Evaluating Stationary Distribution of the Binary GA Markov Chain in Special Cases

Boris Mitavskiy * and Chris Cannings

School of Medicine, University of Sheffield, Sheffield, England, S10 2RX
B.Mitavskiy@sheffield.ac.uk, C.Cannings@sheffield.ac.uk

Abstract. The evolutionary algorithm stochastic process is well-known to be Markovian. These have been under investigation in much of the theoretical evolutionary computing research. When mutation rate is positive, the Markov chain modeling an evolutionary algorithm is irreducible and, therefore, has a unique stationary distribution. Rather little is known about the stationary distribution. In fact, the only quantitative facts established so far, tell us that the stationary distributions of Markov chains modeling evolutionary algorithms concentrate on the *uniform* populations (i.e. these populations consisting of the repeated copy of the same individual). At the same time, knowing the stationary distribution may provide some information about the expected time it takes for the algorithm to reach a certain solution, assessment of the biases due to recombination and *selection* and is of importance in population genetics to assess what's called a "genetic load" (see the introduction for more details). In the recent joint works of the first author, some bounds have been established on the rates at which the stationary distribution concentrates on the uniform populations. The primary tool used in these papers is the "quotient construction" method. It turns out that the "quotient construction" method can be significantly strengthen to yield much more informative results. In the current paper we present a couple of examples where we compute the stationary distribution of a GA Markov chain using the quotient construction method. Furthermore, we show another important asymptotic result which we hope will be of technical importance in the future applications.

1 Introduction

One of the aspects of the theoretical analysis of the evolutionary algorithms is studying the properties of the Markov chains associated with these algorithms. Many research articles in the field of evolutionary computing have been devoted to this subject (see, for instance, [33], [34], [38], [9], [37], [36] and [5] for a survey of known results and open questions). In vast majority of cases, Markov chains modeling evolutionary algorithms are irreducible due to positive mutation rates and, hence, possess unique stationary distributions. Knowing the relative size of the stationary distribution value of one subset comparing to another one (estimating such ratios is what the current paper is devoted to) can be advantageous for a number of reasons. Some of these are described below.

In [35] a heuristic computational approach to estimate the average waiting time for a GA to hit the desired individuals based on the Markov chain model has been offered.

* Supported by EPSRC EP/D003/05/1 Amorphous Computing Grant.

Dagstuhl Seminar Proceedings 08051
Theory of Evolutionary Algorithms
http://drops.dagstuhl.de/opus/volltexte/2008/1484

Due to the super-exponential growth of the state space with respect to the string length and population size, such estimates inevitably involve reducing the size of the Markov chain modeling the genetic algorithm by considering an appropriate aggregation (also known as quotient) Markov chain obtained by lumping some states into one (see, for instance, [15], [7] or [4]). The model of [35] applies to a limited class of Markov chains nevertheless. In [27] a different method has been offered for obtaining the lumping quotients which applies to a wider class of Markov chains modeling EAs. In either case, expressions for the quotient (or aggregated) Markov chain transition probabilities involve the ratios of the form $\frac{w_b}{\sum_{c \in B} w_c}$ where the weights w_b are supposed to measure the relative frequency of occurrence of the states $b \in B$ (here B is a set of states that are lumped into a single state of the quotient chain). In fact, the long-term frequency of occurrence of such states is precisely the stationary distribution value of these states and the current paper is devoted to obtaining rigorous estimates of such ratios (as opposed to the heuristic techniques employed in [27] and [35]). At least, the estimates can be used as reasonable starting points of the heuristic estimation algorithms proposed in [27].

Evolutionary algorithms essentially simulate the natural, or approximate, evolutionary processes, thereby discovering which genes become most frequent in the process of evolution. Thus the theory of evolutionary algorithms implicitly informs us of issues of importance in population genetics. The classical theorems of Geiringer (see [11] for a review of her results, also [3], [10] and [29] for later developments in infinite populations) tells us the frequency of alleles at a set of loci for various genetic systems in the limit under the assumption of no mutation and flat fitness. In [22] a finite population version of the Geiringer theorem has been established which is stated in terms of the stationary distribution of the Markov chain modeling an evolutionary algorithm. This finite population model is more suitable for the purpose of describing evolutionary algorithms (yet it's also closely related to the infinite population model). The Geiringer theorem of [22] applies only in the absence of selection, however. It may be of interest to find out how selection effects the frequencies of occurrence of various genes in the long run and this is precisely the information which the stationary distribution of the Markov chain modeling the EA (together with selection) carries in itself. The procedure for finding the frequencies of specific individuals given the knowledge of the stationary distribution is described in (cite my Geiringer paper).

One particular application of the above would be in assessing the genetic load carried by a (possibly) human population, a concept introduced in [13]. Implicit in the persistence of individuals of reduced fitness, in the equilibrium population is the loss of individuals (or putative) individuals through selection (or reduced fertility); this loss being measured by the genetic load. Assessing the size of this genetic load can be made analytically for simple models of mutation and fitness, but in more complex cases this is not possible. In these latter one could assess the genetic load by estimating the stationary distribution and inferring the load from that, so that the current paper can provide a method for doing this in the multi-locus, multi-allele context. This is of importance to human populations both directly (as in [12]) and through its effects on other species we interact with (as in [14]).

Another application of Geiringer-like theorems as well as other results on the fixed points in the infinite or finite population models, is to assess the sampling biases which

individuals may possess due to a particular representation of the search space. A number of articles are devoted to this rather important practical issue (see, for instance, [31], [17] and [18]). When the estimates are based on the flat fitness landscape only, they don't take into account how the change in fitness values affects the biases. Since Markov chains modeling evolutionary algorithms are irreducible they are biased towards the stationary distribution which tells us the long-term frequency of occurrence of a given population (see section 4 for more details). From that we can often easily deduce the long-term frequency of occurrence of given individuals. In fact, the ratios of the stationary distribution values of various sets of populations may be sufficient. The results of the current paper apply to arbitrary fitness landscapes which, therefore, may provide a tool for a more informative assessment of the biases.

Finally, it seems worth mentioning that the ratio of the stationary distribution values of singleton subsets to one another measure the expected number of visits of the chain started at the state in the denominator to the numerator state before returning back to the denominator state (see chapter 2 of [2]). For example, if $X = \{x, x, \ldots, x\}$ and $Y = \{y, y, \ldots, y\}$ are the populations consisting of a repeated copy of the same individual (an x and a y respectively) then $\frac{\pi(X)}{\pi(Y)}$ measures the expected number of visiting population X starting with the initial population Y before returning back to the population Y. While not related to the expected waiting time until the first visit in an apparent way, this seems like a potentially useful piece information at least for the future theoretical work.

As mentioned briefly before, one difficulty that arises with the Markov chain approach, is the fact that the number of states of this Markov chain grows very fast with respect to the size of the search space and the number of elements in a population. Indeed, if Ω denotes the search space, the number of states of this Markov chain for a population of size m is $|\Omega|^m$. In [25] an elegant and simple method based on the "quotient Markov chain" construction has been introduced to study the stationary distribution of some Markov chains. Asymptotic results about the rate of concentration of the stationary distribution of the Markov chains modeling evolutionary algorithms on the uniform populations (populations consisting of repeated copies of the same individual) have been obtained. Such a notion of a quotient of a Markov chain is frequently referred to as "coarse graining" in the evolutionary computation literature. In probability it's known as "lumping quotient", "aggregation" or "quotient under lumping equivalence relation" (see [4], [15], [7]). We shall discuss the construction of a quotient of an irreducible Markov chain with respect to an arbitrary equivalence relation on the state space. The stationary distribution of the quotient chain is "coherent" with the stationary distribution of the original chain. Although the transition probabilities of the quotient chain depend on the stationary distribution of the original chain, we can still exploit the quotient construction to deduce some estimates on the stationary distribution of the original chain via proposition 15. For the sake of completeness, the quotient construction method and the theory behind it will be presented in sections 4 and 5 respectively. In the current paper we strengthen the quotient construction method (see proposition 13, corollary 16 and corollary 17 which appear for the first time in this paper) and present a few more applications. In particular, we compute exact ratios of the stationary distribution values on some subsets for a couple specific examples (see sections 6 and 8) and

we also establish a rather general asymptotic result about the stationary distributions of Markov chains modeling EAs in section 7. This result is necessary for the application of corollary 17 in section 8. As a matter of fact, we believe that this asymptotic result can be used in conjunction with the quotient construction method to accomplish much more (this will hopefully appear in the sequel papers).

2 A Reminder about the Asymptotic Notation

Throughout the paper we shall make extensive use of the following notation:

If $F : U \to \mathbb{R}$ and $G : U \to \mathbb{R}$ are functions of μ then we write $F = O(G)$ if $\forall\, \mu \in U$ we have $|F(\mu)| \leq k \cdot |G(\mu)|$ for some constant $k \in \mathbb{R}$; $F = \Omega(G)$ if $\forall\, \mu \in U$ we have $|F(\mu)| \geq k \cdot |G(\mu)|$ and $F = \Theta(G)$ if $F = O(G)$ and $F = \Omega(G)$.

Remark 1 It may be worth pointing out that in the definition above U is an arbitrary set. Throughout the current paper we will set $U = (0,\, a]$ for some constant $a < 1$ since the independent variable of the functions we deal with is the mutation rate denoted by μ which takes values in such intervals. All the other parameters (such as population size or string length etc.) are assumed to be fixed constants unless explicitly stated otherwise.

3 Which Algorithms Do We Consider?

In the current paper we consider the classical binary genetic algorithm with the search space $\Omega = \{0,\, 1\}^n$, fitness function $f : \Omega \to \{0,\, 1\}$ and population size 2. The algorithms cycles through the three basic stages, selection, recombination and mutation. By a stage here we mean a probabilistic rule which takes a given population as an input and returns another population as an output with some probability. More precisely, a stage can be described by a Markov transition matrix as follows:

Definition 2 Given a search space Ω and an integer $m > 0$, called *population size*,[1] a *stage* is a Markov transition matrix on the set Ω^m of *populations*[2] of size m where $p^{\text{stage}}_{\mathbf{x}\to\mathbf{y}}$ denotes the probability that if the stage takes population $\mathbf{x}$ as its input it returns the population $\mathbf{y}$ as its output.

Each of these stages is described below in detail for a population of size m (even though in the current paper $m = 2$):

Definition 3 *Selection* is a stage which takes a population of size m, say $\mathbf{x} = (x_1,\, x_2, \ldots x_m)$, and returns another population $\mathbf{y} = (y_1,\, y_2, \ldots y_m)$ as an output where $\forall\, i\, \exists\, j$ with $y_i = x_j$. In the sense of definition 2, this means that $p^{\text{selection}}_{\mathbf{x}\to\mathbf{y}} = 0$ unless $\forall\, i\, \exists\, j$ with $y_i = x_j$.

---

[1] As mentioned earlier, in the current paper we consider only the population size $m = 2$ and $\Omega = \{0,\, 1\}$.

[2] Here we consider *ordered* populations to avoid combinatorial complications. Thus they are elements of Ω^m rather than multi-sets.

In the current paper we will carry out the computations specifically for the stage of fitness-proportional selection, defined below, however one can easily do the same for any type of selection procedure in the sense of definition 3 as a routine exercise.

Definition 4 The stage of *fitness-proportional selection* is such a selection stage (in the sense of definition 3) where the transition probability $p_{\mathbf{x}\to\mathbf{y}}$ can be computed as follows: Let $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_m)$. For every i choose $j(i)$ (say the smallest j) such that $y_i = x_{j(i)}$ (if this is not possible then let $p_{\mathbf{x}\to\mathbf{y}} = 0$ as must be the case for a selection stage according to definition 3). Now let

$$p_{\mathbf{x}\to\mathbf{y}} = \frac{\prod_{i=1}^{m} f(x_{j(i)})}{(\sum_{l=1}^{m} f(x_l))^m}.$$

There is not much we need to assume about the stage of recombination apart from the following property referred to as purity (see [32], [19], [20] and [21]).

Definition 5 A *recombination stage* is any stage in the sense of definition 2 with the property that whenever $\mathbf{x} = (x, x, \ldots, x)$ (i.e. $\mathbf{x}$ consists of a repeated copy of the same individual x), we have $p_{\mathbf{x}\to\mathbf{x}} = 1$.

In the current paper we will be dealing only with the classical binary genetic algorithm where populations and recombination stage can be defined as follows. Since $\Omega = \{0, 1\}^n$, the individuals are binary strings of length n. We can then represent a population of size m as an $m \times n$ matrix where the rows are the individuals.

Example 6 If $\Omega = \{0, 1\}^5$ then the 4×5 matrix $\begin{pmatrix} 0\,1\,1\,1\,0 \\ 0\,0\,1\,0\,1 \\ 0\,1\,0\,1\,0 \\ 0\,1\,1\,1\,0 \end{pmatrix}$ represents the population $\mathbf{x} = (x_1, x_2, x_3, x_4) \in \Omega^4$ of size 4 where $x_1 = (0, 1, 1, 1, 0)$, $x_2 = (0, 0, 1, 0, 1)$, $x_3 = (0, 1, 0, 1, 0)$ and $x_4 = (0, 1, 1, 1, 0)$.

When equipped with the matrix notation it is convenient to describe the class of recombination stages to which our results apply:

Definition 7 A *binary GA recombination stage* is a stage where given any two $m \times n$ matrices $\mathbf{x} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \ldots & x_{mn} \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} y_{11} & y_{12} & \ldots & y_{1n} \\ y_{21} & y_{22} & \ldots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \ldots & y_{mn} \end{pmatrix}$, we have $p_{\mathbf{x}\to\mathbf{y}} > 0 \Longrightarrow \exists$ permutations $\sigma_1, \sigma_2, \ldots, \sigma_n$ on the set of indices $\{1, 2, \ldots, m\}$ such that $y_{ij} = x_{\sigma_j(i)j}$

Less formally, definition 7 describes the class of recombination methods which only reshuffle the genetic material: they neither introduce new alleles nor delete the existing once. This kind of recombination stages includes one-point and masked 2-parent $\to$ 2-children crossovers exploited in binary GAs.

Remark 8 It is easy to see that binary GA recombination stage in the sense of definition 7 is also a recombination stage in the sense of definition 5 since permuting identical alleles (in the columns) does not alter the matrix.

It remains to describe the mutation stage. While this can be done for a rather wide class of algorithms (see, for instance, [26]), it requires introducing an integer-valued metric space structure on the search space (a "hamming distance" analogue). In the current paper we only apply this notion to the binary GA and, to avoid extra definitions, we introduce mutation stage for binary GAs only:

Definition 9 Let $\mu > 0$ and $\mu \leq 1$. Suppose we are given $m \times n$ population matrices $\mathbf{x} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} y_{11} & y_{12} & \dots & y_{1n} \\ y_{21} & y_{22} & \dots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \dots & y_{mn} \end{pmatrix}$. Consider the set of index pairs $N = \{(i,\, j) \,|\, x_{ij} \neq y_{ij}\}$. Then we say that *mutation stage* is a stage (in the sense of definition 2) given by the Markov transition matrix $\{p_{\mathbf{x}\to\mathbf{y}}\}_{\mathbf{x},\,\mathbf{y}\in(\{0,\,1\}^n)^m}$ where $p_{\mathbf{x}\to\mathbf{y}} = \mu^{|N|} \cdot (1-\mu)^{mn-|N|}$ (it is easy to check that this is a well-defined Markov transition matrix).

We have now described all the three basic stages involved in binary GAs. It only remains to introduce the Markov chain that models an EA. In general, if the algorithm cycles through the stages $M_0, M_2, \dots, M_q$ starting with M_1 and finishing with M_q (meaning that the input of the stage M_i is the output of the stage M_{i-1} where all the indices and their subtraction are in $\mathbb{Z}_{q+1}$), its Markov transition matrix is the product $M_q \cdot M_{q-1} \cdot \ldots \cdot M_0$. In most cases, due to some positive mutation rate (certainly the case for binary GAs), the Markov transition matrix for EAs has all positive entries, and, hence, the corresponding Markov chain is irreducible. Such Markov chains are known to have unique global fixed point attractors, called stationary distributions. The current paper is devoted to computing the stationary distribution of a binary GA Markov chain in a couple special cases. We also establish a rather general asymptotic result about the stationary distributions of GA Markov chains in section 7. We believe that the method can be improved and extended further to compute and estimate the stationary distribution of evolutionary algorithms in many other cases.

4 Quotients of Irreducible Markov Chains.

4.1 Notation and Fundamental Facts from Markov Chain Theory.

Throughout the current section we shall be dealing with an irreducible Markov chain $\mathcal{M}$ over a finite state space $\mathcal{X}$. Let $\{p_{\mathbf{x}\to\mathbf{y}}\}$ denote the Markov transition matrix with the convention that $p_{\mathbf{x}\to\mathbf{y}}$ is the probability of getting $\mathbf{y}$ in the next stage given $\mathbf{x}$. In the content of the current paper, saying that $\mathcal{M}$ is *irreducible* means that there exists a natural number $k \in \mathbb{N}$ such that the k^{th} power of the Markov transition matrix of $\mathcal{M}$ consists of only nonzero entries (i.e. $\forall\, \mathbf{x}$ and $\mathbf{y} \in \mathcal{X}$ we have $p^k_{\mathbf{x}\to\mathbf{y}} > 0$ where $p^k_{\mathbf{x}\to\mathbf{y}}$ denotes the probability of reaching the state $\mathbf{y}$ starting with the state $\mathbf{x}$ after exactly k time steps). It

is well-known (see, for instance, [30]) that irreducible Markov chains possess a unique *stationary distribution*, i.e. a probability distribution π on $\mathcal{X}$ such that $\forall\, \mathbf{y} \in \mathcal{X}$ we have $\pi(\mathbf{y}) = \sum_{\mathbf{x}\in\mathcal{X}} \pi(\mathbf{x}) p_{\mathbf{x}\to\mathbf{y}}$. [3] Moreover, the stationary distribution π is a global attractor of the Markov transition matrix in the sense that for every initial probability distribution ρ on $\mathcal{X}$ we have and for any $\mathbf{y} \in \mathcal{X}$ we have $\lim_{k\to\infty} \sum_{\mathbf{x}\in\mathcal{X}} \rho(\mathbf{x}) p^k_{\mathbf{x}\to\mathbf{y}} = \pi(\mathbf{y})$ where where $p^k_{\mathbf{x}\to\mathbf{y}}$ denotes the probability of reaching the state $\mathbf{y}$ starting with the state $\mathbf{x}$ after exactly k time steps. Finally, it is also worth reminding the reader that the ergodic theorem (theorem 1.10.2 of [30]) tells us the following important long-term behavior property. Let $N(\mathbf{x},\, t)$ denote the number of visits to state $\mathbf{x}$ before time t. Then $\lim_{t\to\infty} \frac{N(\mathbf{x},\, t)}{t} = \pi(\mathbf{x})$ i.e. the average number of occurrences of the state $\mathbf{x}$ in the long run is $\pi(\mathbf{x})$ where π is the unique stationary distribution of the Markov chain $\mathcal{M}$.

4.2 Construction of the Quotients.

Suppose we are given an equivalence relation $\sim$ partitioning the state space $\mathcal{X}$. The main idea of the current section is to construct an irreducible Markov chain over the equivalence classes under $\sim$ (i.e. over the set $\mathcal{X}/\sim$) whose stationary distribution is compatible with that of $\mathcal{M}$. This construction is a slight generalization of the construction in [2]:

Definition 10 Suppose $\mathcal{M}$ is an irreducible Markov chain over a finite state space $\mathcal{X}$ with transition matrix $\{p_{\mathbf{x}\to\mathbf{y}}\}$, π is the unique stationary distribution of the Markov chain $\mathcal{M}$, and $\sim$ is an equivalence relation on $\mathcal{X}$. Define the *quotient* Markov chain $\mathcal{M}/\sim$ over the state space $\mathcal{X}/\sim$ of equivalence classes via $\sim$ to be determined by the transition matrix $\{\tilde{p}_{\mathcal{U}\to\mathcal{V}}\}_{\mathcal{U},\,\mathcal{V}\in\mathcal{X}/\sim}$ given as

$$\tilde{p}_{\mathcal{U}\to\mathcal{V}} = \frac{1}{\pi(\mathcal{U})} \sum_{\mathbf{x}\in\mathcal{U}} \pi(\mathbf{x}) \cdot p_{\mathbf{x}\to\mathcal{V}} = \frac{1}{\pi(\mathcal{U})} \sum_{\mathbf{x}\in\mathcal{U}} \sum_{\mathbf{y}\in\mathcal{V}} \pi(\mathbf{x}) \cdot p_{\mathbf{x}\to\mathbf{y}}.$$

Here $p_{\mathbf{x}\to\mathcal{V}}$ denotes the transition probability of getting somewhere inside of $\mathcal{V}$ given $\mathbf{x}$. Since $\mathcal{V} = \bigcup_{y\in\mathcal{V}}\{y\}$ it follows that $p_{\mathbf{x}\to\mathcal{V}} = \sum_{y\in\mathcal{V}} p_{\mathbf{x}\to\mathbf{y}}$ and hence the equation above holds.

Intuitively, the quotient Markov chain $\mathcal{M}/\sim$ is obtained by running the original chain $\mathcal{M}$ starting with the stationary distribution and computing the transition probabilities conditioned with respect to the stationary input. If one starts with an arbitrary distribution and runs the process for a long period of time then the transition probabilities in definition 10 serve as a good approximation to the transition probabilities induced by the corresponding stochastic process. Thus, the following fact should not be a surprise:

Theorem 11 *Let π denote the stationary distribution of an irreducible Markov chain $\mathcal{M}$ determined by the transition matrix $\{p_{\mathbf{x}\to\mathbf{y}}\}_{\mathbf{x},\,\mathbf{y}\in\mathcal{X}}$. Suppose we are given an equivalence relation $\sim$ partitioning the state space $\mathcal{X}$. Then the quotient Markov chain $\mathcal{M}/\sim$ is irreducible and its unique stationary distribution $\tilde{\pi}$ is compatible with π in the sense that for every $\mathcal{O} \in \mathcal{X}/\sim$, we have $\tilde{\pi}(\{\mathcal{O}\}) = \pi(\mathcal{O})$.*

[3] It is worth pointing out straight away that we use a slight abuse of notation here (which is rather common in Markov chain theory): we write $\pi(\mathbf{x})$ in place of $\pi(\{\mathbf{x}\})$ (since π is a probability distribution its arguments are subsets of $\mathcal{X}$ rather than its individual elements).

Proof: Since the original chain $\mathcal{M}$ is assumed to be irreducible, it follows that there exists an $n \in \mathbb{N}$ such that for all $\mathbf{x}$, $\mathbf{y} \in \mathcal{X}$ we have $p^n_{\mathbf{x}\to\mathbf{y}} > 0$ where $p^n_{\mathbf{x}\to\mathbf{y}}$ denotes the probability that $\mathbf{y}$ is reached from $\mathbf{x}$ after exactly n time steps. This, in turn, is equivalent to saying that there exists a sequence of states $\mathbf{x}_1 = \mathbf{x}, \mathbf{x}_2, \ldots, \mathbf{x}_n = y$ such that $p_{\mathbf{x_i}\to\mathbf{x_{i+1}}} > 0$. Let $\mathcal{O}_i$ denote the equivalence class of x_i under $\sim$. Now we see that

$$\tilde{p}_{\mathcal{O}_i\to\mathcal{O}_{i+1}} = \frac{1}{\pi(\mathcal{O}_i)} \sum_{\mathbf{x}\in\mathcal{O}_i} \sum_{\mathbf{z}\in\mathcal{O}_{i+1}} \pi(\mathbf{x}) \cdot p_{\mathbf{x}\to\mathbf{z}} \geq \frac{1}{\pi(\mathcal{O}_i)} \cdot \pi(\mathbf{x}_i) \cdot p_{\mathbf{x}_i\to\mathbf{x}_{i+1}} > 0.$$

This shows that $\tilde{p}^n_{\mathcal{O}_1\to\mathcal{O}_n} > 0$. Since the equivalence classes are nonempty and the choices of $\mathbf{x}$ and $\mathbf{y}$ are arbitrary, it follows that $\tilde{p}^n_{\mathcal{U}\to\mathcal{V}} > 0 \ \forall \mathcal{U}, \mathcal{V} \in \mathcal{X}/\sim$. This shows that the Markov chain $\mathcal{M}/\sim$ is irreducible and, hence, has a unique stationary distribution $\tilde{\pi}$. The fact that $\tilde{\pi}(\{\mathcal{O}\}) = \pi(\mathcal{O})$ is the stationary distribution of $\mathcal{M}/\sim$ can now be verified by direct computation. Indeed, we obtain

$$\sum_{\mathcal{O}\in\mathcal{X}/\sim} \tilde{\pi}(\{\mathcal{O}\}) \cdot \tilde{p}_{\mathcal{O}\to\mathcal{U}} = \sum_{\mathcal{O}\in\mathcal{X}/\sim} \pi(\mathcal{O}) \cdot \frac{1}{\pi(\mathcal{O})} \sum_{\mathbf{x}\in\mathcal{O}} \sum_{\mathbf{z}\in\mathcal{U}} \pi(\mathbf{x}) \cdot p_{\mathbf{x}\to\mathbf{z}} =$$

$$= \sum_{\mathbf{x}\in\mathcal{X}} \sum_{\mathbf{z}\in\mathcal{U}} \pi(\mathbf{x}) \cdot p_{\mathbf{x}\to\mathbf{z}} = \sum_{\mathbf{z}\in\mathcal{U}} \sum_{\mathbf{x}\in\mathcal{X}} \pi(\mathbf{x}) \cdot p_{\mathbf{x}\to\mathbf{z}} \overset{\text{by stationarity of } \pi}{=} \sum_{\mathbf{z}\in\mathcal{U}} \pi(\mathbf{z}) = \pi(\mathcal{U}) = \tilde{\pi}(\{\mathcal{U}\}).$$

This establishes the stationarity of $\tilde{\pi}$ and theorem 11 now follows. □

5 What is the "Quotient Construction" Method and How Does it Work?

Theorem 11 lies in the heart of the "quotient construction method" which we develop in this paper. The next result, which is a simple consequence of theorem 11, will allow us to establish many bounds related to the stationary distribution of the Markov chains modeling evolutionary algorithms. For the applications one does not even have to worry about the quotient Markov chain "behind the scene" and therefore it is convenient to introduce the following definition:

Definition 12 Given an irreducible Markov chain $\mathcal{M}$ with state space $\mathcal{X}$, for any two subsets A and $B \subseteq \mathcal{X}$, we define $p_{A\to B} = \sum_{a\in A} \frac{\pi(a)}{\pi(A)} p_{a\to B}$ where $p_{a\to B} = \sum_{b\in B} p_{a\to b}$ and π is the unique stationary distribution of the Markov chain $\mathcal{M}$.

The next result in conjunction with proposition 15 are the main tools we are going to use for estimating the ratios of the stationary distribution values on various subsets of the state space of a Markov chain.

Proposition 13 *Let $\mathcal{M}$ denote an irreducible Markov chain on a finite state space $\mathcal{X}$. Let π denote its unique stationary distribution. Given any two disjoint nonempty subsets A and B, let $C = (A \cup B)^c$ (here U^c denotes the complement of U in $\mathcal{X}$). Then we have $\frac{\pi(A)}{\pi(B)} = \frac{p_{B\to A}}{p_{A\to A^c}} + \frac{\pi(C)}{\pi(B)} \cdot \frac{p_{C\to A}}{p_{A\to A^c}}$. In particular, when $A \cup B = \mathcal{X}$ we have $\frac{p_{B\to A}}{p_{A\to A^c}}$. Moreover, if $C \neq \emptyset$ (i.e. $A \cup B \neq \mathcal{X}$), we can also write*

$$\frac{\pi(A)}{\pi(B)} = \frac{\beta}{1-\alpha} \text{ where } \beta = \frac{p_{B\to A}}{p_{A\to A^c}} + \frac{p_{B\to C} \cdot p_{C\to A}}{p_{C\to C^c} \cdot p_{A\to A^c}} \text{ and } \alpha = \frac{p_{A\to C} \cdot p_{C\to A}}{p_{C\to C^c} \cdot p_{A\to A^c}}.$$

Proof: Consider the quotient Markov chain with 3 states A, B and C. According to the definition of a stationary distribution we have

$$\pi(A)p_{A\to A} + \pi(B)p_{B\to A} + \pi(C)p_{C\to A} = \pi(A)$$

which is equivalent to

$$(1 - p_{A\to A})\pi(A) = \pi(B)p_{B\to A} + \pi(C)p_{C\to A}.$$

Since the quotient Markov chain is irreducible we must have $\pi(B) > 0$. Observing that $1 - p_{A\to A} = p_{A\to A^c}$, we can also rewrite the equation above as

$$p_{A\to A^c}\frac{\pi(A)}{\pi(B)} = p_{B\to A} + \frac{\pi(C)}{\pi(B)}p_{C\to A}$$

which entails

$$\frac{\pi(A)}{\pi(B)} = \frac{p_{B\to A}}{p_{A\to A^c}} + \frac{\pi(C)}{\pi(B)} \cdot \frac{p_{C\to A}}{p_{A\to A^c}},$$

i.e. the first desired conclusion. It may be worth noting that $p_{A\to A^c} > 0$ due to the fact that our Markov chain is irreducible. When $A \cup B = \mathcal{X}$, we have $C = \emptyset$ so that $\pi(C) = 0$ and $\frac{\pi(A)}{\pi(B)} = \frac{p_{B\to A}}{p_{A\to A^c}}$. Now assume $C \neq \emptyset$. Then, interchanging the roles of A and C, we have already shown that

$$\frac{\pi(C)}{\pi(B)} = \frac{p_{B\to C}}{p_{C\to C^c}} + \frac{\pi(A)}{\pi(B)} \cdot \frac{p_{A\to C}}{p_{C\to C^c}}.$$

Plugging the equation for $\frac{\pi(C)}{\pi(B)}$ into the equation for $\frac{\pi(A)}{\pi(B)}$ we obtain

$$\frac{\pi(A)}{\pi(B)} = \frac{p_{B\to A}}{p_{A\to A^c}} + \left(\frac{p_{B\to C}}{p_{C\to C^c}} + \frac{\pi(A)}{\pi(B)} \cdot \frac{p_{A\to C}}{p_{C\to C^c}}\right) \cdot \frac{p_{C\to A}}{p_{A\to A^c}}.$$

Solving for $\frac{\pi(A)}{\pi(B)}$ then gives the desired conclusion. □

Corollary 14 *Let $\mathcal{M}$ denote an irreducible Markov chain on a finite state space $\mathcal{X}$. Let π denote its unique stationary distribution. Given any two disjoint nonempty subsets A and B, let $C = (A \cup B)^c$. Then we have $\frac{\pi(A)}{\pi(B)} \geq \frac{p_{B\to A}}{p_{A\to A^c}} + \frac{p_{B\to C} \cdot p_{C\to A}}{p_{C\to C^c} \cdot p_{A\to A^c}}$ and, in particular, $\frac{\pi(A)}{\pi(B)} \geq \frac{p_{B\to A}}{p_{A\to A^c}}$ and $\frac{\pi(A)}{\pi(B)} \geq \frac{p_{B\to C} \cdot p_{C\to A}}{p_{C\to C^c} \cdot p_{A\to A^c}}$.*

Proof: Since probabilities are always at most 1, we have $0 \leq \alpha \leq 1$ where α is as in corollary 13. It follows then that $0 \leq 1 - \alpha \leq 1$ which implies the desired conclusions via corollary 13. □

Unfortunately the transition probabilities $p_{A\to B}$ depend on the values of the stationary distribution which is what we are trying to estimate, but in some cases we can still obtain some bounds on these transition probabilities via the following observation:

Proposition 15 *Let $\mathcal{M}$ denote an irreducible Markov chain over the state space $\mathcal{X}$. Let X and $Y \subseteq \mathcal{X}$. Then*

$$\min_{x \in X} p_{x \to Y} \leq \tilde{p}_{X \to Y} \leq \max_{x \in X} p_{x \to Y}.$$

Proof:

$$\min_{x \in X} p_{x \to Y} = \frac{1}{\pi(X)} \sum_{x \in X} \pi(x) \min_{x \in X} p_{x \to Y} \leq \frac{1}{\pi(X)} \sum_{x \in X} \pi(x) p_{x \to Y} = \tilde{p}_{X \to Y} \leq$$

$$\leq \frac{1}{\pi(X)} \sum_{x \in X} \pi(x) \max_{x \in X} p_{x \to Y} = \max_{x \in X} p_{x \to Y}.$$

□

The quality of the bounds in proposition 15 depends largely on the discrepancy among the values $p_{x \to Y}$ for $x \in X$. The situation may be particularly disadvantageous provided $X = X_1 \cup X_2$ with $p_{x \to Y}$ being large for $x \in X_1$ and being small for $x \in X_2$. Luckily, if we know (or can estimate) the ratio of the stationary distribution values of X_1 and X_2, the situation can be repaired via the following observation:

Corollary 16 *Let $\mathcal{M}$ denote an irreducible Markov chain over the state space $\mathcal{X}$. Let X and $Y \subseteq \mathcal{X}$. Suppose $X = X_1 \cup X_2$ where $X_1 \cap X_2 = \emptyset$. Then $p_{X \to Y} = \frac{\pi(X_1)}{\pi(X)} p_{X_1 \to Y} + \frac{\pi(X_2)}{\pi(X)} p_{X_2 \to Y}$. In particular,*

$$\frac{\pi(X_1)}{\pi(X)} \min_{x \in X_1} p_{x \to Y} + \frac{\pi(X_2)}{\pi(X)} \min_{x \in X_2} p_{x \to Y} \leq p_{X \to Y} \leq$$

$$\leq \frac{\pi(X_1)}{\pi(X)} \max_{x \in X_1} p_{x \to Y} + \frac{\pi(X_2)}{\pi(X)} \max_{x \in X_2} p_{x \to Y}.$$

Proof: According to definition 12, we can write

$$p_{X \to Y} = \frac{1}{\pi(X)} \sum_{x \in X} \pi(x) p_{x \to Y} =$$

$$= \frac{1}{\pi(X)} \left(\sum_{x \in X_1} \pi(x) p_{x \to Y} \right) + \frac{1}{\pi(X)} \left(\sum_{x \in X_2} \pi(x) p_{x \to Y} \right) =$$

$$= \frac{1}{\pi(X)} \left(\sum_{x \in X_1} \pi(x) \cdot \frac{\pi(X_1)}{\pi(X_1)} p_{x \to Y} \right) + \frac{1}{\pi(X)} \left(\sum_{x \in X_2} \pi(x) \cdot \frac{\pi(X_2)}{\pi(X_2)} p_{x \to Y} \right) =$$

$$= \frac{\pi(X_1)}{\pi(X)} \left(\frac{1}{\pi(X_1)} \sum_{x \in X_1} \pi(x) p_{x \to Y} \right) + \frac{\pi(X_2)}{\pi(X)} \left(\frac{1}{\pi(X_2)} \sum_{x \in X_2} \pi(x) p_{x \to Y} \right) =$$

$$= \frac{\pi(X_1)}{\pi(X)} p_{X_1 \to Y} + \frac{\pi(X_2)}{\pi(X)} p_{X_2 \to Y}.$$

This shows that $p_{X\to Y} = \frac{\pi(X_1)}{\pi(X)} p_{X_1\to Y} + \frac{\pi(X_2)}{\pi(X)} p_{X_2\to Y}$. The last assertion follows by applying proposition 15 to each of the summands on the right hand side. □

In some situations an estimate for the ratio $\frac{\pi(X_1)}{\pi(X_2)}$ is more readily available than the ratios $\frac{\pi(X_1)}{\pi(X)}$ and $\frac{\pi(X_2)}{\pi(X)}$. Of course, all these are closely related: we can write, for instance, $\frac{\pi(X_1)}{\pi(X)} = \frac{\pi(X_1)}{\pi(X_1)+\pi(X_2)} = \frac{1}{1+\frac{\pi(X_2)}{\pi(X_1)}}$ and, likewise, $\frac{\pi(X_2)}{\pi(X)} = \frac{1}{1+\frac{\pi(X_1)}{\pi(X_2)}}$. These observations allow us to restate corollary 16 as follows.

Corollary 17 *Let $\mathcal{M}$ denote an irreducible Markov chain over the state space $\mathcal{X}$. Let X and $Y \subseteq \mathcal{X}$. Suppose $X = X_1 \cup X_2$ where $X_1 \cap X_2 = \emptyset$. Also let $\lambda = \frac{\pi(X_2)}{\pi(X_1)}$ Then $p_{X\to Y} = \frac{1}{1+\lambda} p_{X_1\to Y} + \frac{1}{1+\frac{1}{\lambda}} p_{X_2\to Y}$. In particular,*

$$\frac{1}{1+\lambda} \min_{x\in X_1} p_{x\to Y} + \frac{1}{1+\frac{1}{\lambda}} \min_{x\in X_2} p_{x\to Y} \leq p_{X\to Y} \leq$$

$$\leq \frac{1}{1+\lambda} \max_{x\in X_1} p_{x\to Y} + \frac{1}{1+\frac{1}{\lambda}} \max_{x\in X_2} p_{x\to Y}.$$

We now proceed applying these observations to compute the stationary distribution of the binary GA Markov chains with population size 2 in the limit of small mutation rate. We believe the method can be further extended to compute and estimate the stationary distribution of Markov chains for a wide class of EAs.

6 First Application: Constant Fitness inside Disjoint Schemata

In this section we consider the binary GA with population size $m = 2$, string length n and fitness function $f : \Omega = \{0, 1\}^n \to \{0, 1\}$ which takes exactly two distinct values. In particular, we assume that $\exists\, i$ with $1 \leq i \leq n$ such that $f(*, *, \ldots, *, 0, *, \ldots, *) = x$ while $f(*, *, \ldots, *, 1, *, \ldots, *) = y$ (here $(*, *, \ldots, *, a, *, \ldots, *)$ for $a = 0$ or 1 represents any individual which has allele a in the specified position and any allele, either 0 or 1, in the $*$ position) where the fixed alleles 0 and 1 appear in the i^{th} position. Since the population size is $m = 2$, the state space $\mathcal{X} = \Omega^2$ of the Markov chain modeling our binary GA consists of $2 \times n$ matrices. We can then partition the state space of this Markov chain into three pairwise disjoint subsets defined as follows:

$$A = \left\{ \begin{pmatrix} * * \ldots * 0 * \ldots * \\ * * \ldots * 0 * \ldots * \end{pmatrix} \right\},\ B = \left\{ \begin{pmatrix} * * \ldots * 1 * \ldots * \\ * * \ldots * 1 * \ldots * \end{pmatrix} \right\} \text{ and}$$

$$C = \left\{ \begin{pmatrix} * * \ldots * 1 * \ldots * \\ * * \ldots * 0 * \ldots * \end{pmatrix}, \begin{pmatrix} * * \ldots * 0 * \ldots * \\ * * \ldots * 1 * \ldots * \end{pmatrix} \right\}.$$

We aim to compute the ratio $\frac{\pi(A)}{\pi(B)}$ via proposition 13. To apply proposition 13 we need to compute the quotient transition probabilities involved in the formula via corollary 15. First, it is worthwhile to remind the reader that a classical binary GA cycles through

the three stages: selection, recombination and mutation. To make our calculations convenient, in this section we will assume that selection stage follows mutation stage. It will not matter for our calculation if recombination is the first or the last stage. In the previous works (see section IX of [26]) it has been shown that we can later drop the requirement on mutation preceding selection in the limit of small mutation rate.

Lemma 18 *Given A, B and C as above, we have $p_{B\to A} = 2\mu(1-\mu)\frac{x^2}{(x+y)^2} + \mu^2$.*

Proof. Fix any population $\mathbf{b} \in B$. In case recombination is the first stage, let $\mathbf{b}'$ denote the output upon completion of recombination of the population $\mathbf{b}$. Otherwise, let $\mathbf{b}' = \mathbf{b}$. Since recombination, as described in definition 7, can not get us out from either A, B or C, $\mathbf{b}' \in B$ with probability 1. Next we need to obtain a population in A upon completion of mutation followed by selection. The only possible ways to do so starting with $\mathbf{b}'$ are as follows:

Option 1: Obtain a population $\mathbf{b}'' \in C$ via mutation starting with $\mathbf{b}'$ first, and then use selection to obtain a population in A.

Option 2: Obtain a population in A right upon completion of mutation (and then one stays in A with probability 1).

The events described in options 1 and 2 are disjoint and, thereby, the transition probability $p_{\mathbf{b}\to A} = p(\text{Event described in option 1}) + p(\text{Event described in option 2})$. Obtaining a population $\mathbf{b}'' \in C$ via mutation starting with $\mathbf{b}'$ amounts to mutating exactly one of the i^{th} alleles of the individuals in $\mathbf{b}'$ (i.e. either mutating the i^{th} allele of the first individual and not mutating the i^{th} allele of the second one, or vise versa: not mutating the i^{th} allele of the first individual and mutating the i^{th} allele of the second one). Thereby, this is disjoint union of two events, each happening with probability $\mu(1-\mu)$ (see definition 9). Thus, the probability of obtaining $\mathbf{b}'' \in C$ starting with $\mathbf{b}'$ upon completion of mutation is $2\mu(1-\mu)$. Once this is done, we need to select the individual with the i^{th} allele 0 from the population $\mathbf{b}''$ twice to obtain a population in A. According to definition 4, this happens with probability $\frac{x^2}{(x+y)^2}$. Therefore we deduce that $p(\text{Event described in option 1}) = 2\mu(1-\mu)\frac{x^2}{(x+y)^2}$. Notice that option 2 requires mutating the i^{th} allele of both individuals in the population $\mathbf{b}''$. This happens with probability μ^2. Once mutation is complete, we obtain an individual in A and neither selection nor recombination can produce an individual outside of A with probability 1. This shows that $p(\text{Event described in option 2}) = \mu^2$ and we deduce that $p_{\mathbf{b}\to A} = p(\text{Event described in option 1}) + p(\text{Event described in option 2}) = 2\mu(1-\mu)\frac{x^2}{(x+y)^2} + \mu^2$. Since the choice of $\mathbf{b} \in B$ is arbitrary, the desired conclusion that $p_{B\to A} = 2\mu(1-\mu)\frac{x^2}{(x+y)^2} + \mu^2$ now follows via proposition 15.

The proofs of the remaining lemmas are analogous to that of lemma 18. We will then present only abbreviated arguments.

Lemma 19 *Given A, B and C as above, we have $p_{A\to A^c} = 2\mu(1-\mu)\frac{y^2+2xy}{(x+y)^2} + \mu^2$.*

Proof. Just as in the proof of lemma 18, in view of proposition 15, it suffices to show that $\forall\, \mathbf{a} \in A$ we have $p_{\mathbf{a}\to A^c} = 2\mu(1-\mu)\frac{y^2+2xy}{(x+y)^2} + \mu^2$. This argument, once again, is

entirely analogous to the one in the proof of lemma 18. There are two alternative paths to take here, either to mutate exactly one of the i^{th} alleles of the individuals comprising the input population (recall this is a population in A), which happens with probability $2\mu(1-\mu)$, and, afterwards, not to select the "non-mutant" twice, which happens with probability $1 - \frac{x^2}{(x+y)^2} = \frac{y^2+2xy}{(x+y)^2}$, or to mutate the i^{th} alleles of both individuals in the input population which happens with probability μ^2. This will result in a population from B and selection won't get us away from there with probability 1. Thus, the total probability $p_{\mathbf{a}\to A^c} = 2\mu(1-\mu)\frac{y^2+2xy}{(x+y)^2} + \mu^2$ which then entails $p_{A\to A^c} = 2\mu(1-\mu)\frac{y^2+2xy}{(x+y)^2} + \mu^2$ via proposition 15.

Lemma 20 *Given A, B and C as above, we have $p_{C\to C^c} = (1-\mu)^2\frac{x^2+y^2}{(x+y)^2} + O(\mu)$.*

Proof: Consider any $\mathbf{c} \in C$. Let

$$\mathbf{c}' = \begin{cases} \mathbf{c} & \text{if selection is the first stage.} \\ \text{Output after recombination with input } \mathbf{c} & \text{if recombination is the first stage.} \end{cases}$$

As we have already seen, $\mathbf{c}' \in C$. One way to obtain a population not in C is not to mutate the i^{th} allele of either individual in $\mathbf{c}'$ which happens with probability $(1-\mu)^2$ and, afterwards, to select either the first or the second individual twice into the final output population which happens with probability $\frac{x^2+y^2}{(x+y)^2}$. The only alternative path is to mutate the i^{th} allele of at least one of the individuals which happens with probability $O(\mu)$. The effect of selection later may only reduce the probability of getting away from C (but only if we mutate the i^{th} allele of both individuals).[4] We then conclude that the total probability $p_{\mathbf{c}\to C^c} = (1-\mu)^2\frac{x^2+y^2}{(x+y)^2} + O(\mu)$ and the desired conclusion that $p_{C\to C^c} = (1-\mu)^2\frac{x^2+y^2}{(x+y)^2} + O(\mu)$ now follows via proposition 15. □

Lemma 21 *Given A, B and C as above, we have $p_{C\to A} = (1-\mu)^2\frac{x^2}{(x+y)^2} + O(\mu)$.*

Proof: The argument is analogous to that in the proof of lemma 20. The only difference is that during selection stage, rather than getting away *anywhere* from $\mathbf{c} \in C$ we actually need to get specifically into A and this requires selecting the individual with fitness x twice (rather than selecting wither one of the individuals twice as in the proof of lemma 20), hence the difference in the final formula. □

Lemma 22 *Given A, B and C as above, we have $p_{B\to C} = p_{A\to C} = 2\mu(1-\mu)\frac{2xy}{(x+y)^2}$.*

Proof: Consider any $\mathbf{b} \in B$. Let

$$\mathbf{b}' = \begin{cases} \mathbf{b} & \text{if selection is the first stage.} \\ \text{Output after recombination with input } \mathbf{b} & \text{if recombination is the first stage.} \end{cases}$$

[4] It is easy to compute this transition probability exactly as a function of μ, x, and y, but in the current paper we will only be concerned with the asymptotic results as $\mu \to 0$ and the current estimate is then sufficient.

As we have already seen, $\mathbf{b}' \in B$. The only way to obtain a population in C starting with $\mathbf{b}'$ is to mutate the i^{th} allele of exactly one of the individuals in $\mathbf{b}'$ (otherwise, if we don't mutate either of them we stay in B with probability 1 upon completion of selection, and if we mutate both, then we get into A and, once again, stay there with probability 1), which, as we have seen before, happens with probability $2\mu(1-\mu)$. Upon completion of mutation we must not get away from C via selection. As we have seen in the proof of lemma 20, getting away from any $\mathbf{c} \in C$ upon completion of selection happens with probability $\frac{x^2+y^2}{(x+y)^2}$ and, therefore, not getting away from $\mathbf{c} \in C$ happens with probability $1 - \frac{x^2+y^2}{(x+y)^2} = \frac{2xy}{(x+y)^2}$. We then deduce that $\forall\, \mathbf{b} \in B$ we have $p_{\mathbf{b}\to C} = 2\mu(1-\mu)\frac{2xy}{(x+y)^2}$ and proposition 15 then tells us that $p_{B\to C} = 2\mu(1-\mu)\frac{2xy}{(x+y)^2}$. The proof that $p_{A\to C} = 2\mu(1-\mu)\frac{2xy}{(x+y)^2}$ is entirely analogous. In fact, one can repeat the previous argument verbatim replacing A with B and B with A throughout. □

We are now in a position to obtain the main result of this section via proposition 13.

Theorem 23 *Given A, B, C and the fitness function f as above, let π_μ denote the unique stationary distribution of the GA Markov chain with mutation rate μ where selection stage follows mutation stage. Then* $\lim_{\mu\to 0} \frac{\pi_\mu(A)}{\pi_\mu(B)} = \left(\frac{x}{y}\right)^2$.

Proof: We simply plug the expressions from lemmas 18, 19, 22, 21 and 20 into the formula in proposition 13 to obtain:

$$\frac{\pi_\mu(A)}{\pi_\mu(B)} = \frac{\beta(\mu)}{1-\alpha(\mu)}$$

so that

$$\lim_{\mu\to 0} \frac{\pi_\mu(A)}{\pi_\mu(B)} = \frac{\lim_{\mu\to 0}\beta(\mu)}{1-\lim_{\mu\to 0}\alpha(\mu)}$$

where

$$\beta(\mu) = \frac{2\mu(1-\mu)\frac{x^2}{(x+y)^2} + \mu^2}{2\mu(1-\mu)\frac{y^2+2xy}{(x+y)^2} + \mu^2} +$$

$$+ \frac{\left(2\mu(1-\mu)\frac{2xy}{(x+y)^2}\right)\cdot\left((1-\mu)^2\frac{x^2}{(x+y)^2} + O(\mu)\right)}{\left((1-\mu)^2\frac{x^2+y^2}{(x+y)^2} + O(\mu)\right)\cdot\left(2\mu(1-\mu)\frac{y^2+2xy}{(x+y)^2} + \mu^2\right)}$$

and

$$\alpha(\mu) = \frac{\left(2\mu(1-\mu)\frac{2xy}{(x+y)^2}\right)\cdot\left((1-\mu)^2\frac{x^2}{(x+y)^2} + O(\mu)\right)}{\left((1-\mu)^2\frac{x^2+y^2}{(x+y)^2} + O(\mu)\right)\cdot\left(2\mu(1-\mu)\frac{y^2+2xy}{(x+y)^2} + \mu^2\right)}.$$

Taking the limit as $\mu \to 0$ we obtain

$$\lim_{\mu\to 0}\beta(\mu) = \frac{x^2}{y^2+2xy} + \frac{2x^3y}{(x^2+y^2)(y^2+2xy)} =$$

$$= \frac{x^2(x^2+y^2)+2x^3y}{(x^2+y^2)(y^2+2xy)} = \frac{x^2(x+y)^2}{(x^2+y^2)(y^2+2xy)}$$

and

$$\lim_{\mu\to 0}\alpha(\mu) = \frac{2x^3y}{(x^2+y^2)(y^2+2xy)}$$

so that

$$1-\lim_{\mu\to 0}\alpha(\mu) = 1-\frac{2x^3y}{(x^2+y^2)(y^2+2xy)} =$$

$$= \frac{(x^2+y^2)(y^2+2xy)-2x^3y}{(x^2+y^2)(y^2+2xy)} = \frac{y^2(x+y)^2}{(x^2+y^2)(y^2+2xy)}$$

and, finally,

$$\lim_{\mu\to 0}\frac{\pi_\mu(A)}{\pi_\mu(B)} = \frac{\lim_{\mu\to 0}\beta(\mu)}{1-\lim_{\mu\to 0}\alpha(\mu)} = \frac{\frac{x^2(x+y)^2}{(x^2+y^2)(y^2+2xy)}}{\frac{y^2(x+y)^2}{(x^2+y^2)(y^2+2xy)}} = \frac{x^2}{y^2} = \left(\frac{x}{y}\right)^2$$

as claimed. □

7 Tight and Rigorous Asymptotic Results for Binary GAs

In this section we consider an arbitrary binary genetic algorithm with string length n arbitrary population size m (not necessarily $m = 2$). We will also assume that the fitness function $f : \Omega = \{0, 1\}^n \to (0, \infty)$ is independent of the mutation rate μ. In previous papers (see [25] and [26]) we have already shown that the *uniform* populations (i.e. these populations which consist of a repeated copy of the same individual only) dominate over the nonuniform ones in the limit of small mutation rate. Furthermore, we have also estimated the rate of domination to be on the order of $\frac{1}{\mu}$ (i.e. if U denotes the set of uniform populations then $\frac{\pi_\mu(U)}{\pi_\mu(U^c)} = \Omega\left(\frac{1}{\mu}\right)$). In this section, apart from this asymptotic result, we will establish another one of similar nature. Although the results in the current section are only asymptotic (rather than evaluating specific stationary distributions) we strongly believe that they will serve as an indispensable tool in reducing the size of the state space and simplifying the nature of the Markov transition matrices for a wide class of EAs (hopefully more on this in the forthcoming papers). Recall that the state space of our Markov Chain consists of $m \times n$ binary matrices where the rows are the individuals. Another important notational concept is that of Holland schemata. Formally, Holland schemata are the elements of the set $\{0, 1, *\}^n$. Each Holland schema $\overrightarrow{t} = (t_1, t_2, \ldots, t_n) \in \{0, 1, *\}^n$ represents the set of individuals $S(\overrightarrow{t}) = \{(a_1, a_2, \ldots, a_n) \,|\, a_i = t_i \text{ if } t_i \neq *\} \subseteq \Omega = \{0, 1\}^n$. We will often abuse the notation by identifying the schema $\overrightarrow{t}$ with either the entire set of individuals it represents or with any of the individuals it represents. In fact, we have already made such informal use of Holland schemata in section 6. Fix any two specified gene indices i and j with $1 \leq i < j \leq n$ and consider the 4 Holland schemata with fixed positions i and j partitioning the search space $\Omega = \{0, 1\}^n$:

$$\overrightarrow{t}_1 = (*, *, \ldots, *, 0, *, \ldots, *, 0, *, \ldots, *),$$

$$\overrightarrow{t}_2 = (*, *, \ldots, *, 0, *, \ldots, *, 1, *, \ldots, *),$$
$$\overrightarrow{t}_3 = (*, *, \ldots, *, 1, *, \ldots, *, 0, *, \ldots, *),$$
$$\overrightarrow{t}_4 = (*, *, \ldots, *, 1, *, \ldots, *, 1, *, \ldots, *)$$

where non-$*$ entries (alleles) appear precisely at the i^{th} and j^{th} positions. Recall the notion of the hamming distance between the schemata (and/or the individuals): Given any two schemata $\overrightarrow{u} = (u_1, u_2, \ldots, u_n)$ and $\overrightarrow{v} = (v_1, v_2, \ldots, v_n) \in \{0, 1, *\}^n$ we define $d(\overrightarrow{u}, \overrightarrow{v}) = |\{l \,|\, u_l \neq v_l\}|$. For example, $d(\overrightarrow{t}_1, \overrightarrow{t}_2) = d(\overrightarrow{t}_1, \overrightarrow{t}_3) = 1$ while $d(\overrightarrow{t}_1, \overrightarrow{t}_4) = 2$. Now consider the following sets of populations: A_l denotes the set of all $m \times n$ matrices where every row fits the schema t_l. We let $A = \bigcup_{i=1}^{4} A_i$. Consider any q and l with $1 \leq q < l \leq 4$. Let $D_{q,l}$ denote the set of all $m \times n$ matrices such that every row of these fits either the schema t_q or the schema t_l (notice that t_l and t_q are disjoint for $l \neq q$ so that the "or" is implied to be a mutually exclusive one). Now let $C = \bigcup_{d(\overrightarrow{t}_q, \overrightarrow{t}_l)=1} D_{q,l} = D_{1,2} \cup D_{1,3} \cup D_{2,4} \cup D_{3,4}$. Finally, let B denote the set of remaining populations (the complement of $A \cup C$ in the set of all binary $m \times n$ matrices). Notice that every matrix in B contains at least one pair of rows such that these rows fit schemata hamming distance 2 apart (i.e. (t_1 and t_4) or (t_2 and t_3)). Notice also that recombination preserves the hamming distance in the sense that if we recombine a pair of individuals fitting the schemata t_1 and t_4 we may get either another pair fitting t_1 and t_4 or a pair fitting t_2 and t_3 (which remain hamming distance 2 apart). This implies, in particular, that B and C remain invariant under recombination. Needless to say, A also remains invariant under recombination. Furthermore, A also remains invariant under selection. Thus, if $\mathbf{a} \in A$ is any individual in A, then obtaining an individual outside of A requires at least one nontrivial mutation which happens with probability $\Theta(\mu)$. Moreover, performing exactly one nontrivial mutation (at either i^{th} or j^{th} position) will produce an individual inside of C. We will then stay inside of C with some constant (independent of the mutation rate) probability $\Theta(1)$. Thus, proposition 15 tells us that $p_{A\to C} = \Theta(\mu)$. Obtaining a population in B requires at least two nontrivial mutations and happens with probability $\Theta(\mu^2)$. Remaining inside of B happens with some constant (independent of the mutation rate) probability $\Theta(1)$. It follows then via proposition 15 that $p_{A\to B} = \Theta(\mu^2)$. Since $B \cap C = \emptyset$ and $B \cup C = A^c$, $p_{A\to A^c} = p_{A\to C} + p_{A\to B} = \Theta(\mu) + \Theta(\mu^2) = \Theta(\mu)$. Moreover, $p_{A\to C} = p_{A\to A^c} - p_{A\to B} = p_{A\to A^c} - \Theta(\mu^2)$. We summarize these observations below:

Lemma 24 *Given A, B and C as above, we have $p_{A\to A^c} = p_{A\to C} = \Theta(\mu)$ and $p_{A\to C} = p_{A\to A^c} - \Theta(\mu^2)$.*

Given any population $\mathbf{c} \in C$, or, likewise, a population $\mathbf{b} \in B$ getting from such population into A can be done via selection only (just select the same individual repeatedly) and doesn't require any nontrivial mutation. This certainly happens with probability $\Theta(1)$. Proposition 15 then tells us that $p_{C\to A} = \Theta(1)$ and $p_{B\to A} = \Theta(1)$. Moreover, getting from $\mathbf{c} \in C$ somewhere into B does require at least one nontrivial mutation and can then be done via selection only. This happens with probability $\Theta(\mu)$ then. Again, proposition 15 tells us that $p_{C\to B} = \Theta(\mu)$. Since $A \cap B = \emptyset$ and $C = (A \cup B)^c$, it follows that $p_{C\to C^c} = p_{C\to A} + p_{C\to B} = \Theta(1) + \Theta(\mu) = \Theta(1)$ and $p_{C\to C^c} - p_{C\to A} = p_{C\to B} = \Theta(\mu)$. These deductions are summarized below:

Lemma 25 *Given A, B and C as above, we have $p_{B\to A} = \Theta(1)$, $p_{C\to A} = \Theta(1)$ and $p_{C\to C^c} = \Theta(1)$. Also $p_{C\to A} = p_{C\to C^c} - \Theta(\mu)$.*

Finally, to get from somewhere in B to anywhere in C we don't have to perform any mutation and only need not to select some of the individuals into the new population. Such an event happens with probability $\Theta(\mu)$, hence the following:

Lemma 26 *Given A, B and C as above, we have $p_{B\to C} = \Theta(1)$.*

We are now in a position to deduce the central result of the current section from proposition 13.

Theorem 27 $\frac{\pi_\mu(A)}{\pi_\mu(B)} = \Theta\left(\frac{1}{\mu^2}\right)$

Proof: Much like in the proof of theorem 23, we simply plug in the constants in lemmas 24, 25 and 26 into proposition13 to obtain $\frac{\pi_\mu(A)}{\pi_\mu(B)} = \frac{\beta(\mu)}{1-\alpha(\mu)}$ where

$$\beta(\mu) = \frac{\Theta(1)}{\Theta(\mu)} + \frac{\Theta(1)\cdot\Theta(1)}{\Theta(1)\cdot\Theta(\mu)} = \Theta\left(\frac{1}{\mu}\right)$$

and

$$\alpha(\mu) = \frac{(p_{A\to A^c} - \Theta(\mu^2))\cdot(p_{C\to C^c} - \Theta(\mu))}{p_{C\to C^c}\cdot p_{A\to A^c}} =$$

$$= \frac{p_{C\to C^c} - \Theta(\mu)}{p_{C\to C^c}} \cdot \frac{p_{A\to A^c} - \Theta(\mu^2)}{p_{A\to A^c}} = \left(1 - \frac{\Theta(\mu)}{\Theta(1)}\right)\cdot\left(1 - \frac{\Theta(\mu^2)}{\Theta(\mu)}\right) =$$

$$= (1-\Theta(\mu))\cdot(1-\Theta(\mu)) = 1 - \Theta(\mu)$$

carefully noticing that both constants within $\Theta(\mu)$ notation are positive. We then obtain

$$\frac{\pi_\mu(A)}{\pi_\mu(B)} = \frac{\beta(\mu)}{1-\alpha(\mu)} = \frac{\Theta\left(\frac{1}{\mu}\right)}{1-(1-\Theta(\mu))} = \Theta\left(\frac{1}{\mu^2}\right)$$

as claimed. □

We can strengthen theorem 27 further but noticing that all the "uniform" populations have probability of the same order of magnitude asymptotically (as $\mu \to 0$) as long as the fitness function is independent of the mutation rate μ. In fact, this is very simple to prove using corollary 14.

Theorem 28 *Given any A_i and A_j as above (recall that A_i is the set of all $m \times n$ matrices which fit the schema t_i as described above), we have $\frac{\pi_\mu(A_i)}{\pi_\mu(A_j)} = \Theta(1)$.*

Proof: First we prove this fact for these i and j where $d(t_i, t_j) = 1$ (where d denotes the hamming distance as described above). For such i and j we apply corollary 14 to obtain $\frac{\pi_\mu(A_i)}{\pi_\mu(A_j)} = \frac{p_{A_j\to A_i}}{p_{A_i\to(A_i)^c}} \geq \frac{\Omega(\mu)}{O(\mu)} = \Omega(1)$. Indeed, getting from anywhere in A_j into somewhere in A_i requires at exactly one mutation, since $d(t_i, t_j) = 1$, and this

happens with probability $\Theta(\mu)$. Then we only need to select the mutated individual repeatedly which happens with probability $\Theta(1)$ so that, according to proposition 15 we have $p_{A_j \to A_i} = \Omega(1)$. Obtaining a population not in A_i starting with any given population in A_i requires at least one nontrivial mutation which happens with probability $O(\mu)$. Thus, we have shown that $\frac{\pi_\mu(A_i)}{\pi_\mu(A_j)} = \Omega(1)$ for these i and j where $d(t_i,\, t_j) = 1$. Interchanging the roles of i and j we deduce that $\frac{\pi_\mu(A_i)}{\pi_\mu(A_j)} = \Theta(1)$ for these i and j where $d(t_i,\, t_j) = 1$. For these i and j where $d(t_i,\, t_j) = 2$ select the "intermediate" k with $d(t_i,\, t_k) = 1$ and $d(t_j,\, t_k) = 1$ (it is easy to verify by inspection that such a k can always be found). We then have $\frac{\pi_\mu(A_i)}{\pi_\mu(A_j)} = \frac{\pi_\mu(A_i)}{\pi_\mu(A_k)} \cdot \frac{\pi_\mu(A_k)}{\pi_\mu(A_j)} = \Theta(1) \cdot \Theta(1) = \Theta(1)$ and the desired conclusion is established. □

Combining theorem 27 with theorem 28 gives us the following:

Corollary 29 *Given A_i for $1 \leq i \leq 4$, $A = \bigcup_{i=1}^{4} A_i$ and B as above, we have $\forall\, i$ with $1 \leq i \leq 4$ $\frac{\pi_\mu(A_i)}{\pi_\mu(A)} = \Theta(1)$ and $\frac{\pi_\mu(A_i)}{\pi_\mu(B)} = \Theta\left(\frac{1}{\mu^2}\right)$.*

Proof: For the first assertion notice that $1 \geq \frac{\pi_\mu(A_i)}{\pi_\mu(A)} \geq \frac{\pi_\mu(A_i)}{4 \max_{1 \leq i \leq 4} \pi_\mu(A_i)} = \frac{1}{4}\Theta(1) = \Theta(1)$. To see the last assertion, write

$$\frac{\pi_\mu(A_i)}{\pi_\mu(B)} = \frac{\pi_\mu(A_i)}{\pi_\mu(A)} \cdot \frac{\pi_\mu(A)}{\pi_\mu(B)} = \Theta(1) \cdot \Theta\left(\frac{1}{\mu^2}\right) = \Theta\left(\frac{1}{\mu^2}\right).$$

□

8 Unimodal Symmetric Functions on $\{0,\, 1\}^2$

In this section we consider a binary GA with string length 2, i.e. the search space is $\Omega = \{0,\, 1\}^2$ and population size 2. We consider a fitness function f of the form $f(0,\, 0) = f(1,\, 1) = x$ and $f(1,\, 0) = f(0,\, 1) = y$. This is a unimodal functions (i.e. a function which depends only on the number of 0s and 1s in its argument) with an extra symmetry condition that $f(0,\, 0) = f(1,\, 1) = x$. For this function we compute the stationary distribution of the corresponding GA Markov chain with a population of size 2 and selection stage following mutation stage within the cycle (the sequence of stages assumption can be dropped later) in the limit of small mutation rate. It may be worth mentioning straight away that the primary value of the results is the not so much the answer itself, since this is only a very special case, but the illustration of the "quotient construction" method which we hope will be further developed and improved to yield more powerful applications. We now consider the following partition of the state space (which consists of 16 populations each represented by a 2×2 binary matrix):

$$A = \left\{ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\},$$

$$B' = \left\{ \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \right\}, B'' = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \right\} \text{ and}$$

$$C = \left\{ \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \right\}.$$

We let $B = B' \cup B''$. Our primary goal here is to compute the ratio $\lim_{\mu \to 0} \frac{\pi(A)}{\pi(B')} = \lim_{\mu \to 0} \frac{\pi(A)}{\pi(B)}$ (as we will see later). As our eventual goal is to apply proposition 13 to the sets of populations A, B and C, we need to evaluate the transition probabilities involved in the formula. Unlike the case in section 6, due to the discrepancies in the transition probability values (getting away from various individuals within the same set B), two of the evaluations will require the use of corollary 17 and theorem 27 will provide us with the necessary estimates of the stationary distribution ratios within the partitions. All this will be illustrated below.

Lemma 30 *Suppose A, C, $B = B' \cup B''$, and the fitness function f are as above. Then* $p_{B \to A} = 4\mu(1-\mu)^3 \cdot \frac{1}{1+O(\mu^2)} \cdot \frac{x^2}{(x+y)^2} + O(\mu^2)$.

Proof: Let $\lambda = \frac{\pi(B'')}{\pi(B')}$. According to theorem 27, $\frac{1}{\lambda} = \Omega\left(\frac{1}{\mu^2}\right)$, which is equivalent to saying that $\lambda = O(\mu^2)$. According to corollary 17 we then have

$$p_{B \to A} = \frac{1}{1+O(\mu^2)} p_{B' \to A} + \frac{1}{1+\Omega\left(\frac{1}{\mu^2}\right)} p_{B'' \to A}.$$

Notice that

$$\frac{1}{1+\Omega\left(\frac{1}{\mu^2}\right)} \leq \frac{1}{1+\frac{q}{\mu^2}} = \frac{\mu^2}{q+\mu^2} \leq \frac{\mu^2}{q} = O(\mu^2)$$

for some constant $q > 0$ independent of μ. This shows that

$$\frac{1}{1+\Omega\left(\frac{1}{\mu^2}\right)} = O(\mu^2).$$

Since $p_{B'' \to A} \leq 1$, we conclude now that

$$p_{B \to A} = \frac{1}{1+O(\mu^2)} p_{B' \to A} + O(\mu^2).$$

It remains to examine $p_{B' \to A}$. If we start with any $\mathbf{b}' \in B'$ then we may perform exactly one mutation in any of the 4 positions which happens with probability $4\mu(1-\mu)^3$ and, afterwards, select the mutated individual into the new population repeatedly twice which happens with probability $\frac{x^2}{(x+y)^2}$. An alternative path requires at least 2 mutations and happens with probability $O(\mu^2)$. We then deduce that

$$p_{\mathbf{b}' \to A} = 4\mu(1-\mu)^3 \cdot \frac{x^2}{(x+y)^2} + O(\mu^2)$$

and proposition 15 then tells us that

$$p_{B' \to A} = 4\mu(1-\mu)^3 \cdot \frac{x^2}{(x+y)^2} + O(\mu^2).$$

Plugging this into the expression for $p_{B\to A}$ we finally obtain

$$p_{B\to A} = \frac{1}{1+O(\mu^2)}\left(4\mu(1-\mu)^3\cdot\frac{x^2}{(x+y)^2}+O(\mu^2)\right)+O(\mu^2) =$$

$$= 4\mu(1-\mu)^3\cdot\frac{1}{1+O(\mu^2)}\cdot\frac{x^2}{(x+y)^2}+O(\mu^2)$$

as claimed. □

The transition probability $p_{B\to C}$ is evaluated in an entirely analogous manner. We provide only a sketchy proof then.

Lemma 31 *Suppose A, C, $B = B' \cup B''$, and the fitness function f are as above. Then $p_{B\to C} = 4\mu(1-\mu)^3\cdot\frac{1}{1+O(\mu^2)}\cdot\frac{2xy}{(x+y)^2}+O(\mu^2)$.*

Proof: The argument goes along exactly the same lines as the proof of lemma 31. The only difference is that upon completion of selection (after a successful single mutation) we need to end up in C rather than in A. This is equivalent to not selecting each of the individuals from the resulting population (which is in C, by the way) twice and happens with probability $1-\frac{x^2+y^2}{(x+y)^2} = \frac{2xy}{(x+y)^2}$. □

Evaluating the remaining transition probabilities is truly routine and involves only proposition 15. We have presented so many of these arguments in the previous sections already that we leave the proofs of all the following lemmas as exercises for the interested reader.

Lemma 32 *Suppose A, C, $B = B' \cup B''$, and the fitness function f are as above. Then $p_{A\to A^c} = 4\mu(1-\mu)^3\cdot\frac{2xy+y^2}{(x+y)^2}+O(\mu^2)$.*

Lemma 33 *Suppose A, C, $B = B' \cup B''$, and the fitness function f are as above. Then $p_{A\to C} = 4\mu(1-\mu)^3\cdot\frac{2xy}{(x+y)^2}+O(\mu^2)$.*

Lemma 34 *Suppose A, C, $B = B' \cup B''$, and the fitness function f are as above. Then $p_{C\to A} = \frac{x^2}{(x+y)^2}+O(\mu)$.*

Lemma 35 *Suppose A, C, $B = B' \cup B''$, and the fitness function f are as above. Then $p_{C\to C^c} = \frac{x^2+y^2}{(x+y)^2}+O(\mu)$.*

We have now evaluated all the transition probabilities involved in proposition 13 and, thereby, set the stage for the following theorem:

Theorem 36 *Suppose A, C, $B = B' \cup B''$, and the fitness function f are as above. Then $\lim_{\mu\to 0}\frac{\pi_\mu(A)}{\pi_\mu(B)} = \left(\frac{x}{y}\right)^2$.*

Proof: Just as in the proofs of theorems 23 and 27 we simply plug in the expressions from lemmas 30, 31, 32, 34 and 35 into the formula in proposition 13 to obtain

$$\lim_{\mu\to 0}\frac{\pi_\mu(A)}{\pi_\mu(B)}=\lim_{\mu\to 0}\frac{\beta(\mu)}{1-\alpha(\mu)}=\frac{\lim_{\mu\to 0}\beta(\mu)}{1-\lim_{\mu\to 0}\alpha(\mu)}$$

where

$$\beta(\mu)=\frac{4\mu(1-\mu)^3\cdot\frac{1}{1+O(\mu^2)}\cdot\frac{x^2}{(x+y)^2}+O(\mu^2)}{4\mu(1-\mu)^3\cdot\frac{2xy+y^2}{(x+y)^2}+O(\mu^2)}+$$

$$+\frac{\left(4\mu(1-\mu)^3\cdot\frac{1}{1+O(\mu^2)}\cdot\frac{2xy}{(x+y)^2}+O(\mu^2)\right)\cdot\left(\frac{x^2}{(x+y)^2}+O(\mu)\right)}{\left(\frac{x^2+y^2}{(x+y)^2}+O(\mu)\right)\cdot\left(4\mu(1-\mu)^3\cdot\frac{2xy+y^2}{(x+y)^2}+O(\mu^2)\right)}$$

and

$$\alpha(\mu)=\frac{\left(4\mu(1-\mu)^3\cdot\frac{2xy}{(x+y)^2}+O(\mu^2)\right)\cdot\left(\frac{x^2}{(x+y)^2}+O(\mu)\right)}{\left(\frac{x^2+y^2}{(x+y)^2}+O(\mu)\right)\cdot\left(4\mu(1-\mu)^3\cdot\frac{2xy+y^2}{(x+y)^2}+O(\mu^2)\right)}.$$

Taking limit as $\mu\to 0$ we compute

$$\lim_{\mu\to 0}\beta(\mu)=\frac{x^2}{y^2+2xy}+\frac{2x^3y}{(x^2+y^2)(y^2+2xy)}=$$

$$=\frac{x^2(x^2+y^2)+2x^3y}{(x^2+y^2)(y^2+2xy)}=\frac{x^2(x+y)^2}{(x^2+y^2)(y^2+2xy)}$$

and

$$\lim_{\mu\to 0}\alpha(\mu)=\frac{2x^3y}{(x^2+y^2)(y^2+2xy)}$$

so that

$$1-\lim_{\mu\to 0}\alpha(\mu)=1-\frac{2x^3y}{(x^2+y^2)(y^2+2xy)}=$$

$$=\frac{(x^2+y^2)(y^2+2xy)-2x^3y}{(x^2+y^2)(y^2+2xy)}=\frac{y^2(x+y)^2}{(x^2+y^2)(y^2+2xy)}$$

and, finally,

$$\lim_{\mu\to 0}\frac{\pi_\mu(A)}{\pi_\mu(B)}=\frac{\lim_{\mu\to 0}\beta(\mu)}{1-\lim_{\mu\to 0}\alpha(\mu)}=\frac{\frac{x^2(x+y)^2}{(x^2+y^2)(y^2+2xy)}}{\frac{y^2(x+y)^2}{(x^2+y^2)(y^2+2xy)}}=\frac{x^2}{y^2}=\left(\frac{x}{y}\right)^2$$

as claimed. □

It's interesting to observe that the ratio in theorem 27 is the same as in theorem 23 (of course, the sets A and B and the fitness function are different).

9 Conclusions

In the current paper we have further strengthened the "quotient construction method" presented in earlier works. In particular we managed to use it to compute exact ratios of the stationary distribution values of GA Markov chains in a couple of simple examples (see sections 6 and 8). Furthermore, we established another, more subtle rigorous asymptotic result which can easily be extended to a wide class of EAs. We believe this result can serve as an important tool in reducing the size of the state space of the EA Markov chains in the limit of small mutation rates. Hopefully more on this topic will appear in sequel papers.

10 Acknowledgements

Boris Mitavskiy is very grateful to the organizers for their invitation to attend Dagstuhl-2008 Evolutionary Computation Theory meeting. Boris Mitavskiy also wants to thank Dr. Jonathan Rowe for the support and encouragement to continue research in the field of Evolutionary Computing Theory.

References

1. Antonisse, J. (1989). A new interpretation of schema notation that overturns the binary encoding constraint. J. Schaffer *et al*, editors, *Procedings of the Third International Conference on Genetic Algorithms*, Morgan Kaufmann, pages 86-97.
2. Aldous, D. and Fill, J. (2001) Reversible Markov Chains and Random Walks on Graphs (unpublished book) http://www.stat.berkeley.edu/ aldous/RWG/book.html
3. Bennett, J. (1953) Panmixia with Tetrasomic and Hexasomic Inheritance. *Genetics*, 39: 150-158.
4. Buchholz, P. (1994) Exact and Ordinary Lumpability in Finite Markov Chains. *Jornal of Applied Probability*, 31(1): 59-75.
5. Coffey, S. (1999) An Applied Probabilist's Guide to Genetic Algorithms. *A Thesis Submitted to The University of Dublin for the degree of Master in Science*.
6. Davis, T. and Principe, J. (1991) A Simulated Annealing Like Convergence Theory for the Simple Genetic Algorithm. In R. Belew, and L. Bookers, editors, *Proceedings of the Fourth International Confernence on Genetic Algorithms*, pages 174-181, Morgan Kaufmann.
7. Dayer, T., Stewart, W. (1997). Quasi-Lumpability, Lower Bounding Coupling Matricies, and nearly Completely Decomposable Markov Chains. *SIAM Journal on Matirx Analysis and Applications*, 18(2): 482-498.
8. De Jong, K., Spears, W., Gordon, D. (1994). "Using Markov Chains to Analize GAFOs" *Proceedings of Foundations of Genetic Algorithms*, pages 115-137.
9. De Silva, C. and Suzuki, J. (2005). "On the Stationary Distribution of GAs with Fixed Crossover Probability". In H. Beyer and U. O'Reilly, editors, *Proceedings of GECCO 2005*, pages 1147-1151, ACM Press.
10. Ellison, B. (1966). Limit Theorems for Random Mating in Infinite Populations. *Jornal of Applied Probability*, 3: 94-114.
11. Geiringer, H. (1949). On some Mathematical Problems Arising in the Development of Mendelian Genetics. *Journal of the American Statistical Association*, 44(248): 526-547.

12. Hackstein J., Hochstenbach P. and Pearson P, (2000) Towards an Understanding of the Genetics of Human Male Infertility; Lesson from Flies. *Trends in Genetics*, 16: 568-572.
13. Haldane J. (1957) The Cost of Natural Selection. *Journal of Genetics*, 55: 511-524.
14. Jian-Long Xu, Jun-Min Wang, Ye-Qing Sun, Li-Jun Wei, Rong-Ting Luo, Ming-Xian Zhang and Zhi-Kang Li, (2006). Heavy Genetic Load Associated with Subspecific Differentiation of Japonica Rice (*Oryza sativa ssp.japonica L.*) *Journal of Experimental Botany*, 57: 2815-2824.
15. Iosifescu, M. (2007). Finite Markov Processes and Their Applications.*Dover Publications*.
16. Mac Lane, S. (1971) Categories for the Working Mathematician. *Graduate Texts in Mathematics 5*, Springer-Verlag.
17. McPhee, F. and Poli, R. (2001). A Schema Theory Analysis of the Evolution of Size in Genetic Programming with Linear Representations. *Genetic Programming, Proceedings of EuroGP*, pages 108-125.
18. McPhee, F. and Poli, R. (2002). Using Schema Theory to Explore Interactions of Multiple Operators. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pages 853-860.
19. Mitavskiy B. (2004). Crossover Invariant Subsets of the Search Space for Evolutionary Algorithms. *Evolutionary Computation*, 12(1): 19-46.
20. Mitavskiy, B. (2003). Comparing Evolutionary Computation Techniques via Their Representation. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, Vol. 1, pages 1196-1209 Springer Verlag.
21. Mitavskiy, B. (2003). A Category Theoretic Method for Comparing Evolutionary Computation Techniques via Their Representation. *Proceedings of the Eighth ESSLLI Student Session*, pages 201 - 210.
22. Mitavskiy, B., Rowe, J. (2006). An Extension of Geiringer Theorem for a wide class of evolutionary search algorithms. *Evolutionary Computation*, 14(1): 87-118.
23. Mitavskiy, B. (2005) "A schema-based version of Geiringer's Theorem for nonlinear Genetic Programming with homologous crossover." *Foundations of Genetic Algorithms 8*, pages 156-175, Springer-Verlag.
24. Mitavskiy, B., Rowe, J. (2006) Some Results about the Markov Chains Associated to GPs and General EAs. *Theoretical Computer Science*, 361(1): 72-110.
25. Mitavskiy, B., Rowe, J., Wright, A., Schmitt, L. (2006) Exploiting Quotients of Markov Chains to Derive Properties of the Stationary Distribution of the Markov Chain Associated to an Evolutionary Algorithm. *Simulated Evolution and Learning, 6th International Conference*, pages 726-733, Springer-Verlag.
26. Mitavskiy, B., Cannings, C. (2007) An Improvement of the "Quotient Construction" Method and Further Asymptotic Results on the Stationary Distribution of the Markov Chains Modeling Evolutionary Algorithms. *2007 IEEE Congress on Evolutionary Computation*, pages 2606-2613.
27. Moey, C., Rowe, J. (2004). "A reduced Markov model of GAs without the exact transition matrix." In Yao, X. et. al. eds., *Proceedings of PPSN VIII, LNCS 3242*. pages 72-80.
28. Moey, C., Rowe, J. (2004). Population Aggregation based on Fitness. *Natural Computing*, 3: 5-19.
29. Rabani, Y., Rabinovich, Y., Sinclair, A. (1995). "A Computational View of Population Genetics." In *Annual ACM Symposium on the Theory of Computing* pages 83-92.
30. Norris, J. (1997) Markov Chains. *Cambridge University Press.*
31. Rowe J., McPhee N. (2001). "The effects of crossover and mutation operators on variable length linear structures." In Spector,L., Goodman,A. at.al. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pages 535-542, Morgan Kaufmann.
32. Radcliffe, N. (1994). The algebra of genetic algorithms. *Annals of Mathematics and Artificial Intelligence*, 10:339-384.

33. Schmitt, L. (2001). Theory of Genetic Algorithms. *Theor. Computer Science*, 259: 1-61.
34. Schmitt, L. (2004). "Theory of Genetic Algorithms II: Models for Genetic Operators over the String-Tensor representation of Populations and Convergence to Global Optima for Arbitrary Fitness Function under Scaling." *Theoretical Computer Science*, 310: 181-231.
35. Spears, W., De Jong, K. (1996). "Analyzing GAs using Markov Chains with Semantically Ordered and Lumped States." In *Foumdations of Genetic Algorithms*, pages 85-100.
36. Suzuki, J. (1997). "A Further Result on the Markov Chain Model", In R. Belew and M. Vose, editors, *Foundations of Genetic Algorithms 4*, pages 53-72, Morgan Kaufmann.
37. Suzuki, J. (1998). A Further Result on the Markov Chain Model of Genetic Algorithm and Its application to Simulated Annealing-Like strategy. *IEEE Transactions on Systems, Man and Cybernatics*, 28(1): 95-102.
38. Vose, M. (1999). The simple genetic algorithm: foundations and theory. *MIT Press* .