Critical population and error threshold on the sharp peak landscape for a Moran model

Raphaël Cerf
Université Paris Sud and IUF

March 20, 2013

Abstract

The goal of this work is to propose a finite population counterpart to Eigen's model, which incorporates stochastic effects. We consider a Moran 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. We deal only with the sharp peak landscape: 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 $\alpha\,\phi(a) = \ln\kappa$ 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. Moreover, in the limit of very small mutations, we obtain a lower bound on the population size allowing the emergence of a quasispecies: if $\alpha < \ln\kappa/\ln\sigma$ then the equilibrium population is totally random, and a quasispecies can be formed only when $\alpha \geq \ln\kappa/\ln\sigma$. Finally, in the limit of very large populations, we recover an error catastrophe reminiscent of Eigen's model: if $\sigma e^{-a} \leq 1$ then the equilibrium population is totally random, and a quasispecies can be formed only when $\sigma e^{-a} > 1$. These results are supported by computer simulations.

I thank an anonymous referee for his careful reading and his remarks, which helped to improve the presentation.

1 Introduction.

In his famous paper [13], Eigen introduced a model for the evolution of a population of macromolecules. In this model, the macromolecules replicate themselves, yet the replication mechanism is subject to errors caused by mutations. These two basic mechanisms are described by a family of chemical reactions. The replication rate of a macromolecule is governed by its fitness. A fundamental discovery of Eigen is the existence of an error threshold on the sharp peak landscape. If the mutation rate exceeds a critical value, called the error threshold, then, at equilibrium, the population is completely random. If the mutation rate is below the error threshold, then, at equilibrium, the population contains a positive fraction of the master sequence (the most fit macromolecule) and a cloud of mutants which are quite close to the master sequence. This specific distribution of individuals is called a quasispecies. This notion has been further investigated by Eigen, McCaskill and Schuster [15] and it had a profound impact on the understanding of molecular evolution [11]. It has been argued that, at the population level, evolutionary processes select quasispecies rather than single individuals. Even more importantly, this theory is supported by experimental studies [12]. Specifically, it seems that some RNA viruses evolve with a rather high mutation rate, which is adjusted to be close to an error threshold. It has been suggested that this is the case for the HIV virus [38]. Some promising antiviral strategies consist of using mutagenic drugs that induce an error catastrophe [2, 8]. A similar error catastrophe could also play a role in the development of some cancers [36].

Eigen's model was initially designed to understand a population of macromolecules governed by a family of chemical reactions. In this setting, the number of molecules is huge, and there is a finite number of types of molecules. From the start, this model is formulated for an infinite population and the evolution is deterministic (mathematically, it is a family of differential equations governing the time evolution of the densities of each type of macromolecule). The error threshold appears when the number of types goes to ∞. This creates a major obstacle if one wishes to extend the notions of quasispecies and error threshold to genetics. Biological populations are finite, and even if they are large so that they might be considered infinite in some approximate scheme, it is not coherent to consider situations where the size of the population is much larger than the number of possible genotypes. Moreover, it has long been recognized that random effects play a major role in the genetic evolution of populations [24], yet they are ruled out from the start in a deterministic infinite population model. Therefore, it is crucial to develop a finite population counterpart to Eigen's model, which incorporates stochastic effects. This problem is already discussed by Eigen, McCaskill and Schuster [15] and more recently by Wilke

[41]. Numerous works have attacked this issue: Demetrius, Schuster and Sigmund [9], McCaskill [27], Gillespie [19], Weinberger [40]. Nowak and Schuster [32] constructed a birth and death model to approximate Eigen's model. This birth and death model plays a key role in our analysis, as we shall see later. Alves and Fontanari [1] study how the error threshold depends on the population in a simplified model. More recently, Musso [29] and Dixit, Srivastava, Vishnoi [10] considered finite population models which approximate Eigen's model when the population size goes to ∞. These models are variants of the classical Wright–Fisher model of population genetics. Although this is an interesting approach, it is already a delicate matter to prove the convergence of these models towards Eigen's model. We adopt here a different strategy. Instead of trying to prove that some finite population model converges in some sense to Eigen's model, we try to prove directly in the finite model an error threshold phenomenon. To this end, we look for the simplest possible model, and we end up with a Moran model [28]. The model we choose here is not particularly original, the contribution of this work is rather to show a way to analyze this kind of finite population model.

Let us describe informally the model (see section 2 for the formal definition). We consider a population of size m of chromosomes of length ℓ over the alphabet $\{ A, T, G, C \}$. We work only with the sharp peak landscape: there is one specific sequence, called the master sequence, whose fitness is $\sigma > 1$, and all the other sequences have fitness equal to 1. The replication rate of a chromosome is proportional to its fitness. When a chromosome replicates, it produces a copy of himself, which is subject to mutations. Mutations occur randomly and independently at each locus with probability q. The offspring of a chromosome replaces a chromosome chosen at random in the population.

The mutations drive the population towards a totally random state, while the replication favors the master sequence. These two antagonistic effects interact in a complicated way in the dynamics and it is extremely difficult to analyze precisely the time evolution of such a model. Let us focus on the equilibrium distribution of the process. A fundamental problem is to determine the law of the number of copies of the master sequence present in the population at equilibrium. If we keep the parameters m, ℓ, q fixed, there is little hope to get useful results. In order to simplify the picture, we consider an adequate asymptotic regime. In Eigen's model, the population size is infinite from the start. The error threshold appears when ℓ goes to ∞ and q goes to 0 in a regime where $\ell q = a$ is kept constant. We wish to understand the influence of the population size m, thus we use a different approach and we consider the following regime. We send simultaneously m, ℓ to ∞ and q to 0 and we try to understand the respective influence of each parameter on the equilibrium law of the master sequence. By the ergodic theorem, the average number of copies of the master sequence at equilibrium is equal to the limit, as the time goes to ∞, of the time average of the number of copies of the master sequence present through the whole evolution of the process. In the finite population model, the number of copies of the master sequence fluctuates with time. Our analysis of these fluctuations relies on the following heuristics. Suppose that the process starts with a population of size m containing exactly one master sequence. The master sequence is likely to invade the whole population and become dominant. Then the master sequence will be present in the population for a very long time without interruption. We call this time the **persistence** time of the master sequence. The destruction of all the master sequences of the population is quite unlikely, nevertheless it will happen and the process will eventually land in the neutral region consisting of the populations devoid of master sequences. The process will wander randomly throughout this region for a very long time. We call this time the **discovery** time of the master sequence. Because the cardinality of the possible genotypes is enormous, the master sequence is difficult to discover, nevertheless the mutations will eventually succeed and the process will start again with a population containing exactly one master sequence. If, on average, the discovery time is much larger than the persistence time, then the equilibrium state will be totally random, while a quasispecies will be formed if the persistence time is much larger than the discovery time. Let us illustrate this idea in a very simple model.

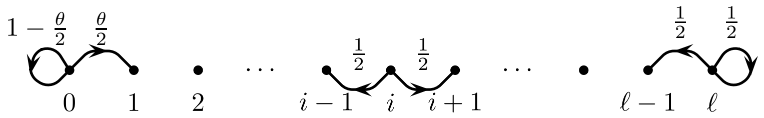

Figure 1: Random walk example

We consider the random walk on $\{0, \dots, \ell\}$ with the transition probabilities depending on a parameter θ given by:

$$p(0,1)=\frac{\theta}{2}\,,\quad p(0,0)=1-\frac{\theta}{2}\,,\quad p(\ell,\ell-1)=p(\ell,\ell)=\frac{1}{2}\,,$$
$$p(i,i-1)=p(i,i+1)=\frac{1}{2}\,,\quad 1\leq i\leq \ell-1\,.$$

The integer ℓ is large and the parameter θ is small. Hence the walker spends its time either wandering in $\{1, \dots, \ell\}$ or being trapped in 0. The state 0 plays the role of the quasispecies while the set $\{1, \dots, \ell\}$ plays the role of the neutral region. With this analogy in mind, the persistence time is the expected time of exit from 0, it is equal to $2/\theta$. The discovery time is the expected time needed to discover 0 starting for instance from 1, it is equal to 2ℓ. The equilibrium law of the walker is the probability measure μ given by

$$\mu(0) \,=\, \frac{1}{1+\theta\ell}\,, \qquad \mu(1) \,=\, \cdots \,=\, \mu(\ell) \,=\, \frac{\theta}{1+\theta\ell}\,.$$

We send ℓ to ∞ and θ to 0 simultaneously. If $\theta\ell$ goes to ∞, the entropy factors win and μ becomes totally random. If $\theta\ell$ goes to 0, the selection drift wins and μ converges to the Dirac mass at 0.

In order to implement the previous heuristics, we have to estimate the persistence time and the discovery time of the master sequence in the Moran model. For the persistence time, we rely on a classical computation from mathematical genetics. Suppose we start with a population containing $m-1$ copies of the master sequence and another non master sequence. The non master sequence is very unlikely to invade the whole population, yet it has a small probability to do so, called the fixation probability. If we neglect the mutations, standard computations yield that, in a population of size m, if the master sequence has a selective advantage of σ, the fixation probability of the non master sequence is roughly of order $1/\sigma^m$ (see for instance [31], section 6.3). Now the persistence time can be viewed as the time needed for non master sequences to invade the population. This time is approximately equal to the inverse of the fixation probability of the non master sequence, that is of order σ^m. For the discovery time, there is no miracle: before discovering the master sequence, the process is likely to explore a significant portion of the genotype space, hence the discovery time should be of order

$$\mathrm{card}\,\{\,A,T,G,C\,\}^\ell \,=\, 4^\ell\,.$$

These simple heuristics indicate that the persistence time depends on the selection drift, while the discovery time depends on the spatial entropy. Suppose that we send m,ℓ to ∞ simultaneously. If the discovery time is much larger than the persistence time, then the population will be neutral most of the time and the fraction of the master sequence at equilibrium will be null. If the persistence time is much larger than the discovery time, then the population will be invaded by the master sequence most of the time and the fraction of the master sequence at equilibrium will be positive. Thus the master sequence vanishes in the regime

$$m,\ell\to+\infty\,, \qquad \frac{m}{\ell}\to 0\,,$$

while a quasispecies might be formed in the regime

$$m,\ell\to+\infty\,, \qquad \frac{m}{\ell}\to +\infty\,.$$

This leads to an interesting feature, namely the existence of a critical population size for the emergence of a quasispecies. For chromosomes of length ℓ, a quasispecies can be formed only if the population size m is such that the ratio m/ℓ is large enough. In order to go further, we must put the heuristics on a firmer ground and we should take the mutations into account when estimating the persistence time. The main problem is to obtain finer estimates on the persistence and discovery times. We cannot compute explicitly the laws of these random times, so we will compare the Moran model with simpler processes.

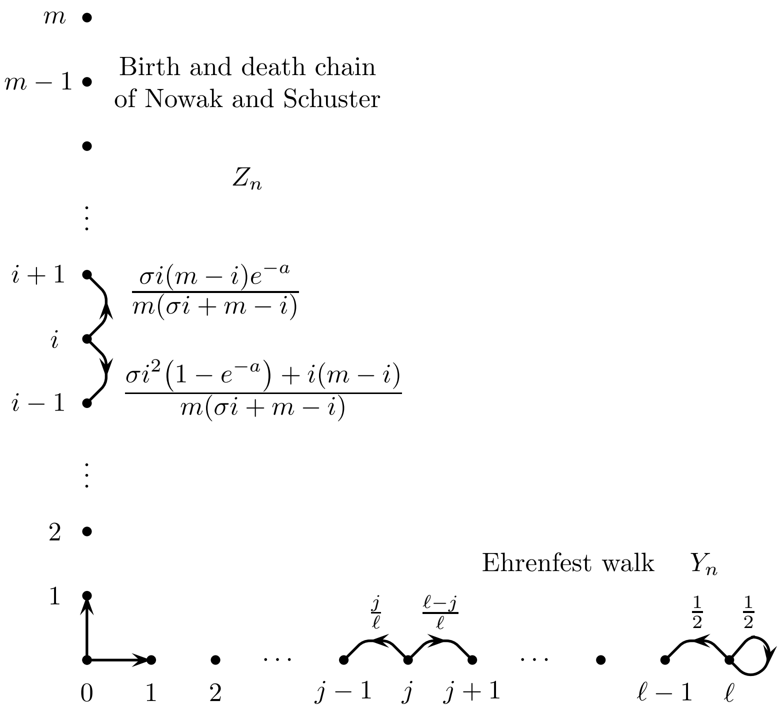

Figure 2: Approximating process

In the non neutral populations, we shall compare the process with a birth and death process $(Z_n)_{n\geq 0}$ on $\{0,\dots,m\}$, which is precisely the one introduced by Nowak and Schuster [32]. The value Z_n approximates the number of copies of the master sequence present in the population. For birth and death processes, explicit formulas are available and we obtain that, if $\ell, m \to +\infty$, $q \to 0$, $\ell q \to a \in]0, +\infty[$, then

$$\textbf{persistence time} \sim \exp\big(m\,\phi(a)\big)\,,$$

where

$$\phi(a) \,=\, \frac{\displaystyle \sigma(1-e^{-a})\ln\frac{\sigma(1-e^{-a})}{\sigma-1} + \ln(\sigma e^{-a})}{(1-\sigma(1-e^{-a}))}\,.$$

In the neutral populations, we shall replace the process by a random walk on $\{\,A,T,G,C\,\}^\ell \,=\, 4^\ell$. The lumped version of this random walk behaves like an Ehrenfest process $(Y_n)_{n\geq 0}$ on $\{\,0,\dots,\ell\,\}$ (see [5] for a nice review). The value Y_n represents the distance of the walker to the master sequence. A celebrated theorem of Kac from 1947 [21], which helped to resolve a famous paradox of statistical mechanics, yields that, when $\ell \to \infty$,

$$\textbf{discovery time} \sim 4^\ell\,.$$

Thus the Moran process is approximated by the process on

$$\Big(\{\,0,\dots,\ell\,\}\times\{\,0\,\}\Big) \cup \Big(\{\,0\,\}\times\{\,0,\dots,m\,\}\Big)$$

described loosely as follows. On $\{\,0,\dots,\ell\,\}\times\{\,0\,\}$, the process follows the dynamics of the Ehrenfest urn. On $\{\,0\,\}\times\{\,0,\dots,m\,\}$, the process follows the dynamics of the birth and death process of Nowak and Schuster [32]. When in $(0,0)$, the process can jump to either axis. With this simple heuristic picture, we recover all the features of our main result. We suppose that

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

in such a way that

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

The critical curve is then defined by the equation

$$\textbf{discovery time} \sim \textbf{persistence time}$$

which can be rewritten as

$$\alpha\,\phi(a) \,=\, \ln 4\,.$$

This way we obtain an equation 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. Moreover, in the regime of very small mutations, we obtain a lower bound on the population size allowing the emergence of a quasispecies: if $\alpha < \ln 4/\ln\sigma$ then the equilibrium population is totally random, and a quasispecies can be formed only when $\alpha \geq \ln 4/\ln\sigma$. Finally, in the limit of very large populations, we recover an error catastrophe reminiscent of Eigen's model: if $\sigma e^{-a} \leq 1$ then the equilibrium population is totally random, and a quasispecies can be formed only when $\sigma e^{-a} > 1$. These results are supported by computer simulations. The good news is that, already for small values of ℓ, the simulations are very conclusive.

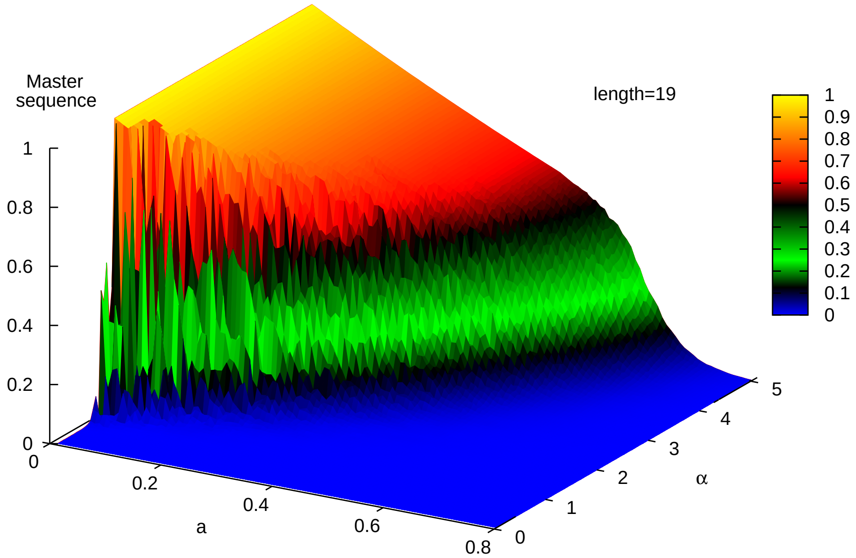

Figure 3: Simulation of the equilibrium density of the Master sequence

It is certainly well known that the population dynamics depends on the population size (see the discussion of Wilke [41]). In a theoretical study [30], Van Nimwegen, Crutchfield and Huynen developed a model for the evolution of populations on neutral networks and they show that an important parameter is the product of the population size and the mutation rate. The nature of the dynamics changes radically depending on whether this product is small or large. Sumedha, Martin and Peliti [37] analyze further the influence of this parameter. In [39], Van Nimwegen and Crutchfield derived analytical expressions for the waiting times needed to increase the fitness, starting from a local optimum. Their scaling relations involve the population size and show the existence of two different barriers, a fitness barrier and an entropy barrier. Although they pursue a different goal than ours, most of the heuristic ingredients explained previously are present in their work, and much more; they observe and discuss also the transition from the quasispecies regime for large populations to the disordered regime for small populations. The dependence on the population size and genome length has been investigated numerically by Elena, Wilke, Ofria and Lenski [16]. Here we show rigorously the existence of a critical population size for the sharp peak landscape in a specific asymptotic regime. The existence of a critical population size for the emergence of a quasispecies is a pleasing result: it shows that, even under the action of selection forces, a form of cooperation is necessary to create a quasispecies. Moreover the critical population size is much smaller than the cardinality of the possible genotypes. In conclusion, even in the very simple framework of the Moran model on the sharp peak landscape, cooperation is necessary to achieve the survival of the master sequence.

As emphasized by Eigen in [14], the error threshold phenomenon is similar to a phase transition in statistical mechanics. Leuthäusser established a formal correspondence between Eigen's model and an anisotropic Ising model [25]. Several researchers have employed tools from statistical mechanics to analyze models of biological evolution, and more specifically the error threshold: see the nice review written by Baake and Gabriel [3]. Baake investigated the so–called Onsager landscape in [4]. This way she could transfer to a biological model the famous computation of Onsager for the two dimensional Ising model. Saakian, Deem and Hu [34] compute the variance of the mean fitness in a finite population model in order to control how it approximates the infinite population model. Deem, Muñoz and Park [33] use a field theoretic representation in order to derive analytical results.

We were also very much inspired by ideas from statistical mechanics, but with a different flavor. We do not use exact computations, rather we rely on softer tools, namely coupling techniques and correlation inequalities. These are the basic tools to prove the existence of a phase transition in classical models, like the Ising model or percolation. We seek large deviation estimates rather than precise scaling relations in our asymptotic regime. Of course the outcome of these techniques is very rough compared to exact computations, yet they are much more robust and their range of applicability is much wider. The model is presented in the next section and the main results in section 3. The remaining sections are devoted to the proofs. In the appendix we recall several classical results of the theory of finite Markov chains.

2 The model.

This section is devoted to the presentation of the model. Let $\mathcal{A}$ be a finite alphabet and let $\kappa = \text{card}\,\mathcal{A}$ 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. We shall study a simple model for the evolution of a finite population of chromosomes on the space $\mathcal{A}^\ell$. An essential feature of the model we consider is that the size of the population is constant throughout the evolution. We denote by m the size of the population. 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 row being the i–th chromosome. The evolution of the population will be random and it will be driven by two antagonistic forces: mutation and replication.

Mutation. We assume that the mutation mechanism is the same for all the loci, and that mutations occur independently. Moreover we choose the most symmetric mutation scheme. We denote by $q \in]0, 1-1/\kappa[$ the probability of the occurrence of a mutation 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 then

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

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

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

The fitness of a chromosome can be interpreted as its reproduction rate. A chromosome u gives birth at random times and the mean time interval between two consecutive births is $1/A(u)$. In the context of Eigen's model, the quantity $A(u)$ is the kinetic constant associated to the chemical reaction for the replication of a macromolecule of type u.

Authorized changes. In our model, the only authorized changes in the population consist of replacing one chromosome of the population by a new one. The new chromosome is obtained by replicating another chromosome, possibly with errors. We introduce a specific notation corresponding to these changes. For a population $x\in\big(\mathcal{A}^\ell\big)^m$, $j\in\{\,1,\dots,m\,\}$, $u\in\mathcal{A}^\ell$, we denote by $x(j\leftarrow u)$ the population x in which the j–th chromosome $x(j)$ has been replaced by u:

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

We make this modeling choice in order to build a very simple model. This type of model is in fact classical in population dynamics, they are called Moran models [17].

The mutation–replication scheme. Several further choices have to be done to define the model precisely. We have to decide how to combine the mutation and the replication processes. There exist two main schemes in the literature. In the first scheme, mutations occur at any time of the life cycle and they are caused by radiations or thermal fluctuations. This leads to a decoupled Moran model. In the second scheme, mutations occur at the same time as births and they are caused by replication errors. This is the case of the famous Eigen model and it leads to the Moran model we study here. This Moran model can be described loosely as follows. Births occur at random times. The rates of birth are given by the fitness function A. There is at most one birth at each instant. When an individual gives birth, it produces an offspring through a replication process. Errors in the replication process induce mutations. The offspring replaces an individual chosen randomly in the population (with the uniform probability).

We build next a mathematical model for the evolution of a finite population of size m on the space $\mathcal{A}^\ell$, driven by mutation and replication as described above. We will end up with a stochastic process on the population space $\big(\mathcal{A}^\ell\big)^m$. Since the genetic composition of a population contains all the necessary information to describe its future evolution, our process will be Markovian.

Discrete versus continuous time. We can either build a discrete time Markov chain or a continuous time Markov process. Although the mathematical construction of a discrete time Markov chain is simpler, a continuous time process seems more adequate as a model of evolution for a population: births, deaths and mutations can occur at any time. In addition, the continuous time model is mathematically more appealing. We will build both types of models, in continuous and discrete time. Continuous time models are conveniently defined by their infinitesimal generators, while discrete time models are defined by their transition matrices (see the appendix). It should be noted, however, that the discrete time and the continuous time processes are linked through a standard stochastization procedure and they have the same stationary distribution. Therefore the asymptotic results we present here hold in both frameworks.

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

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

Transition matrix. The discrete time Moran model is the Markov chain $(X_n)_{n\in\mathbb{N}}$ whose transition matrix is given by

$$\forall n\in\mathbb{N}\quad \forall x\in\big(\mathcal{A}^\ell\big)^m\quad \forall j\in\{\,1,\dots,m\,\}\quad \forall u\in\mathcal{A}^\ell\setminus\{\,x(j)\,\}$$
$$P\big(X_{n+1} = x(j\leftarrow u)\,|\,X_n = x\big) = \frac{1}{m^2\lambda}\sum_{1\leq i\leq m} A(x(i))M(x(i),u)\,,$$

where $\lambda > 0$ is a constant such that

$$\lambda \,\geq\, \max\,\big\{\, A(u) : u \in \mathcal{A}^\ell \,\big\}\,.$$

The other non diagonal coefficients of the transition matrix are zero. The diagonal terms are chosen so that the sum of each row is equal to one. Notice that the continuous time formulation is more concise and elegant: it does not require the knowledge of the maximum of the fitness function A in its definition.

Loose description of the dynamics. We explain first the discrete time dynamics of the Markov chain $(X_n)_{n\in\mathbb{N}}$. Suppose that $X_n = x$ for some $n \in \mathbb{N}$ and let us describe loosely the transition mechanism to $X_{n+1} = y$. An index i in $\{\,1,\dots,m\,\}$ is selected randomly with the uniform probability. With probability $1 - A(x(i))/\lambda$, nothing happens and $y = x$. With probability $A(x(i))/\lambda$, the chromosome $x(i)$ enters the replication process and it produces an offspring u according to the law $M(x(i),\cdot)$ given by the mutation matrix. Another index j is selected randomly with uniform probability in $\{\,1,\dots,m\,\}$. The population y is obtained by replacing the chromosome $x(j)$ in the population x by a chromosome u.

We consider next the continuous time dynamics of the Markov process $(X_t)_{t\in\mathbb{R}^+}$. The dynamics is governed by a clock that rings randomly. The time interval τ between each of the clock ringing is exponentially distributed with parameter $m^2\lambda$:

$$\forall t \in \mathbb{R}^+ \qquad P(\tau > t) \,=\, \exp\big(-m^2\lambda t\big)\,.$$

Suppose that the clock rings at time t and that the process was in state x just before the time t. The population x is transformed into the population y following the same scheme as for the discrete time Markov chain $(X_n)_{n\in\mathbb{N}}$ described previously. At time t, the process jumps to the state y.

3 Main results.

This section is devoted to the presentation of the main results.

Convention. The results hold for both the discrete time and the continuous time models, so we do not make separate statements. The time variable is denoted by t throughout this section, it is either discrete with values in $\mathbb{N}$ or continuous with values in $\mathbb{R}^+$.

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,

$$\text{Master}(\sigma, \ell, m, q) \,=\, \lim_{t \to \infty} E\Big(\frac{1}{m} N(X_t)\Big)\,,$$

as well as the variance

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

The limits exist because the transition mechanism of the Markov process $(X_t)_{t \geq 0}$ is irreducible (and aperiodic for the discrete time case) as soon as the mutation probability is strictly between 0 and 1. Since the state space is finite, the Markov process $(X_t)_{t \geq 0}$ admits a unique invariant probability measure, which describes the steady state of the process. The ergodic theorem for Markov chains implies that the law of $(X_t)_{t \geq 0}$ converges towards this invariant probability measure, hence the above expectations converge. The limits depend on the parameters of the model, that is σ, ℓ, m, q. Our choices for the infinitesimal generator and the transition matrix imply that the discrete time version and the continuous time version have exactly the same invariant probability measure. In order to exhibit a sharp transition phenomenon, we send ℓ, m to ∞ and q to 0. Let $\phi : \mathbb{R}^+ \to \mathbb{R}^+ \cup \{ +\infty \}$ be the function defined by

$$\forall a < \ln\sigma \qquad \phi(a) \;=\; \frac{\sigma(1-e^{-a})\ln\dfrac{\sigma(1-e^{-a})}{\sigma-1} + \ln(\sigma e^{-a})}{(1-\sigma(1-e^{-a}))}$$

and $\phi(a) = 0$ if $a \geq \ln\sigma$.

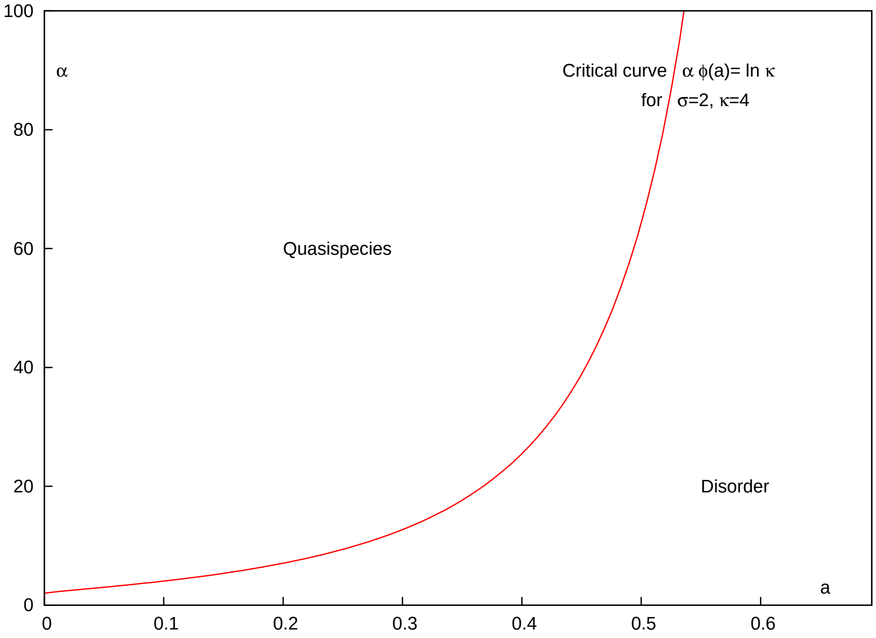

Figure 4: Critical curve

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 $\alpha\,\phi(a) < \ln\kappa$ then $\text{Master}\big(\sigma,\ell,m,q\big) \;\to\; 0$.
- If $\alpha\,\phi(a) > \ln\kappa$ then $\text{Master}\big(\sigma,\ell,m,q\big) \;\to\; \dfrac{\sigma e^{-a}-1}{\sigma-1}$.

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

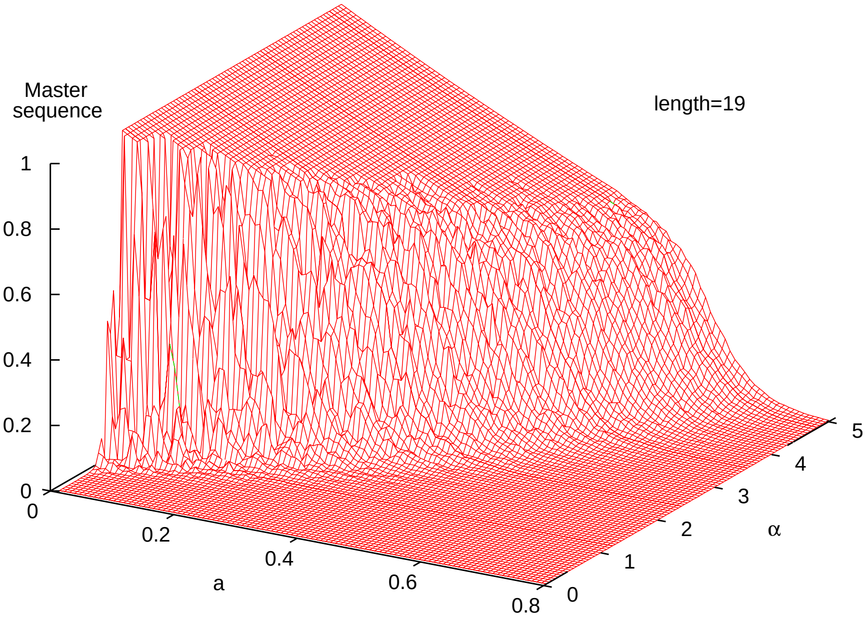

Figure 5: Master sequence at equilibrium

These results are supported by computer simulations (see figure 5). On the simulations, which are of course done for small values of ℓ, the transition associated to the critical population size seems even sharper than the transition associated to the error threshold. The programs are written in C with the help of the GNU scientific library and the graphical output is generated with the help of the Gnuplot program. To increase the efficiency of the simulations, we simulated the occupancy process obtained by lumping the original Moran model. The number of generations in a simulation run was adjusted empirically in order to stabilize the output within a reasonable amount of time. Twenty years ago, Nowak and Schuster could perform simulations with $\ell = 10$ and $m = 100$ for 20 000 generations [32]. Today's computer powers allow to simulate easily models with $\ell = 20$ and $m = 100$ for 10 000 000 000 generations. The good news is that, already for small values of ℓ, the simulations are very conclusive. Figure 6 presents three pictures corresponding to simulations with $\ell = 4, 8, 16$, as well as the theoretical shape for $\ell = \infty$ in the last picture. Notice that the statement of the theorem holds also in the case where α is null or infinite. This yields

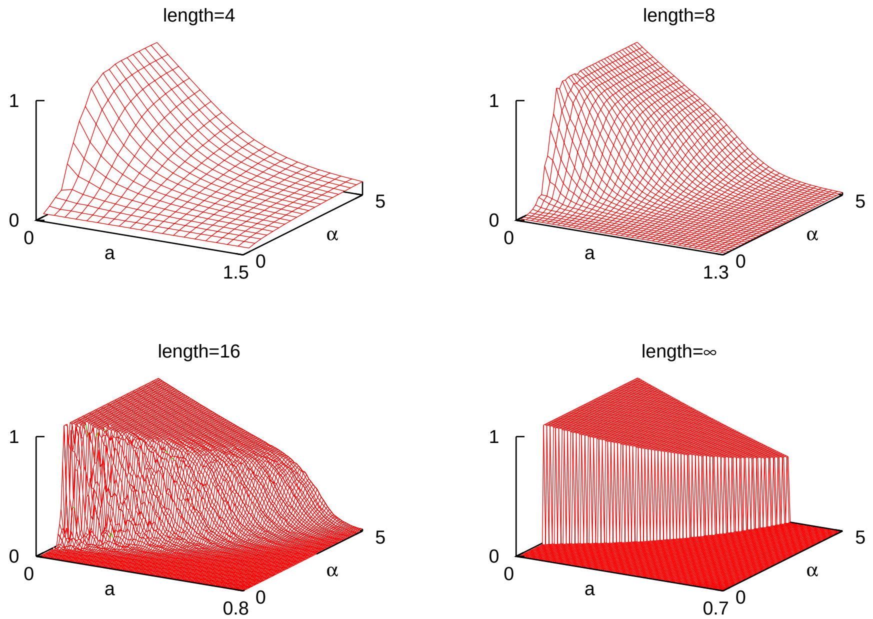

Figure 6: Varying the length ℓ

the following results:

Small populations. If $\ell, m \to +\infty$, $q \to 0$, $\ell q \to a \in]0, +\infty[$, $\frac{m}{\ell} \to 0$, then $\text{Master}\big(\sigma, \ell, m, q\big) \to 0$.

Large populations. Suppose that

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

If $a \geq \ln \sigma$, then $\text{Master}\big(\sigma, \ell, m, q\big) \to 0$. If $a < \ln \sigma$, then

$$\text{Master}\big(\sigma, \ell, m, q\big) \ \to \ \frac{\sigma e^{-a} - 1}{\sigma - 1}\,.$$

Interestingly, the large population regime is reminiscent of Eigen's model. A slightly more restrictive formulation consists of sending ℓ to ∞, m to ∞ and q to 0 in such a way that m/ℓ and ℓq are kept constant. We might then take q and m as functions of ℓ. Let $a, \alpha \in]0, +\infty[$. We take $q = a/\ell$ and $m = \alpha\ell$ and we have

$$\lim_{\ell \to \infty} \text{Master}\big(\sigma, \ell, \alpha\ell, a/\ell\big) \ = \ \begin{cases} \quad 0 & \text{if } \alpha\,\phi(a) < \ln\kappa \\ \dfrac{\sigma e^{-a} - 1}{\sigma - 1} & \text{if } \alpha\,\phi(a) > \ln\kappa \end{cases}$$

Notice that $\alpha\,\phi(a) > \ln\kappa$ implies that $a < \ln\sigma$ and therefore $\sigma e^{-a} > 1$. The critical curve

$$\left\{\,(a,\alpha)\in\mathbb{R}^+\times\mathbb{R}^+ : \alpha\,\phi(a) = \ln\kappa\,\right\}$$

corresponds to parameters (a,α) which are exactly at the error threshold and the critical population size. We are able to compute explicitly the critical curve and the limiting density because we consider a toy model. We did not examine here what happens on the critical curve. It is expected that the limiting density of the master sequence still fluctuates so that $\text{Variance}\big(\sigma,\ell,\alpha\ell,a/\ell\big)$ does not converge to 0 whenever $\alpha\,\phi(a) = \ln\kappa$. An important observation is that the critical scaling should be the same for similar Moran models. In contrast, the critical curve seems to depend strongly on the specific dynamics of the model. However, in the limit where a goes to 0, the function $\phi(a)$ converges towards $\ln\sigma$. This yields the minimal population size allowing the emergence of a quasispecies.

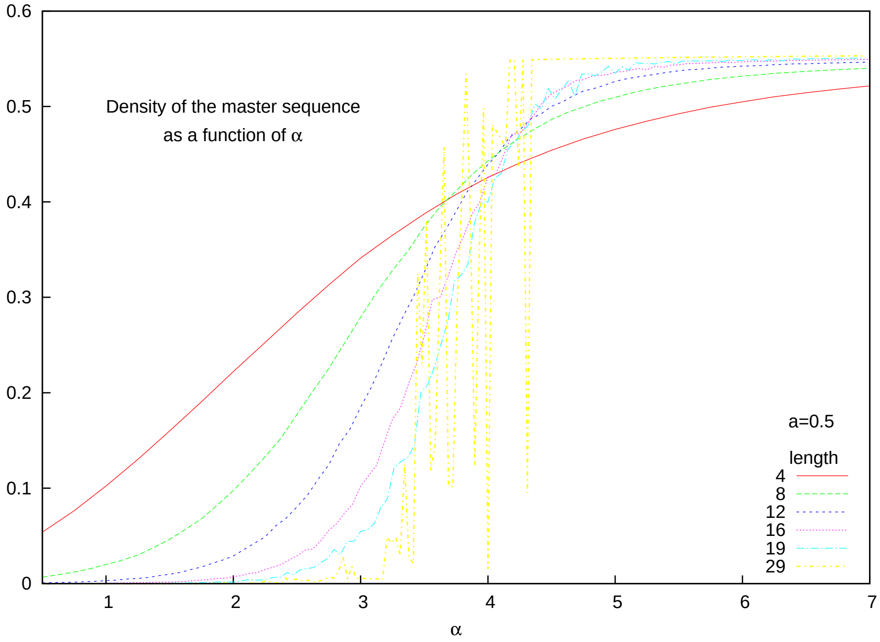

Figure 7: Critical population size

Corollary 3.2 If $\alpha < \ln\kappa/\ln\sigma$ then

$$\forall a > 0 \qquad \lim_{\ell\to\infty} \text{Master}\big(\sigma,\ell,\alpha\ell,a/\ell\big) \;=\; 0\,.$$

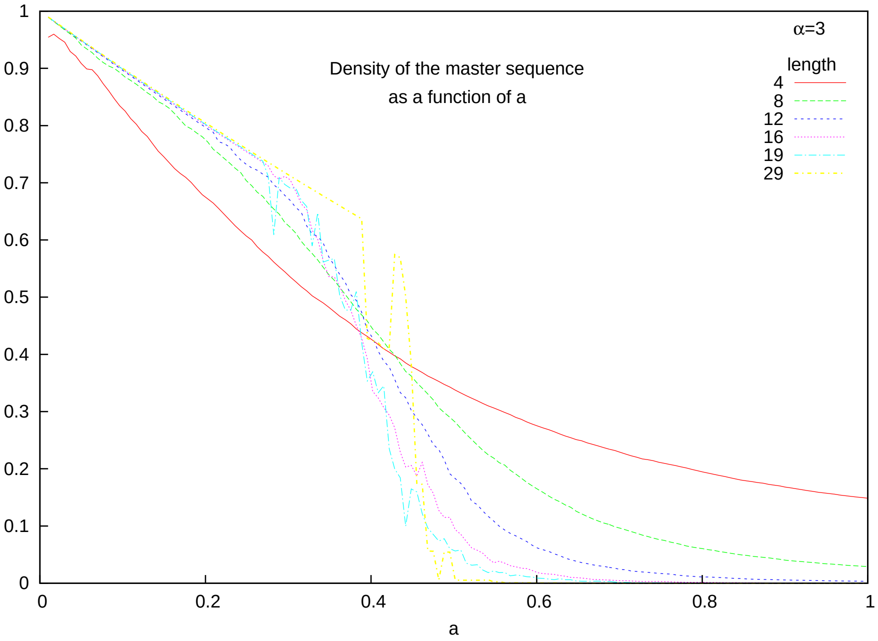

Figure 8: Error threshold

If $\alpha > \ln \kappa / \ln \sigma$ then

$$\exists a > 0 \qquad \lim_{\ell \to \infty} \text{Master}\big(\sigma, \ell, \alpha\ell, a/\ell\big) \,>\, 0\,.$$

We can also compute the maximal mutation rate permitting the emergence of a quasispecies. Interestingly, this maximal mutation rate is reminiscent of the error catastrophe in Eigen's model.

Corollary 3.3 If $a > \ln \sigma$ then

$$\forall \alpha > 0 \qquad \lim_{\ell \to \infty} \text{Master}\big(\sigma, \ell, \alpha\ell, a/\ell\big) \,=\, 0\,.$$

If $a < \ln \sigma$ then

$$\exists \alpha > 0 \qquad \lim_{\ell \to \infty} \text{Master}\big(\sigma, \ell, \alpha\ell, a/\ell\big) \,>\, 0\,.$$

In conclusion, on the sharp peak landscape, a quasispecies can emerge only if

$$m \,>\, \frac{\ln \kappa}{\ln \sigma}\,\ell\,, \qquad q \,<\, \frac{\ln \sigma}{\ell}\,.$$

The heuristic ideas behind theorem 3.1 were explained in the introduction. These heuristics are quite simple, however, the corresponding proofs are rather delicate and technical. There is very little hope to do a proof entirely based on exact computations. Our strategy consists of comparing the original Moran process with simpler processes in order to derive adequate lower and upper bounds. To this end, we couple the various processes starting with different initial conditions (section 4). Unfortunately, the natural coupling for the Moran model we wish to study is not monotone. Therefore we consider an almost equivalent model, which we call the normalized Moran model. This model is obtained by normalizing the reproduction rates so that the total reproduction rate of any population is one (section 5). A main difficulty is that the initial state space of the Moran process has no order structure and it is huge. We use a classical technique, called lumping, in order to reduce the state space (section 6). This way we obtain two lumped processes: the distance process $(D_t)_{t\geq 0}$ which records the Hamming distances between the chromosomes of the population and the Master sequence and the occupancy process $(O_t)_{t\geq 0}$ which records the distribution of these Hamming distances. The distance process is monotone in the neutral case $\sigma = 1$, while the occupancy process is monotone for any value $\sigma \geq 1$ (section 7). Therefore we construct lower and upper processes to bound the occupancy process (section 8). These processes have the same dynamics as the original process in the neutral region and they evolve as a birth and death process as soon as the population contains a master sequence. We use then the ergodic theorem for Markov chains and a renewal argument to estimate the invariant probability measures of these processes. The behavior of the lower and upper bounds depends mainly on the persistence time and the discovery time of the master sequence. We rely on the explicit formulas available for birth and death processes to estimate the persistence time (section 9). To estimate the discovery time, we rely on rough estimates for the mutation dynamics and correlation inequalities (section 10). The mutation dynamics is quite similar to the Ehrenfest urn, however it is more complicated because several mutations can occur simultaneously and exact formulas are not available. The proof is concluded in section 11.

Warning. From section 6 onwards, we work with the normalized Moran model defined in section 5. This model is denoted by $(X_t)_{t\geq 0}$ and its transition matrix by p, like the initial Moran model. We deal only with discrete time processes in the proofs. The time is denoted by t or n.

4 Coupling

The definition of the processes through infinitesimal generator is not very intuitive at first sight. We will provide here a direct construction of the processes, which does not make appeal to a general existence result. This construction is standard and it is the formal counterpart of the loose description of the dynamics given in section 2. Moreover it provides a useful coupling of the processes with different initial conditions and different control parameters σ, q. 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:

- a Poisson process $(\tau(t))_{t\geq 0}$ with intensity $m^2\lambda$ where $\lambda > 0$ is a constant such that $\lambda \,\geq\, \max\{\, A(u) : u \in \mathcal{A}^\ell \,\}$.

- two sequences of random variables I_n, J_n, $n \geq 1$, with uniform law on the index set $\{\, 1, \dots, m \,\}$.

- a family of random variables $U_{n,l}$, $n \geq 1$, $1 \leq l \leq \ell$, with uniform law on the interval $[0,1]$.

- a sequence of random variables S_n, $n \geq 1$, with uniform law on the interval $[0,1]$.

We denote by τ_n the n–th arrival time of the Poisson process $(\tau(t))_{t\geq 0}$, i.e.,

$$\forall n \geq 1 \qquad \tau_n \,=\, \inf\{\, t \geq 0 : \tau(t) = n \,\}\,.$$

The random variables I_n, J_n, $U_{n,l}$, $1 \leq l \leq \ell$, and S_n will be used to decide which move occurs at time τ_n. To build the coupling, it is more convenient to replace the mutation probability q by the parameter p given by

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

We define a Markov chain $(X_n)_{n\in\mathbb{N}}$ with the help of the previous random ingredients, whose law is the law of the Moran model. The process starts at time 0 from an arbitrary population x_0. Let $n \geq 1$, suppose that the process has been defined up to time $n-1$ and that $X_{n-1} = x$. We explain how to build $X_n = y$. Let us set $i = I_n$. If $S_n > A(x(i))/\lambda$, then $y = x$. Suppose next that $S_n \leq A(x(i))/\lambda$. We define y as follows. We index the elements of the alphabet $\mathcal{A}$ in an arbitrary way:

$$\mathcal{A} \,=\, \big\{\, a_1, \dots, a_\kappa \,\big\}\,.$$

Let $j = J_n$. We set

$$\forall l \in \{\,1,\dots,\ell\,\}\qquad y(j,l)\;=\;\begin{cases} a_1 & \text{if}\quad U_{n,l} < \dfrac{p}{\kappa} \\ \vdots & \\ a_r & \text{if}\quad (r-1)\dfrac{p}{\kappa} < U_{n,l} < r\dfrac{p}{\kappa} \\ \vdots & \\ a_\kappa & \text{if}\quad (\kappa-1)\dfrac{p}{\kappa} < U_{n,l} < p \\ x(i,l) & \text{if}\quad U_{n,l} \geq p \end{cases}$$

For $k \neq j$ we set $y(k) = x(k)$. Finally we define $X_n = y$.

We define also a Markov process $(X_t)_{t\in\mathbb{R}^+}$ with right continuous trajectories. The process starts at time 0 from an arbitrary population x_0 and it moves only when there is an arrival in the Poisson process $(\tau(t))_{t\geq 0}$. Let $t>0$ and suppose that $\tau_n = t$ for some $n \geq 1$. Suppose that just before t the process was in state x:

$$\lim_{\substack{s\to t\\ s<t}} X_s \;=\; x\,.$$

We proceed as in the construction of the discrete time process at step n to build the new population y starting from x and we set $X_t = y$. Therefore we have

$$\forall n \geq 0\quad \forall t \in [\tau_n, \tau_{n+1}[\qquad X_t \;=\; X_n\,.$$

5 Normalized model

The Moran model defined previously is difficult to analyze for several reasons. A major problem is that the natural coupling constructed in section 4 is not monotone. We define next a related Moran model which is simpler to study. This model is obtained by normalizing the reproduction rates so that the total reproduction rate of any population is one. The continuous time normalized Moran model is the Markov process $(X_t)_{t\in\mathbb{R}^+}$ whose infinitesimal generator L is defined as follows: for ϕ a function from $\big(\mathcal{A}^\ell\big)^m$ to $\mathbb{R}$ and for any $x\in\big(\mathcal{A}^\ell\big)^m$,

$$L\phi(x)\,=\,\sum_{1\le i,j\le m}\sum_{u\in\mathcal{A}^\ell}\frac{A(x(i))M(x(i),u)}{A(x(1))+\cdots+A(x(m))}\Big(\phi\big(x(j\leftarrow u)\big)-\phi(x)\Big)\,.$$

The discrete time normalized Moran model is the Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition matrix p given by

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

The other non diagonal coefficients of the transition matrix are zero. In the remainder of the paper, we shall work with this Markov chain $(X_n)_{n\in\mathbb{N}}$ and the transition matrix p. We shall prove the main theorem 3.1 of section 3 for this process. In fact, we shall even prove the following stronger result. Let ν be the image of the invariant probability measure of $(X_n)_{n\ge 0}$ through the map

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

The probability measure ν is a measure on the interval $[0,1]$ describing the equilibrium density of the master sequence in the population. Indeed,

$$\forall i\in\{\,0,\dots,m\,\}\qquad\nu\Big(\frac{i}{m}\Big)\,=\,\lim_{n\to\infty}\,P\big(N(X_n)=i\big)\,.$$

The probability ν depends on the parameters σ,ℓ,m,q of the model. Let $\phi(a)$ be the function defined before theorem 3.1, i.e.,

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

and $\phi(a)=0$ if $a\ge\ln\sigma$. Let

$$\rho^*\,=\,\frac{\sigma e^{-a}-1}{\sigma-1}\,.$$

Theorem 5.1 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 $\alpha\,\phi(a) < \ln\kappa$ then ν converges towards the Dirac mass at 0:

$$\forall\varepsilon>0\qquad \nu([0,\varepsilon])\;\to\;1\,.$$

- If $\alpha\,\phi(a) > \ln\kappa$ then ν converges towards the Dirac mass at ρ^*:

$$\forall\varepsilon>0\qquad \nu([\rho^*-\varepsilon,\rho^*+\varepsilon])\;\to\;1\,.$$

We shall prove this theorem for the normalized Moran model $(X_n)_{n\in\mathbb{N}}$. Let us show how this implies theorem 3.1 for the initial model. In the remainder of this argument, we denote by $(X'_n)_{n\in\mathbb{N}}$ the Moran model described in section 2 and by p' its transition matrix. The transition matrices p and p' are related by the simple relation

$$\forall x,y\in\big(\mathcal{A}^\ell\big)^m\,,\quad x\neq y\,,\qquad p'(x,y)\;=\;\beta(x)\,p(x,y)$$

where

$$\forall x\in\big(\mathcal{A}^\ell\big)^m\qquad \beta(x)\;=\;\frac{1}{m\lambda}\big(A(x(1))+\cdots+A(x(m))\big)\,.$$

Let μ and μ' be the invariant probability measures of the processes $(X_t)_{t\geq0}$ and $(X'_t)_{t\geq0}$. The probability μ is the unique solution of the system of equations

$$\forall x\in\big(\mathcal{A}^\ell\big)^m\qquad \mu(x)\;=\;\sum_{y\in(\mathcal{A}^\ell)^m}\mu(y)\,p(y,x)\,.$$

We rewrite these equations as:

$$\forall x\in\big(\mathcal{A}^\ell\big)^m\qquad \mu(x)\sum_{\substack{y\in(\mathcal{A}^\ell)^m\\ y\neq x}}p(x,y)\;=\;\sum_{\substack{y\in(\mathcal{A}^\ell)^m\\ y\neq x}}\mu(y)\,p(y,x)\,.$$

Replacing p by p', we get

$$\forall x\in\big(\mathcal{A}^\ell\big)^m\qquad \frac{\mu(x)}{\beta(x)}\sum_{\substack{y\in(\mathcal{A}^\ell)^m\\ y\neq x}}p'(x,y)\;=\;\sum_{\substack{y\in(\mathcal{A}^\ell)^m\\ y\neq x}}\frac{\mu(y)}{\beta(y)}\,p'(y,x)\,.$$

Using the uniqueness of the invariant probability measure associated to p', we conclude that

$$\forall x \in \big(\mathcal{A}^\ell\big)^m \qquad \mu'(x) \,=\, \frac{\displaystyle\frac{\mu(x)}{\beta(x)}}{\displaystyle\sum_{y\in(\mathcal{A}^\ell)^m} \frac{\mu(y)}{\beta(y)}}\,.$$

In the case of the sharp peak landscape, the function $\beta(x)$ can be rewritten as

$$\forall x \in \big(\mathcal{A}^\ell\big)^m \qquad \beta(x) \,=\, \frac{1}{m\lambda}\big((\sigma-1)N(x)+m\big)\,.$$

Let us denote by ν and ν' the images of μ and μ' through the map

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

We can thus rewrite

$$\sum_{y\in(\mathcal{A}^\ell)^m} \frac{\mu(y)}{\beta(y)} \,=\, \sum_{y\in(\mathcal{A}^\ell)^m} \frac{\lambda\mu(y)}{(\sigma-1)\displaystyle\frac{N(y)}{m}+1} \,=\, \int_{[0,1]} \frac{\lambda\, d\nu(t)}{(\sigma-1)t+1}\,.$$

For any function $f:[0,1]\to\mathbb{R}$, we have then

$$\begin{aligned}\int_{[0,1]} f\, d\nu' \,&=\, \lim_{t\to\infty} E\bigg(f\Big(\frac{1}{m}N(X'_t)\Big)\bigg) \,=\, \sum_{x\in(\mathcal{A}^\ell)^m} f\Big(\frac{1}{m}N(x)\Big)\,\mu'(x)\\ &=\, \frac{\displaystyle\sum_{x\in(\mathcal{A}^\ell)^m} f\Big(\frac{N(x)}{m}\Big)\frac{\mu(x)}{\beta(x)}}{\displaystyle\sum_{y\in(\mathcal{A}^\ell)^m}\frac{\mu(y)}{\beta(y)}} \,=\, \frac{\displaystyle\int_{[0,1]}\frac{\lambda f(t)\, d\nu(t)}{(\sigma-1)t+1}}{\displaystyle\int_{[0,1]}\frac{\lambda\, d\nu(t)}{(\sigma-1)t+1}}\,.\end{aligned}$$

We suppose that we are in the asymptotic regime for ℓ, m, q considered in theorems 3.1 and 5.1. By theorem 5.1, away from the critical curve $\alpha\,\phi(a) \,=\, \ln\kappa$, the probability ν converges towards a Dirac mass. If ν converges towards a Dirac mass at ρ, then we conclude from the above formula that ν' converges towards the same Dirac mass and

$$\begin{aligned}\text{Master}\big(\sigma,\ell,m,q\big) \,&\to\, \rho\,,\\ \text{Variance}\big(\sigma,\ell,m,q\big) \,&\to\, 0\,.\end{aligned}$$

This way we obtain the statements of theorem 3.1. From now onwards, in the proofs, we work exclusively with the normalized Moran process, and we denote it by $(X_t)_{t\geq 0}$.

6 Lumping

The state space of the process $(X_t)_{t\geq 0}$ is huge, it has cardinality $\kappa^{\ell m}$. We will rely on a classical technique to reduce the state space called lumping (see the appendix). We consider here only the sharp peak landscape. In this situation, the fitness of a chromosome is a function of its distance to the master sequence. A close look at the mutation mechanism reveals that chromosomes which are at the same distance from the Master sequence are equivalent for the dynamics, hence they can be lumped together in order to build a simpler process on a reduced space. For simplicity, we consider only the discrete time process. However similar results hold in continuous time.

6.1 Distance process

We denote by d_H the Hamming distance between two chromosomes:

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

We will keep track of the distances of the chromosomes to the master sequence w^*. We define a function $H:\mathcal{A}^\ell\to\{\,0,\dots,\ell\,\}$ by setting

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

The map H induces a partition of $\mathcal{A}^\ell$ into Hamming classes

$$H^{-1}(\{\,b\,\})\,,\qquad b\in\{\,0,\dots,\ell\,\}\,.$$

We prove first that the mutation matrix is lumpable with respect to the function H.

Lemma 6.1 (Lumped mutation matrix) 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

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

Proof. Let $b, c \in \{0, \dots, \ell\}$ and let $u \in \mathcal{A}^\ell$ such that $H(u) = b$. We will compute the law of $H(w)$ whenever w follows the law $M(u, \cdot)$ given by the row of M associated to u. For any $w \in \mathcal{A}^\ell$, we have

$$\begin{aligned} H(w) \,&=\, \sum_{1 \le l \le \ell} 1_{w(l) \neq w^*(l)} \\ &= \sum_{1 \le l \le \ell} \Big(1_{w(l) \neq w^*(l), u(l) = w^*(l)} + 1_{w(l) \neq w^*(l), u(l) \neq w^*(l)} \Big) \\ &= H(u) + \sum_{1 \le l \le \ell} \Big(1_{w(l) \neq w^*(l), u(l) = w^*(l)} - 1_{w(l) = w^*(l), u(l) \neq w^*(l)} \Big) \,. \end{aligned}$$

According to the mutation kernel M, for indices l such that $u(l) = w^*(l)$, the variable $1_{w(l) \neq w^*(l)}$ is Bernoulli with parameter $p(1 - 1/\kappa)$, while for indices l such that $u(l) \neq w^*(l)$, the variable $1_{w(l) = w^*(l)}$ is Bernoulli with parameter p/κ. Moreover these Bernoulli variables are independent. Thus the law of $H(w)$ under the kernel $M(u, w)$ is given by

$$H(u) + \text{Binomial}\big(\ell - H(u), p(1 - 1/\kappa)\big) - \text{Binomial}\big(H(u), p/\kappa\big)$$

where the two binomial random variables are independent. This law depends only on $H(u)$, therefore the sum

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

is a function of $b = H(u)$ and $c = H(w)$ only, which we denote by $M_H(b, c)$. The formula for the lumped matrix M_H is obtained by computing the law of the difference of the two independent binomial laws appearing above. $\square$

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

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

then we have

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

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

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

The partition of $\big(\mathcal{A}^\ell\big)^m$ induced by the map $\mathbb{H}$ is

$$\mathbb{H}^{-1}(\{\,d\,\})\,,\qquad d\in\{\,0,\dots,\ell\,\}^m\,.$$

We define finally the distance process $(D_t)_{t\geq 0}$ by

$$\forall t\geq 0\qquad D_t\,=\,\mathbb{H}\big(X_t\big)\,.$$

Our next goal is to prove that the process $(X_t)_{t\geq 0}$ 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 $(D_t)_{t\geq 0}$ is a genuine Markov process.

Proposition 6.2 ($\mathbb{H}$ Lumpability) Let p be the transition matrix of the Moran model. We have

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

Proof. For the process $(X_t)_{t\geq 0}$, the only transitions having positive probability are the transitions of the form

$$x\quad\longrightarrow\quad x(j\leftarrow u)\,,\qquad 1\leq j\leq m,\quad u\in\mathcal{A}^\ell\,.$$

Let $e\in\{\,0,\dots,\ell\,\}^m$ and let $x,y\in\big(\mathcal{A}^\ell\big)^m$ be such that $\mathbb{H}(x)=\mathbb{H}(y)$. We set $d=\mathbb{H}(x)=\mathbb{H}(y)$. If the vectors d,e differ for more than two components, then the sums appearing in the statement of the proposition are equal to zero. Suppose first that the vectors d,e differ in exactly one component, so that there exist $j\in\{\,1,\dots,m\,\}$ and $c\in\{\,0,\dots,\ell\,\}$ such that $e=d(j\leftarrow c)$ and $d(j)\neq c$. Naturally, $d(j\leftarrow c)$ is the vector d in which the j–th component $d(j)$ has been replaced by c:

$$d(j\leftarrow c)\,=\,\begin{pmatrix}d(1)\\ \vdots\\ d(j-1)\\ c\\ d(j+1)\\ \vdots\\ d(m)\end{pmatrix}$$

We have then

$$\sum_{\substack{z\in(\mathcal{A}^\ell)^m\\ \mathbb{H}(z)=e}}p(x,z)\,=\,\sum_{\substack{w\in\mathcal{A}^\ell\\ H(w)=c}}p\big(x,x(j\leftarrow w)\big)\,.$$

Using lemma 6.1, we have

$$\sum_{\substack{w\in\mathcal{A}^\ell\\ H(w)=c}} p\big(x,x(j\leftarrow w)\big) \,=\, \sum_{\substack{w\in\mathcal{A}^\ell\\ H(w)=c}} \frac{1}{m}\sum_{1\leq i\leq m}\frac{A(x(i))M(x(i),w)}{A(x(1))+\cdots+A(x(m))}$$

$$=\,\frac{1}{m}\sum_{1\leq i\leq m}\frac{A_H\big(H(x(i))\big)M_H(H(x(i)),c)}{A_H\big(H(x(1))\big)+\cdots+A_H\big(H(x(m))\big)}\,.$$

This sum is a function of $\mathbb{H}(x)$ and c only. Since $\mathbb{H}(x)=\mathbb{H}(y)$, the sums are the same for x and y. Suppose next that $d=e$. Then

$$\sum_{\substack{z\in(\mathcal{A}^\ell)^m\\ \mathbb{H}(z)=e}} p(x,z) \,=\, p(x,x)\,+\sum_{1\leq j\leq m}\ \sum_{\substack{w\in\mathcal{A}^\ell\setminus\{\,x(j)\,\}\\ H(w)=H(x(j))}} p\big(x,x(j\leftarrow w)\big)$$

$$=\,1-\sum_{\substack{1\leq j\leq m\\ w\in\mathcal{A}^\ell\setminus\{\,x(j)\,\}}} p\big(x,x(j\leftarrow w)\big)+\sum_{1\leq j\leq m}\ \sum_{\substack{w\in\mathcal{A}^\ell\setminus\{\,x(j)\,\}\\ H(w)=H(x(j))}} p\big(x,x(j\leftarrow w)\big)$$

$$=\,1-\sum_{1\leq j\leq m}\ \sum_{\substack{w\in\mathcal{A}^\ell\setminus\{\,x(j)\,\}\\ H(w)\neq H(x(j))}} p\big(x,x(j\leftarrow w)\big)$$

$$=\,1-\sum_{1\leq j\leq m}\ \sum_{\substack{c\in\{\,0,\dots,\ell\,\}\\ c\neq H(x(j))}}\ \sum_{\substack{w\in\mathcal{A}^\ell\\ H(w)=c}} p\big(x,x(j\leftarrow w)\big)\,.$$

We have seen in the previous case that the last sum is a function of $\mathbb{H}(x)$ and c only. The second sum as well depends only on $\mathbb{H}(x)$. Therefore the above quantity is the same for x and y. □

We apply the classical lumping result (see theorem A.3) to conclude that the distance process $(D_t)_{t\geq 0}$ is a Markov process. From the previous computations, we see that its transition matrix p_H is given by

$$\forall d\in\{\,0,\dots,\ell\,\}^m\quad\forall j\in\{\,1,\dots,m\,\}\quad\forall c\in\{\,0,\dots,\ell\,\}\setminus\{\,d(j)\,\}$$

$$p_H\big(d,d(j\leftarrow c)\big)\,=\,\frac{1}{m}\sum_{1\leq i\leq m}\frac{A_H(d(i))M_H(d(i),c)}{A_H(d(1))+\cdots+A_H(d(m))}\,.$$

6.2 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 $(O_t)_{t\geq0}$ by setting

$$\forall t\geq0\qquad O_t\,=\,\mathcal{O}(X_t)\,=\,\mathcal{O}_H(D_t)\,.$$

For the process $(D_t)_{t\geq0}$, the only transitions having positive probability are the transitions of the form

$$d\quad\longrightarrow\quad d(j\leftarrow c)\,,\qquad1\leq j\leq m,\quad c\in\{\,0,\dots,\ell\,\}\,.$$

Therefore the only possible transitions for the process $(O_t)_{t\geq0}$ are

$$o\quad\longrightarrow\quad o(k\to l)\,,\qquad0\leq k,l\leq\ell\,,$$

where $o(k\to l)$ is the partition obtained by moving a chromosome from the class k to the class l, i.e.,

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

Proposition 6.3 ($\mathcal{O}$ Lumpability) Let p_H be the transition matrix of the distance process. We have

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

Proof. 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:

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

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

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

as requested. $\square$

We apply the classical lumping result (see theorem A.3) to conclude that the occupancy process $(O_t)_{t\geq 0}$ is a Markov process. Let us compute its transition probabilities. Let $o\in\mathcal{P}^m_{\ell+1}$ and $d\in\{\,0,\dots,\ell\,\}^m$ be such that $\mathcal{O}_H(d)\neq o$. Let us consider the sum

$$\sum_{\substack{f\in\{\,0,\dots,\ell\,\}^m\\ \mathcal{O}_H(f)=o}}p_H(d,f)\,.$$

The terms in the sum vanish unless

$$\exists\,j\in\{\,1,\dots,m\,\}\quad\exists\,c\in\{\,0,\dots,\ell\,\}\,,\quad c\neq d(j)\,,\qquad f=d(j\leftarrow c)\,.$$

Suppose that it is the case. If in addition f is such that $\mathcal{O}_H(f)=o$, then

$$o\,=\,\mathcal{O}_H(d)(d(j)\to c)\,.$$

Setting $k=d(j)$ and $l=c$, we conclude that

$$\exists\,k,l\in\{\,0,\dots,\ell\,\}\qquad o\,=\,\mathcal{O}_H(d)(k\to l)\,.$$

The two indices k,l satisfying the above condition are distinct and unique. We have then

$$\begin{aligned}\sum_{\substack{f\in\{\,0,\dots,\ell\,\}^m\\ \mathcal{O}_H(f)=o}}p_H(d,f)\,&=\,\sum_{\substack{j\in\{\,1,\dots,m\,\}\\ d(j)=k}}p_H(d,d(j\leftarrow l))\\&=\,\frac{\mathcal{O}_H(d)(k)\displaystyle\sum_{0\leq h\leq\ell}\mathcal{O}_H(d)(h)\,A_H(h)\,M_H(h,l)}{m\displaystyle\sum_{0\leq h\leq\ell}\mathcal{O}_H(d)(h)\,A_H(h)}\,.\end{aligned}$$

This fraction is a function of $\mathcal{O}_H(d)$, k and l, thus it depends only on $\mathcal{O}_H(d)$ and o as requested. We conclude that the transition matrix of the occupancy process is given by

$$\forall o\in\mathcal{P}^m_{\ell+1}\quad\forall k,l\in\{\,0,\dots,\ell\,\}\,,\quad k\neq l\,,$$

$$p_O\big(o,o(k\to l)\big)\,=\,\frac{\displaystyle o(k)\sum_{h=0}^{\ell}o(h)\,A_H(h)\,M_H(h,l)}{\displaystyle m\sum_{h=0}^{\ell}o(h)\,A_H(h)}\,.$$

6.3 Invariant probability measures

There are several advantages in working with the lumped processes. The main advantage is that the state space is considerably smaller. For the process $(X_t)_{t\geq 0}$, the cardinality of the state space is

$$\operatorname{card}\big(\mathcal{A}^\ell\big)^m\,=\,\kappa^{\ell m}\,.$$

For the distance process $(D_t)_{t\geq 0}$, it becomes

$$\operatorname{card}\{\,0,\dots,\ell\,\}^m\,=\,(\ell+1)^m\,.$$

Finally for the occupancy process, the cardinality is the number of ordered partitions of m into at most $\ell+1$ parts. This number is quite complicated to compute, but in any case

$$\operatorname{card}\mathcal{P}^m_{\ell+1}\,\leq\,(\ell+1)^m\,.$$

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 $f:[0,1]\to\mathbb{R}$,

$$\int_{[0,1]}f\,d\nu\,=\,\lim_{t\to\infty}E\Big(\,f\Big(\frac{1}{m}N(X_t)\Big)\Big)\,=\,\int_{\left(\mathcal{A}^\ell\right)^m}f\Big(\frac{1}{m}N(x)\Big)\,d\mu(x)$$

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

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

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

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

Yet $N(x)$ is also lumpable with respect to $\mathbb{H}$, i.e., it can be written as a function of $\mathbb{H}(x)$:

$$\forall x \in \big(\mathcal{A}^\ell\big)^m \qquad N(x) \,=\, N_H(\mathbb{H}(x))\,,$$

where N_H is the lumped function defined by

$$\forall d \in \{\,0,\dots,\ell\,\}^m \qquad N_H(d) \,=\, \text{card}\big\{\, i : 1 \le i \le m,\, d(i) = 0\,\big\}\,.$$

Let μ_H be the invariant probability measure of the process $(D_t)_{t\geq 0}$. For $d \in \{\,0,\dots,\ell\,\}^m$, we have

$$\begin{aligned}\mu_H(d) \,=\, \lim_{t\to\infty} P\big(D_t = d\big) \,=\, \lim_{t\to\infty} P\big(\mathbb{H}(X_t) = d\big)\\ =\, \lim_{t\to\infty} P\big(X_t \in \mathbb{H}^{-1}(d)\big) \,=\, \mu\big(\mathbb{H}^{-1}(d)\big)\,.\end{aligned}$$

Thus, as it was naturally expected, the probability measure μ_H is the image of the probability measure μ through the map $\mathbb{H}$. It follows that, for any function $f : [0,1] \to \mathbb{R}$,

$$\begin{aligned}\int_{[0,1]} f\, d\nu \,&=\, \int_{(\mathcal{A}^\ell)^m} f\Big(\frac{1}{m}N(x)\Big)\, d\mu(x)\\ &=\, \int_{(\mathcal{A}^\ell)^m} f\Big(\frac{1}{m}N_H\big(\mathbb{H}(x)\big)\Big)\, d\mu(x)\\ &=\, \int_{\{\,0,\dots,\ell\,\}^m} f\Big(\frac{1}{m}N_H(d)\Big)\, d\mu_H(d)\,.\end{aligned}$$

Similarly, the invariant probability measure μ_O of the process $(O_t)_{t\geq 0}$ is the image measure of μ through the map $\mathcal{O}$, and also the image measure of μ_H through the map $\mathcal{O}_H$. We have also, for any function $f : [0,1] \to \mathbb{R}$,

$$\int_{[0,1]} f\, d\nu \,=\, \int_{\mathcal{P}^m_{\ell+1}} f\Big(\frac{1}{m}o(0)\Big)\, d\mu_O(o)\,.$$

Another advantage of the lumped processes is that the spaces $\{\,0,\dots,\ell\,\}^m$ and $\mathcal{P}^m_{\ell+1}$ are naturally endowed with a partial order. Since we cannot deal directly with the distance process $(D_t)_{t\geq 0}$ or the occupancy process $(O_t)_{t\geq 0}$, we shall compare them with auxiliary processes whose dynamics is much simpler.

7 Monotonicity

A crucial property for comparing the Moran model with other processes is monotonicity. We will realize a coupling of the lumped Moran processes with different initial conditions and we will deduce the monotonicity from the coupling construction.

7.1 Coupling of the lumped processes

We build here a coupling of the lumped processes, on the same probability space as the coupling for the process $(X_t)_{t\geq 0}$ described in section 4. We set

$$\forall n \geq 1 \qquad R_n \,=\, \big(S_n, I_n, J_n, U_{n,1}, \dots, U_{n,\ell}\big)\,.$$

The vector R_n is the random input which is used to perform the n–th step of the Markov chain $(X_t)_{t\geq 0}$. By construction the sequence $(R_n)_{n\geq 1}$ is a sequence of independent identically distributed random vectors with values in

$$\mathcal{R} \,=\, [0,1] \times \{\,1, \dots, m\,\}^2 \times [0,1]^\ell\,.$$

We first define two maps $\mathcal{M}_H$ and $\mathcal{S}_H$ in order to couple the mutation and the selection mechanisms.

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 row 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

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

in order to realize the coupling between distance processes with various initial conditions and different parameters σ or p. The coupling map Φ_H is defined by

$$\begin{aligned} &\forall r = (s, i, j, u_1, \dots, u_\ell) \in \mathcal{R}, \quad \forall d \in \{ 0, \dots, \ell \}^m \\ &\qquad\qquad \Phi_H(d,r) = d\big(j \leftarrow \mathcal{M}_H(d(\mathcal{S}_H(d,s)), u_1, \dots, u_\ell)\big) . \end{aligned}$$

Notice that the index i is not used in the map Φ_H. 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 Φ_H. Let $d \in \{ 0, \dots, \ell \}^m$ be the starting point of the process. We build the distance process $(D_t)_{t \geq 0}$ by setting $D(0) = d$ and

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

A routine check shows that the process $(D_t)_{t \geq 0}$ is a Markov chain starting from d with the adequate transition matrix. This way we have coupled the distance processes with various initial conditions and different parameters σ or p.

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 $o \in \mathcal{P}^m_{\ell+1}$ 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 $o \in \mathcal{P}^m_{\ell+1}$, 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

$$\Phi_O : \mathcal{P}^m_{\ell+1} \times \mathcal{R} \to \mathcal{P}^m_{\ell+1}$$

in order to realize the coupling between occupancy processes with various initial conditions and different parameters σ or p. The coupling map Φ_O is defined as follows. Let $r=(s,i,j,u_1,\dots,u_\ell)\in\mathcal{R}$. Let $o \in \mathcal{P}^m_{\ell+1}$, let us set $l=\mathcal{S}_O(o,s)$ and let k be the unique index in $\{\, 0,\dots,\ell \,\}$ satisfying

$$o(0)+\cdots+o(k-1) \,<\, j \,\leq o(0)+\cdots+o(k)\,.$$

The coupling map Φ_O is defined by

$$\Phi_O(o,r) \,=\, o\big(k \to \mathcal{M}_H(l,u_1,\dots,u_\ell)\big)\,.$$

Notice that the index i is not used in the map Φ_O. Let $o \in \mathcal{P}^m_{\ell+1}$ be the starting point of the process. We build the occupancy process $(O_t)_{t\geq 0}$ by setting $O(0)=o$ and

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

A routine check shows that the process $(O_t)_{t\geq 0}$ is a Markov chain starting from o with the adequate transition matrix. This way we have coupled the occupancy processes with various initial conditions and different parameters σ or p.

7.2 Monotonicity of the model

The space $\{\, 0,\dots,\ell \,\}^m$ is naturally endowed with a partial order:

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

Lemma 7.1 The map $\mathcal{M}_H$ is non–decreasing with respect to the Hamming class, i.e.,

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

Proof. We simply use the definition of $\mathcal{M}_H$ (see section 7.1) and we compute the difference

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

Since $\kappa \geq 2$, the absolute value of the sum is at most $c-b$ and the above difference is non–negative. ☐

Lemma 7.2 In the neutral case $\sigma = 1$, the map $\mathcal{S}_H$ is non–decreasing with respect to the Hamming class, i.e.,

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

Proof. In fact, when $\sigma = 1$, the map $\mathcal{S}_H$ depends only on the second variable s:

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

It follows that if $d,e\in\{\,0,\dots,\ell\,\}^m$ are such that $d\leq e$, then

$$\forall s\in[0,1] \qquad d\big(\lfloor ms\rfloor+1\big) \,\leq\, e\big(\lfloor ms\rfloor+1\big)$$

as requested. ☐

Lemma 7.3 In the neutral case $\sigma = 1$, the map Φ_H is non–decreasing with respect to the distances, i.e.,

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

Proof. Let $r=(s,i,j,u_1,\dots,u_\ell)\in\mathcal{R}$ and let $d,e\in\{\,0,\dots,\ell\,\}^m$, $d\leq e$. By lemma 7.2, we have

$$d\big(\mathcal{S}_H(d,s)\big) \,\leq\, e\big(\mathcal{S}_H(e,s)\big)\,.$$

This inequality and lemma 7.1 imply that

$$\mathcal{M}_H\big(d(\mathcal{S}_H(d,s)),u_1,\dots,u_\ell)\big) \,\leq\, \mathcal{M}_H\big(e(\mathcal{S}_H(e,s)),u_1,\dots,u_\ell\big)\,,$$

so that

$$d\big(j\leftarrow\mathcal{M}_H(d(\mathcal{S}_H(d,s)),u_1,\dots,u_\ell)\big) \\ \leq\, e\big(j\leftarrow\mathcal{M}_H(e(\mathcal{S}_H(e,s)),u_1,\dots,u_\ell)\big)\,,$$

whence $\Phi_H(d,r) \,\leq\, \Phi_H(e,r)$ as requested. $\square$

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

$$\kappa = 3\,, \quad \sigma = 2\,, \quad m = 3\,, \quad \frac{2}{3} < s < \frac{3}{4}\,,$$
$$u_1, \dots, u_\ell \in \Big[\frac{p}{3}, 1 - \frac{2p}{3}\Big]\,, \quad j = i = 1\,,$$

then

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

This creates a serious complication. This is why we perform a second lumping and we work with the occupancy process rather than with the distance process. 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

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

Lemma 7.4 The map $\mathcal{S}_O$ is non–increasing with respect to the occupancy distribution, i.e.,

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

Proof. Let $o \preceq o'$. Let $l \in \{\,0, \dots, \ell\,\}$. We have

$$o(0)A_H(0) + \cdots + o(l)A_H(l) \,=\, o(0)(\sigma - 1) + o(0) + \cdots + o(l)\,.$$

Thus

$$\frac{o(0)A_H(0) + \cdots + o(l)A_H(l)}{o(0)A_H(0) + \cdots + o(\ell)A_H(\ell)} \,=\, \psi\big(o(0), o(0) + \cdots + o(l)\big)\,,$$

where ψ is the function defined by

$$\forall \eta, \xi \in [0,m] \qquad \psi(\eta, \xi) \,=\, \frac{\eta(\sigma - 1) + \xi}{\eta(\sigma - 1) + m}\,.$$

The map ψ is non–decreasing in η and ξ on $[0,m]^2$, therefore

$$\psi\big(o(0), o(0) + \cdots + o(l)\big) \,\leq\, \psi\big(o'(0), o'(0) + \cdots + o'(l)\big)\,,$$

i.e.,

$$\frac{o(0)A_H(0)+\cdots+o(l)A_H(l)}{o(0)A_H(0)+\cdots+o(\ell)A_H(\ell)} \leq \frac{o'(0)A_H(0)+\cdots+o'(l)A_H(l)}{o'(0)A_H(0)+\cdots+o'(\ell)A_H(\ell)}\,.$$

It follows that $\mathcal{S}_O(o,s) \geq \mathcal{S}_O(o',s)$ for any $s \in [0,1]$. $\square$

Lemma 7.5 The map Φ_O is non–decreasing with respect to the occupancy distributions, i.e.,

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

Proof. Let $r=(s,i,j,u_1,\dots,u_\ell) \in \mathcal{R}$ and let $o,o' \in \mathcal{P}^m_{\ell+1}$ be such that $o \preceq o'$. Let us set $l=\mathcal{S}_O(o,s)$, $l'=\mathcal{S}_O(o',s)$ and let k,k' be the unique indices in $\{\,0,\dots,\ell\,\}$ satisfying

$$\begin{aligned} o(0)+\cdots+o(k-1) \,&<\, j \,\leq o(0)+\cdots+o(k)\,,\\ o'(0)+\cdots+o'(k'-1) \,&<\, j \,\leq o'(0)+\cdots+o'(k')\,. \end{aligned}$$

Since $o \preceq o'$, we have $k \geq k'$. Let us set

$$b=\mathcal{M}_H(l,u_1,\dots,u_\ell)\,, \qquad b'=\mathcal{M}_H(l',u_1,\dots,u_\ell)\,.$$

Since $l \geq l'$ by lemma 7.4, we have $b \geq b'$ by lemma 7.1. We must now compare

$$\Phi_O(o,r) \,=\, o(k \to b)\,, \qquad \Phi_O(o',r) \,=\, o'(k' \to b')\,.$$

Let $h \in \{\,0,\dots,\ell\,\}$. We have

$$o(k\to b)(0)+\cdots+o(k\to b)(h) \,=\, o(0)+\cdots+o(h) - 1_{k\leq h} + 1_{b\leq h}\,.$$

Since $o \preceq o'$, we have $o(0)+\cdots+o(h) \leq o'(0)+\cdots+o'(h)$. Since $b \geq b'$, we have $1_{b\leq h} \leq 1_{b'\leq h}$. The problem comes from the indicator function $1_{k\leq h}$. We consider several cases:

- $k \leq h$. Then

$$\begin{aligned} o(0)+\cdots+o(h)-1_{k\leq h}+1_{b\leq h} \,&\leq\, o'(0)+\cdots+o'(h)-1+1_{b\leq h}\\ &\leq\, o'(0)+\cdots+o'(h)-1_{k'\leq h}+1_{b'\leq h}\,. \end{aligned}$$

- $k' \leq h < k$. The definition of k,k' implies that

$$o(0)+\cdots+o(h) \,<\, j \,\leq o'(0)+\cdots+o'(h)$$

whence

$$o(0)+\cdots+o(h) \,\leq o'(0)+\cdots+o'(h)-1\,.$$

It follows that

$$o(0)+\cdots+o(h)+1_{b\leq h}\ \leq\ o'(0)+\cdots+o'(h)-1_{k'\leq h}+1_{b'\leq h}\,.$$

- $h<k'$. Then

$$o(0)+\cdots+o(h)+1_{b\leq h}\ \leq\ o'(0)+\cdots+o'(h)+1_{b'\leq h}\,.$$

In each case, we have

$$o(k\to b)(0)+\cdots+o(k\to b)(h)\ \leq\ o'(k'\to b')(0)+\cdots+o'(k'\to b')(h)\,.$$

Therefore $\Phi_O(o,r)\ \preceq\ \Phi_O(o',r)$ as requested. $\square$

Let us try to see the implications of the previous results for the monotonicity of the model (see the appendix for the definition of a monotone process). There is not much to do with the original Moran model, because its state space is not partially ordered. So we examine the distance process and the occupancy process.

Corollary 7.6 In the neutral case $\sigma=1$, the distance process $(D_t)_{t\geq 0}$ is monotone.

Indeed, by lemma 7.3, the map Φ_H is non–decreasing in the neutral case $\sigma=1$, hence the coupling is monotone. Unfortunately, we did not manage to reach the same conclusion in the non neutral case. The main point of lumping further the distance process is to get a process which is monotone even in the non neutral case.

Corollary 7.7 The occupancy process $(O_t)_{t\geq 0}$ is monotone.

By lemma 7.5, the coupling for the occupancy process is monotone.

8 Stochastic bounds

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

8.1 Lower and upper processes

We shall construct a lower process $(O^\ell_t)_{t\geq 0}$ and an upper process $(O^1_t)_{t\geq 0}$ satisfying

$$\forall t \geq 0 \qquad O^\ell_t \,\preceq\, O_t \,\preceq\, O^1_t \,.$$

Loosely speaking, the upper process evolves as follows. As long as there is no master sequence present in the population, the process $(O^1_t)_{t\geq 0}$ evolves exactly as the initial process $(O_t)_{t\geq 0}$. 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 lower 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 Φ_O be the coupling map defined in section 7.1. We define a lower map Φ^ℓ_O by setting, for $o \in \mathcal{P}^m_{\ell+1}$ and $r \in \mathcal{R}$,

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

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

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

A direct application of lemma 7.5 yields that the map Φ^ℓ_O is below the map Φ_O and the map Φ^1_O is above the map Φ_O in the following sense:

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

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

$$\forall n \geq 1 \qquad O^\ell_n \,=\, \Phi^\ell_O\big(O^\ell_{n-1}, R_n\big)\,, \qquad O^1_n \,=\, \Phi^1_O\big(O^1_{n-1}, R_n\big)\,.$$

Proposition 8.1 Suppose that the processes $(O^\ell_t)_{t\geq 0}$, $(O_t)_{t\geq 0}$, $(O^1_t)_{t\geq 0}$, start from the same occupancy distribution o. We have

$$\forall t \geq 0 \qquad O^\ell_t \,\preceq\, O_t \,\preceq\, O^1_t\,.$$

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

$$O^\ell_{n+1} \,=\, \Phi^\ell_O\big(O^\ell_n, R_n\big)\,, \quad O_{n+1} \,=\, \Phi_O\big(O_n, R_n\big)\,, \quad O^1_{n+1} \,=\, \Phi^1_O\big(O^1_n, R_n\big)\,.$$

We use the induction hypothesis and we apply lemma 7.5 to get

$$\Phi_O\big(O^\ell_n, R_n\big) \,\preceq\, \Phi_O\big(O_n, R_n\big) \,\preceq \Phi_O\big(O^1_n, R_n\big)\,.$$

Yet the map Φ^ℓ_O is below the map Φ_O and the map Φ^1_O is above the map Φ_O, thus

$$\Phi^\ell_O\big(O^\ell_n, R_n\big) \,\preceq\, \Phi_O\big(O^\ell_n, R_n\big)\,, \qquad \Phi_O\big(O^1_n, R_n\big) \,\preceq\, \Phi^1_O\big(O^1_n, R_n\big).$$

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

8.2 Dynamics of the bounding processes

We study next the dynamics of the processes $(O_t^\ell)_{t\geq 0}$ and $(O_t^1)_{t\geq 0}$ in $\mathcal{W}^*$. The computations are the same for both processes. Throughout the section, we let θ be either 1 or ℓ and we denote by $(O_t^\theta)_{t\geq 0}$ the corresponding process. For the process $(O_t^\theta)_{t\geq 0}$, 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. Let o^θ_{enter} be the occupancy distribution having one master sequence and $m-1$ chromosomes in the Hamming class θ:

$$\forall l\in\{\,0,\dots,\ell\,\}\qquad o^\theta_{\text{enter}}(l)\;=\;\begin{cases}1 & \text{if } l=0\\ m-1 & \text{if } l=\theta\\ 0 & \text{otherwise}\end{cases}\,.$$

The process $(O_t^\theta)_{t\geq 0}$ always enters the set $\mathcal{W}^*\setminus\mathcal{T}^\theta$ at o^θ_{enter}. The only possible transitions for the first occupancy number of the process $(O_t^\theta)_{t\geq 0}$ starting from a point in $\mathcal{W}^*\setminus\mathcal{T}^\theta$ are

$$\begin{aligned} o(0)\quad&\longrightarrow\quad o(0)-1\,, &\qquad 1\leq o(0)\leq m\,,\\ o(0)\quad&\longrightarrow\quad o(0)+1\,, &\qquad 0\leq o(0)\leq m-1\,.\end{aligned}$$

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 $(O_t^\theta)_{t\geq 0}$ always exits $\mathcal{W}^*\setminus\mathcal{T}^\theta$ at o^θ_{exit}. From the previous observations, we conclude that, whenever $(O_t^\theta)_{t\geq 0}$ starts in $\mathcal{W}^*\setminus\mathcal{T}^\theta$, the dynamics of $(O_t^\theta(0))_{t\geq 0}$ is the one of a standard birth and death process, until the time of exit from $\mathcal{W}^*\setminus\mathcal{T}^\theta$. We denote by $(Z_t^\theta)_{t\geq 0}$ a birth and death process on $\{\,0,\dots,m\,\}$ starting at $Z_0^\theta\;=\;1$ with the following transition probabilities:

- Transitions to the left. For $i \in \{ 1, \dots, m \}$,

$$\begin{aligned} P\big(Z^\theta_{t+1} = i-1 \,|\, Z^\theta_t = i\big) \,&=\, P\big(O^\theta_{t+1}(0) = i-1 \,|\, O^\theta_t(0) = i\big) \\ &= \frac{\sigma i^2\big(1 - M_H(0,0)\big) + i(m-i)\big(1 - M_H(\theta,0)\big)}{m(\sigma i + m - i)}\,. \end{aligned}$$

- Transitions to the right. For $i \in \{ 0, \dots, m-1 \}$,

$$\begin{aligned} P\big(Z^\theta_{t+1} = i+1 \,|\, Z^\theta_t = i\big) \,&=\, P\big(O^\theta_{t+1}(0) = i+1 \,|\, O^\theta_t(0) = i\big) \\ &= \frac{\sigma i(m-i) M_H(0,0) + (m-i)^2 M_H(\theta,0)}{m(\sigma i + m - i)}\,. \end{aligned}$$

8.3 A renewal argument

We prove here a formula for the invariant measure involving two stopping times. The proof is based on a standard renewal argument and the formula is a variant of other well known formulas. Because this formula is a key of the whole analysis, we include its proof here. Let $(X_t)_{t\geq 0}$ 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 $(X_t)_{t\geq 0}$.

Proposition 8.2 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

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

We have

$$\int_{\mathcal{E}} f(x)\, d\mu(x) \,=\, \frac{1}{E(\tau \,|\, X_0 = e)}\, E\bigg(\int_{\tau^*}^{\tau} f(X_s)\, ds \,\Big|\, X_0 = e \bigg)\,.$$

Proof. We define two sequences $(\tau^*_k)_{k\geq 1}$, $(\tau_k)_{k\geq 0}$ of stopping times by setting $\tau_0 = 0$ and

$$\begin{aligned} \tau^*_1 \,&=\, \inf \big\{ t \geq 0 : X_t \in \mathcal{W}^* \big\}\,, & \tau_1 \,&=\, \inf \big\{ t \geq \tau^*_1 : X_t = e \big\}\,, \\ &\vdots & &\vdots \\ \tau^*_k \,&=\, \inf \big\{ t \geq \tau_{k-1} : X_t \in \mathcal{W}^* \big\}\,, & \tau_k \,&=\, \inf \big\{ t \geq \tau^*_k : X_t = e \big\}\,, \\ &\vdots & &\vdots \end{aligned}$$

Our first goal is to evaluate the asymptotic behavior of τ_k as k goes to ∞. For any $k \geq 1$, by the strong Markov property, the trajectory $(X_t)_{t\geq \tau_k}$ of the process after time τ_k is independent of the trajectory $(X_t)_{t\leq\tau_k}$ of the process until time τ_k, and its law is the same as the law of the whole process $(X_t)_{t\geq 0}$ starting from e. As a consequence, the successive excursions

$$\big(X_t,\, \tau_k \leq t \leq \tau_{k+1}\big)\,, \qquad k \geq 1\,,$$

are independent identically distributed. In particular, the sequence

$$\big(\tau_{k+1}-\tau_k\big)_{k\geq 1}$$

is a sequence of i.i.d. random variables, having the same law as the random time τ_1 whenever the process $(X_t)_{t\geq 0}$ starts from e. For $k\geq 1$, we decompose τ_k as the sum

$$\tau_k \,=\, \tau_1 + \sum_{h=1}^{k-1}\big(\tau_{h+1}-\tau_h\big)\,.$$

We denote by $E_e(\cdot)$ the expectation for the process $(X_t)_{t\geq 0}$ starting from e. Since the state space $\mathcal{E}$ is finite, the random time τ_1 is finite with probability one, and it is also integrable. Applying the classical law of large numbers, we get

$$\lim_{k\to\infty}\frac{\tau_k}{k} \,=\, E_e(\tau_1) \qquad \text{with probability 1.}$$

Whenever the process $(X_t)_{t\geq 0}$ starts from e, the random times τ_1^*, τ_1 satisfy $\tau_1^*\geq 1$, $\tau_1\geq 2$, therefore the expected mean $E_e(\tau_1)$ is strictly positive and we conclude that

$$\lim_{k\to\infty}\tau_k \,=\, +\infty \qquad \text{with probability 1.}$$

We define next

$$\forall t\geq 0 \qquad K(t) \,=\, \max\big\{\,k\geq 0 : \tau_k\leq t\,\big\}\,.$$

From the previous discussion, we see that, with probability one, $K(t)$ is finite for any $t\geq 0$. From the very definition of $K(t)$, we have

$$\forall t\geq 0 \qquad \tau_{K(t)} \,\leq\, t \,<\, \tau_{K(t)+1}\,,$$

and since τ_k goes to ∞ with k, then

$$\lim_{t\to\infty}K(t) \,=\, +\infty \qquad \text{with probability 1.}$$

We rewrite the previous double inequality as

$$\frac{\tau_{K(t)}}{K(t)} \,\leq\, \frac{t}{K(t)} \,<\, \frac{\tau_{K(t)+1}}{K(t)+1}\times\frac{K(t)+1}{K(t)}\,.$$

Sending t to ∞, we conclude that

$$\lim_{t\to\infty} \frac{K(t)}{t} = \frac{1}{E_e(\tau_1)} \qquad \text{with probability 1.}$$

We suppose that the process $(X_t)_{t\geq 0}$ starts from e. Let f be a map from $\mathcal{E}$ to $\mathbb{R}$ which vanishes on $\mathcal{E}\setminus\mathcal{W}^*$. By the ergodic theorem A.2, we have

$$\lim_{t\to\infty} E_e\big(f(X_t)\big) = \lim_{t\to\infty} \frac{1}{t}\int_0^t f(X_s)\,ds\,.$$

We decompose the last integral as follows:

$$\int_0^t f(X_s)\,ds = \sum_{k=1}^{K(t)} \int_{\tau_k^*}^{\tau_k} f(X_s)\,ds + \int_{\tau^*_{K(t)+1}\wedge t}^{t} f(X_s)\,ds\,,$$

where $\tau^*_{K(t)+1}\wedge t$ stands for $\min(\tau^*_{K(t)+1},t)$. For $k\geq 1$, the integral

$$N_k = \int_{\tau_k^*}^{\tau_k} f(X_s)\,ds$$

is a deterministic function of the excursion $\big(X_t,\, \tau_{k-1}\leq t\leq\tau_k\big)$, hence the random variables $(N_k, k\geq 1)$ are independent identically distributed. With probability one, $K(t)$ goes to ∞ as t goes to ∞, thus by the classical law of large numbers, we have

$$\lim_{t\to\infty} \frac{1}{K(t)}\sum_{k=1}^{K(t)} N_k = E_e(N_1) \qquad \text{with probability 1.}$$

We write

$$\frac{1}{t}\int_0^t f(X_s)\,ds = \frac{K(t)}{t}\times\frac{1}{K(t)}\sum_{k=1}^{K(t)} N_k + \frac{1}{t}\int_{\tau^*_{K(t)+1}\wedge t}^{t} f(X_s)\,ds\,.$$

We have

$$\begin{aligned}\left|\frac{1}{t}\int_{\tau^*_{K(t)+1}\wedge t}^{t} f(X_s)\,ds\right| &\leq \big(\sup_{[0,1]}|f|\big)\frac{1}{t}\big(\tau_{K(t)+1}-\tau_{K(t)}\big)\\ &\leq \big(\sup_{[0,1]}|f|\big)\frac{K(t)}{t}\Big(\frac{K(t)+1}{K(t)}\,\frac{\tau_{K(t)+1}}{K(t)+1}-\frac{\tau_{K(t)}}{K(t)}\Big)\end{aligned}$$

and the righthand side goes to 0 with probability 1 when t goes to ∞. We conclude that

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t f(X_s)\,ds = \frac{E_e(N_1)}{E_e(\tau_1)} \qquad \text{with probability 1}\,.$$

This yields the desired formula. □

8.4 Bounds on ν

We denote by μ_O^ℓ, μ_O, μ_O^1 the invariant probability measures of the processes $(O_t^\ell)_{t\geq 0}$, $(O_t)_{t\geq 0}$, $(O_t^1)_{t\geq 0}$. From section 6.3, the probability ν is 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 $f : [0,1] \to \mathbb{R}$,

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

We fix now a non–decreasing function $f : [0,1] \to \mathbb{R}$ such that $f(0) = 0$. Proposition 8.1 yields the inequalities

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

Taking the expectation and sending t to ∞, we get

$$\int_{\mathcal{P}_{\ell+1}^m} f\Big(\frac{o(0)}{m}\Big)\, d\mu_O^\ell(o) \,\leq\, \int_{[0,1]} f\, d\nu \,\leq\, \int_{\mathcal{P}_{\ell+1}^m} f\Big(\frac{o(0)}{m}\Big)\, d\mu_O^1(o)\,.$$

We seek next estimates on the above integrals. The strategy is the same for the lower and the upper integral. Thus we fix θ to be either 1 or ℓ and we study the invariant probability measure μ_O^θ. For the process $(O_t^\theta)_{t\geq 0}$, 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. We apply the renewal result of proposition 8.2 to the process $(O_t^\theta)_{t\geq 0}$ 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_{exit}^θ and the function $o \mapsto f(o(0)/m)$. Setting

$$\begin{aligned} \tau^* &\,=\, \inf\big\{\, t \geq 0 : O_t^\theta \in \mathcal{W}^* \,\big\}\,,\\ \tau &\,=\, \inf\big\{\, t \geq \tau^* : O_t^\theta = o_{\text{exit}}^\theta \,\big\}\,, \end{aligned}$$

we have

$$\int_{\mathcal{P}_{\ell+1}^m} f\Big(\frac{o(0)}{m}\Big)\, d\mu_O^\theta(o) \,=\, \frac{E\bigg(\displaystyle\int_{\tau^*}^{\tau} f\Big(\frac{O_s^\theta(0)}{m}\Big)\, ds \,\Big|\, O_0^\theta = o_{\text{exit}}^\theta\bigg)}{E\big(\tau \,|\, O_0^\theta = o_{\text{exit}}^\theta\big)}\,.$$

Yet, whenever the process $(O_t^\theta)_{t\geq 0}$ is in $\mathcal{W}^*\setminus\mathcal{T}^\theta$, the dynamics of $(O_t^\theta(0))_{t\geq 0}$ is the same as the birth and death process $(Z_t^\theta)_{t\geq 0}$ defined at the end of section 8.2. We suppose that $(Z_t^\theta)_{t\geq 0}$ starts from $Z_0^\theta = 1$. 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 $(O^\theta_t)_{t\geq 0}$ always enters $\mathcal{W}^*$ at o^θ_{enter} and it always exits $\mathcal{W}^*\setminus\mathcal{T}^\theta$ at o^θ_{exit}. In particular τ_0 coincides with the exit time of $\mathcal{W}^*\setminus\mathcal{T}^\theta$ after τ^*. From the previous elements, we see that $\big(O^\theta_t(0),\,\tau^*\leq t\leq\tau\big)$ has the same law as $\big(Z^\theta_t\,, 0\leq t\leq\tau_0\big)$, whence

$$E\bigg(\int_{\tau^*}^{\tau} f\Big(\frac{O^\theta_s(0)}{m}\Big)\,ds\,\Big|\,O^\theta_0=o^\theta_{\text{exit}}\bigg)\;=\;E\bigg(\int_0^{\tau_0} f\Big(\frac{Z^\theta_s}{m}\Big)\,ds\,\Big|\,Z^\theta_0=1\bigg)\,.$$

Moreover, using the Markov property, we have

$$E\Big(\tau-\tau^*\,\big|\,O^\theta_0=o^\theta_{\text{exit}}\Big)\;=\;E\Big(\tau\,\big|\,O^\theta_0=o^\theta_{\text{enter}}\Big)\;=\;E\Big(\tau_0\,\big|\,Z^\theta_0=1\Big)\,.$$

Reporting back in the formula for the invariant probability measure μ^θ_O, we get

$$\int_{\mathcal{P}^m_{\ell+1}} f\Big(\frac{o(0)}{m}\Big)\,d\mu^\theta_O(o)\;=\;\frac{E\bigg(\displaystyle\int_0^{\tau_0} f\Big(\frac{Z^\theta_s}{m}\Big)\,ds\,\Big|\,Z^\theta_0=1\bigg)}{E\big(\tau^*\,|\,O^\theta_0=o^\theta_{\text{exit}}\big)+E\big(\tau_0\,\big|\,Z^\theta_0=1\big)}\,.$$

In order to reinterpret this formula, we apply the renewal result stated in proposition 8.2 to the process $(Z^\theta_t)_{t\geq0}$, the set $\{\,1,\dots,m\,\}$, the point 0 and the map $f(\cdot/m)$. Setting

$$\tau_1\;=\;\inf\,\big\{\,t\geq 0: Z^\theta_t=1\,\big\}\,,$$

and denoting by ν^θ the invariant probability measure of $(Z^\theta_t)_{t\geq0}$, we have, with the help of the Markov property,

$$\sum_{i=1}^m f\Big(\frac{i}{m}\Big)\,\nu^\theta(i)\;=\;\frac{E\bigg(\displaystyle\int_0^{\tau_0} f\Big(\frac{Z^\theta_s}{m}\Big)\,ds\,\Big|\,Z^\theta_0=1\bigg)}{E\big(\tau_1\,|\,Z^\theta_0=0\big)+E\big(\tau_0\,\big|\,Z^\theta_0=1\big)}\,.$$

Yet

$$E\big(\tau_1\,|\,Z^\theta_0=0\big)\;=\;\frac{1}{P\big(Z^\theta_1=1\,|\,Z^\theta_0=0\big)}\;=\;\frac{1}{M_H(\theta,0)}\,.$$

We conclude finally that

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

To estimate the integral, we must estimate each term appearing on the right–hand side. In section 9, we deal with the terms involving the birth and death processes. In section 10, we deal with the discovery time τ^*.

9 Birth and death processes

We first give explicit formulas for a birth and death Markov chain that are well adapted to our situation. The formula for the invariant probability measure can be found in classical books, for instance [22].

9.1 General formulas

We consider a birth and death Markov chain $(Z_n)_{n\geq 0}$ on the finite set $\{\,0,\dots,m\,\}$ with transition probabilities given by

$$\begin{aligned} P(Z_{n+1}=i+1\,|\,Z_n=i) &= \delta_i\,, \quad 0\leq i\leq m-1\,,\\ P(Z_{n+1}=i-1\,|\,Z_n=i) &= \gamma_i\,, \quad 1\leq i\leq m\,, \end{aligned}$$

for any $n\geq 0$. We define

$$\pi(0)=1\,,\qquad \pi(i)\,=\,\frac{\delta_1\cdots\delta_i}{\gamma_1\cdots\gamma_i}\,,\qquad 1\leq i\leq m-1\,.$$

Let τ_0 be the hitting time of 0, defined by

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

We have the following explicit formula for the expected value of τ_0:

$$E(\tau_0\,|\,Z_0=1)\,=\,\sum_{i=1}^{m}\frac{1}{\gamma_i}\pi(i-1)\,.$$

Let ν be the invariant probability measure of $(Z_n)_{n\geq 0}$. We have the following explicit formula for ν:

$$\begin{aligned} \nu(0)\,&=\,\frac{1}{1+\delta_0\,E(\tau_0\,|\,Z_0=1)}\,,\\ \forall i\in\{\,1,\dots,m\,\}\qquad \nu(i)\,&=\,\frac{\dfrac{\delta_0}{\gamma_i}\pi(i-1)}{1+\delta_0\,E(\tau_0\,|\,Z_0=1)}\,. \end{aligned}$$

9.2 The case of $(Z^\theta_t)_{t\geq 0}$

We will now apply these formula to the birth and death chains $(Z^\theta_t)_{t\geq 0}$ introduced at the end of section 8.2. For these two processes, we have the following explicit formula for the transition probabilities:

$$\begin{aligned} \gamma_i\,&=\,\frac{\sigma i^2\big(1-M_H(0,0)\big)+i(m-i)\big(1-M_H(\theta,0)\big)}{m(\sigma i+m-i)}\,,\quad 1\leq i\leq m\,,\\ \delta_i\,&=\,\frac{\sigma i(m-i)M_H(0,0)+(m-i)^2M_H(\theta,0)}{m(\sigma i+m-i)}\,,\quad 0\leq i\leq m-1\,. \end{aligned}$$

The transition probabilities δ_i, γ_i depend on the parameters σ, ℓ, m, q as well as θ. We seek estimates of the expected value of τ_0 and of the asymptotic behavior of ν in the regime where

$$m, \ell \to +\infty \,, \quad q \to 0 \,.$$

For this reason, we choose the above specific forms of the formulas, which are well suited for our purposes. Since the results are the same for $\theta = 1$ and $\theta = \ell$, we drop the superscript θ from the notation, and we write simply Z_n, ν instead of Z_n^θ, ν^θ. Our first goal is to estimate the products $\pi(i)$. We start by studying the ratio δ_i/γ_i. We have

$$\forall i \in \{\, 1, \dots, m-1 \,\} \qquad \frac{\delta_i}{\gamma_i} \,=\, \phi\Big(M_H(0,0), M_H(\theta,0), \frac{i}{m}\Big) \,,$$

where $\phi :]0,1] \times [0,1[\times]0,1[\to]0,+\infty[$ is the function defined by

$$\phi(\beta,\varepsilon,\rho) \,=\, \frac{(1-\rho)\big(\sigma\beta\rho + (1-\rho)\varepsilon\big)}{\rho\big(\sigma(1-\beta)\rho + (1-\rho)(1-\varepsilon)\big)} \,.$$

What matters for the behavior of the products $\pi(i)$ is whether the values of ϕ are larger or smaller than 1. The equation $\phi(\beta,\varepsilon,\rho) = 1$ can be rewritten as

$$(\sigma-1)\rho^2 + (1 - \sigma\beta + \varepsilon)\rho - \varepsilon \,=\, 0 \,.$$

This equation admits one positive root, given by

$$\rho(\beta,\varepsilon) \,=\, \frac{1}{2(\sigma-1)}\Big(\sigma\beta - 1 - \varepsilon + \sqrt{(\sigma\beta - 1 - \varepsilon)^2 + 4\varepsilon(\sigma-1)}\Big) \,.$$

Therefore we have

$$\begin{aligned} &\phi(\beta,\varepsilon,\rho) > 1 \quad \text{if} \quad \rho < \rho(\beta,\varepsilon) \,,\\ &\phi(\beta,\varepsilon,\rho) < 1 \quad \text{if} \quad \rho > \rho(\beta,\varepsilon) \,. \end{aligned}$$

This readily implies that

$$\begin{gathered} 1 \le i \le j \le \lfloor \rho(\beta,\varepsilon) m \rfloor \qquad \Longrightarrow \qquad \pi(i) \,\le \pi(j) \,,\\ \lfloor \rho(\beta,\varepsilon) m \rfloor \le i \le j \le m \qquad \Longrightarrow \qquad \pi(i) \,\ge \pi(j) \,,\\ \pi\big(\lfloor \rho(\beta,\varepsilon) m \rfloor\big) \,\ge \pi\big(\lfloor \rho(\beta,\varepsilon) m \rfloor + 1\big) \,. \end{gathered}$$

It follows that the product $\pi(i)$ is maximal for $i = \lfloor \rho(\beta,\varepsilon) m \rfloor$:

$$\max_{1 \le i \le m} \pi(i) \,=\, \pi\big(\lfloor \rho(\beta,\varepsilon) m \rfloor\big) \,.$$

We notice in addition that $\phi(\beta,\varepsilon,\rho)$ is continuous and non–decreasing with respect to the first two variables β,ε. In the next two sections, we compute the relevant asymptotic estimates on the birth and death process. Lemma 6.1 yields

$$M_H(0,0)\,=\,\Big(1-p\big(1-\frac{1}{\kappa}\big)\Big)^{\ell}\,,$$

$$M_H(1,0)\,=\,\Big(1-p\big(1-\frac{1}{\kappa}\big)\Big)^{\ell-1}\frac{p}{\kappa}\,,\quad M_H(\ell,0)\,=\,\Big(\frac{p}{\kappa}\Big)^{\ell}\,.$$

As in theorem 3.1, we suppose that

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

in such a way that

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

In this regime, we have

$$\lim_{\substack{\ell\to\infty,\,q\to0\\ \ell q\to a}} M_H(0,0)\,=\,\exp(-a)\,,$$

$$\lim_{\substack{\ell\to\infty,\,q\to0\\ \ell q\to a}} M_H(1,0)\,=\,\lim_{\substack{\ell\to\infty,\,q\to0\\ \ell q\to a}} M_H(\ell,0)\,=\,0\,.$$

9.3 Persistence time

In this section, we will estimate the expected hitting time $E(\tau_0\,|\,Z_0=1)$. This quantity approximates the persistence time of the master sequence w^*. We estimate first the products $\pi(i)$.

Proposition 9.1 Let $a\in]0,+\infty[$. For $\rho\in[0,1]$, we have

$$\lim_{\substack{\ell,m\to\infty\\ q\to0,\,\ell q\to a}}\frac{1}{m}\ln\pi(\lfloor\rho m\rfloor)\,=\,\int_0^\rho\ln\phi(e^{-a},0,s)\,ds\,.$$

Proof. Let $\rho\in[0,1]$. For $m\geq1$, we have

$$\frac{1}{m}\ln\pi(\lfloor\rho m\rfloor)\,=\,\frac{1}{m}\sum_{i=1}^{\lfloor\rho m\rfloor}\ln\phi\Big(M_H(0,0),M_H(\theta,0),\frac{i}{m}\Big)\,.$$

Let $\varepsilon\in]0,e^{-a}[$. For ℓ,m large enough and q small enough, we have

$$\big|M_H(0,0)-e^{-a}\big|\,<\,\varepsilon\,,\qquad 0\,<\,M_H(\theta,0)\,<\,\varepsilon\,,$$

therefore, using the monotonicity properties of ϕ,

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^{\lfloor\rho m\rfloor}\ln\phi\Big(e^{-a}-\varepsilon,0,\frac{i}{m}\Big)\ \le\ &\frac{1}{m}\ln\pi(\lfloor\rho m\rfloor)\\ &\le\ \frac{1}{m}\sum_{i=1}^{\lfloor\rho m\rfloor}\ln\phi\Big(e^{-a}+\varepsilon,\varepsilon,\frac{i}{m}\Big)\,.\end{aligned}$$

These sums are Riemann sums. Letting ℓ, m go to ∞ and q go to 0, we get

$$\liminf_{\substack{\ell,m\to\infty\\ q\to0,\ \ell q\to a}}\ \frac{1}{m}\ln\pi(\lfloor\rho m\rfloor)\ \ge\ \int_0^\rho \ln\phi(e^{-a}-\varepsilon,0,s)\,ds\,,$$

$$\limsup_{\substack{\ell,m\to\infty\\ q\to0,\ \ell q\to a}}\ \frac{1}{m}\ln\pi(\lfloor\rho m\rfloor)\ \le\ \int_0^\rho \ln\phi(e^{-a}+\varepsilon,\varepsilon,s)\,ds\,.$$

We send ε to 0 to obtain the result stated in the proposition. □

We define

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

Since $\phi(e^{-a},0,s)>1$ for $s<\rho^*(a)$ and $\phi(e^{-a},0,s)<1$ for $s>\rho^*(a)$, the integral

$$\int_0^\rho \ln\phi(e^{-a},0,s)\,ds$$

is maximal for $\rho=\rho^*(a)$.

Corollary 9.2 Let $a\in]0,+\infty[$. The expected hitting time of 0 starting from 1 satisfies

$$\lim_{\substack{\ell,m\to\infty\\ q\to0,\ \ell q\to a}}\ \frac{1}{m}\ln E(\tau_0\,|\,Z_0=1)\ =\ \int_0^{\rho^*(a)}\ln\phi(e^{-a},0,s)\,ds\,.$$

Proof. We have the explicit formula

$$E(\tau_0\,|\,Z_0=1)\ =\ \sum_{i=1}^m\frac{1}{\gamma_i}\pi(i-1)$$

and the following bounds on γ_i:

$$\forall i\in\{\,1,\dots,m\,\}\qquad \frac{1-M_H(0,0)}{m^2}\ \le\ \gamma_i\ \le\ 2\sigma\,.$$

Let $\varepsilon \in]0, e^{-a}[$. For ℓ, m large enough and q small enough, we have

$$\big|M_H(0,0) - e^{-a}\big| \,<\, \varepsilon\,, \qquad 0 \,<\, M_H(\theta,0) \,<\, \varepsilon\,.$$

We first compute an upper bound:

$$\begin{aligned} E(\tau_0 \,|\, Z_0 = 1) \,&\leq\, \frac{m^3}{1 - M_H(0,0)} \max_{1 \leq i \leq m} \pi(i) \\ &\leq\, \frac{m^3}{1 - M_H(0,0)} \pi\Big(\big\lfloor \rho\big(M_H(0,0), M_H(\theta,0)\big) m \big\rfloor\Big)\,. \end{aligned}$$

Using the monotonicity properties of ϕ, we get

$$\begin{aligned} &\pi\big(\big\lfloor \rho\big(M_H(0,0), M_H(\theta,0)\big) m \big\rfloor\big) \\ &\qquad = \prod_{i=1}^{\lfloor \rho(M_H(0,0), M_H(\theta,0)) m \rfloor} \phi\Big(M_H(0,0), M_H(\theta,0), \frac{i}{m}\Big) \\ &\qquad\quad \leq \prod_{i=1}^{\lfloor \rho(M_H(0,0), M_H(\theta,0)) m \rfloor} \phi\Big(e^{-a} + \varepsilon, \varepsilon, \frac{i}{m}\Big) \\ &\qquad\qquad\qquad \leq \prod_{i=1}^{\lfloor \rho(e^{-a}+\varepsilon,\varepsilon) m \rfloor} \phi\Big(e^{-a} + \varepsilon, \varepsilon, \frac{i}{m}\Big)\,. \end{aligned}$$

The last inequality holds because the product $\pi(i)$ corresponding to the parameters $e^{-a} + \varepsilon$, ε is maximal for $i = \lfloor \rho(e^{-a} + \varepsilon, \varepsilon) m \rfloor$. We obtain that

$$E(\tau_0 \,|\, Z_0 = 1) \,\leq\, \frac{m^3}{1 - M_H(0,0)} \prod_{i=1}^{\lfloor \rho(e^{-a}+\varepsilon,\varepsilon) m \rfloor} \phi\Big(e^{-a} + \varepsilon, \varepsilon, \frac{i}{m}\Big)\,.$$

Taking logarithms, we recognize a Riemann sum, hence

$$\limsup_{\substack{\ell, m \to \infty \\ q \to 0,\, \ell q \to a}} \frac{1}{m} \ln E(\tau_0 \,|\, Z_0 = 1) \,\leq\, \int_0^{\rho(e^{-a}+\varepsilon,\varepsilon)} \ln \phi(e^{-a} + \varepsilon, \varepsilon, s)\, ds\,.$$

Conversely,

$$\begin{aligned} E(\tau_0 \,|\, Z_0 = 1) \,&\geq\, \frac{1}{2\sigma} \prod_{i=1}^{\lfloor \rho(e^{-a},0) m \rfloor} \phi\Big(M_H(0,0), M_H(\theta,0), \frac{i}{m}\Big) \\ &\qquad\qquad \geq\, \frac{1}{2\sigma} \prod_{i=1}^{\lfloor \rho(e^{-a},0) m \rfloor} \phi\Big(e^{-a} - \varepsilon, 0, \frac{i}{m}\Big)\,. \end{aligned}$$

Taking logarithms, we recognize a Riemann sum, hence

$$\liminf_{\substack{\ell,m\to\infty\\ q\to 0,\,\ell q\to a}} \frac{1}{m}\ln E(\tau_0\,|\,Z_0=1)\;\geq\;\int_0^{\rho(e^{-a},0)}\ln\phi(e^{-a}-\varepsilon,0,s)\,ds\,.$$

We let ε go to 0 in the upper bound and in the lower bound to obtain the desired conclusion. ☐

9.4 Invariant probability measure

In this section, we estimate the invariant probability measure of the process $(Z_t^\theta)_{t\geq 0}$, or rather the numerator of the last formula of section 8.4. 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 corollary 9.2,

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

Let $f:[0,1]\to\mathbb{R}$ be a non–decreasing function such that $f(0)=0$. We have the formula

$$\sum_{1\leq i\leq m} f\Big(\frac{i}{m}\Big)\nu(i)\;=\;\frac{\delta_0\displaystyle\sum_{1\leq i\leq m} f\Big(\frac{i}{m}\Big)\frac{1}{\gamma_i}\pi(i-1)}{1+\delta_0 E(\tau_0\,|\,Z_0=1)}\,.$$

Moreover $\delta_0=M_H(\theta,0)$, thus the numerator of the last formula of section 8.4 can be rewritten as

$$\Big(\frac{1}{\delta_0}+E\big(\tau_0\,\big|\,Z_0=1\big)\Big)\sum_{i=1}^m f\Big(\frac{i}{m}\Big)\,\nu(i)\quad=\quad\sum_{1\leq i\leq m} f\Big(\frac{i}{m}\Big)\frac{1}{\gamma_i}\pi(i\;-\;1)\,.$$

Our goal is to estimate the asymptotic behavior of the right–hand quantity.

Proposition 9.3 Let $f:[0,1]\to\mathbb{R}$ be a continuous non–decreasing function such that $f(0)=0$. Let $a\in]0,+\infty[$. We have

$$\lim_{\substack{\ell,m\to\infty\\ q\to 0,\,\ell q\to a}} \frac{\displaystyle\sum_{1\leq i\leq m} f\Big(\frac{i}{m}\Big)\frac{1}{\gamma_i}\pi(i-1)}{E(\tau_0\,|\,Z_0=1)}\;=\;f\big(\rho^*(a)\big)\,.$$

Proof. Throughout the proof, we write simply ρ^* instead of $\rho^*(a)$. Let $\eta > 0$. For ℓ, m large enough and q small enough, we have

$$\big|\rho^* - \rho\big(M_H(0,0), M_H(\theta,0)\big)\big| \,<\, \eta\,,$$

whence

$$\begin{aligned}
&\sum_{1\le i\le m} f\Big(\frac{i}{m}\Big)\frac{1}{\gamma_i}\pi(i-1)\\
&\qquad = \sum_{\substack{1\le i\le m\\ |i/m-\rho^*|\le\eta}} f\Big(\frac{i}{m}\Big)\frac{1}{\gamma_i}\pi(i-1) + \sum_{\substack{1\le i\le m\\ |i/m-\rho^*|>\eta}} f\Big(\frac{i}{m}\Big)\frac{1}{\gamma_i}\pi(i-1)\\
&\qquad \le \sum_{\substack{1\le i\le m\\ |i/m-\rho^*|\le\eta}} f(\rho^*+\eta)\frac{1}{\gamma_i}\pi(i-1) + \sum_{\substack{1\le i\le m\\ |i/m-\rho^*|>\eta}} f(1)\frac{1}{\gamma_i}\pi(i-1)\\
&\qquad\qquad \le\, f(\rho^*+\eta)E(\tau_0\,|\,Z_0=1) +\\
&\qquad\qquad\qquad \frac{m^3 f(1)}{1-M_H(0,0)}\Big(\pi\big(\lfloor(\rho^*-\eta)m\rfloor\big)+\pi\big(\lfloor(\rho^*+\eta)m\rfloor\big)\Big)\,.
\end{aligned}$$

To obtain the last inequality, we have used the monotonicity properties of $\pi(i)$ and the bounds on γ_i given at the beginning of the proof of corollary 9.2. The properties of ϕ and the definition of ρ^* imply that

$$\begin{aligned}
&\int_0^{\rho^*} \ln\phi(e^{-a},0,\rho)\,d\rho \,>\\
&\qquad\qquad \max\bigg(\int_0^{\rho^*-\eta} \ln\phi(e^{-a},0,\rho)\,d\rho\,, \int_0^{\rho^*+\eta} \ln\phi(e^{-a},0,\rho)\,d\rho\bigg)\,,
\end{aligned}$$

so that, using proposition 9.1 and corollary 9.2, for m large enough,

$$\frac{m^3 f(1)}{1-M_H(0,0)}\Big(\pi\big(\lfloor(\rho^*-\eta)m\rfloor\big)+\pi\big(\lfloor(\rho^*+\eta)m\rfloor\big)\Big) \,\le\, \eta E(\tau_0\,|\,Z_0=1)\,.$$

Adding together the previous inequalities, we arrive at

$$\sum_{1\le i\le m} f\Big(\frac{i}{m}\Big)\frac{1}{\gamma_i}\pi(i-1) \,\le\, \big(f(\rho^*+\eta)+\eta\big)E(\tau_0\,|\,Z_0=1)\,.$$

Passing to the limit, we obtain that

$$\limsup_{\substack{\ell,m\to\infty\\ q\to0,\,\ell q\to a}} \frac{\displaystyle\sum_{1\le i\le m} f\Big(\frac{i}{m}\Big)\frac{1}{\gamma_i}\pi(i-1)}{E(\tau_0\,|\,Z_0=1)} \,\le\, f(\rho^*+\eta)+\eta\,.$$

We seek next a complementary lower bound. If $\sigma e^{-a} \leq 1$, then $\rho^* = 0$, and obviously

$$\sum_{1\leq i\leq m} f\Big(\frac{i}{m}\Big)\frac{1}{\gamma_i}\pi(i-1) \,\geq\, f(0) \,=\, 0\,.$$

If $\sigma e^{-a} > 1$, then $\rho^* > 0$ and

$$\int_0^{\rho^*} \ln\phi(e^{-a},0,\rho)\,d\rho \,>\, \int_0^{\rho^*-\eta} \ln\phi(e^{-a},0,\rho)\,d\rho\,.$$

By corollary 9.2, for ℓ, m large enough and q small enough,

$$\sum_{\substack{1\leq i\leq m\\ i/m-\rho^*<-\eta}} \frac{1}{\gamma_i}\pi(i-1) \,\leq\, \frac{m^3}{1-M_H(0,0)}\pi\big(\lfloor(\rho^*-\eta)m\rfloor\big) \,\leq\, \eta E(\tau_0\,|\,Z_0=1)\,.$$

Combining these inequalities, we obtain

$$\begin{aligned}\sum_{1\leq i\leq m} f\Big(\frac{i}{m}\Big)\frac{1}{\gamma_i}\pi(i-1) \,\geq\, &\sum_{\substack{1\leq i\leq m\\ i/m-\rho^*\geq-\eta}} f(\rho^*-\eta)\frac{1}{\gamma_i}\pi(i-1)\\ = f(\rho^*-\eta)\Bigg(&\sum_{1\leq i\leq m}\frac{1}{\gamma_i}\pi(i-1) - \sum_{\substack{1\leq i\leq m\\ i/m-\rho^*<-\eta}}\frac{1}{\gamma_i}\pi(i-1)\Bigg)\\ &\geq\, f(\rho^*-\eta)E(\tau_0\,|\,Z_0=1)(1-\eta)\,.\end{aligned}$$

Passing to the limit, we obtain that

$$\liminf_{\substack{\ell,m\to\infty\\ q\to0,\,\ell q\to a}} \frac{\displaystyle\sum_{1\leq i\leq m} f\Big(\frac{i}{m}\Big)\frac{1}{\gamma_i}\pi(i-1)}{E(\tau_0\,|\,Z_0=1)} \,\geq\, f(\rho^*-\eta)(1-\eta)\,.$$

We finally let η go to 0 in the lower and the upper bounds to obtain the claim of the proposition. $\square$

10 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 Moran 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:

$$\tau_* \,=\, \inf\big\{\, n \geq 0 : X_n \notin \mathcal{N} \,\big\} .$$

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

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

10.1 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.

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:

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

Let μ_0 be an exchangeable probability distribution on $\big(\mathcal{A}^\ell\big)^m$, i.e., which satisfies

$$\forall \rho \in \mathfrak{S}_m \quad \forall x \in \big(\mathcal{A}^\ell\big)^m \quad \mu_0(\rho \cdot x) \,=\, \mu_0(x) .$$

Let $(X_n)_{n\geq 0}$ be the normalized neutral Moran 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. Let $(\varepsilon_n)_{n\geq 1}$ be a sequence of i.i.d. Bernoulli random variables with parameter $1/m$:

$$\forall n \geq 1 \qquad P(\varepsilon_n = 0) \,=\, 1 - \frac{1}{m} , \qquad P(\varepsilon_n = 1) \,=\, \frac{1}{m}$$

and let us set

$$\forall n \geq 1 \qquad N(n) \,=\, \varepsilon_1 + \cdots + \varepsilon_n \,.$$

We suppose also that the sequence $(\varepsilon_n)_{n\geq 1}$ and the Markov chain $(W_n)_{n\geq 0}$ are independent.

Proposition 10.1 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 $W_{N(n)}$.

Proof. We start by computing the transition matrix of the process $(W_{N(n)})_{n\geq 0}$. For $u \in \mathcal{A}^\ell$ and $n \geq 0$,

$$\begin{aligned} P(W_{N(n+1)} = u) \,&=\, P(W_{N(n+1)} = u,\, \varepsilon_{n+1} = 0) \\ &\qquad\qquad + P(W_{N(n+1)} = u,\, \varepsilon_{n+1} = 1) \\ &= \Big(1 - \frac{1}{m}\Big) P(W_{N(n)} = u) + \frac{1}{m} P(W_{N(n)+1} = u) \,. \end{aligned}$$

Moreover

$$\begin{aligned} P(W_{N(n)+1} = u) \,&=\, \sum_{v\in\mathcal{A}^\ell} P(W_{N(n)+1} = u,\, W_{N(n)} = v) \\ &= \sum_{v\in\mathcal{A}^\ell} P(W_{N(n)+1} = u \,|\, W_{N(n)} = v) P(W_{N(n)} = v) \\ &= \sum_{v\in\mathcal{A}^\ell} P(W_{N(n)} = v) M(v,u) \,. \end{aligned}$$

Therefore the transition matrix of the process $(W_{N(n)})_{n\geq 0}$ is

$$\Big(1 - \frac{1}{m}\Big) I + \frac{1}{m} M \,,$$

where I is the identity matrix. We do now 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

$$P(X_{n+1}(i) = u \,|\, X_n = x) \,=\, \Big(1 - \frac{1}{m}\Big) 1_{x(i)=u} + \frac{1}{m^2} \sum_{1\leq j\leq m} M(x(j), u) \,.$$

Thus

$$\begin{aligned} P(X_{n+1}(i)=u) \,&=\, \sum_{x\in(\mathcal{A}^\ell)^m} \Big(1-\frac{1}{m}\Big) 1_{x(i)=u} P(X_n=x) \\ &\qquad\qquad + \sum_{x\in(\mathcal{A}^\ell)^m} \frac{1}{m^2} \sum_{1\leq j\leq m} M(x(j),u) P(X_n=x) \\ &= \Big(1-\frac{1}{m}\Big) P(X_n(i)=u) + \sum_{v\in\mathcal{A}^\ell} \frac{1}{m^2} \sum_{1\leq j\leq m} M(v,u) P(X_n(j)=v)\,. \end{aligned}$$

By the induction hypothesis,

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

whence

$$\begin{aligned} P(X_{n+1}(i)=u) \,&=\, \\ \Big(1-\frac{1}{m}\Big) &P(W_{N(n)}=u) + \frac{1}{m} \sum_{v\in\mathcal{A}^\ell} P(W_{N(n)}=v) M(v,u) \\ &\qquad\qquad = \, P(W_{N(n+1)}=u)\,. \end{aligned}$$

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 $n\geq 1$, 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. Let us make its value explicit. Let $i\in\{\,1,\dots,m\,\}$. We set first $\mathcal{I}(i,n,n)=i$. If $X_{n-1}=x$ and $X_n=y$ with $y=x(j\leftarrow u)$, where the chromosome u has been obtained by replicating the k–th chromosome of x, then

$$\mathcal{I}(i,n,n-1) \,=\, \begin{cases} i & \text{if } i\neq j \\ k & \text{if } i=j \end{cases}$$

For $s\leq n$, 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,

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

Proposition 10.2 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 $(W_{N(0)}, \dots, W_{N(n)})$.

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

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

Since we deal with the neutral process, we have

$$\begin{aligned}
&P\left(\begin{matrix} X_{n+1}(i) = u_{n+1} \\ \mathcal{I}(i, n+1, n) = j \end{matrix}\ \middle|\ \begin{matrix} \text{ancestor}(j, n, s) = u_s \\ 0 \leq s \leq n,\ X_n = x \end{matrix}\right)\\
&\qquad\qquad = P\left(\begin{matrix} X_{n+1}(i) = u_{n+1} \\ \mathcal{I}(i, n+1, n) = j \end{matrix}\ \middle|\ X_n = x\right)\\
&\qquad\qquad = \begin{cases} \Big(1 - \dfrac{1}{m}\Big) 1_{x(i) = u_{n+1}} + \dfrac{1}{m^2} M(x(i), u_{n+1}) & \text{if } j = i \\ \dfrac{1}{m^2} M(x(j), u_{n+1}) & \text{if } j \neq i \end{cases}
\end{aligned}$$

Reporting in the previous equality, we get

$$\begin{aligned}
&P\big(\text{ancestor}(i, n+1, s) = u_s,\ 0 \leq s \leq n+1\big) =\\
&\sum_{x \in (\mathcal{A}^\ell)^m} \Big(1 - \frac{1}{m}\Big) 1_{x(i) = u_{n+1}}\, P\begin{pmatrix} \text{ancestor}(i, n, s) = u_s \\ 0 \leq s \leq n,\ X_n = x \end{pmatrix}\\
&\qquad + \sum_{x \in (\mathcal{A}^\ell)^m} \sum_{1 \leq j \leq m} \frac{1}{m^2} M(x(j), u_{n+1})\, P\begin{pmatrix} \text{ancestor}(j, n, s) = u_s \\ 0 \leq s \leq n,\ X_n = x \end{pmatrix}\\
&= \Big(1 - \frac{1}{m}\Big) 1_{u_n = u_{n+1}} P\big(\text{ancestor}(i, n, s) = u_s,\ 0 \leq s \leq n\big)\\
&\qquad + \sum_{1 \leq j \leq m} \frac{1}{m^2} M(u_n, u_{n+1})\, P\big(\text{ancestor}(j, n, s) = u_s,\ 0 \leq s \leq n\big).
\end{aligned}$$

By the induction hypothesis, we have, for any $j \in \{1, \cdots, m\}$,

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

Therefore

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

and the induction step is completed. □

10.2 Mutation dynamics

Throughout the section, 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. By lemma 6.1, for $b,c\in\{\,0,\dots,\ell\,\}$, the coefficient $M_H(b,c)$ of the matrix M_H is equal to

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

Such a Markov chain can be realized on our common probability space. Its construction requires only the family of random variables

$$(U_{n,l},\,n\geq 1,\,1\leq l\leq \ell)$$

with uniform law on the interval $[0,1]$. Let $b\in\{\,0,\dots,\ell\,\}$ be the starting point of the chain. We set $Y_0=b$ and we define inductively for $n\geq 1$

$$\begin{aligned}
Y_n \;&=\; Y_{n-1}-\sum_{k=1}^{Y_{n-1}}1_{U_{n,k}<p/\kappa}+\sum_{k=Y_{n-1}+1}^{\ell}1_{U_{n,k}>1-p(1-1/\kappa)}\\
&=\;\mathcal{M}_H(Y_{n-1},U_{n,1},\dots,U_{n,\ell})\,.
\end{aligned}$$

By lemma 7.1, the map $\mathcal{M}_H$ is non–decreasing with respect to its first argument. Thus the above construction provides a monotone coupling of the processes starting with different initial conditions and we conclude that the Markov chain $(Y_n)_{n\geq 0}$ is monotone.

Proposition 10.3 The matrix M_H is reversible with respect to the binomial law $\mathcal{B}(\ell,1-1/\kappa)$ with parameters ℓ and $1-1/\kappa$. This binomial law is the invariant probability measure of the Markov chain $(Y_n)_{n\geq 0}$.

Notation. We denote simply by $\mathcal{B}$ the binomial law $\mathcal{B}(\ell,1-1/\kappa)$. Thus

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

Proof. We check that the matrix M_H is reversible with respect to $\mathcal{B}$. Let $b,c\in\{\,0,\dots,\ell\,\}$. We use the identity

$$\binom{\ell}{b}\binom{\ell-b}{k}\binom{b}{h} \;=\; \frac{\ell!}{k!\,h!\,(\ell-b-k)!\,(b-h)!}$$

to write

$$\begin{aligned}\mathcal{B}(b)\,M_H(b,c) \;=\; &\binom{\ell}{b}\Big(1-\frac{1}{\kappa}\Big)^{b}\Big(\frac{1}{\kappa}\Big)^{\ell-b}\times\\ \sum_{\substack{0\leq k\leq \ell-b\\ 0\leq h\leq b\\ k-h=c-b}}&\binom{\ell-b}{k}\binom{b}{h}\Big(p\Big(1-\frac{1}{\kappa}\Big)\Big)^{k}\Big(1-p\Big(1-\frac{1}{\kappa}\Big)\Big)^{\ell-b-k}\Big(\frac{p}{\kappa}\Big)^{h}\Big(1-\frac{p}{\kappa}\Big)^{b-h}\\ =\sum_{\substack{0\leq k\leq \ell-b\\ 0\leq h\leq b\\ k-h=c-b}}&\frac{\ell!\,p^{k+h}\Big(1-\dfrac{1}{\kappa}\Big)^{b+k}\Big(\dfrac{1}{\kappa}\Big)^{\ell-b+h}}{k!\,h!\,(\ell-b-k)!\,(b-h)!}\Big(1-p\Big(1-\frac{1}{\kappa}\Big)\Big)^{\ell-b-k}\Big(1-\frac{p}{\kappa}\Big)^{b-h}.\end{aligned}$$

We eliminate the variable $h=k+b-c$ in this formula:

$$\begin{aligned}\mathcal{B}(b)\,M_H(b,c) \;=\; &\sum_{\substack{0\leq k\leq \ell-b\\ c-b\leq k\leq c}}\frac{\ell!}{k!\,(k+b-c)!\,(\ell-b-k)!\,(c-k)!}\times\\ &p^{2k+b-c}\Big(1-\frac{1}{\kappa}\Big)^{b+k}\Big(\frac{1}{\kappa}\Big)^{\ell-c+k}\Big(1-p\Big(1-\frac{1}{\kappa}\Big)\Big)^{\ell-b-k}\Big(1-\frac{p}{\kappa}\Big)^{c-k}.\end{aligned}$$

If we set now $h=k+b-c$ and we eliminate k, we get

$$\begin{aligned}\mathcal{B}(b)\,M_H(b,c) \;=\; &\sum_{\substack{b-c\leq h\leq \ell-c\\ 0\leq h\leq b}}\frac{\ell!}{(h+c-b)!\,h!\,(\ell-c-h)!\,(b-h)!}\times\\ &p^{2h+c-b}\Big(1-\frac{1}{\kappa}\Big)^{c+h}\Big(\frac{1}{\kappa}\Big)^{\ell-b+h}\Big(1-p\Big(1-\frac{1}{\kappa}\Big)\Big)^{\ell-c-h}\Big(1-\frac{p}{\kappa}\Big)^{b-h}\\ &\qquad=\;\mathcal{B}(c)\,M_H(c,b)\,.\end{aligned}$$

We obtain the same expression as before, but with b and c exchanged. Thus the matrix M_H is reversible with respect to $\mathcal{B}$ and $\mathcal{B}$ is the invariant probability measure of $\mathcal{M}_H$. □

When ℓ grows, the law $\mathcal{B}$ concentrates exponentially fast in a neighborhood of its mean

$$\ell_\kappa \;=\; \sum_{l=0}^{\ell} l\,\mathcal{B}(l)\;=\;\ell(1-1/\kappa)\,.$$

We estimate next the probability of the points to the left of ℓ_κ.

Lemma 10.4 For $b \leq \ell/2$, we have

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

Proof. Let $b \leq \ell/2$. Then

$$\mathcal{B}(b) \;=\; \binom{\ell}{b}\Big(1-\frac{1}{\kappa}\Big)^b\Big(\frac{1}{\kappa}\Big)^{\ell-b} \;\geq\; \binom{\ell}{b}\frac{1}{\kappa^\ell} \;\geq\; \frac{(\ell-b)^b}{b^b}\frac{1}{\kappa^\ell} \;\geq\; \left(\frac{\ell}{2b}\right)^b\frac{1}{\kappa^\ell}\,.$$

The upper bound on $\mathcal{B}(b)$ is straightforward. □

The estimates of lemma 10.4 can be considerably enhanced. In the next lemma, we present the fundamental large deviation estimates for the binomial distribution. This is the simplest case of the famous Cramér theorem.

Lemma 10.5 For $\rho \in [0,1]$, we have

$$\lim_{\ell\to\infty}\frac{1}{\ell}\ln\mathcal{B}(\lfloor\rho\ell\rfloor) \;=\; -(1-\rho)\ln\big(\kappa(1-\rho)\big) - \rho\ln\frac{\kappa\rho}{\kappa-1}\,.$$

Proof. We write

$$\begin{aligned}\ln\mathcal{B}(\lfloor\rho\ell\rfloor) \;&=\; \ln\frac{\ell\cdots(\ell-\lfloor\rho\ell\rfloor+1)}{1\cdots\lfloor\rho\ell\rfloor} + (\ell-\lfloor\rho\ell\rfloor)\ln\frac{1}{\kappa} + \lfloor\rho\ell\rfloor\ln\big(1-\frac{1}{\kappa}\big)\\ &=\; \sum_{k=0}^{\lfloor\rho\ell\rfloor-1}\ln\Big(1-\frac{k}{\ell}\Big) - \sum_{k=1}^{\lfloor\rho\ell\rfloor}\ln\frac{k}{\ell} + (\ell-\lfloor\rho\ell\rfloor)\ln\frac{1}{\kappa} + \lfloor\rho\ell\rfloor\ln\big(1-\frac{1}{\kappa}\big)\,.\end{aligned}$$

We recognize Riemann sums for the functions $\ln(1-x)$ and $\ln x$, thus

$$\lim_{\ell\to\infty}\frac{1}{\ell}\ln\mathcal{B}(\lfloor\rho\ell\rfloor) \;=\; \int_0^\rho \ln\frac{1-x}{x}\,dx + (1-\rho)\ln\frac{1}{\kappa} + \rho\ln\big(1-\frac{1}{\kappa}\big)\,.$$

We conclude by performing the integration. □

The minimum of the rate function appearing in lemma 10.5 is ℓ_κ. The typical behavior of the Markov chain $(Y_n)_{n\geq 0}$ is the following. Starting from 1, it very quickly reaches a neighbor of its stable equilibrium ℓ_κ. Then it starts exploring the surrounding space by performing larger and larger excursions outside ℓ_κ. Starting from ℓ_κ, the time needed to hit the point $c \in \{\,0,\dots,\ell\,\}$ is of order $\mathcal{B}(c)^{-1}$. Once the process is close to ℓ_κ, it is unlikely to visit 0 before time $\mathcal{B}(0)^{-1} = \kappa^\ell$. This is why the expected value of the hitting time of 0 starting from 1 is of order κ^ℓ. In the next sections, we derive quantitative bounds on the behavior of the chain $(Y_n)_{n\geq 0}$, starting from 1 or from ℓ. We need only crude bounds, hence we use elementary techniques, namely, we compare the process with a sum of i.i.d. random variables and we use the classical Chebyshev inequality, as well as the exponential Chebyshev inequality. The resulting proofs are somehow clumsy, and better estimates could certainly be derived with more sophisticated tools.

10.3 Falling to equilibrium from the left

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

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

Our first goal is to estimate, for b smaller than $\ell_\kappa = \ell(1 - 1/\kappa)$ and $n \geq 1$, the probability

$$P\big(\tau(b) > n \,|\, Y_0 = 0\big) \,.$$

Rough bound on the drift. Suppose that $\tau(b) > n$. Then $Y_{n-1} < b$ and

$$Y_n \,\geq\, Y_{n-1} - \sum_{k=1}^{b} 1_{U_{n,k} < p/\kappa} + \sum_{k=b+1}^{\ell} 1_{U_{n,k} > 1 - p(1-1/\kappa)} \,.$$

Iterating this inequality, we see that, on the event $\{\, \tau(b) > n \,\}$, we have $Y_n \geq V_n$ where

$$V_n \,=\, \sum_{t=1}^{n} \Big(- \sum_{k=1}^{b} 1_{U_{t,k} < p/\kappa} + \sum_{k=b+1}^{\ell} 1_{U_{t,k} > 1 - p(1-1/\kappa)} \Big) \,.$$

Therefore

$$P\big(\tau(b) > n \,|\, Y_0 = 0\big) \,\leq\, P(V_n < b) \,.$$

We shall bound $P(V_n < b)$ with the help of Chebyshev's inequality. Let us compute the mean and the variance of V_n. Since V_n is a sum of independent Bernoulli random variables, we have

$$\begin{aligned}
E(V_n) \,&=\, n\big(- b\,\frac{p}{\kappa} + (\ell - b)\,p\,(1 - \frac{1}{\kappa})\big) \,=\, np\,\big(\ell_\kappa - b\big) \,, \\
\mathrm{Var}(V_n) \,&=\, n\Big(b\,\frac{p}{\kappa}\big(1 - \frac{p}{\kappa}\big) + (\ell - b)\,p\,\big(1 - \frac{1}{\kappa}\big)\big(1 - p\,\big(1 - \frac{1}{\kappa}\big)\big)\Big) \\
&\leq\, n\big(bp + (\ell - b)p\big) \,=\, n\ell p \,.
\end{aligned}$$

We suppose that n is large enough so that $2b \,<\, E(V_n)$, that is,

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

By Chebyshev's inequality, we have then

$$\begin{aligned}P(V_n < b) \;&=\; P\big(V_n - E(V_n) < b - E(V_n)\big)\\&\leq\; P\Big(\big|V_n - E(V_n)\big| > \frac{1}{2}E(V_n)\Big)\\&\leq\; \frac{4\,\mathrm{Var}(V_n)}{\big(E(V_n)\big)^2} \;\leq\; \frac{4n\ell p}{\Big(np\left(\ell_\kappa - b\right)\Big)^2}\,.\end{aligned}$$

We have thus proved the following estimate.

Lemma 10.6 For n such that

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

we have

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

We derive next a crude lower bound on the descent from 0 to ℓ_κ. This lower bound will be used to derive the upper bound on the discovery time.

Proposition 10.7 We suppose that $\ell \to +\infty$, $q \to 0$ in such a way that

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

For ℓ large enough and q small enough, we have

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

Proof. We decompose

$$\begin{aligned}P\big(\tau(\ell_\kappa) \leq \ell^2 \,|\, Y_0 = 0\big) \;&\geq\; P\big(\tau(\ell_\kappa - \ln \ell) < \ell^2,\, \tau(\ell_\kappa) \leq \ell^2 \,|\, Y_0 = 0\big)\\&= \sum_{t<\ell^2} \sum_{b \geq \ell_\kappa - \ln \ell} P\big(\tau(\ell_\kappa - \ln \ell) = t,\, Y_t = b,\, \tau(\ell_\kappa) \leq \ell^2 \,|\, Y_0 = 0\big)\\&= \sum_{t<\ell^2} \sum_{b \geq \ell_\kappa - \ln \ell} P\big(\tau(\ell_\kappa) \leq \ell^2 \,|\, \tau(\ell_\kappa - \ln \ell) = t,\, Y_t = b,\, Y_0 = 0\big)\\&\qquad\qquad \times P\big(\tau(\ell_\kappa - \ln \ell) = t,\, Y_t = b \,|\, Y_0 = 0\big)\,.\end{aligned}$$

By the Markov property and the monotonicity of the process $(Y_n)_{n\geq 0}$, we have, for $t < \ell^2$ and $b \geq \ell_\kappa - \ln \ell$,

$$\begin{aligned}&P\big(\tau(\ell_\kappa) \leq \ell^2 \,|\, \tau(\ell_\kappa - \ln \ell) = t,\, Y_t = b,\, Y_0 = 0\big)\\&\quad= P\big(\tau(\ell_\kappa) \leq \ell^2 - t \,|\, Y_0 = b\big) \;\geq\; P\big(\tau(\ell_\kappa) \leq \ell^2 - t \,|\, Y_0 = \ell_\kappa - \ln \ell\big)\\&\qquad\geq\; P\big(Y_1 = \ell_\kappa \,|\, Y_0 = \ell_\kappa - \ln \ell\big) \;=\; M_H(\ell_\kappa - \ln \ell, \ell_\kappa)\,.\end{aligned}$$

Reporting this inequality in the previous sum, we get

$$P\big(\tau(\ell_\kappa) \le \ell^2 \,|\, Y_0 = 0\big) \;\ge\; P\big(\tau(\ell_\kappa - \ln \ell) < \ell^2 \,|\, Y_0 = 0\big)\, M_H(\ell_\kappa - \ln \ell, \ell_\kappa)\,.$$

By lemma 10.6 applied with $b = \ell_\kappa - \ln \ell$ and $n = \ell^2 - 1$, we have for ℓ large enough and q small enough

$$P\big(\tau(\ell_\kappa - \ln \ell) \ge \ell^2 \,|\, Y_0 = 0\big) \;\le\; \frac{5}{a(\ln \ell)^2}\,.$$

Moreover, for ℓ large enough and q small enough,

$$M_H(\ell_\kappa - \ln \ell, \ell_\kappa) \;\ge\; \Big(\frac{p}{\kappa}\Big)^{\ln \ell} (1-q)^\ell \;\ge\; \Big(\frac{p}{\kappa}\Big)^{\ln \ell} e^{-2a}\,.$$

Putting the previous inequalities together, we obtain the desired lower bound. ☐

We will need more information in order to derive the lower bound on the discovery time. We wish to control the time and speed at which the Markov chain $(Y_n)_{n \ge 0}$, starting from 1, reaches a neighborhood of its equilibrium ℓ_κ without visiting 0. This will require a stronger inequality than the one stated in lemma 10.6, this is the purpose of the next lemma.

Lemma 10.8 For $n \ge 1$, $b \in \{\, 0, \dots, \ell \,\}$ and $\lambda > 0$, we have

$$P\big(\tau(b) > n \,|\, Y_0 = 0\big) \;\le\; \exp\Big(\lambda b + nb\frac{p}{\kappa}(e^\lambda - 1) + n(\ell - b)p\frac{\kappa - 1}{\kappa}(e^{-\lambda} - 1)\Big)\,.$$

Proof. We obtain this inequality as a consequence of Chebyshev's exponential inequality. Indeed, we have

$$\begin{aligned} P\big(\tau(b) > n \,|\, Y_0 = 0\big) \;&\le\; P(V_n < b) \\ &= P(-\lambda V_n > -\lambda b) \;=\; P\big(\exp(-\lambda V_n) > \exp(-\lambda b)\big) \\ &\le \exp(\lambda b) E\big(\exp(-\lambda V_n)\big) \;=\; \exp(\lambda b)\Big(E\big(\exp(-\lambda V_1)\big)\Big)^n\,. \end{aligned}$$

Yet

$$\begin{aligned} E\big(\exp(-\lambda V_1)\big) \;&=\; E\Big(\exp\Big(\lambda \sum_{k=1}^{b} 1_{U_{1,k} < p/\kappa} - \lambda \sum_{k=b+1}^{\ell} 1_{U_{1,k} > 1 - p(1 - 1/\kappa)}\Big)\Big) \\ &= \Big(1 + \frac{p}{\kappa}(e^\lambda - 1)\Big)^b \Big(1 + p\frac{\kappa - 1}{\kappa}(e^{-\lambda} - 1)\Big)^{\ell - b}\,. \end{aligned}$$

Thus

$$P\big(\tau(b) > n \,|\, Y_0 = 0\big) \;\le\;$$
$$\exp\Big(\lambda b + nb \ln\big(1 + \frac{p}{\kappa}(e^\lambda - 1)\big) + n(\ell - b)\ln\big(1 + p\frac{\kappa - 1}{\kappa}(e^{-\lambda} - 1)\big)\Big)\,.$$

Using the inequality $\ln(1+t) \leq t$, we obtain the desired result. $\square$

We derive next two kinds of estimates: first for the start of the fall, and second for the completion of the fall.

Start of the fall. We show here that, after a time $\sqrt{\ell}$, the Markov chain $(Y_n)_{n\geq 0}$ is with high probability in the interval $[\ln\ell, \ell]$.

Proposition 10.9 We suppose that $\ell \to +\infty\,, q \to 0\,, \ell q \to a \in]0, +\infty[$. For ℓ large enough and q small enough, we have

$$\forall t \geq \sqrt{\ell} \qquad P\big(Y_t \geq \ln\ell \,|\, Y_0 = 0\big) \,\geq\, 1 - \exp\Big(-\frac{1}{2}(\ln\ell)^2\Big)\,.$$

Proof. We write, for $t \geq \sqrt{\ell}$,

$$\begin{aligned}
P\big(Y_t \geq \ln\ell \,|\, Y_0 = 0\big) \,&\geq\, P\big(Y_t \geq \ln\ell,\, \tau(2\ln\ell) \leq \sqrt{\ell} \,|\, Y_0 = 0\big)\\
&= \sum_{n=1}^{\sqrt{\ell}} \sum_{k=2\ln\ell}^{\ell} P\big(Y_t \geq \ln\ell\,, \tau(2\ln\ell) = n,\, Y_n = k \,|\, Y_0 = 0\big)\\
&= \sum_{n=1}^{\sqrt{\ell}} \sum_{k=2\ln\ell}^{\ell} P\big(Y_t \geq \ln\ell \,|\, \tau(2\ln\ell) = n,\, Y_n = k,\, Y_0 = 0\big)\times\\
&\qquad\qquad\qquad P\big(\tau(2\ln\ell) = n,\, Y_n = k \,|\, Y_0 = 0\big)\,.
\end{aligned}$$

Now, for $n \leq \sqrt{\ell}$ and $k \geq 2\ln\ell$, by the Markov property, and thanks to the monotonicity of the process $(Y_t)_{t\geq 0}$,

$$\begin{aligned}
&P\big(Y_t \geq \ln\ell \,|\, \tau(2\ln\ell) = n,\, Y_n = k,\, Y_0 = 0\big)\\
&\qquad = P\big(Y_t \geq \ln\ell \,|\, Y_n = k\big) \,\geq\, P\big(Y_t \geq \ln\ell \,|\, Y_n = \lfloor 2\ln\ell \rfloor\big)\\
&\qquad\qquad = P\big(Y_{t-n} \geq \ln\ell \,|\, Y_0 = \lfloor 2\ln\ell \rfloor\big)\,.
\end{aligned}$$

For $b < \ln\ell$, we have by lemmas A.1 and 10.4,

$$P\big(Y_{t-n} = b \,|\, Y_0 = \lfloor 2\ln\ell \rfloor\big) \,\leq\, \frac{\mathcal{B}(b)}{\mathcal{B}(2\ln\ell)} \,\leq\, \frac{\mathcal{B}(\ln\ell)}{\mathcal{B}(2\ln\ell)} \,\leq\, \left(\frac{(4\kappa\ln\ell)^2}{\ell}\right)^{\ln\ell}\,,$$

whence

$$P\big(Y_{t-n} \geq \ln\ell \,|\, Y_0 = \lfloor 2\ln\ell \rfloor\big) \,\geq\, 1 - \ln\ell\left(\frac{(4\kappa\ln\ell)^2}{\ell}\right)^{\ln\ell}\,.$$

Reporting this inequality in the previous sum, we get

$$\begin{aligned}
&P\big(Y_t \geq \ln\ell \,|\, Y_0 = 0\big) \,\geq\\
&\qquad\qquad \left(1 - \ln\ell\left(\frac{(4\kappa\ln\ell)^2}{\ell}\right)^{\ln\ell}\right) P\big(\tau(2\ln\ell) \leq \sqrt{\ell} \,|\, Y_0 = 0\big)\,.
\end{aligned}$$

By lemma 10.8 applied with $\lambda = \ln 2$, $n = \sqrt{\ell}$, $b = 2\ln\ell$, for ℓ large enough and q small enough,

$$P\big(\tau(2\ln\ell) > \sqrt{\ell}\,|\,Y_0 = 0\big) \,\leq\, \exp -\frac{a\sqrt{\ell}}{4}\,,$$

whence

$$\begin{aligned}P\big(Y_t \geq \ln\ell\,|\,Y_0 = 0\big) \,\geq\, &\left(1 - \ln\ell\left(\frac{(4\kappa\ln\ell)^2}{\ell}\right)^{\ln\ell}\right)\Big(1 - \exp -\frac{a\sqrt{\ell}}{4}\Big)\\ &\geq\, 1 - \exp\Big(-\frac{1}{2}(\ln\ell)^2\Big)\,,\end{aligned}$$

where the last inequality holds for ℓ large enough. $\square$

Completion of the fall. We show here that, for $\varepsilon > 0$, after a time $4\ell/a\varepsilon$, the Markov chain $(Y_n)_{n\geq 0}$ is with high probability in the interval $[\ell_\kappa(1-\varepsilon), \ell]$.

Proposition 10.10 We suppose that $\ell \to +\infty\,, q \to 0\,, \ell q \to a \in]0,+\infty[$. Let $\varepsilon \in]0,1[$. There exists $c(\varepsilon) > 0$ such that, for ℓ large enough and q small enough, we have

$$\forall t \geq \frac{4\ell}{a\varepsilon} \qquad P\big(Y_t \geq \ell_\kappa(1-\varepsilon)\,|\,Y_0 = 0\big) \,\geq\, 1 - \exp(-c(\varepsilon)\ell)\,.$$

Proof. Let $t \geq 4\ell/(a\varepsilon)$. We write

$$\begin{aligned}&P\big(Y_t \geq \ell_\kappa(1-\varepsilon)\,|\,Y_0 = 0\big)\\ &\qquad\geq\, P\Big(Y_t \geq \ell_\kappa(1-\varepsilon),\, \tau(\ell_\kappa(1-\varepsilon/2)) \leq \frac{4\ell}{a\varepsilon}\,|\,Y_0 = 0\Big)\\ &= \sum_{n=1}^{4\ell/(a\varepsilon)}\ \sum_{k\geq\ell_\kappa(1-\varepsilon/2)} P\big(Y_t \geq \ell_\kappa(1-\varepsilon)\,|\,\tau(\ell_\kappa(1-\varepsilon/2)) = n,\, Y_n = k,\, Y_0 = 0\big)\\ &\qquad\qquad\times P\big(\tau(\ell_\kappa(1-\varepsilon/2)) = n,\, Y_n = k\,|\,Y_0 = 0\big)\,.\end{aligned}$$

Now, for $n \leq 4\ell/(a\varepsilon)$ and $k \geq \ell_\kappa(1-\varepsilon/2)$, by the Markov property,

$$\begin{aligned}&P\big(Y_t \geq \ell_\kappa(1-\varepsilon)\,|\,\tau(\ell_\kappa(1-\varepsilon/2)) = n,\, Y_n = k,\, Y_0 = 0\big)\\ &\quad=\, P\big(Y_t \geq \ell_\kappa(1-\varepsilon)\,|\,Y_n = k\big) \,=\, P\big(Y_{t-n} \geq \ell_\kappa(1-\varepsilon)\,|\,Y_0 = k\big)\\ &\qquad\leq\, P\big(Y_{t-n} \geq \ell_\kappa(1-\varepsilon)\,|\,Y_0 = \lfloor\ell_\kappa(1-\varepsilon/2)\rfloor\big)\,.\end{aligned}$$

We have used the monotonicity of the process $(Y_t)_{t\geq 0}$ with respect to the starting point to get the last inequality. For $b < \ell_\kappa(1-\varepsilon)$, we have by lemmas A.1 and 10.4,

$$P\big(Y_{t-n} = b\,|\,Y_0 = \lfloor \ell_\kappa(1-\varepsilon/2)\rfloor\big) \,\leq\, \frac{\mathcal{B}(b)}{\mathcal{B}(\ell_\kappa(1-\varepsilon/2))} \,\leq\, \frac{\mathcal{B}(\ell_\kappa(1-\varepsilon))}{\mathcal{B}(\ell_\kappa(1-\varepsilon/2))}\,,$$

whence

$$P\big(Y_{t-n} \geq \ell_\kappa(1-\varepsilon)\,|\,Y_0 = \lfloor \ell_\kappa(1-\varepsilon/2)\rfloor\big) \,\geq\, 1-\ell_\kappa(1-\varepsilon)\frac{\mathcal{B}(\ell_\kappa(1-\varepsilon))}{\mathcal{B}(\ell_\kappa(1-\varepsilon/2))}\,.$$

Thanks to the large deviation estimates of lemma 10.5, we have

$$\limsup_{\ell\to\infty}\frac{1}{\ell}\ln\left(\ell_\kappa(1-\varepsilon)\frac{\mathcal{B}(\ell_\kappa(1-\varepsilon))}{\mathcal{B}(\ell_\kappa(1-\varepsilon/2))}\right) \,<\, 0\,,$$

thus there exists $c(\varepsilon) > 0$ such that, for ℓ large enough

$$P\big(Y_{t-n} \geq \ell_\kappa(1-\varepsilon)\,|\,Y_0 = \lfloor \ell_\kappa(1-\varepsilon/2)\rfloor\big) \,\geq\, 1-\exp(-c(\varepsilon)\ell)\,.$$

Reporting this inequality in the previous sum, we get

$$P\big(Y_t \geq \ell_\kappa(1-\varepsilon)\,|\,Y_0 = 0\big) \,\geq\, \Big(1-\exp(-c(\varepsilon)\ell)\Big)P\Big(\tau(\ell_\kappa(1-\varepsilon/2)) \leq \frac{4\ell}{a\varepsilon}\,|\,Y_0 = 0\Big)\,.$$

We apply lemma 10.8 with $b = \ell_\kappa(1-\varepsilon/2)$, $\lambda > 0$ and $n = 4\ell/(a\varepsilon)$: for ℓ large enough and q small enough,

$$P\Big(\tau(\ell_\kappa(1-\frac{\varepsilon}{2})) > \frac{4\ell}{a\varepsilon}\,|\,Y_0 = 0\Big) \,\leq\, \exp\Big(\lambda\ell_\kappa(1-\frac{\varepsilon}{2}) + \frac{4\ell}{a\varepsilon}\Big(\ell_\kappa(1-\frac{\varepsilon}{2})\frac{p}{\kappa}(e^\lambda-1) + (\ell-\ell_\kappa(1-\frac{\varepsilon}{2}))p\frac{\kappa-1}{\kappa}(e^{-\lambda}-1)\Big)\Big)\,.$$

We send ℓ to ∞ and q to 0 in such a way that ℓq converges to $a > 0$. We obtain

$$\limsup_{\substack{\ell\to\infty\\ q\to 0,\,\ell q\to a}}\ \frac{1}{\ell}\ln P\Big(\tau(\ell_\kappa(1-\frac{\varepsilon}{2})) > \frac{4\ell}{a\varepsilon}\,|\,Y_0 = 0\Big) \,\leq\, \lambda\frac{\kappa-1}{\kappa}(1-\frac{\varepsilon}{2}) + \frac{4}{\varepsilon}\Big((1-\frac{\varepsilon}{2})\frac{1}{\kappa}(e^\lambda-1) + \big(\frac{1}{\kappa-1}+\frac{\varepsilon}{2}\big)\frac{\kappa-1}{\kappa}(e^{-\lambda}-1)\Big)\,.$$

Expanding the last term as λ goes to 0, we see that it is negative for λ small enough, therefore there exists $c'(\varepsilon) > 0$ such that for ℓ large enough and q small enough,

$$P\Big(\tau(\ell_\kappa(1-\frac{\varepsilon}{2})) > \frac{4\ell}{a\varepsilon}\,|\,Y_0 = 0\Big) \,\leq\, \exp(-c'(\varepsilon)\ell)\,.$$

Reporting in the previous inequality on Y_t, we obtain that

$$P\big(Y_t \geq \ell_\kappa(1-\varepsilon) \,|\, Y_0 = 0\big) \;\geq\; \Big(1-\exp(-c(\varepsilon)\ell)\Big)\Big(1-\exp(-c'(\varepsilon)\ell)\Big)$$

and this yields the desired result. ☐

10.4 Falling to equilibrium from the right

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

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

Proposition 10.11 We suppose that $\ell \to +\infty$, $q \to 0$ in such a way that

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

For ℓ large enough and q small enough, we have

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

Proof. Our first goal is to estimate, for b larger than $\ell_\kappa = \ell(1-1/\kappa)$ and $n \geq 1$, the probability

$$P\big(\theta(b) > n \,|\, Y_0 = \ell\big)\,.$$

Suppose that $\theta(b) > n$. Then $Y_{n-1} > b$ and

$$Y_n \;\leq\; Y_{n-1} - \sum_{k=1}^{b} 1_{U_{n,k} < p/\kappa} + \sum_{k=b+1}^{\ell} 1_{U_{n,k} > 1-p(1-1/\kappa)}\,.$$

Iterating this inequality, we see that, if $Y_0 = \ell$, on the event $\{\,\theta(b) > n\,\}$, we have $Y_n \leq \ell + V_n$, where V_n is the random variable defined in section 10.3. Therefore

$$P\big(\theta(b) > n \,|\, Y_0 = \ell\big) \;\leq\; P(\ell + V_n > b)\,.$$

We suppose that $b - \ell > np\,\big(\ell_\kappa - b\big)$. By Chebyshev's inequality, we have then

$$\begin{aligned} P(\ell + V_n \geq b) \;&=\; P\Big(V_n - E(V_n) \geq b - \ell - np\,\big(\ell_\kappa - b\big)\Big) \\ &\leq\; \frac{\mathrm{Var}(V_n)}{\big(b - \ell - np\,\big(\ell_\kappa - b\big)\big)^2}\,. \end{aligned}$$

We take $n = \ell^2$ and $b = \ln \ell + \ell_\kappa$. Then, for ℓ large enough,

$$b - \ell - np\,\big(\ell_\kappa - b\big) \,=\, \ln \ell + \ell_\kappa - \ell + \ell^2 p \ln \ell \,\sim\, \ell^2 p \ln \ell \,>\, 0\,,$$

whence, by the previous inequalities, for ℓ large enough and q small enough,

$$P\big(\theta(\ln \ell + \ell_\kappa) \geq \ell^2 \,|\, Y_0 = \ell\big) \,\leq\, \frac{1}{a(\ln \ell)^2}\,.$$

We decompose next

$$\begin{aligned}
P\big(\theta(\ell_\kappa) \leq \ell^2 \,|\, Y_0 = \ell\big) \,&\geq\, P\big(\theta(\ell_\kappa + \ln \ell) < \ell^2,\, \theta(\ell_\kappa) \leq \ell^2 \,|\, Y_0 = \ell\big)\\
&= \sum_{t<\ell^2} \sum_{b \leq \ell_\kappa + \ln \ell} P\big(\theta(\ell_\kappa + \ln \ell) = t,\, Y_t = b,\, \theta(\ell_\kappa) \leq \ell^2 \,|\, Y_0 = \ell\big)\\
&= \sum_{t<\ell^2} \sum_{b \leq \ell_\kappa + \ln \ell} P\big(\theta(\ell_\kappa) \leq \ell^2 \,|\, \theta(\ell_\kappa + \ln \ell) = t,\, Y_t = b,\, Y_0 = \ell\big)\\
&\qquad\qquad \times P\big(\theta(\ell_\kappa + \ln \ell) = t,\, Y_t = b \,|\, Y_0 = \ell\big)\,.
\end{aligned}$$

By the Markov property and the monotonicity of the process $(Y_n)_{n\geq 0}$, we have, for $t < \ell^2$ and $b \leq \ell_\kappa + \ln \ell$,

$$\begin{aligned}
&P\big(\theta(\ell_\kappa) \leq \ell^2 \,|\, \theta(\ell_\kappa + \ln \ell) = t,\, Y_t = b,\, Y_0 = \ell\big)\\
&\quad = P\big(\theta(\ell_\kappa) \leq \ell^2 - t \,|\, Y_0 = b\big) \,\geq\, P\big(\theta(\ell_\kappa) \leq \ell^2 - t \,|\, Y_0 = \ell_\kappa + \ln \ell\big)\\
&\qquad\qquad \geq\, P\big(Y_1 = \ell_\kappa \,|\, Y_0 = \ell_\kappa + \ln \ell\big) \,=\, M_H(\ell_\kappa + \ln \ell, \ell_\kappa)\,.
\end{aligned}$$

Reporting this inequality in the previous sum, we get

$$P\big(\theta(\ell_\kappa) \leq \ell^2 \,|\, Y_0 = \ell\big) \,\geq\, P\big(\theta(\ell_\kappa + \ln \ell) < \ell^2 \,|\, Y_0 = \ell\big)\, M_H(\ell_\kappa + \ln \ell, \ell_\kappa)\,.$$

For ℓ large enough and q small enough, we have

$$M_H(\ell_\kappa + \ln \ell, \ell_\kappa) \,\geq\, \Big(\frac{p}{\kappa}\Big)^{\ln \ell} (1-q)^\ell \,\geq\, \Big(\frac{p}{\kappa}\Big)^{\ln \ell} e^{-2a}\,.$$

Using the previous bound on $P\big(\theta(\ln \ell + \ell_\kappa) \geq \ell^2 \,|\, Y_0 = \ell\big)$, we obtain that

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

as requested. □

We derive next a large deviation upper bound for the time needed to go from ℓ to 0. This will yield an upper bound on the discovery time. We define

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

Proposition 10.12 For any $a \in]0,+\infty[$,

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

Proof. We prove that, starting from ℓ, the walker has probability of order $1/\kappa^\ell$ to visit 0 before time ℓ^2. To do this, we decompose the trajectory until time ℓ^2 into two parts: the descent to the equilibrium ℓ_κ, which is very likely to occur, and the ascent to 0, which is very unlikely to occur. We estimate the probability of the ascent with the help of a beautiful technique developed by Schonmann [35] in a different context, namely the study of the metastability of the Ising model. More precisely, we use the reversibility of the process to relate the probability of an ascending path to the probability of a descending path. It turns out that the most likely way to go from ℓ_κ to 0 is obtained as the time–reverse of a typical path going from 0 to ℓ_κ. Thanks to the monotonicity of the process, this estimate yields a lower bound on the hitting time of 0 which is uniform with respect to the starting point. We bound then easily $E(\tau_0\,|\,Y_0=\ell)$ by summing over intervals of length ℓ^2 and using the Markov property.

We should normally work with $\lfloor\ell_\kappa\rfloor$ instead of ℓ_κ. To alleviate the notation, we do as if ℓ_κ was an integer. We write

$$\begin{aligned}
P\big(\tau_0\leq 2\ell^2\,|\,Y_0=\ell\big)\;&\geq\;P\big(\theta(\ell_\kappa)\leq\ell^2,\,\tau_0\leq 2\ell^2\,|\,Y_0=\ell\big)\\
&=\sum_{t\leq\ell^2}\sum_{b\leq\ell_\kappa}P\big(\theta(\ell_\kappa)=t,\,Y_t=b,\,\tau_0\leq 2\ell^2\,|\,Y_0=\ell\big)\\
&=\sum_{t\leq\ell^2}\sum_{b\leq\ell_\kappa}P\big(\tau_0\leq 2\ell^2\,|\,\theta(\ell_\kappa)=t,\,Y_t=b,\,Y_0=\ell\big)\\
&\qquad\qquad\times P\big(\theta(\ell_\kappa)=t,\,Y_t=b\,|\,Y_0=\ell\big)\,.
\end{aligned}$$

By the Markov property and the monotonicity of the process $(Y_n)_{n\geq 0}$, we have, for $t\leq\ell^2$ and $b\leq\ell_\kappa$,

$$\begin{aligned}
&P\big(\tau_0\leq 2\ell^2\,|\,\theta(\ell_\kappa)=t,\,Y_t=b,\,Y_0=\ell\big)\\
&\quad=\;P\big(\exists n\in\{\,t,\dots,2\ell^2\,\}\quad Y_n=0\,|\,\theta(\ell_\kappa)=t,\,Y_t=b,\,Y_0=0\big)\\
&\qquad\qquad=\;P\big(\tau_0\leq 2\ell^2-t\,|\,Y_0=b\big)\;\geq\;P\big(\tau_0\leq\ell^2\,|\,Y_0=\ell_\kappa\big)\,.
\end{aligned}$$

Reporting this inequality in the previous sum, we get

$$P\big(\tau_0\leq 2\ell^2\,|\,Y_0=\ell\big)\;\geq\;P\big(\theta(\ell_\kappa)\leq\ell^2\,|\,Y_0=\ell\big)\,P\big(\tau_0\leq\ell^2\,|\,Y_0=\ell_\kappa\big)\,.$$

We estimate next the probability of the ascending part, i.e., the last probability in the above formula. We start with the estimate of proposition 10.7:

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

Yet

$$\begin{aligned} P\big(\tau(\ell_\kappa) \leq \ell^2 \,|\, Y_0 = 0\big) \;&=\; P\big(\exists t \leq \ell^2 \quad Y_t \geq \ell_\kappa \,|\, Y_0 = 0\big) \\ &\leq\; \sum_{b \geq \ell_\kappa} P\big(\exists t \leq \ell^2 \quad Y_t = b \,|\, Y_0 = 0\big)\,. \end{aligned}$$

From the last inequalities, we see that there exists $b \geq \ell_\kappa$ such that

$$P\big(\exists t \leq \ell^2 \quad Y_t = b \,|\, Y_0 = 0\big) \;\geq\; \frac{1}{\ell}\Big(1 - \frac{5}{a(\ln \ell)^2}\Big)\Big(\frac{p}{\kappa}\Big)^{\ln \ell} e^{-2a}\,.$$

Using the reversibility of M_H with respect to $\mathcal{B}$ (see proposition 10.3), we have

$$\begin{aligned} &\mathcal{B}(b)\, P\big(\tau_0 \leq \ell^2 \,|\, Y_0 = b\big) \\ &\qquad = \sum_{t \leq \ell^2} \sum_{y_1, \dots, y_{t-1} > 0} \mathcal{B}(b)\, M_H(b, y_1)\, \dots\, M_H(y_{t-1}, 0) \\ &\qquad = \sum_{t \leq \ell^2} \sum_{y_1, \dots, y_{t-1} > 0} \mathcal{B}(0) M_H(0, y_{t-1})\, \dots\, M_H(y_1, b) \\ &\qquad\qquad\qquad = \mathcal{B}(0)\, P\big(\exists t \leq \ell^2 \quad Y_t = b \,|\, Y_0 = 0\big)\,. \end{aligned}$$

Thus

$$\begin{aligned} P\big(\tau_0 \leq \ell^2 \,|\, Y_0 = b\big) \;&=\; \frac{\mathcal{B}(0)}{\mathcal{B}(b)} P\big(\exists t \leq \ell^2 \quad Y_t = b \,|\, Y_0 = 0\big) \\ &\geq\; \frac{\kappa^{-\ell}}{\ell}\Big(1 - \frac{5}{a(\ln \ell)^2}\Big)\Big(\frac{p}{\kappa}\Big)^{\ln \ell} e^{-2a}\,. \end{aligned}$$

By monotonicity of the process $(Y_t)_{t \geq 0}$, since $b \geq \ell_\kappa$, then

$$P\big(\tau_0 \leq \ell^2 \,|\, Y_0 = \ell_\kappa\big) \;\geq\; P\big(\tau_0 \leq \ell^2 \,|\, Y_0 = b\big)\,.$$

Using proposition 10.11 and the previous inequalities, we conclude that

$$P\big(\tau_0 \leq 2\ell^2 \,|\, Y_0 = \ell\big) \;\geq\; \frac{\kappa^{-\ell}}{\ell}\bigg(\Big(1 - \frac{5}{a(\ln \ell)^2}\Big)\Big(\frac{p}{\kappa}\Big)^{\ln \ell} e^{-2a}\bigg)^2\,.$$

Let $\varepsilon > 0$. For ℓ large enough and q small enough,

$$P\big(\tau_0 \leq 2\ell^2 \,|\, Y_0 = \ell\big) \;\geq\; \frac{1}{\kappa^{\ell(1+\varepsilon)}}\,.$$

Now, for $n \geq 0$,

$$\begin{aligned}
&P\big(\tau_0 > 2n\ell^2 \,|\, Y_0 = \ell\big)\\
&\qquad = \sum_{b \geq 1} P\big(\tau_0 > 2n\ell^2, Y_{2(n-1)\ell^2} = b,\, \tau_0 > 2(n-1)\ell^2 \,|\, Y_0 = \ell\big)\\
&\qquad = \sum_{b \geq 1} P\big(\tau_0 > 2n\ell^2 \,|\, Y_{2(n-1)\ell^2} = b,\, \tau_0 > 2(n-1)\ell^2,\, Y_0 = \ell\big)\\
&\qquad\qquad\qquad \times P\big(Y_{2(n-1)\ell^2} = b,\, \tau_0 > 2(n-1)\ell^2 \,|\, Y_0 = \ell\big)\,.
\end{aligned}$$

By the Markov property and the monotonicity of the process, we have

$$\begin{aligned}
&P\big(\tau_0 > 2n\ell^2 \,|\, Y_{2(n-1)\ell^2} = b,\, \tau_0 > 2(n-1)\ell^2,\, Y_0 = \ell\big)\\
&\qquad = P\big(\tau_0 > 2n\ell^2 \,|\, Y_{2(n-1)\ell^2} = b\big) \,=\, P\big(\tau_0 > 2\ell^2 \,|\, Y_0 = b\big)\\
&\qquad\qquad \leq\, P\big(\tau_0 > 2\ell^2 \,|\, Y_0 = \ell\big) \,\leq\, 1 - \frac{1}{\kappa^{\ell(1+\varepsilon)}}\,.
\end{aligned}$$

Reporting in the previous sum, we get

$$P\big(\tau_0 > 2n\ell^2 \,|\, Y_0 = \ell\big) \,\leq\, \Big(1 - \frac{1}{\kappa^{\ell(1+\varepsilon)}}\Big) P\big(\tau_0 > 2(n-1)\ell^2 \,|\, Y_0 = \ell\big)\,.$$

Iterating, we obtain

$$P\big(\tau_0 > 2n\ell^2 \,|\, Y_0 = \ell\big) \,\leq\, \Big(1 - \frac{1}{\kappa^{\ell(1+\varepsilon)}}\Big)^n\,.$$

Thus

$$\begin{aligned}
&E(\tau_0 \,|\, Y_0 = \ell) \,=\, \sum_{n \geq 1} P(\tau_0 \geq n \,|\, Y_0 = \ell)\\
&\quad \leq \sum_{n \geq 0} \sum_{t = 2n\ell^2 + 1}^{2(n+1)\ell^2} P(\tau_0 \geq t \,|\, Y_0 = \ell) \,\leq\, \sum_{n \geq 0} 2\ell^2 P(\tau_0 > 2n\ell^2 \,|\, Y_0 = \ell)\\
&\qquad\qquad \leq\, 2\ell^2 \sum_{n \geq 0} \Big(1 - \frac{1}{\kappa^{\ell(1+\varepsilon)}}\Big)^n \,=\, 2\ell^2 \kappa^{\ell(1+\varepsilon)}\,.
\end{aligned}$$

This bound is true for any $\varepsilon > 0$. Sending successively ℓ to ∞ and ε to 0, we obtain the desired upper bound. ☐

10.5 Discovery time

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

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

we have indeed

$$\begin{aligned} E\big(\tau^{*,\ell}\,|\,O_0^\ell=o^\ell_{\text{exit}}\big) \;&=\; E\big(\tau^*\,|\,O_0=(0,0,0,\dots,m)\big)\,,\\ E\big(\tau^{*,1}\,|\,O_0^1=o^1_{\text{exit}}\big) \;&=\; E\big(\tau^*\,|\,O_0=(0,m,0,\dots,0)\big)\,. \end{aligned}$$

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

$$\tau^{*} \;=\; \inf\,\big\{\, t\geq 0 : D_t\in\mathcal{W}^*\,\big\}\,.$$

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

$$\begin{aligned} E\big(\tau^*\,|\,O_0=(0,0,\dots,0,m)\big) \;&=\; E\big(\tau^*\,|\,D_0=(\ell)^m\big)\,,\\ E\big(\tau^*\,|\,O_0=(0,m,0,\dots,0)\big) \;&=\; E\big(\tau^*\,|\,D_0=(1)^m\big)\,. \end{aligned}$$

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 10.14.

Proposition 10.13 Let $a\in]0,+\infty[$ and $\alpha\in[0,+\infty[$. For any $d\in\mathcal{N}$,

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

Proof. Since we are in the neutral case $\sigma=1$, by corollary 7.6, 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 10.2, 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. We consider also a sequence $(\varepsilon_n)_{n\geq 1}$ of i.i.d. Bernoulli random variables with parameter $1/m$ and we set

$$\forall n\geq 1\qquad N(n)\,=\,\varepsilon_1+\cdots+\varepsilon_n\,.$$

We suppose also that the processes $(N(n))_{n\geq 1}$ and $(Y_n)_{n\geq 0}$ are independent. Let us look at the distance process at time n starting from $(\ell)^m$. From proposition 10.1, we know that the law of the i–th chromosome in D_n is the same as the law of $Y_{N(n)}$ 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

$$\widetilde{\tau}_0\,=\,\inf\,\big\{\,n\geq 0:Y_{N(n)}=0\,\big\}$$

and we observe 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 10.2, the ancestral line of any chromosome present at time n has the same law as

$$Y_{N(0)},Y_{N(1)},\dots,Y_{N(n)}\,.$$

From the previous observations, we conclude that

$$\forall n\geq 0\qquad P\big(\tau^*>n\,|\,D_0=(\ell)^m\big)\,\leq\,P(\widetilde{\tau}_0>n\,|\,Y_0=\ell)\,.$$

Summing this inequality over $n\geq 0$, we have

$$E\big(\tau^*\,|\,D_0=(\ell)^m\big)\,\leq\,E(\widetilde{\tau}_0\,|\,Y_0=\ell)\,.$$

For $n\geq 0$, let

$$T_n\,=\,\inf\,\big\{\,t\geq 0:N(t)=n\,\big\}\,.$$

The variables T_n-T_{n-1}, $n\geq 1$, have the same law, therefore

$$\forall n\geq 0\qquad E(T_n)\,=\,nE(T_1)\,=\,nm\,.$$

We will next express the upper bound on τ^* as a function of

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

The following computation is a simple case of Wald's identity (see for instance chapter 5, problem 11 of [6]):

$$\begin{aligned} E(\widetilde{\tau}_0 \,|\, Y_0=\ell) \;&=\; \sum_{t,n\geq 1} P(T_n\geq t,\,\tau_0=n \,|\, Y_0=\ell)\\ &=\sum_{n,t\geq 1} P(T_n\geq t)\,P(\tau_0=n \,|\, Y_0=\ell)\\ &=\sum_{n\geq 1} E(T_n)\,P(\tau_0=n \,|\, Y_0=\ell) \;=\; mE(\tau_0 \,|\, Y_0=\ell)\,. \end{aligned}$$

We obtain the desired conclusion with the help of proposition 10.12. □

In fact, we have derived the following upper bound on the discovery time.

Corollary 10.14 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

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

The harder part is to bound the discovery time τ^* from below. We give first an estimate showing that a visit to 0 becomes very unlikely if the starting point is far from 0.

Lemma 10.15 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

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

By proposition 10.1, for any $t\geq 0$, any $i\in\{\,1,\dots,m\,\}$,

$$P\big(D_t(i)=0 \,|\, D_0=(b)^m\big) \;=\; P\big(Y_{N(t)}=0 \,|\, Y_0=b\big)\,.$$

Using proposition 10.3 and lemma A.1, we have

$$P\big(Y_{N(t)}=0 \,|\, Y_0=b\big) \;\leq\; \frac{\mathcal{B}(0)}{\mathcal{B}(b)}\,.$$

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. $\square$

We derive next the lower bound on τ^*.

Proposition 10.16 Let $a \in]0, +\infty[$ and $\alpha \in [0, +\infty[$. For any $d \in \mathcal{N}$,

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

Proof. 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 of 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 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 $m\ell^{3/4}$, no mutation can recreate the master sequence. This implies that $\tau^* > m\ell^{3/4}$. Let us look at the population at time $m\ell^{3/4}$. Each chromosome present at this time has undergone an evolution whose law is the same as the mutation dynamics studied in section 10.2. The initial drift of the mutation dynamics is quite violent, therefore at time $m\ell^{3/4}$, it is very unlikely that a chromosome 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 the FKG inequality and an exponential estimate. Thus, at time $m\ell^{3/4}$, 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 $m\ell^2$. Because of the mutation drift, a chromosome starting at $\ln \ell$ has a very low probability of hitting 0 before time $m\ell^2$. Thus the process is very unlikely to discover the master sequence before time $m\ell^2$. Arguing again as before, we obtain that, for any $\varepsilon > 0$, at time $m\ell^2$, it is very unlikely that a particle evolving with the mutation dynamics is still in $\{\,0, \cdots, \ell_\kappa(1-\varepsilon)\,\}$. Thus, according to this scenario, we have $\tau^* > m\ell^2$ and

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

Let us make 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

$$\mathcal{E} \,=\, \big\{\, \forall n\leq m\ell^{3/4} \quad \forall l\leq \ln\ell \quad U_{n,l}>p/\kappa \,\big\}\,.$$

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

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

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

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

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

$$\begin{aligned}
P\big(\tau^*>\kappa^{\ell(1-\varepsilon)}\big) \,&\geq\, P\big(\tau^*>\kappa^{\ell(1-\varepsilon)},\,\mathcal{E}\big)\\
&\geq\, P\Big(\forall t\in\{\,m\ell^{3/4},\dots,\kappa^{\ell(1-\varepsilon)}\,\}\quad D_t\in\mathcal{N},\,\mathcal{E}\Big)\\
&=\, \sum_{d\in\mathcal{N}} P\Big(\forall t\in\{\,m\ell^{3/4},\dots,\kappa^{\ell(1-\varepsilon)}\,\}\quad D_t\in\mathcal{N},\,D_{m\ell^{3/4}}=d,\,\mathcal{E}\Big)\\
&\geq\, \sum_{d\geq(\ln\ell)^m} P\Big(\forall t\in\{\,m\ell^{3/4},\dots,\kappa^{\ell(1-\varepsilon)}\,\}\quad D_t\in\mathcal{N}\,|\,D_{m\ell^{3/4}}=d,\,\mathcal{E}\Big)\\
&\qquad\qquad\qquad\qquad \times P(D_{m\ell^{3/4}}=d,\,\mathcal{E})\,. \qquad (\Sigma_1)
\end{aligned}$$

Using the Markov property, we have

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

In the neutral case, by corollary 7.6, 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 (Σ_1), we get

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

Estimation of $P(D_{m\ell^{3/4}} \geq (\ln \ell)^m,\, \mathcal{E})$. The status of the process at time $m\ell^{3/4}$ is a function of the random vectors

$$R_n \,=\, \big(S_n, I_n, J_n, U_{n,1}, \dots, U_{n,\ell}\big)\,, \qquad 1 \leq n \leq m\ell^{3/4}\,.$$

We make an intermediate conditioning with respect to S_n, I_n, J_n:

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

The variables S_n, I_n, J_n, $1 \leq n \leq m\ell^{3/4}$ being fixed, the state of the process at time $m\ell^{3/4}$ is a non–decreasing function of the variables

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

Indeed, the mutation map $\mathcal{M}_H(\cdot, u_1, \dots, u_\ell)$ is non–decreasing with respect to $u_1, \dots, u_\ell$. Thus the events $\mathcal{E}$ and $\{\, D_{m\ell^{3/4}} \geq (\ln \ell)^m \,\}$ are both non–decreasing with respect to these variables. By the FKG inequality for a product measure (see the end of the appendix), we have

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

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

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

Reporting in the conditioning, we obtain

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

Next, we have

$$\begin{aligned} P\big(D_{m\ell^{3/4}} \geq (\ln \ell)^m\big) \,&=\, 1 - P\Big(\exists i \in \{\,1, \dots, m\,\} \quad D_{m\ell^{3/4}}(i) < \ln \ell\Big) \\ &\geq\, 1 - \sum_{1 \leq i \leq m} P\big(D_{m\ell^{3/4}}(i) < \ln \ell\big)\,. \qquad (\Sigma_2) \end{aligned}$$

From proposition 10.1,

$$\forall i\in\{\,1,\dots,m\,\}\qquad P\big(D_{m\ell^{3/4}}(i)<\ln\ell\big)\,=\,P\big(Y_{N(m\ell^{3/4})}<\ln\ell\big)\,,$$

therefore

$$P\big(D_{m\ell^{3/4}}\geq(\ln\ell)^m\big)\,\geq\,1-mP\big(Y_{N(m\ell^{3/4})}<\ln\ell\big)\,,$$

The random variable $N(m\ell^{3/4})$ follows the binomial law with parameter $m\ell^{3/4}$ and $1/m$, therefore

$$P\big(N(m\ell^{3/4})<\sqrt{\ell}\big)\,\leq\,P\Big(\exp\big(-N(m\ell^{3/4})\big)>\exp\big(-\sqrt{\ell}\big)\Big)\,\leq$$

$$\exp\big(\sqrt{\ell}\big)\Big(\frac{1}{em}+1-\frac{1}{m}\Big)^{m\ell^{3/4}}\leq\,\exp\Big(\sqrt{\ell}+\ell^{3/4}\big(\frac{1}{e}-1\big)\Big)\,\leq\,\exp\Big(\sqrt{\ell}-\frac{1}{2}\ell^{3/4}\Big)\,.$$

The previous estimate and proposition 10.9 yield

$$\begin{aligned}P\big(Y_{N(m\ell^{3/4})}\geq\ln\ell\big)\,&\geq\,\sum_{t=\sqrt{\ell}}^{+\infty}P\big(Y_{N(m\ell^{3/4})}\geq\ln\ell,\,N(m\ell^{3/4})=t\big)\\&=\sum_{t=\sqrt{\ell}}^{+\infty}P\big(Y_t\geq\ln\ell\big)\,P\big(N(m\ell^{3/4})=t\big)\end{aligned}$$

$$\geq\,\Big(1-\exp\Big(-\frac{1}{2}(\ln\ell)^2\Big)\Big)\Big(1-\exp\Big(\sqrt{\ell}-\frac{1}{2}\ell^{3/4}\Big)\Big)\,\geq\,1-\exp\Big(-\frac{1}{3}(\ln\ell)^2\Big)\,,$$

the last inequality being valid for ℓ large enough. Plugging this inequality in the sum (Σ_2), we get

$$P\big(D_{m\ell^{3/4}}\geq(\ln\ell)^m\big)\,\geq\,\Big(1-m\exp\Big(-\frac{1}{3}(\ln\ell)^2\Big)\Big)\,.$$

The last estimate, the inequalities (□) and (■) together give

$$P\big(D_{m\ell^{3/4}}\geq(\ln\ell)^m,\,\mathcal{E}\big)\,\geq\,\Big(1-m\exp\Big(-\frac{1}{3}(\ln\ell)^2\Big)\Big)\Big(1-\frac{p}{\kappa}\Big)^{m\ell^{3/4}\ln\ell}.\qquad(\triangledown)$$

Estimation of $P(\tau^*>\kappa^{\ell(1-\varepsilon)}\,|\,D_0=(\ln\ell)^m)$. Let $\varepsilon'>0$. We have

$$\begin{aligned}&P\Big(\tau^*>\kappa^{\ell(1-\varepsilon)}\,\big|\,D_0=(\ln\ell)^m\Big)\\&\quad=\,P\Big(\tau^*>m\ell^2,\,D_t\in\mathcal{N}\text{ for }m\ell^2\leq t\leq\kappa^{\ell(1-\varepsilon)}\,\big|\,D_0=(\ln\ell)^m\Big)\\&\quad=\sum_{d\in\mathcal{N}}P\Big(\begin{matrix}\tau^*>m\ell^2,\,D_{m\ell^2}=d\\D_t\in\mathcal{N}\text{ for }m\ell^2\leq t\leq\kappa^{\ell(1-\varepsilon)}\end{matrix}\,\Big|\,D_0=(\ln\ell)^m\Big)\\&\geq\sum_{d\geq(\ell_\kappa(1-\varepsilon'))^m}P\Big(D_t\in\mathcal{N}\text{ for }m\ell^2\leq t\leq\kappa^{\ell(1-\varepsilon)}\,\big|\,\tau^*>m\ell^2,\,D_{m\ell^2}=d\Big)\\&\qquad\qquad\times P\Big(\tau^*>m\ell^2,\,D_{m\ell^2}=d\,\big|\,D_0=(\ln\ell)^m\Big)\,.\qquad(\Sigma_3)\end{aligned}$$

Using the Markov property and the monotonicity of the process $(D_t)_{t\geq 0}$, we have for $d \geq (\ell_\kappa(1-\varepsilon'))^m$,

$$\begin{aligned}
P\Big(D_t \in \mathcal{N} \text{ for } m\ell^2 \leq t \leq \kappa^{\ell(1-\varepsilon)} \,\Big|\, \tau^* > m\ell^2,\, D_{m\ell^2} = d\Big)& \\
= P\Big(\forall t \in \{\,0,\dots,\kappa^{\ell(1-\varepsilon)} - m\ell^2\,\} \quad D_t \in \mathcal{N} \,\Big|\, D_0 = d\Big)& \\
= P\Big(\tau^* > \kappa^{\ell(1-\varepsilon)} - m\ell^2 \,\Big|\, D_0 = d\Big) \geq P\Big(\tau^* > \kappa^{\ell(1-\varepsilon)} \,\Big|\, D_0 = d\Big)& \\
\geq P\big(\tau^* > \kappa^{\ell(1-\varepsilon)} \,\big|\, D_0 = (\ell_\kappa(1-\varepsilon'))^m\big)\,.&
\end{aligned}$$

Reporting in the previous sum (Σ_3), we get

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

Estimation of $P(\tau^* > m\ell^2,\, D_{m\ell^2} \geq (\ell_\kappa(1-\varepsilon'))^m \,|\, D_0 = (\ln\ell)^m)$**.** We write

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

To control the last term, we use the inequality of lemma 10.15 with $n = m\ell^2$ and $b = \ln\ell$:

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

By lemma 10.4, we have

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

whence

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

For the other term, we use the monotonicity of the process $(D_t)_{t\geq 0}$ to get

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

From proposition 10.1, for any $i \in \{ 1, \dots, m \}$,

$$P\Big(D_{m\ell^2}(i) < \ell_\kappa(1-\varepsilon') \,\Big|\, D_0 = (0)^m\Big) \;=\; P\Big(Y_{N(m\ell^2)} < \ell_\kappa(1-\varepsilon') \,\Big|\, Y_0 = 0\Big)\,,$$

therefore

$$\begin{aligned} P\Big(D_{m\ell^2} \geq (\ell_\kappa(1-\varepsilon'))^m \,\Big|\, D_0 = (\ln\ell)^m\Big) & \\ \geq\; 1 - mP\Big(Y_{N(m\ell^2)} < \ell_\kappa(1-\varepsilon') \,\Big|\, Y_0 = 0\Big)\,. & \qquad (\sharp) \end{aligned}$$

The random variable $N(m\ell^2)$ follows the binomial law with parameter $m\ell^2$ and $1/m$, therefore

$$\begin{aligned} P\Big(N(m\ell^2) < \frac{4\ell}{a\varepsilon'}\Big) &\;\leq\; P\Big(\exp\big(-N(m\ell^2)\big) > \exp\big(-\frac{4\ell}{a\varepsilon'}\big)\Big) \\ &\leq\; \exp\big(\frac{4\ell}{a\varepsilon'}\big)\Big(\frac{1}{em} + 1 - \frac{1}{m}\Big)^{m\ell^2} \\ &\quad\leq\; \exp\Big(\frac{4\ell}{a\varepsilon'} + \ell^2\big(\frac{1}{e} - 1\big)\Big) \;\leq\; \exp\Big(\frac{4\ell}{a\varepsilon'} - \frac{1}{2}\ell^2\Big)\,. \end{aligned}$$

The previous estimate and proposition 10.10 yield

$$\begin{aligned} P\Big(Y_{N(m\ell^2)} \geq \ell_\kappa(1-\varepsilon') \,\Big|\, Y_0 = 0\Big) & \\ \geq\; P\Big(Y_{N(m\ell^2)} \geq \ell_\kappa(1-\varepsilon'),\, N(m\ell^2) \geq \frac{4\ell}{a\varepsilon'} \,\Big|\, Y_0 = 0\Big) & \\ \geq\; \sum_{t=4\ell/(a\varepsilon')}^{+\infty} P\Big(Y_{N(m\ell^2)} \geq \ell_\kappa(1-\varepsilon'),\, N(m\ell^2) = t \,\Big|\, Y_0 = 0\Big) & \\ =\; \sum_{t=4\ell/(a\varepsilon')}^{+\infty} P\Big(Y_t \geq \ell_\kappa(1-\varepsilon') \,\Big|\, Y_0 = 0\Big)\, P\big(N(m\ell^2) = t\big) & \\ \geq\; \Big(1 - \exp\big(-c(\varepsilon')\ell\big)\Big)\Big(1 - \exp\Big(\frac{4\ell}{a\varepsilon'} - \frac{1}{2}\ell^2\Big)\Big) \;\geq\; 1 - \exp\big(-\frac{1}{2}c(\varepsilon')\ell\big)\,, & \end{aligned}$$

where $c(\varepsilon') > 0$ and the last inequality is valid for ℓ large enough. Plugging this inequality in the inequality $(\sharp)$, we obtain

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

This inequality, the estimates $(\natural)$ and $(\flat)$ together give

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

Estimation of $P(\tau^* > \kappa^{\ell(1-\varepsilon)} \,|\, D_0 = (\ell_\kappa(1-\varepsilon'))^m)$. We use the inequality of lemma 10.15 with $n = \kappa^{\ell(1-\varepsilon)}$ and $b = \ell_\kappa(1-\varepsilon')$:

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

For ε' small enough, using the large deviation estimates of lemma 10.5, we see that there exists $c(\varepsilon,\varepsilon') > 0$ such that, for ℓ large enough,

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

Plugging the inequalities ($\clubsuit$) and ($\spadesuit$) into ($\heartsuit$), we obtain

$$P\Big(\tau^* > \kappa^{\ell(1-\varepsilon)} \,\big|\, D_0 = (\ln \ell)^m\Big) \,\geq \qquad (\,\triangle\,)$$
$$\times\Big(1-\exp(-c(\varepsilon,\varepsilon')\ell)\Big)\Big(1 - m\exp\big(-\frac{1}{2}c(\varepsilon')\ell\big) \,-\, (m\ell)^2\Big(\frac{2\ln\ell}{\ell}\Big)^{\ln \ell}\Big)\,.$$

Reporting the inequalities ($\bigtriangledown$), ($\triangle$) into ($\circ$), we conclude that, for ℓ large enough,

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

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. ☐

11 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 9 and 10 in order to evaluate the formula for the invariant measure obtained at the end of section 8.4. For $\theta = \ell, 1$, we rewrite this formula as

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

Recall that $M_H(\theta,0) = \delta_0$. By proposition 9.3,

$$\lim_{\substack{\ell,m\to\infty\\ q\to 0,\, \ell q\to a}} \Big(\frac{1}{M_H(\theta,0)E\big(\tau_0 \,\big|\, Z^\theta_0 = 1\big)} + 1\Big) \sum_{i=1}^m f\Big(\frac{i}{m}\Big)\nu^\theta(i) \;=\; f\big(\rho^*(a)\big)\,.$$

By corollary 9.2, and using the function ϕ defined in section 9.2,

$$\lim_{\substack{\ell,m\to\infty\\ q\to 0,\, \ell q\to a}} \frac{1}{m} \ln E(\tau_0 \,|\, Z^\theta_0 = 1) \;=\; \int_0^{\rho^*(a)} \ln \phi(e^{-a},0,s)\, ds\,.$$

By propositions 10.13 and 10.16, 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 10.14 and proposition 10.12,

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

These estimates allow to evaluate the ratio between the discovery time and the persistence time. We define a function $\phi :]0,+\infty[\to [0,+\infty]$ by setting $\phi(a) = 0$ if $a \geq \ln\sigma$ and

$$\forall a < \ln\sigma \qquad \phi(a) \;=\; \int_0^{\rho^*(a)} \ln \phi(e^{-a},0,s)\, ds\,.$$

We have then, for $\alpha \in [0,+\infty[$ or $\alpha = +\infty$,

$$\lim_{\substack{\ell,m\to\infty,\, q\to 0\\ \ell q\to a,\, \frac{m}{\ell}\to\alpha}} \frac{E\big(\tau_0 \,\big|\, Z_0^\theta = 1\big)}{E\big(\tau^* \mid O_0^\theta = o^\theta_{\text{exit}}\big)} = \begin{cases} \;0 & \text{if } \alpha\,\phi(a) < \ln\kappa \\ +\infty & \text{if } \alpha\,\phi(a) > \ln\kappa \end{cases}$$

Notice that the result is the same for $\theta = \ell$ and $\theta = 1$. Putting together the bounds on ν given in section 8.4 and the previous considerations, we conclude that

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

This is valid for any continuous non–decreasing function $f : [0,1] \to \mathbb{R}$ such that $f(0) = 0$. To obtain the statement of theorem 5.1, it remains to compute the integral. For $a < \ln\sigma$,

$$\begin{aligned} \phi(a) \;&=\; \int_0^{\rho^*(a)} \ln\phi(e^{-a},0,s)\,ds \\ &= \int_0^{\rho^*(a)} \ln\frac{\sigma e^{-a}(1-s)}{\sigma(1-e^{-a})s + (1-s)}\,ds \\ &= \frac{\sigma(1-e^{-a})\ln\dfrac{\sigma(1-e^{-a})}{\sigma-1} + \ln(\sigma e^{-a})}{(1-\sigma(1-e^{-a}))} \end{aligned}$$

and we are done.

A Appendix on Markov chains

In this appendix, we recall classical definitions and results from the theory of Markov chains with finite state space. The goal is to clarify the objects involved in the definition of the model, and to state the fundamental general results used in the proofs. This material can be found in any reference book on Markov chains, for instance [7], [18], [22]. The definitions and results on monotonicity, coupling and the FKG inequality are exposed in the books of Liggett [26] and Grimmett [20].

Construction of continuous time Markov processes. The most convenient way to define a continuous time process is to give its infinitesimal generator. The infinitesimal generator of a Markov process $(X_t)_{t\geq 0}$ with values in a finite state space $\mathcal{E}$ is the linear operator L acting on the functions from $\mathcal{E}$ to $\mathbb{R}$ defined as follows. For any function $\phi : \mathcal{E} \to \mathbb{R}$, any $x \in \mathcal{E}$,

$$L\phi(x) \,=\, \lim_{t\to 0} \frac{1}{t}\Big(E\big(\phi(X_t)|X_0 = x\big) - \phi(x)\Big)\,.$$

It turns out that the law of the process $(X_t)_{t\geq 0}$ is entirely determined by the generator L. Therefore all the probabilistic results on the process $(X_t)_{t\geq 0}$ can in principle be derived working only with its infinitesimal generator.

In the case where the state space of the process is finite, the situation is quite simple and it is possible to provide direct constructions of a process having a specific infinitesimal generator. These constructions are not unique, but they provide more insight into the dynamics. Suppose that the generator L is given by

$$\forall x \in \mathcal{E} \qquad L\phi(x) \,=\, \sum_{y\in\mathcal{E}} c(x,y)\big(\phi(y) - \phi(x)\big)\,.$$

The evolution of a process $(X_t)_{t\geq 0}$ having L as infinitesimal generator can loosely be described as follows. Suppose that $X_t = x$. Let

$$c(x) \,=\, \sum_{y\neq x} c(x,y)\,.$$

Let τ be a random variable whose law is exponential with parameter $c(x)$:

$$\forall s \geq 0 \qquad P(\tau \geq s) \,=\, \exp(-c(x)s)\,.$$

The process waits at x until time $t+\tau$. At time $t+\tau$, it jumps to a state $y \neq x$ chosen according to the following law:

$$P\big(X_{t+\tau} = y\big) \,=\, \frac{c(x,y)}{c(x)}\,.$$

The same scheme is then applied starting from y. In this construction, the waiting times τ and the jumps are all independent.

Construction of discrete time Markov chains. To build a discrete time Markov chain, we need only to define its transition mechanism. When the state space $\mathcal{E}$ is finite, this amounts to giving its transition matrix

$$\big(p(x,y),\, x,y \in \mathcal{E}\big)\,.$$

The only requirement on p is that it is a stochastic matrix, i.e., it satisfies

$$\begin{aligned} \forall x,y \in \mathcal{E} \qquad & 0 \leq p(x,y) \leq 1\,, \\ \forall x \in \mathcal{E} \qquad & \sum_{y \in \mathcal{E}} p(x,y) \;=\; 1\,. \end{aligned}$$

In the sequel, we consider a discrete time Markov chain $(X_t)_{t \geq 0}$ with values in a finite state space $\mathcal{E}$ and with transition matrix $(p(x,y))_{x,y \in \mathcal{E}}$.

Invariant probability measure. If the Markov chain is irreducible and aperiodic, then it admits a unique invariant probability measure μ, i.e., the set of equations

$$\mu(y) \;=\; \sum_{x \in \mathcal{E}} \mu(x)\, p(x,y)\,, \qquad y \in \mathcal{E}\,,$$

admits a unique solution. The Markov chain $(X_t)_{t \geq 0}$ is said to be reversible with respect to a probability measure ν if it satisfies the detailed balanced conditions:

$$\forall x,y \in \mathcal{E} \qquad \nu(x)\, p(x,y) \;=\; \nu(y)\, p(y,x)\,.$$

If the Markov chain $(X_t)_{t \geq 0}$ is reversible with respect to a probability measure ν, then ν is an invariant probability measure for $(X_t)_{t \geq 0}$. In case $(X_t)_{t \geq 0}$ is in addition irreducible and aperiodic, then ν is the unique invariant probability measure of the chain.

Lemma A.1 Suppose that μ is an invariant probability measure for the Markov chain $(X_t)_{t \geq 0}$. We have then

$$\forall x,y \in \mathcal{E} \quad \forall t \geq 0 \qquad \mu(x) P\big(X_t = y \,|\, X_0 = x\big) \;\leq\; \mu(y)\,.$$

Proof. The proof is done by induction on t. For $t = 0$, we have

$$P\big(X_0 = y \,|\, X_0 = x\big) \;=\; 0 \quad \text{if} \quad y \neq x\,,$$

and the result holds. Suppose it has been proved until time $t \in \mathbb{N}$. We have then, for $x, y \in \mathcal{E}$,

$$\begin{aligned}\mu(x)\,P\big(X_{t+1}=y\,|\,X_0=x\big) \;&=\; \sum_{z\in\mathcal{E}} \mu(x)\,P\big(X_{t+1}=y,\,X_t=z\,|\,X_0=x\big)\\ &= \sum_{z\in\mathcal{E}} \mu(x)\,P(X_t=z\,|\,X_0=x\big)\,P\big(X_{t+1}=y\,|\,X_t=z)\\ &\leq \sum_{z\in\mathcal{E}} \mu(z)\,p(z,y) \;=\; \mu(y)\end{aligned}$$

and the claim is proved at time $t+1$. $\square$

We state next the ergodic theorem for Markov chains. We consider only the case where the state space $\mathcal{E}$ is finite.

Theorem A.2 Suppose that the Markov chain $(X_t)_{t\geq 0}$ is irreducible aperiodic. Let μ be its invariant probability measure. For any initial distribution μ_0, for any function $f:\mathcal{E}\to\mathbb{R}$, we have, with probability one,

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t f(X_s)\,ds \;=\; \int_{\mathcal{E}} f(x)\,d\mu(x)\,.$$

Lumping. The basic lumping result for Markov chains can be found in section 6.3 of the book of Kemeny and Snell [23]. Let $(E_1,\dots,E_r)$ be a partition of $\mathcal{E}$. Let $f:\mathcal{E}\to\{\,1,\dots,r\,\}$ be the function defined by

$$\forall x\in\mathcal{E}\qquad f(x)\;=\;\begin{cases}1 & \text{if}\quad x\in E_1\\ \vdots & \vdots\\ r & \text{if}\quad x\in E_r\end{cases}\,.$$

The Markov chain $(X_t)_{t\geq 0}$ is said to be lumpable with respect to the partition $(E_1,\dots,E_r)$ if, for every initial distribution μ_0 of X_0, the process $\big(f(X_t)\big)_{t\geq 0}$ is a Markov chain on $\{\,1,\dots,r\,\}$ whose transition probabilities do not depend on μ_0.

Theorem A.3 (Lumping theorem) A necessary and sufficient condition for the Markov chain $(X_t)_{t\geq 0}$ to be lumpable with respect to the partition $(E_1,\dots,E_r)$ is that,

$$\forall i,j\in\{\,1,\dots,r\,\}\quad \forall x,y\in E_i\qquad \sum_{z\in E_j} p(x,z)\;=\;\sum_{z\in E_j} p(y,z)\,.$$

Suppose that this condition holds. For $i,j\in\{\,1,\dots,r\,\}$, let us denote by $p_E(i,j)$ the common value of the above sums. The process $\big(f(X_t)\big)_{t\geq 0}$ is then a Markov chain with transition matrix $(p_E(i,j))_{1\leq i,j\leq r}$.

Monotonicity. We recall some standard definitions concerning monotonicity for stochastic processes. A classical reference is Liggett's book [26], especially for applications to particle systems. We consider a discrete time Markov chain $(X_t)_{t\geq 0}$ 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

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

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

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

is non–decreasing.

Coupling. A natural way to prove monotonicity is to construct an adequate coupling. A coupling is a family of processes $(X_t^x)_{t\geq 0}$ 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 $(X_t^x)_{t\geq 0}$ is the Markov chain starting from $X_0=x$. The coupling is said to be monotone if

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

If there exists a monotone coupling, then the Markov chain is monotone.

FKG inequality. We consider the product space $[0,1]^n$ equipped with the product order. Let μ be a probability measure on $[0,1]$ and let us denote by $\mu^{\otimes n}$ the product probability measure on $[0,1]^n$ whose marginals are equal to μ. The Harris inequality, or the FKG inequality in this context, says that, for any non–decreasing functions $f,g:[0,1]^n\to\mathbb{R}$, we have

$$\int_{[0,1]^n}fg\,d\mu^{\otimes n}\,\geq\int_{[0,1]^n}f\,d\mu^{\otimes n}\int_{[0,1]^n}g\,d\mu^{\otimes n}\,.$$

The case of Bernoulli product measures is exposed in section 2.2 of Grimmett's book [20].

References

[1] Domingos Alves and Jose Fernando Fontanari. Error threshold in finite populations. *Phys. Rev. E*, 57:7008–7013, 1998.

[2] Jon P. Anderson, Richard Daifuku, and Lawrence A. Loeb. Viral error catastrophe by mutagenic nucleosides. *Annual Review of Microbiology*, 58(1):183205, 2004.

[3] Ellen Baake and Wilfried Gabriel. Biological evolution through mutation, selection, and drift: An introductory review. *Ann. Rev. Comp. Phys. VII*, pages 203–264, 2000.

[4] Michael Baake and Ellen Baake. An exactly solved model for mutation, recombination and selection. *Canad. J. Math.*, 55(1):3–41, 2003.

[5] Nicholas Hugh Bingham. Fluctuation theory for the Ehrenfest urn. *Advances in Applied Probability*, 23(3):598–611, 1991.

[6] Leo Breiman. *Probability.* Addison-Wesley Publishing Company, Reading, Mass., 1968.

[7] Leo Breiman. *Probability*, volume 7 of *Classics in Applied Mathematics.* SIAM, Philadelphia, 1992. Corrected reprint of the 1968 original.

[8] Shane Crotty, Craig E. Cameron, and Raul Andino. RNA virus error catastrophe: Direct molecular test by using ribavirin. *Proceedings of the National Academy of Sciences*, 98(12):68956900, 2001.

[9] Lloyd Demetrius, Peter Schuster, and Karl Sigmund. Polynucleotide evolution and branching processes. *Bulletin of Mathematical Biology*, 47(2):239 – 262, 1985.

[10] Narendra M. Dixit, Piyush Srivastava, and Nisheeth K. Vishnoi. A finite population model of molecular evolution: Theory and computation. *preprint, available on arXiv*, 2012.

[11] Esteban Domingo. Quasispecies theory in virology. *Journal of Virology*, 76(1):463–465, 2002.

[12] Esteban Domingo, Christof Biebricher, Manfred Eigen, and John J. Holland. *Quasispecies and RNA virus evolution: principles and consequences.* Landes Bioscience, Austin, Tex., 2001.

[13] Manfred Eigen. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523, 1971.

[14] Manfred Eigen. Natural selection: a phase transition? *Biophysical Chemistry*, 85(2–3):101–123, 2000.

[15] Manfred Eigen, John McCaskill, and Peter Schuster. The molecular quasi-species. *Advances in Chemical Physics*, 75:149–263, 1989.

[16] Santiago F. Elena, Claus O. Wilke, Charles Ofria, and Richard E. Lenski. Effects of population size and mutation rate on the evolution of mutational robustness. *Evolution*, 61(3):666–74, 2007.

[17] Warren J. Ewens. *Mathematical population genetics. I*, volume 27 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, 2004.

[18] William Feller. *An introduction to probability theory and its applications. Vol. I and II.* John Wiley & Sons Inc., New York, 1968 and 1971.

[19] 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.

[20] Geoffrey Grimmett. *Percolation*, volume 321 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, second edition, 1999.

[21] Mark Kac. Random walk and the theory of brownian motion. *American Mathematical Monthly*, 54(7):369–391, 1947.

[22] Samuel Karlin and Howard M. Taylor. *A first course in stochastic processes.* Academic Press, New York-London, second edition, 1975.

[23] John G. Kemeny and J. Laurie Snell. *Finite Markov chains*. Springer-Verlag, New York, 1976. Undergraduate Texts in Mathematics.

[24] Motoo Kimura. *The Neutral Theory of Molecular Evolution.* Cambridge University Press, 1985 (reprint).

[25] Ira Leuthäusser. Statistical mechanics of Eigen's evolution model. *J. Statist. Phys.*, 48(1-2):343–360, 1987.

[26] Thomas M. Liggett. *Interacting particle systems.* Classics in Mathematics. Springer-Verlag, 2005. Reprint of the 1985 original.

[27] John McCaskill. A stochastic theory of macromolecular evolution. *Biological Cybernetics*, 50:63–73, 1984.

[28] Pat Moran. Random processes in genetics. *Proc. Cambridge Philos. Soc.*, 54:60–71, 1958.

[29] Fabio Musso. A stochastic version of the Eigen model. *Bull. Math. Biol.*, 73(1):151–180, 2011.

[30] Erik Van Nimwegen, James P. Crutchfield, and Martijn Huynen. Neutral evolution of mutational robustness. *Proc. Natl . Acad. Sci . USA*, 96:9716–9720, 1999.

[31] Martin A. Nowak. *Evolutionary dynamics.* The Belknap Press of Harvard University Press, Cambridge, MA, 2006.

[32] 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.

[33] Jeong-Man Park, Enrique Muñoz, and Michael W. Deem. Quasispecies theory for finite populations. *Phys. Rev. E*, 81:011902, 2010.

[34] 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.

[35] Roberto H. Schonmann. The pattern of escape from metastability of a stochastic Ising model. *Comm. Math. Phys.*, 147(2):231–240, 1992.

[36] Ricard V. Solé and Thomas S. Deisboeck. An error catastrophe in cancer? *Journal of Theoretical Biology*, 228:47–54, 2004.

[37] Sumedha, Olivier C Martin, and Luca Peliti. Population size effects in evolutionary dynamics on neutral networks and toy landscapes. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(05):P05011, 2007.

[38] Kushal Tripathi, Rajesh Balagam, Nisheeth K. Vishnoi, and Narendra M. Dixit. Stochastic simulations suggest that HIV–1 survives close to its error threshold. *preprint*, 2012.

[39] Erik van Nimwegen and James Crutchfield. Metastable evolutionary dynamics: Crossing fitness barriers or escaping via neutral paths? *Bulletin of Mathematical Biology*, 62:799–848, 2000.

[40] 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.

[41] Claus Wilke. Quasispecies theory in the context of population genetics. *BMC Evolutionary Biology*, 5:1–8, 2005.

Contents