0,\dots,m\,\}$,

$$\forall n\geq 0\qquad P\big(Z^\theta_{n+1}=k\,|\,Z^\theta_n=h\big)\;=\;P\big(O^\theta_{n+1}(0)=k\,|\,O^\theta_n(0)=h\big)\,,$$

and for $h=0$ and $k\in\{\,0,\dots,m\,\}$,

$$\forall n\geq 0\qquad P\big(Z^\theta_{n+1}=k\,|\,Z^\theta_n=0\big)\;=\;\binom{m}{k}\,M_H(\theta,0)^k\big(1-M_H(\theta,0)\big)^{m-k}\,.$$

Let us denote by $p^\theta(h,k)$ the above transition probability and let us compute its value. We use the definition of the transition mechanism of $(O^\theta_n)_{n\geq 0}$ to get

$$p^\theta(h,k)\;=\;\sum_{i\in\{\,0,\dots,m\,\}}\sum_{j=0}^{i}p^\theta(h,i,j,k)$$

where $p^\theta(h,i,j,k)$ is given by

$$\begin{aligned}p^\theta(h,i,j,k)\;&=\;P\left(\begin{array}{c}i\text{ master sequences are selected}\,,\\ j\text{ master sequences do not mutate,}\\ k-j\text{ non master sequences}\\ \text{mutate into a master sequence}\end{array}\,\middle|\,Z^\theta_n=h\right)\\
&=\;\binom{m}{i}\frac{(\sigma h)^i(m-h)^{m-i}}{\big((\sigma-1)h+m\big)^m}\binom{i}{j}\,M_H(0,0)^j\big(1-M_H(0,0)\big)^{i-j}\\
&\qquad\qquad\times\binom{m-i}{k-j}\,M_H(\theta,0)^{k-j}\big(1-M_H(\theta,0)\big)^{m-i-k+j}\,.\end{aligned}$$

The Markov chain $(Z^\theta_n)_{n\geq 0}$ corresponds to the evolution of the number of master sequences in a Wright–Fisher model with two types, the master type having fitness σ and the other type having fitness 1, and with the following mutation matrix between the two types:

$$\begin{aligned}P(\text{the master type mutates into the non master type})\;&=\;1-M_H(0,0)\,,\\ P(\text{the non master type mutates into the master type})\;&=\;M_H(\theta,0)\,.\end{aligned}$$

We can also realize the Markov chain $(Z^\theta_n)_{n\geq 0}$ on our common probability space. We define two maps $\Xi^\ell,\Xi^1:\{\,0,\dots,m\,\}\to\mathcal{P}^m_{\ell+1}$ by setting

$$\forall i\in\{\,0,\dots,m\,\}\qquad \Xi^\ell(i)\;=\;(i,0,\dots,0,m-i\,)\,,\quad \Xi^1(i)\;=\;(i,m-i,0,\dots,0)\,.$$

Let $i \in \{ 0, \dots, m \}$ be the starting point of the process. We set $Z_0^\theta = i$ and

$$\forall n \geq 1 \qquad Z_n^\theta \,=\, \Psi_O^\theta\big(\Xi^\theta(Z_{n-1}^\theta), R_n\big)(0)\,.$$

This construction yields a Markov chain $(Z_n^\theta)_{n\geq 0}$ starting from i with the adequate transition matrix. Moreover the maps Ξ^ℓ, Ξ^1 are non–decreasing. By lemma 5.5, the map Ψ_O is also non–decreasing with respect to the occupancy distribution. We conclude that the above coupling is monotone.

Corollary 6.2 The Markov chain $(Z_n^\theta)_{n\geq 0}$ is monotone.

6.3 Invariant probability measures

Our goal is to estimate the law ν of the fraction of the master sequence in the population at equilibrium. The probability measure ν is the probability measure on the interval $[0,1]$ satisfying the following identities. For any function $f : [0,1] \to \mathbb{R}$,

$$\int_{[0,1]} f\,d\nu \,=\, \lim_{n\to\infty} f\Big(\frac{1}{m}N(X_n)\Big) \,=\, \sum_{x\in(\mathcal{A}^\ell)^m} f\Big(\frac{1}{m}N(x)\Big)\,\mu(x)$$

where μ is the invariant probability measure of the Markov chain $(X_n)_{n\geq 0}$ and $N(x)$ is the number of copies of the master sequence w^* present in the population x:

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

In fact, the probability measure ν is the image of μ through the map

$$x \in \big(\mathcal{A}^\ell\big)^m \mapsto \frac{1}{m}N(x) \in [0,1]\,.$$

We denote by μ_O^ℓ, μ_O, μ_O^1 the invariant probability measures of the Markov chains $(O_n^\ell)_{n\geq 0}$, $(O_n)_{n\geq 0}$, $(O_n^1)_{n\geq 0}$. The probability ν is also the image of μ_O through the map

$$o \in \mathcal{P}_{\ell+1}^m \mapsto \frac{1}{m}o(0) \in [0,1]\,.$$

Thus, for any function $f : [0,1] \to \mathbb{R}$,

$$\int_{[0,1]} f\,d\nu \,=\, \sum_{o\in\mathcal{P}_{\ell+1}^m} f\Big(\frac{o(0)}{m}\Big)\,\mu_O(o) \,=\, \lim_{n\to\infty} f\Big(\frac{1}{m}O_n(0)\Big)\,.$$

We fix now a non–decreasing function $f : [0,1] \to \mathbb{R}$ such that $f(0) = 0$. Proposition 6.1 yields the inequalities

$$\forall n \geq 0 \qquad f\Big(\frac{1}{m}O_n^\ell(0)\Big) \,\leq\, f\Big(\frac{1}{m}O_n(0)\Big) \,\leq\, f\Big(\frac{1}{m}O_n^1(0)\Big)\,.$$

Taking the expectation and sending n to ∞, we get

$$\sum_{o\in\mathcal{P}^m_{\ell+1}} f\Big(\frac{o(0)}{m}\Big)\,\mu^\ell_O(o)\,\leq\,\int_{[0,1]} f\,d\nu\,\leq\,\sum_{o\in\mathcal{P}^m_{\ell+1}} f\Big(\frac{o(0)}{m}\Big)\,\mu^1_O(o)\,.$$

We seek next estimates on the above sums. The strategy is the same for the lower and the upper sum. Thus we fix θ to be either 1 or ℓ and we study the invariant probability measure μ^θ_O. For the Markov chain $(O^\theta_n)_{n\geq 0}$, the states of $\mathcal{T}^\theta$ are transient, while the populations in $\mathcal{N}\cup\big(\mathcal{W}^*\setminus\mathcal{T}^\theta\big)$ form a recurrent class. Let o^θ_{exit} be the occupancy distribution having m chromosomes in the Hamming class θ:

$$\forall l\in\{\,0,\dots,\ell\,\}\qquad\qquad o^\theta_{\text{exit}}(l)\,=\,\begin{cases} m & \text{if } l=\theta\\ 0 & \text{otherwise}\end{cases}.$$

The process $(O^\theta_n)_{n\geq 0}$ always exits $\mathcal{W}^*\setminus\mathcal{T}^\theta$ at o^θ_{exit}. This allows us to estimate the invariant measure with the help of the following renewal result.

Proposition 6.3 Let $(X_n)_{n\geq 0}$ be a discrete time Markov chain with values in a finite state space $\mathcal{E}$ which is irreducible and aperiodic. Let μ be the invariant probability measure of the Markov chain $(X_n)_{n\geq 0}$. Let $\mathcal{W}^*$ be a subset of $\mathcal{E}$ and let e be a point of $\mathcal{E}\setminus\mathcal{W}^*$. Let f be a map from $\mathcal{E}$ to $\mathbb{R}$ which vanishes on $\mathcal{E}\setminus\mathcal{W}^*$. Let

$$\tau^*\,=\,\inf\,\big\{\,n\geq 0: X_n\in\mathcal{W}^*\,\big\}\,,\qquad \tau\,=\,\inf\,\big\{\,n\geq\tau^*: X_n=e\,\big\}\,.$$

We have

$$\sum_{x\in\mathcal{E}} f(x)\,\mu(x)\,=\,\frac{1}{E(\tau\,|\,X_0=e)}\,E\bigg(\sum_{n=\tau^*}^{\tau} f(X_n)\,\Big|\,X_0=e\bigg)\,.$$

This result is proved in detail in [3]. We apply the renewal result of proposition 6.3 to the process $(O^\theta_n)_{n\geq 0}$ restricted to $\mathcal{N}\cup\big(\mathcal{W}^*\setminus\mathcal{T}^\theta\big)$, the set $\mathcal{W}^*\setminus\mathcal{T}^\theta$, the occupancy distribution o^θ_{exit} and the function $o\mapsto f(o(0)/m)$. Setting

$$\begin{aligned}\tau^*\,&=\,\inf\,\big\{\,n\geq 0: O^\theta_n\in\mathcal{W}^*\,\big\}\,,\\ \tau\,&=\,\inf\,\big\{\,n\geq\tau^*: O^\theta_n=o^\theta_{\text{exit}}\,\big\}\,,\end{aligned}$$

we have

$$\sum_{o\in\mathcal{P}^m_{\ell+1}} f\Big(\frac{o(0)}{m}\Big)\,\mu^\theta_O(o)\,=\,\frac{E\bigg(\displaystyle\sum_{n=\tau^*}^{\tau} f\Big(\frac{O^\theta_n(0)}{m}\Big)\,\Big|\,O^\theta_0=o^\theta_{\text{exit}}\bigg)}{E\big(\tau\,|\,O^\theta_0=o^\theta_{\text{exit}}\big)}\,.$$

Yet, whenever the process $(O_n^\theta)_{n\geq 0}$ is in $\mathcal{W}^*\setminus\mathcal{T}^\theta$, the dynamics of the number of master sequences $(O_n^\theta(0))_{n\geq 0}$ is the same as the dynamics of the Markov chain $(Z_n^\theta)_{n\geq 0}$ defined at the end of section 6.2. Let τ_0 be the hitting time of 0, defined by

$$\tau_0 \,=\, \inf\big\{\, n\geq 0 : Z_n^\theta = 0\,\big\}\,.$$

The process $(O_n^\theta)_{n\geq 0}$ always exits $\mathcal{W}^*\setminus\mathcal{T}^\theta$ at o_{exit}^θ. Therefore τ coincides with the exit time of $\mathcal{W}^*\setminus\mathcal{T}^\theta$ after τ^*. Let $i\in\{\,1,\dots,m\,\}$. From the previous elements, we see that, conditionally on the event $\big\{\,O_{\tau^*}^\theta(0)=i\,\big\}$, the trajectory $\big(O_n^\theta(0),\,\tau^*\,\leq\,n\,\leq\,\tau\big)$ has the same law as the trajectory $\big(Z_n^\theta\,,0\leq n\leq\tau_0\big)$ starting from $Z_0^\theta=i$, whence

$$E\Big(\tau-\tau^*\,\big|\,O_{\tau^*}^\theta(0)=i\Big)\;=\;E\Big(\tau_0\,\big|\,Z_0^\theta=i\Big)\,,$$

$$E\bigg(\sum_{n=\tau^*}^{\tau} f\Big(\frac{O_n^\theta(0)}{m}\Big)\,\Big|\,O_{\tau^*}^\theta(0)=i\bigg)\;=\;E\bigg(\sum_{n=0}^{\tau_0} f\Big(\frac{Z_n^\theta}{m}\Big)\,\Big|\,Z_0^\theta=i\bigg)\,.$$

Conditioning with respect to $O_{\tau^*}^\theta(0)$ and reporting in the formula for the invariant probability measure μ_O^θ, we get

$$\sum_{o\in\mathcal{P}_{\ell+1}^m} f\Big(\frac{o(0)}{m}\Big)\,\mu_O^\theta(o)\;=$$
$$\frac{\displaystyle\sum_{i=1}^m E\bigg(\sum_{n=0}^{\tau_0} f\Big(\frac{Z_n^\theta}{m}\Big)\,\Big|\,Z_0^\theta=i\bigg)P\big(O_{\tau^*}^\theta(0)=i\,|\,O_0^\theta=o_{\text{exit}}^\theta\big)}{\displaystyle E\big(\tau^*\,|\,O_0^\theta=o_{\text{exit}}^\theta\big)+\sum_{i=1}^m E\big(\tau_0\,\big|\,Z_0^\theta=i\big)P\big(O_{\tau^*}^\theta(0)=i\,|\,O_0^\theta=o_{\text{exit}}^\theta\big)}\,.$$

We must next estimate these expectations. In section 7, we deal with the terms involving the Markov chain $(Z_n^\theta)_{n\geq 0}$. In section 8, we deal with the discovery time τ^*.

7 Approximating processes

This section is devoted to the study of the dynamics of the Markov chains $(Z_n^\ell)_{n\geq 0}$ and $(Z_n^1)_{n\geq 0}$. The estimates are carried out exactly in the same way for both Markov chains. As we said before, the Markov chain $(Z_n^\theta)_{n\geq 0}$ corresponds to the evolution of the number of master sequences in a Wright-Fisher model with two types. Throughout the section, we fix $\theta=1$ or $\theta=\ell$ and we remove θ from the notation in most places, writing simply p, Z_n instead of p^θ, Z_n^θ.

7.1 Large deviations for the transition matrix

We first recall a basic estimate for the binomial coefficients.

Lemma 7.1 For any $n \geq 1$, any $k \in \{0, \dots, n\}$, we have

$$\Big| \ln \frac{n!}{k!(n-k)!} + k \ln \frac{k}{n} + (n-k) \ln \frac{n-k}{n} \Big| \leq 2 \ln n + 3 .$$

Proof. The proof of this estimate is standard (see for instance [7]). Setting, for $n \in \mathbb{N}$, $\phi(n) = \ln n! - n \ln n + n$, we have

$$\ln \frac{n!}{k!(n-k)!} = \ln n! - \ln k! - \ln(n-k)!$$
$$= n \ln n - n + \phi(n) - \big(k \ln k - k + \phi(k)\big) - \big((n-k)\ln(n-k) - (n-k) + \phi(n-k)\big)$$
$$= -k \ln \frac{k}{n} - (n-k) \ln \frac{n-k}{n} + \phi(n) - \phi(k) - \phi(n-k) .$$

Comparing the discrete sum

$$\ln n! = \sum_{1 \leq k \leq n} \ln k$$

to the integral

$$\int_1^n \ln x \, dx ,$$

we see that $1 \leq \phi(n) \leq \ln n + 2$ for all $n \geq 1$. On one hand,

$$\phi(n) - \phi(k) - \phi(n-k) \leq \ln n ,$$

on the other hand,

$$\phi(n) - \phi(k) - \phi(n-k) \geq 1 - (\ln k + 2) - (\ln(n-k) + 2) \geq -3 - 2 \ln n$$

and we have the desired inequalities. $\square$

For $p, t \in [0,1]$, we define

$$I(p,t) = t \ln \frac{t}{p} + (1-t) \ln \frac{1-t}{1-p} .$$

For $t \notin [0,1]$, we set $I(p,t) = +\infty$. For $p = 0$, we have

$$\forall t \in [0,1] \qquad I(0,t) = \begin{cases} 0 & \text{if } t = 0 \\ +\infty & \text{if } t > 0 \end{cases} .$$

The function $I(p, \cdot)$ is the rate function governing the large deviations of the binomial distribution $\mathcal{B}(n,p)$ with parameters n and p. This is the simplest case of the famous Cramér theorem, which we recall next.

Lemma 7.2 Let X be a random variable whose law is the binomial distribution $\mathcal{B}(n,p)$ with parameters n and p. For $t\in[0,1]$, we have

$$\lim_{n\to\infty}\frac{1}{n}\ln P(X=\lfloor nt\rfloor)\,=\,-I(p,t)\,.$$

Using the estimate of lemma 7.1, we have even

$$\forall k\in\{\,0,\dots,n\,\}\qquad \ln P(X=k)\,=\,-nI\Big(p,\frac{k}{n}\Big)+O(2\ln n+3)\,.$$

We define two functions $f,F:[0,1]\to[0,1]$ by

$$f(r)\,=\,\frac{\sigma r}{(\sigma-1)r+1}\,,\qquad F(r)=e^{-a}f(r)\,.$$

We define a function $I_\ell:[0,1]^4\to[0,+\infty]$ by

$$I_\ell(r,s,\beta,t)\,=\,I\big(f(r),s\big)+s\,I\Big(M_H(0,0),\frac{\beta}{s}\Big)+(1-s)\,I\Big(M_H(\theta,0),\frac{t-\beta}{1-s}\Big)\,.$$

This function depends on ℓ through the mutation probabilities $M_H(0,0)$ and $M_H(\theta,0)$. Using the estimate provided by lemma 7.1 and the expression of p^θ, we see that

$$\begin{aligned}
&\forall h,i,j,k\in\{\,0,\dots,m\,\}\\
&\qquad\ln p(h,i,j,k)\,=\,-mI\Big(f\Big(\frac{h}{m}\Big),\frac{i}{m}\Big)-iI\Big(M_H(0,0),\frac{j}{i}\Big)\\
&\qquad\qquad\qquad\qquad-(m-i)I\Big(M_H(\theta,0),\frac{k-j}{m-i}\Big)+\Phi(h,i,j,k,m)\\
&\qquad\qquad\qquad\qquad=\,-mI_\ell\Big(\frac{h}{m},\frac{i}{m},\frac{j}{m},\frac{k}{m}\Big)+\Phi(h,i,j,k,m)\,,
\end{aligned}$$

where the error term $\Phi(h,i,j,k,m)$ satisfies

$$\forall h,i,j,k\in\{\,0,\dots,m\,\}\qquad \big|\Phi(h,i,j,k,m)\big|\,\le\,6\ln m+9\,.$$

We consider the regime where

$$\ell\to+\infty\,,\qquad m\to+\infty\,,\qquad q\to 0\,,\qquad \ell q\to a\in]0,+\infty[\,.$$

In this regime, we have for $\theta=1$ or $\theta=\ell$,

$$M_H(0,0)\to e^{-a}\,,\qquad M_H(\theta,0)\to 0\,,$$

so that, for $r,s,\beta,t\in[0,1]^4$,

$$I_\ell(r,s,\beta,t)\,\to\,\begin{cases}I(r,s,t) & \text{if }\beta=t\\ +\infty & \text{if }\beta\neq t\end{cases}\,,$$

where the function $I(r,s,t)$ is given by

$$\forall r,s,t\in[0,1]^3\qquad I(r,s,t)\,=\,I\big(f(r),s\big)+s\,I\Big(e^{-a},\frac{t}{s}\Big)\,.$$

Proposition 7.3 We define a function V_1 on $[0,1]\times[0,1]$ by

$$\forall r,t\in[0,1]\qquad V_1(r,t)\,=\,\inf\,\Big\{\,I\big(r,s,t\big):s\in[0,1]\,\Big\}\,.$$

The one step transition probabilities of $(Z_n)_{n\geq 0}$ satisfy the large deviation principle governed by V_1: for any subset U of $[0,1]$, we have, for any $n\geq 0$,

$$-\inf\,\big\{\,V_1(r,t):t\in\overset{o}{U}\,\big\}\,\leq\,\liminf_{\substack{\ell,m\to\infty,\,q\to 0\\ \ell q\to a}}\frac{1}{m}\ln P\big(Z_{n+1}\in mU\,|\,Z_n=\lfloor rm\rfloor\big)\,,$$

$$\limsup_{\substack{\ell,m\to\infty,\,q\to 0\\ \ell q\to a}}\frac{1}{m}\ln P\big(Z_{n+1}\in mU\,|\,Z_n=\lfloor rm\rfloor\big)\,\leq\,-\inf\,\big\{\,V_1(r,t):t\in\overline{U}\,\big\}\,.$$

Proof. Let $r\in[0,1]$ and let U be a subset of $[0,1]$. We have, for any $n\geq 0$,

$$\begin{aligned}P\big(Z_{n+1}\in mU\,|\,Z_n=\lfloor rm\rfloor\big)\,&=\,\sum_{k\in mU\cap\{\,0,\dots,m\,\}}p\big(\lfloor rm\rfloor,k\big)\\&=\,\sum_{\substack{k\in\{\,0,\dots,m\,\}\\ k\in mU}}\,\sum_{i=0}^{m}\sum_{j=0}^{i}p\big(\lfloor rm\rfloor,i,j,k\big)\,.\end{aligned}$$

From the previous inequalities, we have

$$\begin{aligned}&P\big(Z_{n+1}\in mU\,|\,Z_n=\lfloor rm\rfloor\big)\\&\quad\leq\,(m+1)^3\max\Big\{\,p\big(\lfloor rm\rfloor,i,j,k\big):0\leq i\leq m,\,0\leq j\leq i,\,k\in mU\,\Big\}\\&\qquad\leq\,m^{11}\exp\Big(-m\min\Big\{\,I_\ell\Big(\frac{\lfloor rm\rfloor}{m},\frac{i}{m},\frac{j}{m},\frac{k}{m}\Big):\\&\qquad\qquad\qquad 0\leq i\leq m,\,0\leq j\leq i,\,k\in mU\,\Big\}\Big)\,.\end{aligned}$$

For each $m\geq 1$, let i_m,j_m,k_m be three integers in $\{\,0,\dots,m\,\}$ which realize the above minimum. By compactness of $[0,1]$, up to the extraction of a subsequence, we can suppose that, as m goes to ∞,

$$\frac{i_m}{m}\to s\,,\qquad\frac{j_m}{m}\to\beta\,,\qquad\frac{k_m}{m}\to t\,.$$

If $\beta < t$ then

$$\limsup_{\substack{\ell,m\to\infty,\,q\to 0\\ \ell q\to a}} -I_\ell\Big(\frac{\lfloor rm\rfloor}{m},\frac{i_m}{m},\frac{j_m}{m},\frac{k_m}{m}\Big) \le$$

$$\limsup_{\substack{\ell,m\to\infty,\,q\to 0\\ \ell q\to a}} -\Big(1-\frac{i_m}{m}\Big)\, I\Bigg(M_H(\theta,0),\frac{\dfrac{k_m}{m}-\dfrac{j_m}{m}}{1-\dfrac{i_m}{m}}\Bigg) = -\infty\,,$$

because

$$\limsup_{\substack{\ell,m\to\infty,\,q\to 0\\ \ell q\to a}} -\frac{k_m-j_m}{m}\ \ln\frac{\dfrac{k_m-j_m}{m}}{\Big(1-\dfrac{i_m}{m}\Big)M_H(\theta,0)} = -\infty\,.$$

Thus we need only to consider the case where $\beta = t$. We have then

$$\limsup_{\substack{\ell,m\to\infty,\,q\to 0\\ \ell q\to a}} -I_\ell\Big(\frac{\lfloor rm\rfloor}{m},\frac{i_m}{m},\frac{j_m}{m},\frac{k_m}{m}\Big) \le -I\big(f(r),s\big) - s\,I\Big(e^{-a},\frac{t}{s}\Big)\,.$$

This implies the large deviation upper bound:

$$\limsup_{\substack{\ell,m\to\infty,\,q\to 0\\ \ell q\to a}} \frac{1}{m}\ln P\big(Z_{n+1}\in mU\,|\,Z_n=\lfloor rm\rfloor\big)$$

$$\le -\inf\Big\{\,I\big(r,s,t\big) : s\in[0,1], t\in\overline{U}\,\Big\}\,.$$

Conversely, let $s,t\in[0,1]$. We have

$$\begin{aligned}
P\big(Z_{n+1}=\lfloor tm\rfloor\,|\,Z_n=\lfloor rm\rfloor\big) &\ge p\big(\lfloor rm\rfloor,\lfloor sm\rfloor,\lfloor tm\rfloor,\lfloor tm\rfloor\big)\\
&\ge \frac{1}{m^7}\exp\Big(-mI_\ell\Big(\frac{\lfloor rm\rfloor}{m},\frac{\lfloor sm\rfloor}{m},\frac{\lfloor tm\rfloor}{m},\frac{\lfloor tm\rfloor}{m}\Big)\Big)\\
\ge \frac{1}{m^7}\exp\Bigg(&-mI\Big(f\Big(\frac{\lfloor rm\rfloor}{m}\Big),\frac{\lfloor sm\rfloor}{m}\Big) - \lfloor sm\rfloor\, I\Big(M_H(0,0),\frac{\lfloor tm\rfloor}{\lfloor sm\rfloor}\Big)\\
&\qquad - \big(m-\lfloor sm\rfloor\big)\,\ln\frac{1}{1-M_H(\theta,0)}\Bigg)\,.
\end{aligned}$$

Taking ln and sending m,ℓ to ∞, we obtain

$$\liminf_{\substack{\ell,m\to\infty,\,q\to 0\\ \ell q\to a}} \frac{1}{m}\ln P\big(Z_{n+1}=\lfloor tm\rfloor\,|\,Z_n=\lfloor rm\rfloor\big) \ge -I\big(r,s,t\big)\,.$$

Suppose now that t belongs to $\overset{o}{U}$, the interior of U. For m large enough, the integer $\lfloor tm \rfloor$ belongs to mU. From the previous estimate, we have

$$\liminf_{\substack{\ell,m\to\infty,\,q\to 0\\ \ell q\to a}} \frac{1}{m}\ln P\big(Z_{n+1}\in mU \,|\, Z_n=\lfloor rm\rfloor\big) \;\geq\; -I\big(r,s,t\big)\,.$$

Optimizing over s,t, we get the large deviation lower bound:

$$\begin{aligned}\liminf_{\substack{\ell,m\to\infty,\,q\to 0\\ \ell q\to a}} \frac{1}{m}\ln P\big(Z_{n+1}\in mU \,|\, Z_n=\lfloor rm\rfloor\big)&\\ \geq -\inf\,\Big\{\,I\big(r,s,t\big) : s\in[0,1], t\in\overset{o}{U}\Big\}\,.&\end{aligned}$$

This finishes the proof of the large deviation principle. □

Proceeding in the same way, we can prove that the l–step transition probabilities satisfy a large deviation principle. For $l\geq 1$, we define a function V_l on $[0,1]\times[0,1]$ by

$$\begin{aligned}V_l(r,t) \;=\; \inf\,\Big\{\,\sum_{k=0}^{l-1} I\big(\rho_k,\gamma_k,\rho_{k+1}\big) : \rho_0=r,\ \rho_l=t,&\\ \rho_k,\gamma_k\in[0,1] \text{ for } 0\leq k<l\,\Big\}\,.&\end{aligned}$$

Corollary 7.4 For $l\geq 1$, the l–step transition probabilities of $(Z_n)_{n\geq 0}$ satisfy the large deviation principle governed by V_l: for any subset U of $[0,1]$, any $r\in[0,1]$, we have, for any $n\geq 0$,

$$-\inf\,\big\{\,V_l(r,t) : t\in\overset{o}{U}\,\big\} \;\leq\; \liminf_{\substack{\ell,m\to\infty,\,q\to 0\\ \ell q\to a}} \frac{1}{m}\ln P\big(Z_{n+l}\in mU \,|\, Z_n=\lfloor rm\rfloor\big)\,,$$

$$\limsup_{\substack{\ell,m\to\infty,\,q\to 0\\ \ell q\to a}} \frac{1}{m}\ln P\big(Z_{n+l}\in mU \,|\, Z_n=\lfloor rm\rfloor\big) \;\leq\; -\inf\,\big\{\,V_l(r,t) : t\in\overline{U}\,\big\}\,.$$

The rate function for the one step transition probabilities is given by

$$I\big(r,s,t\big) \;=\; I\big(f(r),s\big)+s\,I\Big(e^{-a},\frac{t}{s}\Big)\,,$$

we see that

$$I\big(r,s,t\big) \;=\; 0 \qquad\Longleftrightarrow\qquad s=f(r)\,,\quad e^{-a}=\frac{t}{s}\,.$$

Therefore the Markov chain $(Z_n)_{n\geq 0}$ can be considered as a random perturbation of the dynamical system associated to the map F:

$$z_0 \in [0,1]\,, \qquad \forall n\geq 1 \quad z_n \,=\, F(z_{n-1})\,.$$

Let us set

$$\rho^*(a) \,=\, \rho(e^{-a},0) \,=\, \begin{cases} \dfrac{\sigma e^{-a}-1}{\sigma-1} & \text{if } \sigma e^{-a}>1 \\ 0 & \text{if } \sigma e^{-a}\leq 1 \end{cases}$$

If $\sigma e^{-a}\leq 1$ the function F admits only one fixed point, O, which is stable. If $\sigma e^{-a}>1$ the function F admits two fixed points, O, which is unstable, and $\rho^*(a)$ which is stable.

The natural strategy to study the Markov chain $(Z_n)_{n\geq 0}$ is to use the Freidlin–Wentzell theory [9]. This theory has been initially developed in a continuous setting, for brownian perturbations of differential equations. This theory is quite robust and it can be applied well beyond this initial framework. A discrete version has been worked out by Kifer [13, 14]. Unfortunately, our model does not satisfy the initial hypothesis employed by Kifer. Indeed we have a degeneracy at 0 and the convergence in the large deviation principle for the transition probabilities p is not uniform with respect to the starting point. In addition, the rate function might be infinite, for instance

$$P\big(Z_{n+1}>\frac{m}{2}\,|\,Z_n=0\big) \,=\, \sum_{k>m/2}\binom{m}{k}\,M_H(\theta,0)^k\big(1-M_H(\theta,0)\big)^{m-k}\,,$$

and since $M_H(\theta,0)$ goes to 0 as ℓ goes to ∞, we have

$$\lim_{\substack{\ell,m\to\infty,\,q\to 0\\ \ell q\to a}} \frac{1}{m}\ln P\big(Z_{n+1}>\frac{m}{2}\,|\,Z_n=0\big) \,=\, -\infty\,.$$

For these reasons, the Markov chain $(Z_n)_{n\geq 0}$ does not fit into the model studied by Kifer. Undoubtedly, it is possible to relax somehow the hypothesis of Kifer's model in order to include our case. However this would require a full check of the proofs and the final presentation would not be convenient. Moreover we do not need the general results concerning the perturbed dynamics in the presence of a finite number of attractors. In our case, we have one stable fixed point and one unstable fixed point. Our situation is similar to the case studied by Darden [4] and by Morrow and Sawyer [18]. In Darden's case, the mutations are not taken into account, so the model is slightly too restrictive. In their model, Morrow and Sawyer assume a lower bound on the variance of the process which is uniform with respect to the starting point. This condition is violated here, because the variance vanishes at the unstable fixed point. In fact, we could possibly use the result of Morrow and Sawyer on a subset of the form $\{\lfloor \delta m \rfloor, \dots, m\}$, where $\delta > 0$ is fixed, but we would still need to study the process in the vicinity of 0. The other drawback is also that, apart from the hypothesis on the variance, their model is a bit too general for our purposes (they consider a Markov chain evolving in a bounded subset of $\mathbb{R}^d$). It appears that, in any case, if we try to apply the results of Kifer or of Morrow and Sawyer, we have to make a specific study of our process in the vicinity of the unstable fixed point 0, because we must control the hitting time of 0, uniformly over the starting point. Therefore we choose to make a self–contained proof of the estimates we need. We take advantage of the specific dynamics to make a simpler proof, namely we use the discrete structure of $\{0, \dots, m\}$ and the monotonicity of the model, instead of relying on compactness and a condition of uniform convergence. The crucial quantity to analyze the dynamics is the following cost function V. We define, for $s, t \in [0,1]$,

$$V(s,t) = \inf_{l \geq 1} V_l(s,t) =$$

$$\inf_{l \geq 1} \inf \Big\{ \sum_{k=0}^{l-1} I\big(\rho_k, \gamma_k, \rho_{k+1}\big) : \rho_0 = s,\ \rho_l = t,\ \rho_k, \gamma_k \in [0,1] \text{ for } 0 \leq k < l \Big\}.$$

Lemma 7.5 Suppose that $\sigma e^{-a} > 1$. For $s, t \in [0,1]$, we have $V(s,t) = 0$ if and only if
- either $s = t = 0$,
- or there exists $l \geq 1$ such that $t = F^l(s)$,
- or $s \neq 0, t = \rho^*(a)$.

Proof. Throughout the proof we write ρ^* instead of $\rho^*(a)$. Let $s, t \in [0,1]$ be such that $V(s,t) = 0$. Suppose first that $s = 0$. Since $I(0, \gamma, \rho) = +\infty$ unless $\gamma = \rho = 0$, then any sequence $(\rho_0, \gamma_0, \dots, \gamma_l)$ such that $\rho_0 = s = 0$ and

$$\sum_{k=0}^{l-1} I\big(\rho_k, \gamma_k, \rho_{k+1}\big) < +\infty$$

has to be the null sequence, so that necessarily $t = 0$. We suppose next that $s > 0$. For each $n \geq 1$, let $(\rho_0^n, \gamma_0^n, \dots, \rho_{l(n)}^n)$ be a sequence of length $l(n)$ in $[0,1]$ such that

$$\rho_0^n = s,\ \rho_{l(n)}^n = t, \qquad \sum_{k=0}^{l(n)-1} I\big(\rho_k^n, \gamma_k^n, \rho_{k+1}^n\big) \leq \frac{1}{n}.$$

We consider two cases. If the sequence $(l(n))_{n\geq 1}$ is bounded, then we can extract a subsequence

$$\big(\rho_0^{\phi(n)},\gamma_0^{\phi(n)},\dots,\rho_{l(\phi(n))}^{\phi(n)}\big)$$

such that $l(\phi(n))=l$ does not depend on n and for any $k\in\{\,0,\dots,l-1\,\}$, the following limits exist:

$$\lim_{n\to\infty}\rho_k^{\phi(n)}\,=\,\rho_k\,,\qquad\lim_{n\to\infty}\gamma_k^{\phi(n)}\,=\,\gamma_k\,.$$

The map I being continuous, we have then

$$\forall k\in\{\,0,\dots,l-1\,\}\qquad I\big(\rho_k,\gamma_k,\rho_{k+1}\big)\,=\,0\,,$$

whence

$$\forall k\in\{\,0,\dots,l\,\}\qquad\rho_k\,=\,F^k(\rho_0)\,.$$

Since in addition $\rho_0=s$ and $\rho_l=t$, we conclude that $t=F^l(s)$. Suppose next that the sequence $(l(n))_{n\geq 1}$ is not bounded. Our goal is to show that $t=\rho^*$. Using Cantor's diagonal procedure, we can extract a subsequence

$$\big(\rho_0^{\phi(n)},\gamma_0^{\phi(n)},\dots,\rho_{l(\phi(n))}^{\phi(n)}\big)$$

such that, for any $k\geq 0$, the following limits exist:

$$\lim_{n\to\infty}\rho_k^{\phi(n)}\,=\,\rho_k\,,\qquad\lim_{n\to\infty}\gamma_k^{\phi(n)}\,=\,\gamma_k\,.$$

The map I being continuous, we have then

$$\forall k\geq 0\qquad I\big(\rho_k,\gamma_k,\rho_{k+1}\big)\,=\,0\,,$$

whence

$$\forall k\geq 0\qquad\rho_k\,=\,F^k(\rho_0)\,.$$

Let $\varepsilon>0$. We have

$$I\big(\rho^*,f(\rho^*),\rho^*\big)\,=\,0\,.$$

The map I being continuous, there exists a neighborhood U of ρ^* such that

$$\forall\rho\in U\qquad V_1(\rho^*,\rho)\,\leq\,I\big(\rho^*,f(\rho^*),\rho\big)\,<\,\varepsilon\,.$$

Since $s>0$,

$$\lim_{n\to\infty}F^n(s)\,=\,\rho^*$$

and

$$\exists h\geq 1\qquad F^h(s)\in U\,.$$

In particular,

$$\lim_{n\to\infty} \rho_h^{\phi(n)} \;=\; F^h(s) \;\in\; U\,,$$

so that, for n large enough,

$$\rho_h^{\phi(n)} \;\in\; U\,,$$

and

$$V(\rho^*,t) \;\leq\; V_1\big(\rho^*,\rho_h^{\phi(n)}\big)+V\big(\rho_h^{\phi(n)},t\big) \;\leq\; \varepsilon+\frac{1}{n}\,.$$

Letting successively n go to ∞ and ε go to 0 we obtain that $V(\rho^*,t)=0$. Let $\delta\in]0,\rho^*/2[$ and let U be the neighborhood of ρ^* given by

$$U\;=]\rho^*-\delta,\rho^*+\delta[\,.$$

Let α be the infimum

$$\alpha\;=\;\inf\big\{\,I\big(\rho_0,\gamma_0,\rho_1\big):\rho_0\in\overline{U},\gamma_0\in[0,1],\rho_1\not\in U\,\big\}\,.$$

The function I is continuous on the compact set $\overline{U}\times[0,1]\times\big([0,1]\setminus U\big)$, hence

$$\exists(\rho_0^*,\gamma_0^*,\rho_1^*)\in\overline{U}\times[0,1]\times\big([0,1]\setminus U\big)\qquad \alpha\;=\;I\big(\rho_0^*,\gamma_0^*,\rho_1^*\big)\,.$$

Since F is non–decreasing and continuous, we have

$$F\big(\overline{U}\big)\;=\;F\big([\rho^*-\delta,\rho^*+\delta]\big)\;=\;\big[F(\rho^*-\delta),F(\rho^*+\delta)\big]\,.$$

Moreover we have

$$\rho^*-\delta\;<\;F(\rho^*-\delta)\;\leq\;F(\rho^*+\delta)\;<\;\rho^*+\delta\,.$$

Thus $F(\overline{U})\;\subset\;U$ and necessarily $\rho_1^*\neq F(\rho_0^*)$ and $\alpha>0$. It follows that any sequence $(\rho_0,\gamma_0,\dots,\rho_l)$ such that

$$\rho_0\in U\,,\qquad \sum_{k=0}^{l-1}I\big(\rho_k,\gamma_k,\rho_{k+1}\big)\;<\;\alpha$$

is trapped in U. As a consequence, a point t satisfying $V(\rho^*,t)=0$ must belong to U. This is true for any $\delta>0$, hence for any neighborhood of ρ^*, thus $t=\rho^*$. □

We shall derive estimates in the regime where

$$\ell\to+\infty\,,\qquad m\to+\infty\,,\qquad q\to0\,,\qquad \ell q\to a\in]0,+\infty[\,.$$

Several inequalities will be valid only when the parameters are sufficiently close to their limits. We will say that a property holds asymptotically to express that it holds for ℓ,m large enough and q small enough.

7.2 Persistence time

In this section, we will estimate the expected hitting time τ_0 starting from a point of $\{ 1,\dots,m \}$. This quantity approximates the persistence time of the master sequence w^*. We recall that

$$\rho^*(a) \;=\; \rho(e^{-a},0) \;=\; \begin{cases} \dfrac{\sigma e^{-a}-1}{\sigma-1} & \text{if } \sigma e^{-a} > 1 \\ 0 & \text{if } \sigma e^{-a} \leq 1 \end{cases}$$

and that the functions $f,F:[0,1]\to[0,1]$ are given by

$$f(r) \;=\; \frac{\sigma r}{(\sigma-1)r+1}\,, \qquad F(r)=e^{-a}f(r)\,.$$

Proposition 7.6 Let $a\in]0,+\infty[$ and let $i\in\{ 1,\dots,m \}$. The expected hitting time τ_0 of 0 starting from i satisfies

$$\lim_{\substack{\ell,m\to\infty\\ q\to 0,\,\ell q\to a}} \frac{1}{m}\ln E(\tau_0\,|\,Z_0=i) \;=\; V(\rho^*(a),0)\,.$$

Proof. Before proceeding to the proof, let us explain the general strategy, which comes directly from the theory of Freidlin and Wentzell. To obtain the upper bound on the persistence time, we show that, starting from any point in $\{ 1,\dots,m \}$, the probability to reach a neighborhood of 0 in a finite number of steps is larger than

$$\exp\Big(-mV(\rho^*,0)-m\varepsilon\Big)\,.$$

This way we can bound from above τ_0 by a geometric law with this parameter (see lemma 7.7). To obtain the lower bound on the persistence time, we first show in lemma 7.8 that, starting from any point, the process has a reasonable probability of reaching any neighborhood of ρ^* before visiting 0. This estimate is quite tedious, because the process might start from $Z_0=1$, which is close to the unstable fixed point of F. Since we need to control the hitting time of 0 starting from any point, such an estimate seems to be indispensable and it cannot be done in the more general situations considered by Kifer [13] or Morrow and Sawyer [18] without adding some extra assumptions. So we give a lower bound on the probability of following the iterates of a discrete approximation of F. With a Poisson fluctuation, the process jumps away from 0, then, because F is expanding in the neighborhood of 0, it reaches the point ηm after $\ln m$ steps, for some $\eta>0$, and with a finite number of steps, it lands in a neighborhood of ρ^*. We then study the excursions of the process outside a neighborhood of 0 and ρ^*. Whenever the process is outside such a neighborhood, it reenters the neighborhood in a finite number of steps with probability larger than $1-\exp(-cm)$ for some $c>0$ depending on the neighborhood. Thus the process is very unlikely to stay a long time outside a neighborhood of the two attractors $\{\,0,\rho^*\,\}$. In fact, the length of these excursions is bounded by a constant, up to negligible events. We consider the hitting time τ_δ of the δ–neighborhood of 0. Obviously we have $\tau_0 \geq \tau_\delta$. We look then at the portion of the trajectory starting at the last visit to the neighborhood of ρ^* before reaching a neighborhood of 0. Such an excursion occurs at a given time with probability of order

$$\exp\Big(-mV(\rho^*,0)+m\varepsilon\Big)\,,$$

therefore it is unlikely to occur before time $\exp\big(mV(\rho^*,0)-m\varepsilon\big)$. We start now with the implementation of this scheme. Throughout the proof we write ρ^* instead of $\rho^*(a)$. We start by proving an upper bound on the hitting time. The next argument works in both cases $\sigma e^{-a}\leq 1$ and $\sigma e^{-a}>1$. In the case $\sigma e^{-a}\leq 1$, we have $\rho^*=0$ and $V(\rho^*,0)=0$, and the proof becomes simpler, there is no need to consider a path from ρ^* to 0. Let $\varepsilon>0$. We have

$$I\big(\rho^*,f(\rho^*),\rho^*\big)\,=\,0\,.$$

The map I being continuous, there exists $\delta>0$ such that

$$\forall \rho\in]\rho^*-\delta,\rho^*+\delta[\qquad I\big(\rho,f(\rho),\rho^*\big)\,<\,\varepsilon\,.$$

Moreover

$$\lim_{n\to\infty}F^n(1)\,=\,\rho^*\,.$$

Thus

$$\exists h\geq 1\qquad F^h(1)\in]\rho^*-\delta,\rho^*+\delta[\,.$$

Let $l\geq 1$ and let $(\rho_0,\gamma_0,\dots,\rho_l)$ be a sequence in $[0,1]$ such that

$$\rho_0=\rho^*,\ \rho_l=0,\qquad \sum_{k=0}^{l-1}I\big(\rho_k,\gamma_k,\rho_{k+1}\big)\,\leq\,V(\rho^*,0)+\varepsilon\,.$$

We consider the sequence obtained by concatenating the two previous sequences:

$$\begin{gathered}t_0=1,s_0=f(1),t_1=F(1),\dots,t_{h-1}=F^{h-1}(1),s_{h-1}=f(t_{h-1}),\\ t_h=F^h(1),s_h=f(t_h),t_{h+1}=\rho^*,s_{h+1}=\gamma_0,\\ t_{h+2}=\rho_1,\cdots,t_{h+l}=\rho_{l-1},s_{h+l}=\gamma_{l-1},t_{h+l+1}=\rho_l=0\,.\end{gathered}$$

We set $j = h + l + 1$. This sequence satisfies

$$t_0 = 1,\, t_j = 0, \qquad \sum_{k=0}^{j-1} I\big(t_k, s_k, t_{k+1}\big) \;\leq\; V(\rho^*, 0) + 3\varepsilon\,.$$

We have then

$$P(Z_j = 0 \,|\, Z_0 = m) \;\geq\; \prod_{k=0}^{j-1} p\big(\lfloor mt_k \rfloor, \lfloor ms_k \rfloor, \lfloor mt_{k+1} \rfloor, \lfloor mt_{k+1} \rfloor\big)\,.$$

Taking ln, sending m to ∞ and using the estimate on the transition probabilities obtained in the proof of proposition 7.3, we have

$$\liminf_{\substack{\ell, m \to \infty,\, q \to 0 \\ \ell q \to a}} \frac{1}{m} \ln P(Z_j = 0 \,|\, Z_0 = m) \;\geq\; -\sum_{k=0}^{j-1} I\big(t_k, s_k, t_{k+1}\big)$$
$$\geq\; -V(\rho^*, 0) - 3\varepsilon\,.$$

Thus, asymptotically, we have

$$P(Z_j = 0 \,|\, Z_0 = m) \;\geq\; \exp\Big(- mV(\rho^*, 0) - 4m\varepsilon \Big)\,.$$

Using the monotonicity of the Markov chain $(Z_n)_{n \geq 0}$, we conclude that, asymptotically,

$$\forall i \in \{\, 1, \dots, m \,\} \qquad P(Z_j = 0 \,|\, Z_0 = i) \;\geq\; \exp\Big(- mV(\rho^*, 0) - 4m\varepsilon \Big)\,.$$

We have thus a lower bound on the probability of reaching 0 in j steps starting from any point in $\{\, 1, \dots, m \,\}$. For any $n \geq 0$, we have, using the Markov property,

$$\begin{aligned}
P\big(\tau_0 > (n+1)j \,|\, Z_0 = m\big) \;&=\; \sum_{h=1}^{m} P\big(\tau_0 > (n+1)j,\, Z_{nj} = h \,|\, Z_0 = m\big) \\
&= \sum_{h=1}^{m} P\big(\tau_0 > nj,\, Z_{nj} = h,\, Z_{nj+1} \neq 0,\, \dots Z_{(n+1)j} \neq 0 \,|\, Z_0 = m\big) \\
&= \sum_{h=1}^{m} P\big(Z_{nj+1} \neq 0,\, \dots Z_{(n+1)j} \neq 0 \,|\, \tau_0 > nj,\, Z_{nj} = h,\, Z_0 = m\big) \\
&\qquad\qquad \times P\big(\tau_0 > nj,\, Z_{nj} = h \,|\, Z_0 = m\big) \\
&= \sum_{h=1}^{m} P\big(\tau_0 > j \,|\, Z_0 = h\big)\, P\big(\tau_0 > nj,\, Z_{nj} = h \,|\, Z_0 = m\big) \\
&\leq\; \Big(1 - \exp\big(- mV(\rho^*, 0) - 4m\varepsilon\big)\Big)\, P\big(\tau_0 > nj \,|\, Z_0 = m\big)\,.
\end{aligned}$$

Iterating this inequality, we obtain the following result.

Lemma 7.7 For any $\varepsilon > 0$, there exists $j \geq 1$ such that

$$\forall n \geq 0 \qquad P\big(\tau_0 > nj \,|\, Z_0 = m\big) \;\leq\; \Big(1 - \exp\big(-mV(\rho^*, 0) - 4m\varepsilon\big)\Big)^n .$$

It follows that

$$\begin{aligned} E(\tau_0 \,|\, Z_0 = m) \;&=\; \sum_{k\geq 1} P\big(\tau_0 \geq k \,|\, Z_0 = m\big) \\ &= \sum_{n\geq 0} \sum_{k=nj+1}^{(n+1)j} P\big(\tau_0 \geq k \,|\, Z_0 = m\big) \\ &\leq \sum_{n\geq 0} j\, P\big(\tau_0 > nj \,|\, Z_0 = m\big) \;\leq\; j \exp\big(mV(\rho^*, 0) + 4m\varepsilon\big) \end{aligned}$$

whence

$$\limsup_{\substack{\ell, m\to\infty,\, q\to 0 \\ \ell q \to a}} \frac{1}{m} \ln E(\tau_0 \,|\, Z_0 = m) \;\leq\; V(\rho^*, 0) + 4\varepsilon .$$

Letting ε go to 0 yields the desired upper bound. We prove next a lower bound on the hitting time. If $\sigma e^{-a} \leq 1$, then $\rho^* = 0$, $V(\rho^*, 0) = 0$, and obviously

$$\liminf_{\substack{\ell, m\to\infty,\, q\to 0 \\ \ell q \to a}} \frac{1}{m} \ln E(\tau_0 \,|\, Z_0 = m) \;\geq\; V(\rho^*, 0) \;=\; 0 .$$

Thus we need only to consider the case $\sigma e^{-a} > 1$. We start by estimating from below the probability of going from 1 to a neighborhood of ρ^* without visiting 0. Before proceeding with the mathematical details, let us explain the strategy to get this lower bound. When $Z_0 = 1$, the binomial law involved in the replication mechanism can be approximated by a Poisson law of parameter σ, and the process $(Z_n)_{n\geq 0}$ can jump to any fixed $h \in \mathbb{N}$ with a probability larger than a positive quantity independent of m. Using a simple estimate on the central term of the binomial law, we have that

$$P\big(Z_{n+1} = G_m(h) \,|\, Z_n = h\big) \;\geq\; \frac{1}{(m+1)^2}$$

where G_m is a map from $\{0, \dots, m\}$ to $\{0, \dots, m\}$ such that

$$\frac{1}{m} G_m(h) \;\geq\; F\Big(\frac{h}{m}\Big) - \frac{1}{m} .$$

Therefore we study the iterates of the function $F(x) - 1/m$. This function, which is a small perturbation of F, has two fixed points, one unstable close to 0, of order $1/m$, and one stable close to ρ^*. We take h large enough so that h/m is above the unstable fixed point. Then the repulsive dynamics of $F(x)-1/m$ will bring the point h/m close to a value $\eta > 0$ (independent of m) in a number of iterates of order $\ln m$. Once the process $(Z_n)_{n\geq 0}$ is at $\lfloor \eta m \rfloor$, a finite number of iterates leads into the neighborhood of ρ^*. The lower bound is obtained by combining the three steps:

$$P(1 \to h)P(h \to \eta m)P(\eta m \to (\rho^* - \delta)m) \;\geq\; c\bigg(\frac{1}{(m+1)^2}\bigg)^{c \ln m + c},$$

where c is a constant independent of m. This is the idea of the proof of the next lemma.

Lemma 7.8 For any $\delta > 0$, there exist $m_0, c > 0$, depending on δ, such that, for $m \geq m_0$,

$$P\big(Z_1 > 0, \dots, Z_{\lfloor c \ln m \rfloor - 1} > 0,\, Z_{\lfloor c \ln m \rfloor} > m(\rho^* - \delta) \,|\, Z_0 = 1\big) \;\geq\; \frac{1}{m^{c \ln m}}\,.$$

Proof. The binomial law $\mathcal{B}(n,p)$ of parameters $n \geq 0$ and $p < 1$ is maximal at $\lfloor (n+1)p \rfloor$, therefore

$$\binom{n}{\lfloor (n+1)p \rfloor} p^{\lfloor (n+1)p \rfloor}(1-p)^{n - \lfloor (n+1)p \rfloor} \;\geq\; \frac{1}{n+1}\,.$$

See for instance chapter VI in Feller's book [8]. We shall use this inequality to bound from below the transition probabilities of the Markov chain $(Z_n)_{n\geq 0}$. Let us define a map $G_m : \{\,0, \dots, m\,\} \to \{\,0, \dots, m\,\}$ by

$$\forall h \in \{\,0, \dots, m-1\,\} \quad G_m(h) \;=\; \left\lfloor \Big(\Big\lfloor (m+1) f\Big(\frac{h}{m}\Big) \Big\rfloor + 1 \Big) e^{-a} \right\rfloor,$$
$$G_m(m) \;=\; \lfloor (m+1)e^{-a} \rfloor\,.$$

Applying the previous lower bound to the binomial laws involved in the transition step of $(Z_n)_{n\geq 0}$, we obtain

$$\forall n, h \geq 0 \qquad P\big(Z_{n+1} \geq G_m(h) \,|\, Z_n = h\big) \;\geq$$
$$p\Big(h, \Big\lfloor (m+1) f\Big(\frac{h}{m}\Big) \Big\rfloor, G_m(h), G_m(h)\Big) \;\geq\; \frac{1}{(m+1)^2}\,.$$

It follows that for $n, h \geq 0$,

$$
\begin{aligned}
&P\big(Z_1 \geq (G_m)^1(h), \dots, Z_n \geq (G_m)^n(h) \,|\, Z_0 = h\big)\\
&= \sum_{l \geq (G_m)^{n-1}(h)} P\big(Z_1 \geq (G_m)^1(h), \dots, Z_{n-1} = l, Z_n \geq (G_m)^n(h) \,|\, Z_0 = h\big)\\
&= \sum_{l \geq (G_m)^{n-1}(h)} P\big(Z_n \geq (G_m)^n(h) \,|\, Z_{n-1} = l\big)\\
&\qquad\qquad\qquad \times P\big(Z_1 \geq (G_m)^1(h), \dots, Z_{n-1} = l \,|\, Z_0 = h\big)
\end{aligned}
$$

$$
\begin{aligned}
&\geq \; P\big(Z_n \geq (G_m)^n(h) \,|\, Z_{n-1} = (G_m)^{n-1}(h)\big)\\
&\qquad \times P\big(Z_1 \geq (G_m)^1(h), \dots, Z_{n-1} \geq (G_m)^{n-1}(h) \,|\, Z_0 = h\big)\\
&\quad \geq \; \frac{1}{(m+1)^2} P\big(Z_1 \geq (G_m)^1(h), \dots, Z_{n-1} \geq (G_m)^{n-1}(h) \,|\, Z_0 = h\big)\,.
\end{aligned}
$$

Iterating this inequality, we obtain, for $n, h \geq 0$,

$$
P\big(Z_1 \geq (G_m)^1(h), \dots, Z_n \geq (G_m)^n(h) \,|\, Z_0 = h\big) \;\geq\; \frac{1}{(m+1)^{2n}}\,.
$$

The map $G_m : \{\,0, \dots, m\,\} \to \{\,0, \dots, m\,\}$ is non–decreasing. Moreover

$$
\forall h \in \{\,0, \dots, m\,\} \qquad G_m(h) \;\geq\; \left\lfloor (m+1) f\Big(\frac{h}{m}\Big) e^{-a} \right\rfloor \;\geq\; m f\Big(\frac{h}{m}\Big) e^{-a} - 1\,.
$$

Let us define a map $H_m : [0,1] \to [0,1]$ by

$$
\forall x \in [0,1] \qquad H_m(x) \;=\; F(x) - \frac{1}{m}\,.
$$

We can rewrite the previous inequality as

$$
\forall h \in \{\,0, \dots, m\,\} \qquad G_m(h) \;\geq\; m\, H_m\Big(\frac{h}{m}\Big)\,.
$$

Iterating this inequality, we get, thanks to the fact that both G_m and H_m are non–decreasing,

$$
\forall n \geq 0 \quad \forall h \in \{\,0, \dots, m\,\} \qquad (G_m)^n(h) \;\geq\; m\, (H_m)^n\Big(\frac{h}{m}\Big)\,.
$$

The map H_m, which is a small perturbation of the map F, has two fixed points $\rho'_m < \rho''_m$, whose expansion as m goes to ∞ is given by

$$
\begin{aligned}
\rho'_m \;&=\; \frac{1}{m(\sigma e^{-a} - 1)} + o\Big(\frac{1}{m}\Big)\,,\\
\rho''_m \;&=\; \frac{\sigma e^{-a} - 1}{\sigma - 1} - \frac{\sigma e^{-a}}{m(\sigma e^{-a} - 1)} + o\Big(\frac{1}{m}\Big)\,.
\end{aligned}
$$

Let $\eta > 0$. If $x \le \eta$, we have $F(x) \ge \alpha x$, where

$$\alpha \,=\, \frac{\sigma e^{-a}}{(\sigma-1)\eta+1}\,.$$

For η sufficiently small, we have $\alpha > 1$ and the map F restricted to $[0,\eta]$ is expanding. Let us study the iterates of x through the map H_m. We set

$$N \,=\, \inf\,\big\{\, n\ge 0 : (H_m)^n(x) > \eta \,\big\}\,.$$

For $1 \le n < N$, we have

$$(H_m)^n(x) \,=\, H_m\big((H_m)^{n-1}(x)\big) \,\ge\, \alpha (H_m)^{n-1}(x) - \frac{1}{m}\,,$$

which we rewrite as

$$\frac{1}{\alpha^n}\,(H_m)^n(x) \,\ge\, \frac{1}{\alpha^{n-1}}\,(H_m)^{n-1}(x) - \frac{1}{m\alpha^{n-1}}\,.$$

Summing from $n = 1$ to $N-1$, we get

$$(H_m)^{N-1}(x) \,\ge\, \alpha^{N-1}\left(x - \frac{1}{m}\sum_{n=0}^{N-2}\frac{1}{\alpha^n}\right) \,\ge\, \alpha^{N-1}\Big(x - \frac{\alpha}{m(\alpha-1)}\Big)\,.$$

Let h be an integer such that

$$h \,\ge\, 2\frac{\alpha}{\alpha-1}\,.$$

Notice that this condition does not depend on m. We suppose that $m > h$. We take $x = h/m$, and we denote by $N(h)$ the associated integer. We have then

$$\eta \,>\, (H_m)^{N(h)-1}\Big(\frac{h}{m}\Big) \,\ge\, \alpha^{N(h)-1}\frac{h}{2m}\,.$$

Thus $N(h)$ satisfies

$$N(h) \,<\, 1 + \frac{1}{\ln\alpha}\ln\frac{2m\eta}{h}$$

and we have

$$\begin{aligned}
&P\big(Z_1 > 0,\dots,Z_{N(h)-1} > 0,\, Z_{N(h)} > m\eta \,|\, Z_0 = h\big)\\
&\quad\ge\, P\Big(Z_1 \ge m(H_m)^1\Big(\frac{h}{m}\Big),\dots,Z_{N(h)} \ge m(H_m)^{N(h)}\Big(\frac{h}{m}\Big) \,|\, Z_0 = h\Big)\\
&\qquad\ge\, P\big(Z_1 \ge (G_m)^1(h),\dots,Z_{N(h)} \ge (G_m)^{N(h)}(h) \,|\, Z_0 = h\big)\\
&\qquad\qquad\ge\, \frac{1}{(m+1)^{2N(h)}}\,.
\end{aligned}$$

We control next the probability to go from 1 to h. We have

$$P(Z_1 \geq h \,|\, Z_0 = 1) \;\geq\; \binom{m}{h} \frac{\sigma^h (m-1)^{m-h}}{\big(\sigma - 1 + m\big)^m} M_H(0,0)^h \,.$$

In this regime, where h is fixed and m is large, the Binomial law involved in the replication mechanism can be approximated by a Poisson law of parameter σ, whence, for m large enough,

$$P(Z_1 \geq h \,|\, Z_0 = 1) \;\geq\; \frac{1}{2} \exp(-\sigma) \frac{\sigma^h}{h!} \exp(-ah) \,.$$

We control finally the probability to go from ηm to the neighborhood of ρ^*. We do this by following the iterates of F starting from η, and by controlling the error term with respect to the iterates of H_m.

Lemma 7.9 We suppose that $\sigma e^{-a} > 1$. For any $m \geq 1$, $n \geq 0$, $x \in [0,1]$, we have

$$(H_m)^n(x) \;\geq\; F^n(x) \;-\; \frac{1}{m} \frac{(\sigma e^{-a})^{n+1}}{\sigma e^{-a} - 1} \,.$$

Proof. We have

$$\forall x \in [0,1] \qquad \big|F'(x)\big| \;\leq\; \sigma e^{-a} \,,$$

and, for any $n \geq 0$,

$$(H_m)^{n+1}(x) \;=\; H_m\big((H_m)^n(x)\big) \;=\; F\big((H_m)^n(x)\big) - \frac{1}{m} \,.$$

We shall prove the following inequality by induction on n:

$$(H_m)^n(x) \;\geq\; F^n(x) \;-\; \frac{1}{m} \sum_{k=1}^{n} (\sigma e^{-a})^k \,.$$

The inequality is true for $n = 0, 1$. Suppose that the inequality holds for some $n \geq 0$. Since F is non–decreasing, we deduce from the inequality on F' and the mean value theorem that

$$\begin{aligned}
(H_m)^{n+1}(x) \;&\geq\; F\Big(F^n(x) - \frac{1}{m} \sum_{k=1}^{n} (\sigma e^{-a})^k\Big) - \frac{1}{m} \\
&\geq\; F^{n+1}(x) - \sigma e^{-a} \frac{1}{m} \sum_{k=1}^{n} (\sigma e^{-a})^k - \frac{1}{m} \\
&\geq\; F^{n+1}(x) - \frac{1}{m} \sum_{k=1}^{n+1} (\sigma e^{-a})^k
\end{aligned}$$

and the inequality is proved at rank $n+1$. Summing the geometric series, we obtain the inequality stated in the lemma. $\square$

Let $\delta > 0$. Now we have

$$\lim_{n\to\infty} F^n(\eta) \,=\, \rho^*\,,$$

therefore

$$\exists t \geq 1 \qquad F^t(\eta) \,>\, \rho^* - \delta\,.$$

For m large enough, we have

$$F^t(\eta) \,-\, \frac{1}{m}\frac{(\sigma e^{-a})^{t+1}}{\sigma e^{-a}-1} \,>\, \rho^* - \delta\,,$$

so that, by lemma 7.9,

$$(H_m)^t(\eta) \,>\, \rho^* - \delta\,.$$

Let i be an integer strictly larger than ηm. We have

$$\begin{aligned}
&P\big(Z_1 > 0, \dots, Z_{t-1} > 0,\, Z_t > m(\rho^* - \delta) \,|\, Z_0 = i\big)\\
&\qquad\qquad \geq\, P\big(Z_1 \geq m(H_m)^1(\eta), \dots, Z_t \geq m(H_m)^t(\eta) \,|\, Z_0 = i\big)\\
&\qquad\quad \geq\, P\Big(Z_1 \geq m(H_m)^1\Big(\frac{i}{m}\Big), \dots, Z_t \geq m(H_m)^t\Big(\frac{i}{m}\Big) \,|\, Z_0 = i\Big)\\
&\qquad\quad \geq\, P\big(Z_1 \geq (G_m)^1(i), \dots, Z_t \geq (G_m)^t(i) \,|\, Z_0 = i\big) \,\geq\, \frac{1}{(m+1)^{2t}}\,.
\end{aligned}$$

To conclude, we use the monotonicity of $(Z_n)_{n\geq 0}$ and we combine the three previous estimates. The values h, t do not depend on m, and there exists a positive constant c depending on η, h such that, for m large enough,

$$N(h) + t + 1 \,<\, c\ln m\,,$$

$$\Big(\frac{1}{2}\exp(-\sigma)\frac{\sigma^h}{h!}\exp(-ah)\Big)^{c\ln m}\frac{1}{(m+1)^{2N(h)+2t}} \,\geq\, \frac{1}{m^{c\ln m}}\,.$$

Let us set

$$s \,=\, \lfloor c\ln m\rfloor - \big(N(h)+t\big)\,.$$

We have

$$
\begin{aligned}
&P(Z_1>0,\dots,Z_{\lfloor c\ln m\rfloor-1}>0,Z_{\lfloor c\ln m\rfloor}>m(\rho^*-\delta)\,|\,Z_0=1)\ \geq\\
&\quad\sum_{j\geq h}\sum_{i>m\eta}P(Z_1\geq h,\dots,Z_{s-1}\geq h,Z_s=j,Z_{s+1}>0,\dots,Z_{s+N(h)-1}>0,\\
&\qquad\qquad Z_{s+N(h)}=i,\ Z_{s+N(h)+1}>0,\dots,Z_{s+N(h)+t-1}>0,\\
&\qquad\qquad\qquad\qquad Z_{s+N(h)+t}>m(\rho^*-\delta)\,|\,Z_0=1\big)\\
&\geq\sum_{j\geq h}\sum_{i>m\eta}P(Z_1\geq h,\dots,Z_{s-1}\geq h,Z_s=j\,|\,Z_0=1)\times\\
&\quad P\big(Z_{s+1}>0,\dots,Z_{s+N(h)-1}>0,\ Z_{s+N(h)}=i\,|\,Z_s=j\big)P\big(Z_{s+N(h)+1}>0,\\
&\qquad\qquad\dots,Z_{s+N(h)+t-1}>0,\ Z_{s+N(h)+t}>m(\rho^*-\delta)\,|\,Z_{s+N(h)}=i\big)\\
&\geq P(Z_1\geq h,\dots,Z_s\geq h\,|\,Z_0=1)\sum_{i>m\eta}P\big(Z_1>0,\dots,Z_{N(h)-1}>0,\\
&\qquad Z_{N(h)}=i\,|\,Z_0=h\big)P\big(Z_1>0,\dots,Z_{t-1}>0,\ Z_t>m(\rho^*-\delta)\,|\,Z_0=i\big)\\
&\quad\geq\ \Big(P(Z_1\geq h\,|\,Z_0=1)\Big)^s\times\\
&\qquad\qquad P\big(Z_1>0,\dots,Z_{N(h)-1}>0,\ Z_{N(h)}>m\eta\,|\,Z_0=h\big)\frac{1}{(m+1)^{2t}}\\
&\quad\geq\ \Big(\frac{1}{2}\exp(-\sigma)\frac{\sigma^h}{h!}\exp(-ah)\Big)^s\frac{1}{(m+1)^{2N(h)+2t}}\ \geq\ \frac{1}{m^{c\ln m}}\,.
\end{aligned}
$$

This is the required lower bound. □

Whenever the starting point is away from 0, the estimate of lemma 7.9 can be considerably enhanced, as shown in the next lemma.

Lemma 7.10 We suppose that $\sigma e^{-a}>1$. For any $\delta>0$, there exist $h\geq 1$ and $c>0$, depending on δ, such that

$$P\big(Z_1>0,\dots,Z_{h-1}>0,\ Z_h>m(\rho^*-\delta)\,|\,Z_0=\lfloor m\delta\rfloor\big)\ \geq\ 1-\exp(-cm)\,.$$

Proof. Let $\delta>0$. Since

$$\lim_{n\to\infty}F^n(\delta)\ =\ \rho^*\,,$$

then there exists $h\geq 1$ such that $F^h(\delta)>\rho^*-\delta$. By continuity of the map F, there exist $\rho_0,\rho_1,\dots,\rho_h>0$ such that $\rho_0=\delta$, $\rho_h>\rho^*-\delta$ and

$$\forall k\in\{\,1,\dots,h\,\}\qquad F(\rho_{k-1})\ >\ \rho_k\,.$$

Now,

$$
\begin{aligned}
&P\big(Z_1>0,\dots,Z_{h-1}>0,\, Z_h>m(\rho^*-\delta)\,|\,Z_0=\lfloor m\delta\rfloor\big)\ \geq\\
&\qquad\qquad\qquad P\Big(\forall k\in\{\,1,\dots,h\,\}\quad Z_k\geq m\rho_k\,|\,Z_0=\lfloor m\delta\rfloor\Big)\,.
\end{aligned}
$$

Passing to the complementary event, we have

$$
\begin{aligned}
&P\big(\exists k\in\{\,1,\dots,h-1\,\}\quad Z_k=0\quad\text{or}\quad Z_h\leq m(\rho^*-\delta)\,|\,Z_0=\lfloor m\delta\rfloor\big)\\
&\qquad\leq\ P\Big(\exists k\in\{\,1,\dots,h\,\}\quad Z_k<m\rho_k\,|\,Z_0=\lfloor m\delta\rfloor\Big)\\
&\leq\sum_{1\leq k\leq h}P\Big(Z_1\geq m\rho_1,\dots,Z_{k-1}\geq m\rho_{k-1},\,Z_k<m\rho_k\,|\,Z_0=\lfloor m\delta\rfloor\Big)\\
&\qquad\leq\sum_{1\leq k\leq h}\sum_{i\geq m\rho_{k-1}}P\Big(Z_{k-1}=i,\,Z_k<m\rho_k\,|\,Z_0=\lfloor m\delta\rfloor\Big)\\
&\leq\sum_{1\leq k\leq h}\sum_{i\geq m\rho_{k-1}}P\Big(Z_k<m\rho_k\,|\,Z_{k-1}=i\Big)\,P\Big(Z_{k-1}=i\,|\,Z_0=\lfloor m\delta\rfloor\Big)\\
&\qquad\qquad\qquad\leq\sum_{1\leq k\leq h}P\Big(Z_1<m\rho_k\,|\,Z_0=\lfloor m\rho_{k-1}\rfloor\Big)\,.
\end{aligned}
$$

The large deviation principle for the transition probabilities of the Markov chain $(Z_n)_{n\geq 0}$ stated in proposition 7.3 implies that for $k\in\{\,1,\dots,h\,\}$,

$$
\begin{aligned}
&\limsup_{\substack{\ell,m\to\infty,\,q\to 0\\ \ell q\to a}}\ \frac{1}{m}\ln P\Big(Z_1<m\rho_k\,|\,Z_0=\lfloor m\rho_{k-1}\rfloor\Big)\\
&\qquad\qquad\leq\ -\inf\,\Big\{\,I\big(\rho_{k-1},s,t\big):s\in[0,1],t\leq\rho_k\,\Big\}\ <\ 0\,.
\end{aligned}
$$

Since h is fixed, we conclude that

$$
\limsup_{\substack{\ell,m\to\infty,\,q\to 0\\ \ell q\to a}}\ \frac{1}{m}\ln P\left(\begin{matrix}\exists k\in\{\,1,\dots,h-1\,\}\quad Z_k=0\\ \text{or}\quad Z_h\leq m(\rho^*-\delta)\end{matrix}\ \Big|\ Z_0=\lfloor m\delta\rfloor\right)\ <\ 0
$$

and this yields the desired estimate. □

With the estimate of lemma 7.10, we show that the process is very unlikely to stay a long time in $[m\delta,m(\rho^*-\delta)]$.

Corollary 7.11 We suppose that $\sigma e^{-a}>1$. Let $\delta>0$. There exist $h\geq 1$ and $c>0$ such that

$$
\begin{aligned}
&\forall k\in[m\delta,m(\rho^*-\delta)]\quad\forall n\geq 0\\
&\quad P\Big(m\delta\leq Z_t\leq m(\rho^*-\delta)\text{ for }0\leq t\leq n\,|\,Z_0=k\Big)\ \leq\ \exp\Big(-cm\Big\lfloor\frac{n}{h}\Big\rfloor\Big)\,.
\end{aligned}
$$

Proof. Let $k \in [m\delta, m(\rho^* - \delta)]$. Let $\delta > 0$ and let $h \geq 1$ and $c > 0$ be associated to δ as in lemma 7.10. We divide the interval $\{0, \dots, n\}$ into subintervals of length h and we use repeatedly the estimate of lemma 7.10. Let $i \geq 0$. We write

$$
\begin{aligned}
&P\big(m\delta \leq Z_t \leq m(\rho^* - \delta) \text{ for } 0 \leq t \leq (i+1)h \,|\, Z_0 = k\big) \;= \\
&\sum_{\delta m \leq j \leq (\rho^* - \delta)m} P\big(m\delta \leq Z_t \leq m(\rho^* - \delta) \text{ for } 0 \leq t \leq (i+1)h,\, Z_{ih} = j \,|\, Z_0 = k\big) \\
&= \sum_{\delta m \leq j \leq (\rho^* - \delta)m} P\big(m\delta \leq Z_t \leq m(\rho^* - \delta) \text{ for } 0 \leq t \leq ih,\, Z_{ih} = j \,|\, Z_0 = k\big) \\
&\qquad\qquad \times P\big(m\delta \leq Z_t \leq m(\rho^* - \delta) \text{ for } ih \leq t \leq (i+1)h \,|\, Z_{ih} = j\big) \\
&\leq \sum_{\delta m \leq j \leq (\rho^* - \delta)m} P\big(m\delta \leq Z_t \leq m(\rho^* - \delta) \text{ for } 0 \leq t \leq ih,\, Z_{ih} = j \,|\, Z_0 = k\big) \\
&\qquad\qquad \times P\big(Z_h \leq m(\rho^* - \delta) \,|\, Z_0 = \lfloor m\delta \rfloor\big) \\
&\leq P\big(m\delta \leq Z_t \leq m(\rho^* - \delta) \text{ for } 0 \leq t \leq ih \,|\, Z_0 = k\big) \exp(-cm)\,.
\end{aligned}
$$

Iterating this inequality, we obtain

$$
\forall i \geq 0 \qquad P\big(m\delta \leq Z_t \leq m(\rho^* - \delta) \text{ for } 0 \leq t \leq ih \,|\, Z_0 = k\big) \;\leq\; \exp(-cmi)\,.
$$

The claim of the corollary follows by applying this inequality with i equal to the integer part of n/h. □

We have computed the relevant estimates to reach the neighborhood of ρ^*. Our next goal is to study the hitting time τ_0 starting from a neighborhood of ρ^*. Since we need only a lower bound, we shall study the hitting time of a neighborhood of 0. For $\delta > 0$, we define

$$
\tau_\delta \;=\; \inf\,\big\{\, n \geq 0 : Z_n < m\delta \,\big\}\,.
$$

Let $i > (\rho^* - \delta)m$. We shall estimate the expectation of τ_δ starting from i. The strategy consists in looking at the portion of the trajectory starting at the last visit to the neighborhood of ρ^* before reaching the neighborhood of 0. Accordingly, we define

$$
S \;=\; \max\,\big\{\, n \leq \tau_\delta : Z_n > (\rho^* - \delta)m \,\big\}\,.
$$

We write, for $n, k \geq 1$,

$$
\begin{aligned}
P\big(\tau_\delta \leq n \,|\, Z_0 = i\big) \;&=\; \sum_{1 \leq s < t \leq n} P\big(\tau_\delta = t,\, S = s \,|\, Z_0 = i\,\big) \\
= \sum_{\substack{1 \leq s < t \leq n \\ s < t \leq s+k}} P\big(\tau_\delta = t,\, S = s \,|\, Z_0 = i\,\big) &+ \sum_{\substack{1 \leq s < n \\ s+k < t \leq n}} P\big(\tau_\delta = t,\, S = s \,|\, Z_0 = i\,\big)\,.
\end{aligned}
$$

Let $h \geq 1$ and $c > 0$ be associated to δ as in corollary 7.11. For $1 \leq s < n$ and $t > s + k$,

$$
\begin{aligned}
&P\big(\tau_\delta = t,\, S = s \,|\, Z_0 = i\,\big) \\
&\qquad = \sum_{m\delta \leq j \leq (\rho^* - \delta)m} P\big(\tau_\delta = t,\, S = s,\, Z_{s+1} = j \,|\, Z_0 = i\,\big) \\
&\qquad \leq \sum_{m\delta \leq j \leq (\rho^* - \delta)m} P\Bigg(\begin{matrix} \delta m \leq Z_r \leq (\rho^* - \delta)m \\ \text{for } s+1 \leq r \leq t-1 \end{matrix} \,\Bigg|\, Z_{s+1} = j\,\Bigg) \\
&\qquad\qquad \leq m \exp\Big(- cm \Big\lfloor \frac{t-s-2}{h} \Big\rfloor \Big)\,,
\end{aligned}
$$

whence

$$
\sum_{\substack{1 \leq s < n \\ s+k < t \leq n}} P\big(\tau_\delta = t,\, S = s \,|\, Z_0 = i\,\big) \;\leq\; n \sum_{t \geq k} m \exp\Big(- cm \Big\lfloor \frac{t-1}{h} \Big\rfloor \Big)\,.
$$

For $1 \leq s < t \leq n$ and $t \leq s + k$,

$$
\begin{aligned}
&P\big(\tau_\delta = t,\, S = s \,|\, Z_0 = i\,\big) \\
&\qquad \leq \sum_{j > (\rho^* - \delta)m} P\big(\tau_\delta = t,\, S = s,\, Z_s = j \,|\, Z_0 = i\,\big) \\
&\qquad\qquad \leq \sum_{j > (\rho^* - \delta)m} P\big(Z_t < \delta m \,|\, Z_s = j\,\big) \\
&\qquad\qquad\qquad \leq mP\big(Z_{t-s} < \delta m \,|\, Z_0 = \lfloor (\rho^* - \delta)m \rfloor\,\big)\,,
\end{aligned}
$$

whence

$$
\begin{aligned}
&\sum_{\substack{1 \leq s < n \\ s < t \leq s+k}} P\big(\tau_\delta = t,\, S = s \,|\, Z_0 = i\,\big) \;\leq \\
&\qquad\qquad n \sum_{1 \leq t \leq k} mP\big(Z_t < \delta m \,|\, Z_0 = \lfloor (\rho^* - \delta)m \rfloor\,\big)\,.
\end{aligned}
$$

Putting together the previous inequalities, we obtain

$$
\begin{aligned}
P\big(\tau_\delta \leq n \,|\, Z_0 = i\big) \;\leq\;& n \sum_{t \geq k} m \exp\Big(- cm \Big\lfloor \frac{t-1}{h} \Big\rfloor \Big) \\
&+ n \sum_{1 \leq t \leq k} mP\big(Z_t < \delta m \,|\, Z_0 = \lfloor (\rho^* - \delta)m \rfloor\,\big)\,.
\end{aligned}
$$

We choose k large enough so that

$$\limsup_{\substack{\ell,m\to\infty,\, q\to 0\\ \ell q\to a}} \frac{1}{m}\ln\Big(\sum_{t\geq k} m\exp\Big(-cm\Big\lfloor\frac{t}{h}\Big\rfloor\Big)\Big) \,<\, -V(\rho^*-\delta,\delta)\,,$$

and we use the large deviation principle stated in corollary 7.4 to estimate the second sum:

$$\limsup_{\substack{\ell,m\to\infty,\, q\to 0\\ \ell q\to a}} \frac{1}{m}\ln\Big(\sum_{1\leq t\leq k} mP\big(Z_t<\delta m\,|\,Z_0=\lfloor(\rho^*-\delta)m\rfloor\,\big)\Big)$$
$$\leq\,-\min_{1\leq t\leq k} V_t(\rho^*-\delta,\delta)\,\leq\,-V(\rho^*-\delta,\delta)\,.$$

Applying the previous inequalities with $n=\exp(mV(\rho^*-\delta,\delta)-m\delta)$, we conclude that

$$\lim_{\substack{\ell,m\to\infty,\, q\to 0\\ \ell q\to a}} P\big(\tau_\delta\leq\exp(mV(\rho^*-\delta,\delta)-m\delta)\,|\,Z_0=i\big)\,=\,0$$

and therefore

$$\liminf_{\substack{\ell,m\to\infty,\, q\to 0\\ \ell q\to a}} \frac{1}{m}\ln E(\tau_\delta\,|\,Z_0=i)\,\geq\,V(\rho^*-\delta,\delta)-\delta\,.$$

To derive a lower bound on the expectation of τ_0 starting from 1, we combine the previous estimates as follows. By lemma 7.8, asymptotically,

$$P\big(Z_1>0,\dots,Z_{\lfloor c\ln m\rfloor-1}>0,\,Z_{\lfloor c\ln m\rfloor}>m(\rho^*-\delta)\,|\,Z_0=1\big)\,\geq\,\frac{1}{m^{c\ln m}}\,.$$

Thus, letting $i=\lfloor(\rho^*-\delta)m\rfloor+1$, for any $n\geq\lfloor c\ln m\rfloor$,

$$\begin{aligned}
&P(\tau_0>n\,|\,Z_0=1)\,\geq\\
&\quad\sum_{j\geq i} P\big(Z_1>0,\dots,Z_{\lfloor c\ln m\rfloor-1}>0,\,Z_{\lfloor c\ln m\rfloor}=j,\,\tau_0>n\,|\,Z_0=1\big)\\
&\qquad\geq\sum_{j\geq i} P\big(Z_1>0,\dots,Z_{\lfloor c\ln m\rfloor-1}>0,\,Z_{\lfloor c\ln m\rfloor}=j\,|\,Z_0=1\big)\\
&\qquad\qquad\times P\big(Z_{\lfloor c\ln m\rfloor+1}>0,\dots,Z_n>0\,|\,Z_{\lfloor c\ln m\rfloor}=j\big)\\
&\quad\geq\,P\big(Z_1>0,\dots,Z_{\lfloor c\ln m\rfloor-1}>0,\,Z_{\lfloor c\ln m\rfloor}>m(\rho^*-\delta)\,|\,Z_0=1\big)\\
&\qquad\qquad\times P\big(Z_{\lfloor c\ln m\rfloor+1}>0,\dots,Z_n>0\,|\,Z_{\lfloor c\ln m\rfloor}=i\big)\\
&\qquad\qquad\geq\,\frac{1}{m^{c\ln m}}\,P\big(\tau_0>n-\lfloor c\ln m\rfloor\,|\,Z_0=i\big)\,.
\end{aligned}$$

Summing from $n = \lfloor c \ln m \rfloor$ to $+\infty$, we get

$$E(\tau_0 \,|\, Z_0 = 1) \;\geq\; \frac{1}{m^{c \ln m}} E(\tau_0 \,|\, Z_0 = i)\,.$$

The very definition of τ_δ implies that $\tau_0 \geq \tau_\delta$, whence

$$E(\tau_0 \,|\, Z_0 = i) \;\geq\; E(\tau_\delta \,|\, Z_0 = i)\,.$$

From the lower bound on τ_δ and the previous inequalities, we deduce that

$$\liminf_{\substack{\ell, m \to \infty,\, q \to 0 \\ \ell q \to a}} \frac{1}{m} \ln E(\tau_0 \,|\, Z_0 = 1) \;\geq\; V(\rho^* - \delta, \delta) - \delta\,.$$

The conclusion follows by letting δ go to 0. $\square$

7.3 Concentration near ρ^*

In this section, we estimate the numerator of the last formula of section 6.3. As usual, we drop the superscript θ from the notation when it is not necessary, and we put it back when we need to emphasize the differences between the cases $\theta = \ell$ and $\theta = 1$. We define, as before proposition 7.6,

$$\rho^*(a) \;=\; \rho(e^{-a}, 0) \;=\; \begin{cases} \dfrac{\sigma e^{-a} - 1}{\sigma - 1} & \text{if } \sigma e^{-a} > 1 \\ 0 & \text{if } \sigma e^{-a} \leq 1 \end{cases}$$

Let $f : [0,1] \to \mathbb{R}$ be a non–decreasing continuous function such that $f(0) = 0$. Our goal here is to estimate the expected value of the sum

$$\sum_{n=0}^{\tau_0} f\Big(\frac{Z_n}{m}\Big)\,.$$

The Markov chain $(Z_n)_{n \geq 0}$ is a perturbation of the dynamical system associated to the map F, therefore it spends most of its time in the neighborhood of the stable fixed point ρ^*. On very large time intervals, the process visits points far away from ρ^* and then it returns quickly to ρ^*. Thus the fraction of the time spent away from ρ^* is negligible. We will show that the above sum is, on average, comparable to $f(\rho^*)\tau_0$.

Proposition 7.12 Uniformly over $i \in \{\,1, \ldots, m\,\}$, we have

$$\lim_{\substack{\ell, m \to \infty \\ q \to 0,\, \ell q \to a}} \frac{E\bigg(\displaystyle\sum_{n=0}^{\tau_0} f\Big(\frac{Z_n}{m}\Big) \,\Big|\, Z_0 = i\bigg)}{E\big(\tau_0 \,\big|\, Z_0 = i\big)} \;=\; f(\rho^*)\,.$$

Proof. Before proceeding to the proof, let us explain the general strategy, which comes directly from the theory of Freidlin and Wentzell. Let us denote by $U(\delta)$ the δ–neighborhood of ρ^*. We choose δ small enough, so that when the process is in the neighborhood $U(2\delta)$, the value $f(Z_n/m)$ is approximated by $f(\rho^*)$. When the process is outside of $\{ 0 \} \cup U(2\delta)$, it reenters $U(\delta)$ in $\lfloor c \ln m \rfloor$ steps with probability at least $m^{-c \ln m}$, for some $c > 0$ (see lemma 7.15). Therefore the average length of an excursion is bounded by $m^{c \ln m}$. At a given time, the probability to start an excursion from $U(\delta)$ reaching the outside of $U(2\delta)$ is less than $\exp(-cm)$. With this estimate we can control the number of excursions (see lemma 7.14) and we show that, typically, their total length until the time τ_0 is negligible compared to τ_0.

Let $\varepsilon > 0$. Since f is continuous, there exists $\delta > 0$ such that

$$\forall \rho \in]\rho^* - 2\delta, \rho^* + 2\delta[\qquad \big| f(\rho) - f(\rho^*) \big| < \varepsilon \,.$$

We define then a sequence of stopping times to follow the excursions of $(Z_n)_{n \geq 0}$ outside $]\rho^* - \delta, \rho^* + \delta[$. More precisely, we set $T_0 = 0$ and

$$\begin{aligned}
T_1^* &= \inf \Big\{ n \geq 0 : \frac{Z_n}{m} \in]\rho^* - \delta, \rho^* + \delta[\Big\} \,,\\
T_1 &= \inf \Big\{ n \geq T_1^* : \frac{Z_n}{m} \notin]\rho^* - 2\delta, \rho^* + 2\delta[\Big\} \,,\\
&\ \vdots \\
T_k^* &= \inf \Big\{ n \geq T_{k-1} : \frac{Z_n}{m} \in]\rho^* - \delta, \rho^* + \delta[\Big\} \,,\\
T_k &= \inf \Big\{ n \geq T_k^* : \frac{Z_n}{m} \notin]\rho^* - 2\delta, \rho^* + 2\delta[\Big\} \,,\\
&\ \vdots
\end{aligned}$$

We decompose the sum over the intervals $[T_{k-1}, T_k^*[$, $[T_k^*, T_k[$, $k \geq 1$. Denoting by $s \wedge t$ the minimum $\min(s,t)$, we have

$$\sum_{n=0}^{\tau_0} f\Big(\frac{Z_n}{m}\Big) - f(\rho^*)\,\tau_0 =$$
$$\sum_{k \geq 1} \sum_{n = T_{k-1} \wedge \tau_0}^{T_k^* \wedge \tau_0 - 1} \bigg(f\Big(\frac{Z_n}{m}\Big) - f(\rho^*) \bigg) + \sum_{k \geq 1} \sum_{n = T_k^* \wedge \tau_0}^{T_k \wedge \tau_0 - 1} \bigg(f\Big(\frac{Z_n}{m}\Big) - f(\rho^*) \bigg) .$$

We bound next the absolute value as follows:

$$\bigg| \sum_{n=0}^{\tau_0} f\Big(\frac{Z_n}{m}\Big) - f(\rho^*)\,\tau_0 \bigg| \leq 2f(1) \sum_{k \geq 1} \Big(T_k^* \wedge \tau_0 - T_{k-1} \wedge \tau_0 \Big) + \varepsilon\, \tau_0 \,.$$

It remains to deal with the sum. We define, for $n \geq 0$,

$$K(n) \,=\, \max\left\{\, k \geq 1 : T_{k-1} < n \,\right\},$$

and the sum becomes

$$\sum_{k \geq 1} \Big(T_k^* \wedge \tau_0 - T_{k-1} \wedge \tau_0 \Big) \,=\, \sum_{k=1}^{K(\tau_0)} \Big(T_k^* \wedge \tau_0 - T_{k-1} \Big)\,.$$

Let $\eta > 0$. We set

$$t_m^\eta \,=\, \exp\big(m(V(\rho^*, 0) + \eta) \big)\,.$$

We decompose the sum as follows:

$$\sum_{k=1}^{K(\tau_0)} \Big(T_k^* \wedge \tau_0 - T_{k-1} \Big) \,\leq\, \tau_0 1_{\tau_0 > t_m^\eta} + 1_{\tau_0 \leq t_m^\eta} \sum_{k=1}^{K(\tau_0)} \Big(T_k^* \wedge \tau_0 - T_{k-1} \Big)\,.$$

We suppose that the process starts from $i \in \{\,1, \dots, m\,\}$. The estimates are carried out exactly in the same way for any value of i, therefore, to alleviate the notation, we remove the starting point from the notation. Throughout the proof the expectation E and the probability P are meant with respect to the initial condition $Z_0 = i$. Taking expectation in the previous inequalities, we get

$$\left| E\left(\sum_{n=0}^{\tau_0} f\Big(\frac{Z_n}{m}\Big) \right) - f(\rho^*)\, E\big(\tau_0\big) \right|$$

$$\leq\, \varepsilon E\big(\tau_0\big) + 2f(1) E\big(\tau_0 1_{\tau_0 > t_m^\eta}\big) + 2f(1) E\left(1_{\tau_0 \leq t_m^\eta} \sum_{k=1}^{K(\tau_0)} \Big(T_k^* \wedge \tau_0 - T_{k-1} \Big) \right).$$

Next, we take care of the second term.

Lemma 7.13 For any $N, j \geq 1$,

$$E\big(\tau_0 1_{\tau_0 > Nj}\big) \,\leq\, NjP\big(\tau_0 > Nj\big) + \sum_{n \geq N} jP\big(\tau_0 > nj\big)\,.$$

Proof. We compute

$$\begin{aligned} E\big(\tau_0 1_{\tau_0 > Nj}\big) \,&=\, \sum_{k > Nj} kP\big(\tau_0 = k\big) \,=\, \sum_{k > Nj} \sum_{n \geq 0} 1_{n < k}\, P\big(\tau_0 = k\big) \\ &= \sum_{n \geq 0} \sum_{\substack{k > Nj \\ k > n}} P\big(\tau_0 = k\big) \,=\, \sum_{n \geq 0} P\big(\tau_0 > \max(Nj, n)\big) \\ &\qquad\qquad \leq\, NjP\big(\tau_0 > Nj\big) + \sum_{n \geq Nj} P\big(\tau_0 > n\big)\,. \end{aligned}$$

Next,

$$\sum_{n\geq Nj} P\big(\tau_0>n\big) \,=\, \sum_{n\geq N}\sum_{k=0}^{j-1} P\big(\tau_0>nj+k\big) \,\leq\, \sum_{n\geq N} jP\big(\tau_0>nj\big)$$

and we have the desired inequality. □

We apply lemma 7.7 with $\varepsilon=\eta/8$: there exists $j\geq 1$ such that

$$\forall n\geq 0 \qquad P\big(\tau_0>nj\,|\,Z_0=m\big) \,\leq\, \Big(1-\exp\big(-mV(\rho^*,0)-m\eta/2\big)\Big)^n .$$

We apply lemma 7.13 with this j and

$$N \,=\, \lfloor t_m^\eta/j\rfloor \,=\, \Big\lfloor \frac{1}{j}\exp\big(mV(\rho^*,0)+m\eta\big)\Big\rfloor$$

and we use the previous inequality:

$$E\big(\tau_0 1_{\tau_0>t_m^\eta}\big) \,\leq\, E\big(\tau_0 1_{\tau_0>Nj}\big) \,\leq\, NjP\big(\tau_0>Nj\big)+\sum_{n\geq N} jP\big(\tau_0>nj\big)$$

$$\leq\, \Big(Nj+j\exp\big(mV(\rho^*,0)+m\eta/2\big)\Big)\Big(1-\exp\big(-mV(\rho^*,0)-m\eta/2\big)\Big)^N$$

$$\leq\, (1+j)\exp\big(mV(\rho^*,0)+m\eta\big)\exp\bigg(-N\exp\big(-mV(\rho^*,0)-m\eta/2\big)\bigg) .$$

Thanks to the choice of N, this last quantity goes to 0 as m goes to ∞. We deal now with the last sum in the inequality before lemma 7.13. We give first an upper bound on K.

Lemma 7.14 There exists $c>0$, depending on δ, such that, asymptotically,

$$\forall k,n\geq 0 \qquad P\big(K(n)>k\big) \,\leq\, \frac{n^k}{k!}\exp(-cmk) .$$

Proof. We denote by $U(\delta)$ the δ–neighborhood of ρ^*:

$$U(\delta) \,=]\rho^*-\delta,\rho^*+\delta[.$$

For $k\geq 0$, we define

$$S_k^* \,=\, \sup\Big\{\, T_k^*\leq n<T_k : \frac{Z_n}{m}\in U(\delta)\,\Big\} .$$

For $k,n\geq 0$, we have

$$P\big(K(n)>k\big) \,=\, P\big(T_k<n\big) \,=\, \sum_{t^*\leq s<t<n} P\big(T_k^*=t^*,S_k=s,T_k=t\big) .$$

Let $h \geq 1$ and $c > 0$ be associated to δ as in corollary 7.11. We can suppose that $h \geq 2$. For given values of t^* and s, we split the sum over t in two parts:

$$\sum_{t:s<t<n} P\big(T_k^* = t^*, S_k = s, T_k = t\big) \;=\; \sum_{t:t>s+h} \cdots \quad + \sum_{t:s<t\leq s+h} \cdots$$

We study next the first sum, when $t > s+h$. We condition on the state at time $s+1$:

$$\begin{aligned}
\sum_{t:t>s+h} \cdots &= \sum_{\substack{t:t>s+h \\ j\in mU(2\delta)\setminus mU(\delta)}} P\big(T_k^* = t^*, S_k = s, Z_{s+1} = j, T_k = t\big) \\
&= \sum_{\substack{t:t>s+h \\ j\in mU(2\delta)\setminus mU(\delta)}} P\begin{pmatrix} T_k^* = t^*, Z_{s+1} = j, Z_t \notin mU(2\delta) \\ Z_{s+1}, \dots, Z_{t-1} \in mU(2\delta)\setminus mU(\delta) \end{pmatrix}
\end{aligned}$$

$$= \sum_{\substack{t:t>s+h \\ j\in mU(2\delta)\setminus mU(\delta)}} P\big(Z_{s+1}, \dots, Z_{t-1} \in mU(2\delta)\setminus mU(\delta), Z_t \notin mU(2\delta) \,\big|\, Z_{s+1} = j\big)$$

$$\times P\big(T_k^* = t^*, S_k = s, Z_{s+1} = j\big)\,.$$

For $0 \leq s < n$ and $t > s+h$,

$$\begin{aligned}
P\big(Z_{s+1}, \dots, Z_{t-1} \in mU(2\delta)\setminus mU(\delta), Z_t \notin mU(2\delta) \,\big|\, Z_{s+1} = j\big) \\
\leq P\left(\begin{matrix} \delta m \leq Z_r \leq (\rho^*-\delta)m \\ \text{for } s+1 \leq r \leq t-1 \end{matrix} \,\middle|\, Z_{s+1} = j\right) \\
\leq \exp\Big(-cm\Big\lfloor \frac{t-s-2}{h} \Big\rfloor\Big)\,.
\end{aligned}$$

Thus

$$\sum_{t:t>s+h} \cdots \;\leq\; \bigg(\sum_{t\geq h} \exp\Big(-cm\Big\lfloor \frac{t-1}{h} \Big\rfloor\Big)\bigg) P\big(T_k^* = t^*, S_k = s\big)\,.$$

Let us focus on the second sum. We condition on the state at time s:

$$\begin{aligned}
\sum_{t:s<t\leq s+h} \cdots &= \sum_{\substack{t:s<t\leq s+h \\ j\in mU(\delta)}} P\big(T_k^* = t^*, S_k = s, Z_s = j, T_k = t\big) \\
&\leq \sum_{\substack{t:s<t\leq s+h \\ j\in mU(\delta)}} P\big(Z_t \notin mU(2\delta) \,\big|\, Z_s = j\big) P\big(T_k^* = t^*, S_k = s, Z_s = j\big) \\
&\leq \sum_{\substack{t:1\leq t\leq h \\ j\in mU(\delta)}} P\big(Z_t \notin mU(2\delta) \,\big|\, Z_0 = j\big) P\big(T_k^* = t^*, S_k = s, Z_s = j\big)\,.
\end{aligned}$$

For any $j \in mU(\delta)$, using the monotonicity of $(Z_n)_{n\geq 0}$,

$$\begin{aligned}
P\big(Z_t \notin mU(2\delta) \,\big|\, Z_0 = j\big) \;\leq\; & P\big(Z_t \leq m(\rho^* - 2\delta) \,\big|\, Z_0 = j\big) \\
&\quad + P\big(Z_t \geq m(\rho^* + 2\delta) \,\big|\, Z_0 = j\big) \\
\leq\; & P\big(Z_t \leq m(\rho^* - 2\delta) \,\big|\, Z_0 = \lfloor(\rho^* - \delta)m\rfloor\big) \\
&\quad + P\big(Z_t \geq m(\rho^* + 2\delta) \,\big|\, Z_0 = \lfloor(\rho^* + \delta)m\rfloor\big)\,.
\end{aligned}$$

We use the large deviation principle stated in corollary 7.4 to estimate the last two terms. For any $t \in \{\,1,\dots,h\,\}$,

$$\begin{aligned}
\limsup_{\substack{\ell,m\to\infty,\, q\to 0\\ \ell q\to a}} \frac{1}{m} \ln P\big(Z_t \leq m(\rho^* - 2\delta) \,\big|\, Z_0 = \lfloor(\rho^* - \delta)m\rfloor\big) \\
\leq -\inf\big\{\, V_t(\rho^* - \delta, \rho) : \rho \leq \rho^* - 2\delta) \,\big\}\,, \\
\limsup_{\substack{\ell,m\to\infty,\, q\to 0\\ \ell q\to a}} \frac{1}{m} \ln P\big(Z_t \geq m(\rho^* + 2\delta) \,\big|\, Z_0 = \lfloor(\rho^* + \delta)m\rfloor\big) \\
\leq -\inf\big\{\, V_t(\rho^* + \delta, \rho) : \rho \geq \rho^* + 2\delta) \,\big\}\,.
\end{aligned}$$

By compactness, the infima are realized. Because of the constraints on ρ, the point ρ realizing the infimum

$$\inf\big\{\, V_t(\rho^* - \delta, \rho) : \rho \leq \rho^* - 2\delta) \,\big\}$$

is not an iterate of $\rho^* - \delta$ through F, hence by lemma 7.5, the above infimum is positive. We argue in the same way for the second infimum and we conclude that there exists $c' > 0$, depending on δ, such that, asymptotically,

$$\forall j \in mU(\delta) \qquad \sum_{t:1\leq t\leq h} P\big(Z_t \notin mU(2\delta) \,\big|\, Z_0 = j\big) \;\leq\; \exp(-c'm)\,,$$

whence

$$\sum_{t:s<t\leq s+h} \cdots \;\leq\; \exp(-c'm)\, P\big(T_k^* = t^*, S_k = s\big)\,.$$

Let $c'' > 0$ be such that, asymptotically,

$$\sum_{t\geq h} \exp\Big(-cm\Big\lfloor \frac{t-1}{h} \Big\rfloor\Big) + \exp(-c'm) \;\leq\; \exp(-c''m)\,.$$

Reporting in the initial equality, we obtain that, asymptotically, for any $n, k \geq 0$,

$$\begin{aligned}
P\big(T_k < n\big) \;\leq\; & \sum_{t^* \leq s < n} \exp(-c''m) P\big(T_k^* = t^*, S_k = s\big) \\
\leq\; & \sum_{t^* < n} P\big(T_k^* = t^*\big) \exp(-c''m) \;\leq\; \sum_{t^* < n} P\big(T_{k-1} < t^*\big) \exp(-c''m)\,.
\end{aligned}$$

Iterating this inequality, we obtain

$$P\big(T_k < n\big) \le \sum_{0 \le n_0 < \cdots < n_{k-1} < n} \exp(-c''mk)\,,$$

whence

$$P\big(T_k < n\big) \le \frac{n^k}{k!}\exp(-c''mk)$$

as required. $\square$

We estimate now the last sum. By the Cauchy–Schwarz inequality, we have

$$\begin{aligned}E\bigg(1_{\tau_0 \le t_m^\eta}\sum_{k=1}^{K(\tau_0)}\Big(T_k^* \wedge \tau_0 - T_{k-1}\Big)\bigg) \\ = \sum_{k\ge 1} E\bigg(1_{\tau_0 \le t_m^\eta} 1_{k \le K(\tau_0)}\Big(T_k^* \wedge \tau_0 - T_{k-1}\Big)\bigg) \\ \le \sum_{k \ge 1} P\big(\tau_0 \le t_m^\eta,\, K(\tau_0) \ge k\big)^{1/2}\bigg(E\Big(1_{k \le K(\tau_0)}\big(T_k^* \wedge \tau_0 - T_{k-1}\big)^2\Big)\bigg)^{1/2}.\end{aligned}$$

If $1 \le k \le K(\tau_0)$, then $T_{k-1} < \tau_0$ and $Z_{T_{k-1}} > 0$, so that, using the Markov property,

$$\begin{aligned}E\Big(1_{k \le K(\tau_0)}\big(T_k^* \wedge \tau_0 - T_{k-1}\big)^2\Big) = \\ \sum_{1 \le j \le m} E\Big(1_{k \le K(\tau_0)}\big(T_k^* \wedge \tau_0 - T_{k-1}\big)^2 \,\Big|\, Z_{T_{k-1}} = j\Big)P\big(Z_{T_{k-1}} = j\big) \\ \le \sum_{1 \le j \le m} E\big((T_1^* \wedge \tau_0)^2 \,|\, Z_0 = j\big)P\big(Z_{T_{k-1}} = j\big)\,.\end{aligned}$$

We will next bound the time $T_1^* \wedge \tau_0$, starting from $j \in \{1, \dots, m\}$.

Lemma 7.15 We suppose that $\sigma e^{-a} > 1$. For any $\delta > 0$, there exist $m_0, c > 0$, depending on δ, such that, for $m \ge m_0$, for $j \in \{1, \dots, m\}$,

$$P\big(m(\rho^* - \delta) < Z_{\lfloor c \ln m \rfloor} < m(\rho^* + \delta) \,|\, Z_0 = j\big) \ge \frac{1}{m^{c \ln m}}\,.$$

Proof. Using lemma 7.8, there exists $c > 0$ such that, for m large enough,

$$P\big(Z_{\lfloor c \ln m \rfloor} \le m(\rho^* - \delta) \,|\, Z_0 = 1\big) \le 1 - \frac{1}{m^{c \ln m}}\,.$$

Proceeding as in lemma 7.10, we obtain that there exist $h, c' > 0$ such that

$$P\big(Z_h \geq m(\rho^* + \delta) \,|\, Z_0 = m\big) \,\leq\, \exp(-c'm)\,.$$

We have then

$$\begin{aligned} P\big(Z_{\lfloor c\ln m\rfloor} \geq m(\rho^* + \delta) \,|\, Z_0 = m\big) \,=\, & \\ \sum_{j\in\{\,1,\dots,m\,\}} & P\big(Z_{\lfloor c\ln m\rfloor} \geq m(\rho^* + \delta),\; Z_{\lfloor c\ln m\rfloor - h} = j \,|\, Z_0 = m\big) \\ = \sum_{j\in\{\,1,\dots,m\,\}} & P\big(Z_h \geq m(\rho^* + \delta) \,|\, Z_0 = j\big) P\big(Z_{\lfloor c\ln m\rfloor - h} = j \,|\, Z_0 = m\big) \\ & \qquad\qquad\qquad \leq\, \exp(-c'm)\,. \end{aligned}$$

Using the monotonicity of $(Z_n)_{n\geq 0}$, we have then

$$\begin{aligned} & P\big(Z_{\lfloor c\ln m\rfloor} \notin]m(\rho^* - \delta), m(\rho^* + \delta)[\,|\, Z_0 = j\big) \\ & \;\leq\, P\big(Z_{\lfloor c\ln m\rfloor} \leq m(\rho^* - \delta) \,|\, Z_0 = j\big) \,+\, P\big(Z_{\lfloor c\ln m\rfloor} \geq m(\rho^* + \delta) \,|\, Z_0 = j\big) \\ & \;\leq\, P\big(Z_{\lfloor c\ln m\rfloor} \leq m(\rho^* - \delta) \,|\, Z_0 = 1\big) \,+\, P\big(Z_{\lfloor c\ln m\rfloor} \geq m(\rho^* + \delta) \,|\, Z_0 = m\big) \\ & \qquad\qquad\qquad\qquad \leq\, 1 - \frac{1}{m^{c\ln m}} + \exp(-c'm)\,. \end{aligned}$$

This estimate is uniform with respect to $j \in \{\,1, \dots, m\,\}$. □

Corollary 7.16 We suppose that $\sigma e^{-a} > 1$. For any $\delta > 0$, there exist $m_0, c > 0$, depending on δ, such that, for $m \geq m_0$, for $j \in \{\,1, \dots, m\,\}$,

$$\forall n \geq 0 \qquad P\big(T_1^* \wedge \tau_0 \geq n\lfloor c\ln m\rfloor \,\big|\, Z_0 = j\big) \,\leq\, \Big(1 - \frac{1}{m^{c\ln m}}\Big)^n\,.$$

Proof. We proceed as in corollary 7.11 to obtain this inequality. We divide the interval $\{\,0, \dots, n\lfloor c\ln m\rfloor\,\}$ into subintervals of length $\lfloor c\ln m\rfloor$ and we use repeatedly the estimate of lemma 7.15. □

By corollary 7.16, we have, for any $j \in \{\,1, \dots, m\,\}$,

$$\begin{aligned} E\big((T_1^* \wedge \tau_0)^2 \,\big|\, Z_0 = j\big) \,&=\, \sum_{k\geq 1} P\big(T_1^* \wedge \tau_0 \geq \sqrt{k} \,\big|\, Z_0 = j\big) \\ &\leq \sum_{k\geq 1} \Big(1 - \frac{1}{m^{c\ln m}}\Big)^{\Big\lfloor \frac{\sqrt{k}}{\lfloor c\ln m\rfloor} \Big\rfloor}\,. \end{aligned}$$

Let us set

$$\alpha = 1 - \frac{1}{m^{c\ln m}}\,, \qquad t = \lfloor c\ln m\rfloor\,.$$

We have

$$\sum_{k\geq 1}\alpha^{\lfloor\sqrt{k}/t\rfloor}\ \leq\ \sum_{k\geq 1}\alpha^{\sqrt{k}/t-1}\ \leq\ \int_0^\infty \alpha^{\sqrt{x}/t-1}\,dx\ =\ \frac{t^2}{\alpha(\ln\alpha)^2}\int_0^\infty e^{\sqrt{y}}\,dy\,,$$

therefore, asymptotically, for any j,

$$E\big((T_1^*\wedge\tau_0)^2\,\big|\,Z_0=j\big)\ \leq m^{3c\ln m}\,.$$

Reporting in the previous inequalities, we have

$$\forall k\geq 1\qquad E\Big(1_{k\leq K(\tau_0)}\big(T_k^*\wedge\tau_0-T_{k-1}\big)^2\Big)\ \leq\ m^{3c\ln m}$$

and, using the estimate of lemma 7.14,

$$E\Bigg(1_{\tau_0\leq t_m^\eta}\sum_{k=1}^{K(\tau_0)}\Big(T_k^*\wedge\tau_0-T_{k-1}\Big)\Bigg)\ \leq\ m^{3c\ln m}\sum_{k\geq 0}P\big(K(t_m^\eta)>k\big)^{1/2}$$

$$\leq\ m^{3c\ln m}\Big(t_m^\eta\,\exp(-c'm/3)+\sum_{k\geq t_m^\eta\,\exp(-c'm/3)}\Big(\frac{(t_m^\eta)^k}{k!}\exp(-c'mk)\Big)^{1/2}\Big)$$

$$\leq\ m^{3c\ln m}\Big(t_m^\eta\,\exp(-c'm/3)+\sum_{k\geq 1}\exp\Big(\frac{k}{2}-c'm\frac{k}{3}\Big)\Big)\,.$$

To get the last inequality, we have used that $k!\geq(k/e)^k$, whence, for $k\geq t_m^\eta\,\exp(-c'm/3)$,

$$\frac{(t_m^\eta)^k}{k!}\ \leq\ \Big(\frac{et_m^\eta}{k}\Big)^k\ \leq\ \exp(k+c'mk/3)\,.$$

We choose η such that $\eta<c'/3$. The above inequality implies that

$$\limsup_{\substack{\ell,m\to\infty,\,q\to 0\\ \ell q\to a}}\frac{1}{m}\ln E\Bigg(1_{\tau_0\leq t_m^\eta}\sum_{k=1}^{K(\tau_0)}\Big(T_k^*\wedge\tau_0-T_{k-1}\Big)\Bigg)\ \leq\ V(\rho^*,0)+\eta-\frac{c'}{3}\,.$$

All these estimates, together with proposition 7.6, imply that, asymptotically, uniformly with respect to $i\in\{\,1,\dots,m\,\}$,

$$\Bigg|E\Bigg(\sum_{n=0}^{\tau_0}f\Big(\frac{Z_n}{m}\Big)\,\Big|\,Z_0=i\Bigg)-f(\rho^*)\,E\big(\tau_0\,|\,Z_0=i\big)\Bigg|\ \leq\ 3\varepsilon E\big(\tau_0\,|\,Z_0=i\big)\,.$$

This achieves the proof of proposition 7.12. $\square$

8 The neutral phase

We denote by $\mathcal{N}$ the set of the populations which do not contain the master sequence w^*, i.e.,

$$\mathcal{N} = \Big(\mathcal{A}^\ell \setminus \{\, w^* \,\}\Big)^m .$$

Since we deal with the sharp peak landscape, the transition mechanism of the process restricted to the set $\mathcal{N}$ is neutral. We consider a Wright–Fisher process $(X_n)_{n\geq 0}$ starting from a population of $\mathcal{N}$. We wish to evaluate the first time when a master sequence appears in the population:

$$\tau_* \,=\, \inf\big\{\, n\geq 0 : X_n \not\in \mathcal{N}\,\big\} .$$

We call the time τ_* the discovery time. Until the time τ_*, the process evolves in $\mathcal{N}$ and the dynamics of the Wright–Fisher model in $\mathcal{N}$ does not depend on σ. In particular, the law of the discovery time τ_* is the same for the Wright–Fisher model with $\sigma > 1$ and the neutral Wright–Fisher model with $\sigma = 1$. Therefore, we compute the estimates for the latter model.

Neutral hypothesis. Throughout this section, we suppose that $\sigma = 1$.

8.1 Ancestral lines

It is a classical fact that neutral evolutionary processes are much easier to analyze than evolutionary processes with selection. The main reason is that the mutation mechanism and the sampling mechanism can be decoupled. For instance, it is possible to compute explicitly the law of a chromosome in the population at time n. Let μ_0 be an exchangeable probability distribution on $\big(\mathcal{A}^\ell\big)^m$. Let $(X_n)_{n\geq 0}$ be the neutral Wright–Fisher process with mutation matrix M and initial law μ_0. Let ν_0 be the component marginal of μ_0:

$$\forall u\in\mathcal{A}^\ell \qquad \nu_0(u) \,=\, \mu_0\big(\{\, x\in \big(\mathcal{A}^\ell\big)^m : x(1)=u\,\}\big) .$$

Let $(W_n)_{n\geq 0}$ be a Markov chain with state space $\mathcal{A}^\ell$, having for transition matrix the mutation matrix M and with initial law ν_0.

Proposition 8.1 Let $i\in\{\,1,\dots,m\,\}$. For any $n\geq 0$, the law of the i–th chromosome of X_n is equal to the law of W_n.

Proof. We do the proof by induction over n. The result holds for $n=0$. Suppose that it has been proved until time n. Let $i\in\{\,1,\dots,m\,\}$. We have, for any $u \in \mathcal{A}^\ell$,

$$\begin{aligned} P(X_{n+1}(i) = u) \;&=\; \sum_{x \in (\mathcal{A}^\ell)^m} P(X_{n+1}(i) = u,\, X_n = x) \\ &=\; \sum_{x \in (\mathcal{A}^\ell)^m} P(X_{n+1}(i) = u \,|\, X_n = x) P(X_n = x)\,. \end{aligned}$$

Yet we have

$$P(X_{n+1}(i) = u \,|\, X_n = x) \;=\; \frac{1}{m} \sum_{1 \le j \le m} M(x(j), u)\,.$$

Thus

$$\begin{aligned} P(X_{n+1}(i) = u) \;&=\; \sum_{x \in (\mathcal{A}^\ell)^m} \frac{1}{m} \sum_{1 \le j \le m} M(x(j), u) P(X_n = x) \\ &=\; \sum_{v \in \mathcal{A}^\ell} \frac{1}{m} \sum_{1 \le j \le m} M(v, u) P(X_n(j) = v)\,. \end{aligned}$$

By the induction hypothesis,

$$\forall v \in \mathcal{A}^\ell \quad \forall j \in \{\, 1, \dots, m \,\} \qquad P(X_n(j) = v) \;=\; P(W_n = v)\,,$$

whence

$$P(X_{n+1}(i) = u) \;=\; \sum_{v \in \mathcal{A}^\ell} P(W_n = v) M(v, u) \;=\; P(W_{n+1} = u)\,.$$

The result still holds at time $n + 1$. □

We perform next a similar computation to obtain the law of an ancestral line. Let us first define an ancestral line. For $i \in \{\, 1, \dots, m \,\}$ and $n \ge 1$, we denote by $\mathcal{I}(i, n, n-1)$ the index of the ancestor at time $n - 1$ of the i–th chromosome at time n. More precisely, if the i–th chromosome of the population at time n has been obtained by replicating the j–th chromosome of the population at time $n - 1$, then $\mathcal{I}(i, n, n-1) = j$. For $s \le n$, the index $\mathcal{I}(i, n, s)$ of the ancestor at time s of the i–th chromosome at time n is then defined recursively with the help of the following formula:

$$\mathcal{I}(i, n, s) \;=\; \mathcal{I}(\mathcal{I}(i, n, n-1), n-1, s)\,.$$

The ancestor at time s of the i–th chromosome at time n is the chromosome

$$\text{ancestor}(i, n, s) \;=\; X_s(\mathcal{I}(i, n, s))\,.$$

The ancestral line of the i–th chromosome at time n is the sequence of its ancestors until time 0,

$$(\text{ancestor}(i, n, s),\, 0 \le s \le n) \;=\; (X_s(\mathcal{I}(i, n, s)),\, 0 \le s \le n)\,.$$

Proposition 8.2 Let $i \in \{1,\dots,m\}$. For any $n \geq 0$, the law of the ancestral line $(\text{ancestor}(i,n,s),\, 0 \leq s \leq n)$ of the i–th chromosome of X_n is equal to the law of $(W_0,\dots,W_n)$.

Proof. We do the proof by induction over n. The result is true at rank $n=0$. Suppose it has been proved until time n. Let $i \in \{1,\dots,m\}$ and let $u_0,\dots,u_{n+1} \in \mathcal{A}^\ell$. We compute

$$
\begin{aligned}
&P\big(\text{ancestor}(i,n+1,s) = u_s,\, 0 \leq s \leq n+1\big)\\
&\quad = \sum_{x \in (\mathcal{A}^\ell)^m} \sum_{1 \leq j \leq m} P\begin{pmatrix} X_{n+1}(i) = u_{n+1},\, \mathcal{I}(i,n+1,n) = j \\ X_n = x,\, \text{ancestor}(j,n,s) = u_s,\, 0 \leq s \leq n \end{pmatrix}\\
&\quad = \sum_{x \in (\mathcal{A}^\ell)^m} \sum_{1 \leq j \leq m} P\left(\begin{matrix} X_{n+1}(i) = u_{n+1} \\ \mathcal{I}(i,n+1,n) = j \end{matrix} \,\middle|\, \begin{matrix} \text{ancestor}(j,n,s) = u_s \\ 0 \leq s \leq n,\, X_n = x \end{matrix} \right)\\
&\qquad\qquad \times P\begin{pmatrix} \text{ancestor}(j,n,s) = u_s \\ 0 \leq s \leq n,\, X_n = x \end{pmatrix}.
\end{aligned}
$$

Since we deal with the neutral process, we have

$$
\begin{aligned}
&P\left(\begin{matrix} X_{n+1}(i) = u_{n+1} \\ \mathcal{I}(i,n+1,n) = j \end{matrix} \,\middle|\, \begin{matrix} \text{ancestor}(j,n,s) = u_s \\ 0 \leq s \leq n,\, X_n = x \end{matrix} \right)\\
&\qquad = P\left(\begin{matrix} X_{n+1}(i) = u_{n+1} \\ \mathcal{I}(i,n+1,n) = j \end{matrix} \,\middle|\, X_n = x \right) = \frac{1}{m} M(x(j), u_{n+1}).
\end{aligned}
$$

Reporting in the previous inequality, we get

$$
\begin{aligned}
&P\big(\text{ancestor}(i,n+1,s) = u_s,\, 0 \leq s \leq n+1\big)\\
&\qquad = \sum_{x \in (\mathcal{A}^\ell)^m} \sum_{1 \leq j \leq m} \frac{1}{m} M(x(j), u_{n+1})\, P\begin{pmatrix} \text{ancestor}(j,n,s) = u_s \\ 0 \leq s \leq n,\, X_n = x \end{pmatrix}\\
&\qquad = \sum_{1 \leq j \leq m} \frac{1}{m} M(u_n, u_{n+1})\, P\big(\text{ancestor}(i,n,s) = u_s,\, 0 \leq s \leq n\big).
\end{aligned}
$$

By the induction hypothesis, we have

$$
P\big(\text{ancestor}(i,n,s) = u_s,\, 0 \leq s \leq n\big) = P\big(W_0 = u_0,\dots,W_n = u_n\big).
$$

Therefore

$$
\begin{aligned}
&P\big(\text{ancestor}(i,n+1,s) = u_s,\, 0 \leq s \leq n+1\big) =\\
&\qquad P\big(W_{n+1} = u_{n+1} \,|\, W_n = u_n\big) P\big(W_0 = u_0,\dots,W_n = u_n\big)\\
&\qquad\qquad = P\big(W_0 = u_0,\dots,W_{n+1} = u_{n+1}\big)
\end{aligned}
$$

and the induction step is completed. □

8.2 Mutation dynamics

We consider a Markov chain $(Y_n)_{n\geq 0}$ with state space $\{\,0,\dots,\ell\,\}$ and having for transition matrix the lumped mutation matrix M_H. In this section, we recall some properties and estimates concerning the Markov chain $(Y_n)_{n\geq 0}$. We refer to the corresponding section of [3] for the detailed proofs. The Markov chain $(Y_n)_{n\geq 0}$ is monotone. We denote by $\mathcal{B}$ the binomial law $\mathcal{B}(\ell, 1-1/\kappa)$ with parameters ℓ and $1-1/\kappa$, i.e.,

$$\forall b\in\{\,0,\dots,\ell\,\}\qquad \mathcal{B}(b)\,=\,\binom{\ell}{b}\Big(1-\frac{1}{\kappa}\Big)^b\Big(\frac{1}{\kappa}\Big)^{\ell-b}.$$

The matrix M_H is reversible with respect to the binomial law $\mathcal{B}$. This binomial law is the invariant probability measure of the Markov chain $(Y_n)_{n\geq 0}$. When ℓ grows, the law $\mathcal{B}$ concentrates exponentially fast in a neighborhood of its mean $\ell_\kappa=\ell(1-1/\kappa)$. We restate next without proofs several inequalities and estimates proved in [3].

Lemma 8.3 For any $b\in\{\,0,\dots,m\,\}$ and $n\geq 0$, we have

$$P\big(Y_n=0\,|\,Y_0=b\big)\,\leq\,\frac{\mathcal{B}(0)}{\mathcal{B}(b)}\,.$$

Lemma 8.4 For $b\leq \ell/2$, we have

$$\frac{1}{\kappa^\ell}\left(\frac{\ell}{2b}\right)^b\,\leq\,\mathcal{B}(b)\,\leq\,\frac{\ell^b}{\kappa^{\ell-b}}\,.$$

For $b\in\{\,0,\dots,\ell\,\}$, we define the hitting time $\tau(b)$ of $\{\,b,\dots,\ell\,\}$ by

$$\tau(b)\,=\,\inf\,\big\{\,n\geq 0: Y_n\geq b\,\big\}\,.$$

Lemma 8.5 For n such that

$$n\,>\,\frac{2b}{p\,\big(\ell_\kappa-b\big)}\,,$$

we have

$$P\big(\tau(b)>n\,|\,Y_0=0\big)\,\leq\,\frac{4\ell}{np\,\big(\ell_\kappa-b\big)^2}\,.$$

Proposition 8.6 We suppose that $\ell\to+\infty$, $q\to 0$, $\ell q\to a\in]0,+\infty[$. Asymptotically, we have

$$P\big(\tau(\ell_\kappa)\leq \ell^2\,|\,Y_0=0\big)\,\geq\,\Big(1-\frac{5}{a(\ln\ell)^2}\Big)\Big(\frac{p}{\kappa}\Big)^{\ln\ell}e^{-2a}\,.$$

Proposition 8.7 We suppose that $\ell \to +\infty\,, q \to 0\,, \ell q \to a \in]0,+\infty[$. Asymptotically, we have

$$\forall n \geq \sqrt{\ell} \qquad P\big(Y_n \geq \ln \ell \,|\, Y_0 = 0\big) \,\geq\, 1 - \frac{6}{a\sqrt{\ell}}\,.$$

Proposition 8.8 We suppose that $\ell \to +\infty\,, q \to 0\,, \ell q \to a \in]0,+\infty[$. Let $\varepsilon \in]0,1/4[$. Asymptotically, we have

$$\forall n \geq \frac{\ell}{a\varepsilon} \qquad P\big(Y_n \geq \ell_\kappa(1-4\varepsilon) \,|\, Y_0 = 0\big) \,\geq\, 1 - \frac{6}{\ell\varepsilon}\,.$$

We define

$$\tau_0 \,=\, \inf\,\big\{\, n \geq 0 : Y_n = 0\,\big\}\,.$$

Proposition 8.9 For any $a \in]0,+\infty[$,

$$\limsup_{\substack{\ell\to\infty,\, q\to 0\\ \ell q\to a}} \,\frac{1}{\ell} \ln E(\tau_0 \,|\, Y_0 = \ell) \,\leq\, \ln \kappa\,.$$

8.3 Discovery time

The dynamics of the processes $(O_n^\ell)_{n\geq 0}$, $(O_n^1)_{n\geq 0}$ in $\mathcal{N}$ are the same as the original process $(O_n)_{n\geq 0}$, therefore we can use the original process to compute their discovery times. Letting

$$\tau^{*,\ell} \,=\, \inf\,\big\{\, n \geq 0 : O_n^\ell \in \mathcal{W}^*\,\big\}\,, \qquad \tau^{*,1} \,=\, \inf\,\big\{\, n \geq 0 : O_n^1 \in \mathcal{W}^*\,\big\}\,,$$
$$\tau^* \,=\, \inf\,\big\{\, n \geq 0 : O_n \in \mathcal{W}^*\,\big\}\,,$$

we have

$$E\big(\tau^{*,\ell} \,|\, O_0^\ell = o_{\text{exit}}^\ell\big) \,=\, E\big(\tau^* \,|\, O_0 = (0,0,0,\dots,m)\big)\,,$$
$$E\big(\tau^{*,1} \,|\, O_0^1 = o_{\text{exit}}^1\big) \,=\, E\big(\tau^* \,|\, O_0 = (0,m,0,\dots,0)\big)\,.$$

In addition, the law of the discovery time τ^* is the same for the distance process and the occupancy process. With a slight abuse of notation, we let

$$\tau^* \,=\, \inf\,\big\{\, n \geq 0 : D_n \in \mathcal{W}^*\,\big\}\,.$$

Notation. For $b \in \{\,0,\dots,\ell\,\}$, we denote by $(b)^m$ the vector column whose components are all equal to b:

$$(b)^m \,=\, \begin{pmatrix} b\\ \vdots\\ b \end{pmatrix}\,.$$

We have

$$E\big(\tau^*\,|\,O_0=(0,0,\dots,0,m)\big) \,=\, E\big(\tau^*\,|\,D_0=(\ell)^m\big)\,,$$
$$E\big(\tau^*\,|\,O_0=(0,m,0,\dots,0)\big) \,=\, E\big(\tau^*\,|\,D_0=(1)^m\big)\,.$$

We will carry out the estimates of τ^* for the distance process $(D_n)_{n\geq 0}$. Notice that the case $\alpha=+\infty$ is not covered by the result of next proposition. This case will be handled separately, with the help of the intermediate inequality of corollary 8.11.

Proposition 8.10 Let $a\in]0,+\infty[$ and $\alpha\in[0,+\infty[$. For any $d\in\mathcal{N}$,

$$\lim_{\substack{\ell,m\to\infty,\,q\to 0\\ \ell q\to a,\,\frac{m}{\ell}\to\alpha}} \frac{1}{\ell}\ln E\big(\tau^*\,|\,D_0=d\big) \,=\, \ln\kappa\,.$$

Proof. Since we are in the neutral case $\sigma=1$, then, by corollary 5.4, the distance process $(D_n)_{n\geq 0}$ is monotone. Therefore, for any $d\in\mathcal{N}$, we have

$$E\big(\tau^*\,|\,D_0=(1)^m\big) \,\leq\, E\big(\tau^*\,|\,D_0=d\big) \,\leq\, E\big(\tau^*\,|\,D_0=(\ell)^m\big)\,.$$

As in the section 8.2, we consider a Markov chain $(Y_n)_{n\geq 0}$ with state space $\{\,0,\dots,\ell\,\}$ and having for transition matrix the lumped mutation matrix M_H. Let us look at the distance process at time n starting from $(\ell)^m$. From proposition 8.1, we know that the law of the i–th chromosome in D_n is the same as the law of Y_n starting from ℓ. The main difficulty is that, because of the replication events, the m chromosomes present at time n are not independent, nor are their genealogical lines. However, this dependence does not improve significantly the efficiency of the search mechanism, as long as the population is in the neutral space $\mathcal{N}$. To bound the discovery time τ^* from above, we consider the time needed for a single chromosome to discover the master sequence w^*, that is

$$\tau_0 \,=\, \inf\,\big\{\,n\geq 0: Y_n=0\,\big\}$$

and we remark that, if the master sequence has not been discovered until time n in the distance process, that is,

$$\forall t\leq n\quad \forall i\in\{\,1,\dots,m\,\}\qquad D_t(i)\geq 1\,,$$

then certainly the ancestral line of any chromosome present at time n does not contain the master sequence. By proposition 8.2, the ancestral line of any chromosome present at time n has the same law as $Y_0,\dots,Y_n$. From the previous observations, we conclude that

$$\forall n\geq 0\qquad P\big(\tau^*>n\,|\,D_0=(\ell)^m\big) \,\leq\, m\,P(\tau_0>n\,|\,Y_0=\ell)\,.$$

Summing this inequality over $n\geq 0$, we obtain the following upper bound on the discovery time.

Corollary 8.11 Let τ_0 be the hitting time of 0 for the process $(Y_n)_{n\geq 0}$. For any $d \in \mathcal{N}$, any $m \geq 1$, we have

$$E\big(\tau^* \,|\, D_0 = d\big) \;\leq\; m\, E(\tau_0 \,|\, Y_0 = \ell)\,.$$

With the help of proposition 8.9, we conclude that

$$\limsup_{\substack{\ell,m\to\infty,\, q\to 0\\ \ell q\to a,\, \frac{m}{\ell}\to\alpha}} \frac{1}{\ell} \ln E\big(\tau^* \,|\, D_0 = d\big) \;\leq\; \ln\kappa\,.$$

The harder part is to bound the discovery time τ^* from below. The main difficulty to obtain the adequate lower bound on τ^* is that the process starts very close to the master sequence, hence the probability of creating quickly a master sequence is not very small. Our strategy consists in exhibiting a scenario in which the whole population is driven into a neighborhood of the equilibrium ℓ_κ. Once the whole population is close to ℓ_κ, the probability to create a master sequence in a short time is of order $1/\kappa^\ell$, thus it requires a time of order κ^ℓ. The key point is to design a scenario whose probability is much larger than $1/\kappa^\ell$. Indeed, the mean discovery time is bounded from below by the probability of the scenario multiplied by κ^ℓ. We rely on the following scenario. First we ensure that until time $\ell^{3/4}$, no mutation can recreate the master sequence. This implies that $\tau^* > \ell^{3/4}$. Let us then look at the population at time $\ell^{3/4}$. Each chromosome present at this time has undergone an evolution whose law is the same as the mutation dynamics studied in section 8.2. The initial drift of the mutation dynamics is quite violent, therefore at time $\ell^{3/4}$, it is very unlikely that a chromosome evolving with the mutation dynamics is still in $\{\,0,\cdots,\ln\ell\,\}$. The problem is that the chromosomes are not independent. We take care of this problem with the help of correlation inequalities. Thus, at time $\ell^{3/4}$, in this scenario, all the chromosomes of the population are at distance larger than $\ln\ell$ from the master sequence. We wait next until time ℓ^2. Because of the mutation drift, a chromosome starting at $\ln\ell$ has a very low probability of hitting 0 before time ℓ^2. Thus the process is very unlikely to discover the master sequence before time ℓ^2. Arguing again as above, we obtain that, for any $\varepsilon > 0$, at time ℓ^2, it is very unlikely that a chromosome evolving with the mutation dynamics is still in $\{\,0,\cdots,\ell_\kappa(1-\varepsilon)\,\}$. Thus, according to this scenario, we have $\tau^* > \ell^2$ and

$$\forall i \in \{\,1,\dots,m\,\} \qquad D_{\ell^2}(i) \;\geq\; \ell_\kappa(1-\varepsilon)\,.$$

Let us precise next the scenario and the corresponding estimates. We suppose that the distance process starts from $(1)^m$ and we will estimate the probability of a specific scenario leading to a discovery time close to κ^ℓ.

Let $\mathcal{E}$ be the event

$$\mathcal{E} \,=\, \Big\{\, \forall n \leq \ell^{3/4} \quad \forall i \in \{\,1,\dots,m\,\} \quad U_n^{i,1} > p/\kappa \,\Big\}\,.$$

If the event $\mathcal{E}$ occurs, then, until time $\ell^{3/4}$, none of the mutation events in the process $(D_n)_{n\geq 0}$ can create a master sequence. Indeed, on $\mathcal{E}$,

$$\begin{aligned}\forall b \in \{\,1,\dots,\ell\,\} \quad \forall n \leq \ell^{3/4} \quad \forall i \in \{\,1,\dots,m\,\}&\\ \mathcal{M}_H(b, U_n^{i,1},\dots,U_n^{i,\ell}) \,\geq\, \mathcal{M}_H(1, U_n^{i,1},\dots,U_n^{i,\ell})&\\ \geq\, 1 + \sum_{l=2}^{\ell} 1_{U_n^{i,l} > 1-p(1-1/\kappa)} \,\geq\, 1\,.&\end{aligned}$$

Thus, on the event $\mathcal{E}$, we have $\tau^* \geq \ell^{3/4}$. The probability of $\mathcal{E}$ is

$$P(\mathcal{E}) \,=\, \Big(1 - \frac{p}{\kappa}\Big)^{m\ell^{3/4}}\,.$$

Let $\varepsilon > 0$. We suppose that the process starts from $(1)^m$ and we estimate the probability

$$\begin{aligned} P\big(\tau^* > \kappa^{\ell(1-\varepsilon)}\big) \,&\geq\, P\big(\tau^* > \kappa^{\ell(1-\varepsilon)},\, \mathcal{E}\big)\\ &\geq\, P\Big(\forall n \in \{\,\ell^{3/4},\dots,\kappa^{\ell(1-\varepsilon)}\,\} \quad D_n \in \mathcal{N},\, \mathcal{E}\Big)\\ &=\, \sum_{d\in\mathcal{N}} P\Big(\forall n \in \{\,\ell^{3/4},\dots,\kappa^{\ell(1-\varepsilon)}\,\} \quad D_n \in \mathcal{N},\, D_{\ell^{3/4}} = d,\, \mathcal{E}\Big)\\ &\geq\, \sum_{d\geq(\ln \ell)^m} P\Big(\forall n \in \{\,\ell^{3/4},\dots,\kappa^{\ell(1-\varepsilon)}\,\} \quad D_n \in \mathcal{N} \,|\, D_{\ell^{3/4}} = d,\, \mathcal{E}\Big)\\ &\qquad\qquad \times P(D_{\ell^{3/4}} = d,\, \mathcal{E})\,.\end{aligned}$$

Using the Markov property, we have

$$\begin{aligned} P\Big(\forall n \in \{\,\ell^{3/4},\dots,\kappa^{\ell(1-\varepsilon)}\,\} \quad D_n \in \mathcal{N} \,|\, D_{\ell^{3/4}} = d,\, \mathcal{E}\Big)&\\ =\, P\Big(\forall n \in \{\,0,\dots,\kappa^{\ell(1-\varepsilon)} - \ell^{3/4}\,\} \quad D_n \in \mathcal{N} \,|\, D_0 = d\Big)&\\ =\, P\Big(\tau^* > \kappa^{\ell(1-\varepsilon)} - \ell^{3/4} \,|\, D_0 = d\Big) \,\geq\, P\Big(\tau^* > \kappa^{\ell(1-\varepsilon)} \,|\, D_0 = d\Big)\,.&\end{aligned}$$

In the neutral case, by corollary 5.4, the distance process is monotone. Therefore, for $d \geq (\ln \ell)^m$,

$$P\big(\tau^* > \kappa^{\ell(1-\varepsilon)} \,|\, D_0 = d\big) \,\geq\, P\big(\tau^* > \kappa^{\ell(1-\varepsilon)} \,|\, D_0 = (\ln \ell)^m\big)\,.$$

Reporting in the previous sum, we get

$$P(\tau^* > \kappa^{\ell(1-\varepsilon)}) \,\geq\, P\big(\tau^* > \kappa^{\ell(1-\varepsilon)} \,|\, D_0 = (\ln \ell)^m\big)\, P\big(D_{\ell^{3/4}} \geq (\ln \ell)^m,\, \mathcal{E}\big)\,.$$

We first study the last term in the above inequality. The status of the process at time $\ell^{3/4}$ is a function of the random matrices

$$R_n \,=\, \big(S_n^i, U_n^{i,1}, \dots, U_n^{i,\ell}\big)_{1\leq i\leq m}\,, \qquad 1 \leq n \leq \ell^{3/4}\,.$$

We make an intermediate conditioning with respect to the variables S_n^i:

$$\begin{aligned} P\big(D_{\ell^{3/4}} &\geq (\ln \ell)^m,\, \mathcal{E}\big) \\ &= E\Big(P\big(D_{\ell^{3/4}} \geq (\ln \ell)^m,\, \mathcal{E} \,\big|\, S_n^i,\, 1 \leq i \leq m,\, 1 \leq n \leq \ell^{3/4}\big)\Big)\,. \end{aligned}$$

The variables S_n^i, $1 \leq i \leq m$, $1 \leq n \leq \ell^{3/4}$ being fixed, all the indices of the chromosomes selected for replication are fixed, and since the mutation map $\mathcal{M}_H(\cdot, u_1, \dots, u_\ell)$ is non–decreasing with respect to $u_1, \dots, u_\ell$, we see that the state of the process at time $\ell^{3/4}$ is a non–decreasing function of the variables

$$U_n^{i,1}, \dots, U_n^{i,\ell}, \qquad 1 \leq i \leq m\,, \quad 1 \leq n \leq \ell^{3/4}\,.$$

Thus the events $\mathcal{E}$ and $D_{\ell^{3/4}} \geq (\ln \ell)^m$ are both non–decreasing with respect to these variables. By the FKG inequality for a product measure (see section 5.3), we have

$$\begin{aligned} P\big(D_{\ell^{3/4}} \geq (\ln \ell)^m,\, \mathcal{E} \,\big|\, S_n^i,\, 1 \leq i \leq m,\, 1 \leq n \leq \ell^{3/4}\big) \,\geq& \\ P\big(D_{\ell^{3/4}} \geq (\ln \ell)^m \,\big|\, S_n^i,\, 1 \leq i \leq m,\, 1 \leq n \leq \ell^{3/4}\big)& \\ \times P\big(\mathcal{E} \,\big|\, S_n^i,\, 1 \leq i \leq m,\, 1 \leq n \leq \ell^{3/4}\big)&\,. \end{aligned}$$

Yet $\mathcal{E}$ does not depend on the variables S_n^i, therefore

$$P\big(\mathcal{E} \,\big|\, S_n^i,\, 1 \leq i \leq m,\, 1 \leq n \leq \ell^{3/4}\big) \,=\, P(\mathcal{E})\,.$$

Reporting in the conditioning, we obtain

$$\begin{aligned} P\big(D_{\ell^{3/4}} \geq (\ln \ell)^m,\, \mathcal{E}\big) \,\geq& \\ E\Big(P\big(D_{\ell^{3/4}} \geq (\ln \ell)^m \,\big|\, S_n^i,\, 1 \leq i \leq m,\, 1 \leq n \leq \ell^{3/4}\big)\, P(\mathcal{E})\Big)& \\ = P\big(D_{\ell^{3/4}} \geq (\ln \ell)^m\big)\, P\big(\mathcal{E}\big)&\,. \end{aligned}$$

By proposition 5.8, the distance process has positive correlations, therefore

$$\begin{aligned} P\big(D_{\ell^{3/4}} \geq (\ln \ell)^m\big) \,&=\, P\big(\forall i \in \{\,1, \dots, m\,\} \quad D_{\ell^{3/4}}(i) \geq \ln \ell\big) \\ &\geq \prod_{1\leq i\leq m} P\big(D_{\ell^{3/4}}(i) \geq \ln \ell\big)\,. \end{aligned}$$

From propositions 8.1 and 8.7,

$$\forall i \in \{\,1,\dots,m\,\} \qquad P\big(D_{\ell^{3/4}}(i) \geq \ln \ell\big) \;=\; P\big(Y_{\ell^{3/4}} \geq \ln \ell\big) \;\geq\; 1 - \frac{6}{a\sqrt{\ell}}\,.$$

Putting the previous estimates together, we have

$$P\big(D_{\ell^{3/4}} \geq (\ln \ell)^m,\, \mathcal{E}\big) \;\geq\; \Big(1 - \frac{6}{a\sqrt{\ell}}\Big)^m \Big(1 - \frac{p}{\kappa}\Big)^{m\ell^{3/4}}\,.$$

We study next

$$P\big(\tau^* > \kappa^{\ell(1-\varepsilon)} \,|\, D_0 = (\ln \ell)^m\big)\,.$$

We give first an estimate showing that a visit to 0 becomes very unlikely if the starting point is far from 0.

Lemma 8.12 For $b \in \{\,1,\dots,\ell\,\}$, we have

$$\forall n \geq 0 \qquad P\big(\tau^* \leq n \,|\, D_0 = (b)^m\big) \;\leq\; nm\frac{\mathcal{B}(0)}{\mathcal{B}(b)}\,.$$

Proof. Let $n \geq 0$ and $b \in \{\,1,\dots,\ell\,\}$. We write

$$\begin{aligned}
P\big(\tau^* \leq n \,|\, D_0 = (b)^m\big) \;=&\\
&P\big(\exists\, t \leq n \quad \exists\, i \in \{\,1,\dots,m\,\} \quad D_t(i) = 0 \,|\, D_0 = (b)^m\big)\\
&\qquad\leq \sum_{1\leq t\leq n}\sum_{1\leq i\leq m} P\big(D_t(i) = 0 \,|\, D_0 = (b)^m\big)\,.
\end{aligned}$$

By proposition 8.1, for any $t \geq 0$, any $i \in \{\,1,\dots,m\,\}$,

$$P\big(D_t(i) = 0 \,|\, D_0 = (b)^m\big) \;=\; P\big(Y_t = 0 \,|\, Y_0 = b\big)\,.$$

Using lemma 8.3, and putting together the previous inequalities, we get

$$P\big(\tau^* \leq n \,|\, D_0 = (b)^m\big) \;\leq\; nm\frac{\mathcal{B}(0)}{\mathcal{B}(b)}$$

as requested. $\square$

Let $\varepsilon' > 0$. Now

$$\begin{aligned}
&P\Big(\tau^* > \kappa^{\ell(1-\varepsilon)} \,\Big|\, D_0 = (\ln \ell)^m\Big)\\
&\qquad\geq\; P\Big(\tau^* > \ell^2,\, D_n \in \mathcal{N} \text{ for } \ell^2 \leq n \leq \kappa^{\ell(1-\varepsilon)} \,\Big|\, D_0 = (\ln \ell)^m\Big)\\
&\qquad=\; \sum_{d\in\mathcal{N}} P\Big(\begin{matrix}\tau^* > \ell^2,\, D_{\ell^2} = d\\ D_n \in \mathcal{N} \text{ for } \ell^2 \leq n \leq \kappa^{\ell(1-\varepsilon)}\end{matrix} \,\Big|\, D_0 = (\ln \ell)^m\Big)\\
&\geq \sum_{d\geq(\ell_\kappa(1-\varepsilon'))^m} P\Big(D_n \in \mathcal{N} \text{ for } \ell^2 \leq n \leq \kappa^{\ell(1-\varepsilon)} \,\Big|\, \tau^* > \ell^2,\, D_{\ell^2} = d\Big)\\
&\qquad\qquad\qquad \times P\Big(\tau^* > \ell^2,\, D_{\ell^2} = d \,\Big|\, D_0 = (\ln \ell)^m\Big)\,.
\end{aligned}$$

Using the Markov property and the monotonicity of the process $(D_n)_{n\geq 0}$, we have for $d \geq (\ell_\kappa(1-\varepsilon'))^m$,

$$
\begin{aligned}
P\Big(D_n \in \mathcal{N} \text{ for } \ell^2 \leq n \leq \kappa^{\ell(1-\varepsilon)} \,\Big|\, \tau^* > \ell^2,\, D_{\ell^2} = d\Big) \\
= P\Big(\forall n \in \{\,0,\dots,\kappa^{\ell(1-\varepsilon)} - \ell^2\,\} \quad D_n \in \mathcal{N} \,\Big|\, D_0 = d\Big) \\
= P\Big(\tau^* > \kappa^{\ell(1-\varepsilon)} - \ell^2 \,\Big|\, D_0 = d\Big) \geq P\Big(\tau^* > \kappa^{\ell(1-\varepsilon)} \,\Big|\, D_0 = d\Big) \\
\geq P\big(\tau^* > \kappa^{\ell(1-\varepsilon)} \,\big|\, D_0 = (\ell_\kappa(1-\varepsilon'))^m\big)\,.
\end{aligned}
$$

Reporting in the previous sum, we get

$$
\begin{aligned}
P\Big(\tau^* > \kappa^{\ell(1-\varepsilon)} \,\Big|\, D_0 = (\ln \ell)^m\Big) \geq P\big(\tau^* > \kappa^{\ell(1-\varepsilon)} \,\big|\, D_0 = (\ell_\kappa(1-\varepsilon'))^m\big) \\
\times P\Big(\tau^* > \ell^2,\, D_{\ell^2} \geq (\ell_\kappa(1-\varepsilon'))^m \,\Big|\, D_0 = (\ln \ell)^m\Big)\,.
\end{aligned}
$$

We first take care of the last probability. We write

$$
\begin{aligned}
P\Big(\tau^* > \ell^2,\, D_{\ell^2} \geq (\ell_\kappa(1-\varepsilon'))^m \,\Big|\, D_0 = (\ln \ell)^m\Big) \geq \\
P\Big(D_{\ell^2} \geq (\ell_\kappa(1-\varepsilon'))^m \,\Big|\, D_0 = (\ln \ell)^m\Big) - P\Big(\tau^* \leq \ell^2 \,\Big|\, D_0 = (\ln \ell)^m\Big)\,.
\end{aligned}
$$

To control the last term, we use the inequality of lemma 8.12 with $n = \ell^2$ and $b = \ln \ell$:

$$
P\big(\tau^* \leq \ell^2 \,|\, D_0 = (\ln \ell)^m\big) \leq \ell^2 m \frac{\mathcal{B}(0)}{\mathcal{B}(\ln \ell)}\,.
$$

By lemma 8.4, we have

$$
\frac{\mathcal{B}(0)}{\mathcal{B}(\ln \ell)} \leq \Big(\frac{2\ln \ell}{\ell}\Big)^{\ln \ell}
$$

whence

$$
P\big(\tau^* \leq \ell^2 \,|\, D_0 = (\ln l)^m\big) \leq \ell^2 m \Big(\frac{2\ln \ell}{\ell}\Big)^{\ln \ell}\,.
$$

For the other term, we use the monotonicity of the process $(D_n)_{n\geq 0}$ and the fact that it has positive correlations (by corollary 5.4) to get

$$
\begin{aligned}
P\Big(D_{\ell^2} \geq (\ell_\kappa(1-\varepsilon'))^m \,\Big|\, D_0 = (\ln \ell)^m\Big) \\
\geq P\Big(D_{\ell^2} \geq (\ell_\kappa(1-\varepsilon'))^m \,\Big|\, D_0 = (0)^m\Big) \\
\geq \prod_{1\leq i\leq m} P\Big(D_{\ell^2}(i) \geq \ell_\kappa(1-\varepsilon') \,\Big|\, D_0 = (0)^m\Big)\,.
\end{aligned}
$$

By proposition 8.1 and 8.8, for $i \in \{ 1,\dots,m \}$,

$$P\Big(D_{\ell^2}(i) \geq \ell_\kappa(1-\varepsilon') \,\Big|\, D_0 = (0)^m\Big) \\ \qquad = \; P\big(Y_{\ell^2} \geq \ell_\kappa(1-\varepsilon') \,|\, Y_0 = 0\big) \;\geq\; 1 - \frac{24}{\ell\varepsilon'}\,.$$

Putting together the previous estimates, we obtain

$$P\Big(\tau^* > \ell^2,\, D_{\ell^2} \geq (\ell_\kappa(1-\varepsilon'))^m \,\Big|\, D_0 = (\ln\ell)^m\Big) \;\geq \\ \Big(1-\frac{24}{\ell\varepsilon'}\Big)^m \;-\; \ell^2 m\Big(\frac{2\ln\ell}{\ell}\Big)^{\ln\ell}\,.$$

It remains to study $P\big(\tau^* > \kappa^{\ell(1-\varepsilon)} \,|\, D_0 = (\ell_\kappa(1-\varepsilon'))^m\big)$. We use the inequality of lemma 8.12 with $n = \kappa^{\ell(1-\varepsilon)}$ and $b = \ell_\kappa(1-\varepsilon')$:

$$P\big(\tau^* \leq \kappa^{\ell(1-\varepsilon)} \,\big|\, D_0 = (\ell_\kappa(1-\varepsilon'))^m\big) \;\leq\; \kappa^{\ell(1-\varepsilon)} m \frac{\mathcal{B}(0)}{\mathcal{B}(\ell_\kappa(1-\varepsilon'))}\,.$$

Using the large deviation estimates of lemma 7.2, we see that, for ε' small enough, there exists $c(\varepsilon) > 0$ such that, for ℓ large enough,

$$P\big(\tau^* \leq \kappa^{\ell(1-\varepsilon)} \,\big|\, D_0 = (\ell_\kappa(1-\varepsilon'))^m\big) \;\leq\; m\exp(-c(\varepsilon)\ell)\,.$$

Collecting all the previous estimates, we conclude that, for ℓ large enough,

$$P\Big(\tau^* > \kappa^{\ell(1-\varepsilon)} \,\Big|\, D_0 = (1)^m\Big) \;\geq\; \Big(1-\frac{6}{a\sqrt{\ell}}\Big)^m \Big(1-\frac{p}{\kappa}\Big)^{m\ell^{3/4}} \\ \times \Big(1 - m\exp(-c(\varepsilon)\ell)\Big)\Big(\Big(1-\frac{24}{\ell\varepsilon'}\Big)^m \;-\; \ell^2 m\Big(\frac{2\ln\ell}{\ell}\Big)^{\ln\ell}\Big)\,.$$

Moreover, by Markov's inequality,

$$E\Big(\tau^* \,|\, D_0 = (1)^m\Big) \;\geq\; \kappa^{\ell(1-\varepsilon)}\, P\Big(\tau^* \geq \kappa^{\ell(1-\varepsilon)} \,\Big|\, D_0 = (1)^m\Big)\,.$$

It follows that

$$\liminf_{\substack{\ell,m\to\infty,\, q\to 0\\ \ell q\to a,\, \frac{m}{\ell}\to\alpha}} \frac{1}{\ell}\ln E\Big(\tau^* \,|\, D_0 = (1)^m\Big) \;\geq\; (1-\varepsilon)\ln\kappa\,.$$

Letting ε go to 0 yields the desired lower bound. $\square$

9 Synthesis

As in theorem 3.1, we suppose that

$$\ell \to +\infty\,, \qquad m \to +\infty\,, \qquad q \to 0\,,$$

in such a way that

$$\ell q \to a \in]0,+\infty[\,, \qquad \frac{m}{\ell} \to \alpha \in [0,+\infty]\,.$$

We put now together the estimates of sections 7 and 8 in order to evaluate the formula for the invariant measure obtained at the end of section 6.3. For $\theta = \ell, 1$, we had

$$\int_{\mathcal{P}^m_{\ell+1}} f\Big(\frac{o(0)}{m}\Big)\, d\mu^\theta_O(o) \;=\;$$
$$\frac{\displaystyle\sum_{i=1}^m E\bigg(\sum_{n=0}^{\tau_0} f\Big(\frac{Z^\theta_n}{m}\Big)\,\Big|\, Z^\theta_0 = i\bigg) P\big(O^\theta_{\tau^*}(0) = i \,|\, O^\theta_0 = o^\theta_{\text{exit}}\big)}{\displaystyle E\big(\tau^* \,|\, O^\theta_0 = o^\theta_{\text{exit}}\big) + \sum_{i=1}^m E\big(\tau_0 \,\big|\, Z^\theta_0 = i\big) P\big(O^\theta_{\tau^*}(0) = i \,|\, O^\theta_0 = o^\theta_{\text{exit}}\big)}\,.$$

Using the monotonicity of $(Z^\theta_n)_{n\geq 0}$, we have

$$E\big(\tau_0 \,\big|\, Z^\theta_0 = 1\big) \;\leq\; \sum_{i=1}^m E\big(\tau_0 \,\big|\, Z^\theta_0 = i\big) P\big(O^\theta_{\tau^*}(0) = i \,|\, O^\theta_0 = o^\theta_{\text{exit}}\big)$$
$$\leq\; E\big(\tau_0 \,\big|\, Z^\theta_0 = m\big)\,.$$

With the help of proposition 7.6, we deduce from the previous inequalities that

$$\lim_{\substack{\ell,m\to\infty\\ q\to 0,\, \ell q\to a}} \frac{1}{m} \ln \sum_{i=1}^m E\big(\tau_0 \,\big|\, Z^\theta_0 = i\big) P\big(O^\theta_{\tau^*}(0) = i \,|\, O^\theta_0 = o^\theta_{\text{exit}}\big) \;=\; V(\rho^*(a), 0)\,.$$

By proposition 8.10, for $\alpha \in [0,+\infty[$,

$$\lim_{\substack{\ell,m\to\infty,\, q\to 0\\ \ell q\to a,\, \frac{m}{\ell}\to\alpha}} \frac{1}{\ell} \ln E\big(\tau^* \,|\, O^\theta_0 = o^\theta_{\text{exit}}\big) \;=\; \ln\kappa\,.$$

For the case $\alpha = +\infty$, by corollary 8.11 and proposition 8.9,

$$\limsup_{\substack{\ell,m\to\infty,\, q\to 0\\ \ell q\to a,\, \frac{m}{\ell}\to\infty}} \frac{1}{\ell} \ln\Big(\frac{1}{m} E\big(\tau^* \,|\, O^\theta_0 = o^\theta_{\text{exit}}\big)\Big) \;\leq\; \ln\kappa\,.$$

These estimates allow to evaluate the ratio between the discovery time and the persistence time. We define a function $\psi :]0,+\infty[\to [0,+\infty[$ by setting

$$\forall a \in]0,+\infty[\qquad \psi(a) \,=\, V(\rho^*(a),0)\,.$$

We first prove the properties of the function Ψ stated before theorem 3.1. If $a \geq \ln\sigma$, then $\rho^*(a)=0$ and $\Psi(a)=0$. If $a<\ln\sigma$, then $\rho^*(a)>0$ and it follows from lemma 7.5 that $\Psi(a)>0$. Finally, for $a>0$, we have that

$$V(\rho^*(a),0)\,\leq\, I(e^{-a},0)\,=\,\ln\frac{1}{1-e^{-a}}\,<\,+\infty\,.$$

For $\alpha\in[0,+\infty[$ or $\alpha=+\infty$, we have

$$\lim_{\substack{\ell,m\to\infty,\, q\to 0\\ \ell q\to a,\, \frac{m}{\ell}\to\alpha}} \frac{\displaystyle\sum_{i=1}^m E\big(\tau_0\,\big|\, Z_0^\theta=i\big)P\big(O^\theta_{\tau^*}(0)=i\,\big|\, O_0^\theta=o^\theta_{\text{exit}}\big)}{E\big(\tau^*\,\big|\,O_0^\theta=o^\theta_{\text{exit}}\big)}$$

$$=\begin{cases} \quad 0 & \text{if } \alpha\,\psi(a)<\ln\kappa\\ +\infty & \text{if } \alpha\,\psi(a)>\ln\kappa\end{cases}$$

Notice that the result is the same for $\theta=\ell$ and $\theta=1$. By proposition 7.12, we have

$$\lim_{\substack{\ell,m\to\infty\\ q\to 0,\, \ell q\to a}} \frac{\displaystyle\sum_{i=1}^m E\bigg(\sum_{n=0}^{\tau_0} f\Big(\frac{Z_n^\theta}{m}\Big)\,\Big|\, Z_0^\theta=i\bigg)P\big(O^\theta_{\tau^*}(0)=i\,\big|\,O_0^\theta=o^\theta_{\text{exit}}\big)}{\displaystyle\sum_{i=1}^m E\big(\tau_0\,\big|\,Z_0^\theta=i\big)P\big(O^\theta_{\tau^*}(0)=i\,\big|\,O_0^\theta=o^\theta_{\text{exit}}\big)}\,=\,f(\rho^*)\,.$$

Putting together the bounds on ν given in section 6.3 and the previous considerations, we conclude that

$$\lim_{\substack{\ell,m\to\infty,\, q\to 0\\ \ell q\to a,\, \frac{m}{\ell}\to\alpha}} \int_{[0,1]} f\,d\nu \,=\, \begin{cases} \quad 0 & \text{if } \alpha\,\psi(a)<\ln\kappa\\ f\big(\rho^*(a)\big) & \text{if } \alpha\,\psi(a)>\ln\kappa\end{cases}$$

This is valid for any continuous non–decreasing function $f:[0,1]\to\mathbb{R}$ such that $f(0)=0$.

References

[1] Domingos Alves and Jose Fernando Fontanari. Error threshold in finite populations. *Phys. Rev. E*, 57:7008–7013, 1998.

[2] Raphaël Cerf. Critical control of a genetic algorithm. *preprint, available on arXiv*, 2010.

[3] Raphaël Cerf. Critical population and error threshold on the sharp peak landscape for a Moran model. *preprint, available on arXiv*, 2012.

[4] Thomas Darden. Asymptotics of fixation under heterosis: a large deviation approach. *unpublished manuscript*, 1983.

[5] Lloyd Demetrius, Peter Schuster, and Karl Sigmund. Polynucleotide evolution and branching processes. *Bulletin of Mathematical Biology*, 47(2):239 – 262, 1985.

[6] Narendra M. Dixit, Piyush Srivastava, and Nisheeth K. Vishnoi. A finite population model of molecular evolution: Theory and computation. *preprint, available on arXiv*, 2012.

[7] Richard S. Ellis. *Entropy, large deviations, and statistical mechanics.* Classics in Mathematics. Springer-Verlag, Berlin, 2006. Reprint of the 1985 original.

[8] William Feller. *An introduction to probability theory and its applications. Vol. I.* John Wiley & Sons Inc., New York, 1968.

[9] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, New York, second edition, 1998. Translated from the 1979 Russian original by Joseph Szücs.

[10] Daniel T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22:403–434, 1976.

[11] Geoffrey Grimmett. *Percolation*, volume 321 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, second edition, 1999.

[12] Theodore E. Harris. A correlation inequality for Markov processes in partially ordered state spaces. *Ann. Probability*, 5(3):451–454, 1977.

[13] Yuri Kifer. *Random perturbations of dynamical systems*, volume 16 of *Progress in Probability and Statistics*. Birkhäuser Boston Inc., Boston, MA, 1988.

[14] Yuri Kifer. A discrete-time version of the Wentzell-Freidlin theory. *Ann. Probab.*, 18(4):1676–1692, 1990.

[15] Thomas M. Liggett. *Interacting particle systems.* Classics in Mathematics. Springer-Verlag, 2005. Reprint of the 1985 original.

[16] John McCaskill. A stochastic theory of macromolecular evolution. *Biological Cybernetics*, 50:63–73, 1984.

[17] Igor Mezić. FKG inequalities in cellular automata and coupled map lattices. *Phys. D*, 103(1-4):491–504, 1997.

[18] Gregory J. Morrow and Stanley Sawyer. Large deviation results for a class of Markov chains arising from population genetics. *Ann. Probab.*, 17(3):1124–1146, 1989.

[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] Gabriela Ochoa. *Error Thresholds and Optimal Mutation Rates in Genetic Algorithms.* PhD thesis, The University of Sussex, Brighton, 2001.

[22] Jeong-Man Park, Enrique Muñoz, and Michael W. Deem. Quasispecies theory for finite populations. *Phys. Rev. E*, 81:011902, 2010.

[23] 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.

[24] Edward D. Weinberger. *A Stochastic Generalization of Eigen's Theory of Natural Selection.* Ph.D. Dissertation. The Courant Institute of Mathematical Sciences, New York University, 1987.

Contents