How fast does the stationary distribution of the Markov chain modelling EAs concentrate on the homogeneous populations for small mutation rate?

Boris Mitavskiy and Jonathan Rowe

School of Computer Science, University of Birmingham, Birmingham B15 2TT, Great Britain
B.S.Mitavskiy@cs.bham.ac.uk

Abstract. One of the main difficulties faced when analyzing Markov chains modelling evolutionary algorithms is that their cardinality grows quite fast. A reasonable way to deal with this issue is to introduce an appropriate notion of a "coarse graining" or, in mathematical language, a quotient of such a Markov chain. The main topic of the current work is the construction of such a notion. We shell introduce a general 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, in some cases we can still exploit the quotient construction to deduce some properties of the stationary distribution of the original chain. As an example we shell establish some inequalities telling us how fast the stationary distribution of Markov chains modelling EAs concentrates on the homogenous populations as mutation rate converges to 0.

1 Introduction

One of the aspects of the theoretical analysis of the evolutionary algorithms is studying the properties of the Markov chains associated to these algorithms. Many research articles in the field of evolutionary computing have been devoted to this subject (see, for instance, [6], [7], [9] and [1] for a survey of known results and open questions). The main difficulty faced with this approach is 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 sates of this Markov chain for a population of size m is $|\Omega|^m$ which grows polynomially with respect to $|\Omega|$ and exponentially with respect to the population size m. In the current paper we introduce a construction which can be viewed as a "quotient" (or, according to a commonly accepted EA terminology, a "coarse graining") of a Markov chain with respect to an equivalence relation. This construction is applicable to all irreducible Markov chains (which is true of Markov chains modelling EAs with positive mutation rate: see, for instance, [9] or [1]). The "quotient chain" will be shown to be irreducible as well and its unique stationary distribution is coherent with that of the original chain. Although the transition probabilities of the quotient chain depend on the stationary distribution of the original chain (which is the subject of our investigation), we can still

Dagstuhl Seminar Proceedings 06061
Theory of Evolutionary Algorithms
http://drops.dagstuhl.de/opus/volltexte/2006/596

use the quotient construction to deduce some interesting properties of the stationary distribution of the quotient chain (and, hence, of the original chain as well). To illustrate this technique, we shell establish some inequalities related to how fast the stationary distribution of a Markov chain modelling an EA concentrates on the homogenous populations (populations consisting of the repeated copies of the same individual only) with respect to the decreasing mutation rate. It should be noted that an inequality analogous to the one in corollary 16 has been obtained in [7] using entirely different methods. To the best of the author's knowledge, the inequality in corollary 17 is completely new. This inequality establishes a connection between the rate of concentration of the stationary distribution of the Markov chain modelling EAs on the homogenous populations and the maximum expected waiting time to reach a homogenous population starting from any other population. Estimating such expectations for various recombination operators remains an open problem for now. We hope that the main contribution of the current paper are not so much the results on the rate of concentration of the stationary distribution on the homogenous populations but the innovative quotient of the Markov chain construction which allows us to deduce these results as simple corollaries for we hope to find other useful applications of this method.

2 Notation

Ω is a finite set, called a *search space*.

$f : \Omega \to (0, \infty)$ is a function, called a *fitness function*. The goal is to find a maximum of the function f.

$\mathcal{F}_q$ is a collection of q-ary operations on Ω. Intuitively $\mathcal{F}_q$ can be thought of as the collection of reproduction operators: some q parents produce one offspring. In nature often $q = 2$, for every child has two parents, but in the artificial setting there seems to be no special reason to assume that every child has no more than two parents. When $q = 1$, the family $\mathcal{F}_1$ can be thought of as asexual reproductions or mutations. The following definitions will be used in section 3 to describe the general evolutionary search algorithm. This approach makes it easy to state the Geiringer Theorem.

Definition 1 A population P of size m is simply an element of Ω^m. (Intuitively it is convenient to think of a population as a "column vector".) The diagonal elements of Ω^m are called homogenous populations (in other words, a population is homogenous if it is of the form $(x, x, \ldots, x)^T$ for some $x \in \Omega$).

Remark 2 There are 2 primary methods for representing populations: multi-sets and ordered multi-sets. Each has advantages, depending upon the particular analytical goals. Lothar Shmitt has published a number of papers which use the ordered multi-set representation to advantage (see, for instance, [6] and [7]). According to definition 1, in the current paper we continue the development of analysis based upon the presentation pioneered by Lothar Schmitt. The following example illustrates an aspect of the representation which the reader would do well to keep in mind:

Example 3 Let $\Omega = \{0, 1\}^3$. Consider the populations $\begin{pmatrix} 0\ 0\ 0 \\ 1\ 1\ 1 \\ 1\ 1\ 1 \end{pmatrix}$, $\begin{pmatrix} 1\ 1\ 1 \\ 0\ 0\ 0 \\ 1\ 1\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1\ 1\ 1 \\ 1\ 1\ 1 \\ 0\ 0\ 0 \end{pmatrix}$. According to definition 1 (the ordered multi-set model which is exploited in the current paper) these are distinct populations despite the fact that they represent the same population under the multi-set model.

An *elementary step* is a probabilistic rule which takes one population as an input and produces another population of the same size as an output. For example, the following elementary step corresponds to the fitness-proportional selection which has been studied in detail by Wright and Fisher (see [10] and [2]).

Definition 4 An elementary step of type 1 (alternatively, of type *selection*) takes a given population $P = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$ with $x_i \in \Omega$ as an input. The individuals of P are evaluated:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \begin{matrix} \to\ f(x_1) \\ \to\ f(x_2) \\ \vdots \quad \vdots \\ \to f(x_m) \end{matrix}$$

A new population

$$P' = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

is obtained where y_i's are chosen independently m times form the individuals of P and $y_i = x_j$ with probability $\frac{f(x_j)}{\Sigma_{l=1}^{m} f(x_l)}$.

In other words, all of the individuals of P' are among those of P, and the expectation of the number of occurrences of any individual of P in P' is proportional to the number of occurrences of that individual in P times the individual's fitness value. In particular, the fitter the individual is, the more copies of that individual are likely to be present in P'. On the other hand, the individuals having relatively small fitness value are not likely to enter into P' at all. This is designed to imitate the natural survival of the fittest principle.

Another two types of elementary steps involved in an evolutionary algorithm are recombination and mutation. Detailed descriptions and some theoretical properties of these elementary steps, such as finite-population versions of Geiringer theorem, have been introduced and studied in [3], [4] and [5]. For the purposes of the current paper we shell not be interested in the details of the elementary step of recombination. The only two features of recombination which are of interest to us will be the following "week" form of purity (see [8] for the original notion) and a rather week technical condition (closely

related to the notion of a non-annihilating Markov transition matrix) which ensures irreducibility of the corresponding Markov chain modelling the EA. Both of these notions are defined below:

Definition 5 We say that an elementary step s is pure if an output of a homogenous population (see definition 1) under the action of s is also homogenous with probability 1.

Definition 6 We say that an elementary step s is non-annihilating if $\forall\ \mathbf{y} \in \mathcal{X}\ \exists\ \mathbf{x} \in \mathcal{X}$ such that $\mathbf{y}$ can be obtained as an output of the elementary step s on the input $\mathbf{x}$ with positive probability.

Definition 7 Mutation is an elementary step s determined by the *mutation tuple* $(\mathcal{M}, \mu)$ where $\mathcal{M}$ is some family of functions from Ω into itself containing the identity map and μ is a probability distribution on $\mathcal{M}$ which assigns a positive probability to every element of $\mathcal{M}$. Given an input population $P = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$, the output of the population P under the elementary step s is the population $P = \begin{pmatrix} F_1(x_1) \\ F_2(x_2) \\ \vdots \\ F_m(x_m) \end{pmatrix}$ where the functions $F_i \in \mathcal{M}$ are sampled independently with respect to the probability distribution μ. We say that mutation elementary step is ergodic if $\forall x$ and $y \in \Omega\ \exists$ a transformation $F \in \mathcal{M}$ such that $F(x) = y$. Let $\mathbf{1} : \Omega \to \Omega$ denote the identity map. Then we shell say that $\epsilon = \mu(\mathcal{M} - \{\mathbf{1}\})$ is the mutation rate.

Definition 8 A cycle is a finite sequence of elementary steps, say $\{s_n\}_{n=1}^{j}$, which are either of type 1 or of type 2 and such that all of the steps in the sequence $\{s_n\}_{n=1}^{j}$ have the same underlying search space and the same arity of input/output.

Remark 9 Intuitively, these steps are linked together in such a way that the output of the step s_i is the input of the step s_{i+1}. This is why all of the steps in the same cycle must have the same underlying search space and the same arity of input/output (otherwise the input/output relationship does not make sense).

We are finally ready to describe a rather wide class of evolutionary heuristic search algorithms.

3 How Does a Heuristic Search Algorithm Work?

A general evolutionary search algorithm works as follows: Fix a *cycle*, say $C = \{s_n\}_{n=1}^{j}$ (see definition 8). Now start the algorithm with an initial population $P = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$ The

initial population P may be selected completely randomly, or it may also be predetermined depending on the circumstances. The actual method of selecting the initial population P is irrelevant for the purposes of the current paper. To run the algorithm with cycle $C = \{s_n\}$, simply input P into s_1, run s_1, then input the output of s_1 into s_2 ... input the output of s_{j-1} into s_j and produce the new output, say P'. Now use P' as an initial population and run the cycle C again. Continue this loop finitely many times depending on the circumstances.

4 The Markov Chain Associated to an Evolutionary Algorithm

In [9] it has been pointed out that heuristic search algorithms give rise to the following Markov process[1] (see also [1], for instance): The state space of this Markov process is the set of all populations of a fixed size m. This set, in our notation, is simply Ω^m. The transition probability $p_{\mathbf{xy}}$ is simply the probability that the population $\mathbf{y} \in \Omega^m$ is obtained from the population $\mathbf{x}$ by going through the cycle once (where the notion of a cycle is described in section 3: see definition 8 and remark 9). The aim of the current paper is to continue the investigation of the properties of the stationary distribution of the Markov chain modelling EAs initiated in [5]. In the current paper we study the rate of concentration of the stationary distribution of the Markov chain modelling EAs on the homogenous populations.

5 Quotients of irreducible Markov Chains

Throughout the current section we shell be dealing with an irreducible Markov chain $\mathcal{M}$ over a finite state space $\mathcal{X}$. $\{p_{\mathbf{x},\mathbf{y}}\}$ denotes the markov transition Matrix with the convention that $p_{\mathbf{x},\mathbf{y}}$ is the probability of getting y in the next stage given x. 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 (site later):

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

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

where $p_{\mathbf{x},\mathcal{V}}$ 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 $p_{\mathbf{x},\mathcal{V}} = \sum_{y\in\mathcal{V}} p_{\mathbf{x},\mathbf{y}}$ and hence the equation above holds.

[1] In the current paper we use the ordered multi-set representation.

Intuitively, the quotient Markov chain $M/\sim$ is obtained by running the original chain 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. The following fact should not appear surprising then:

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

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 $p^n_{\mathbf{x},\,\mathbf{y}}>0$ where $p^n_{\mathbf{x},\,\mathbf{y}}$ 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 $p_{\mathbf{x_i},\,\mathbf{x_{i+1}}}>0$. Let $\mathcal{O}_i$ denote the equivalence class of x_i under $\sim$. Now we see that

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

This shows that $\tilde{p}^n_{\mathcal{O}_1,\,\mathcal{O}_n}>0$. Since the equivalence classes are nonempty and and the choices of $\mathbf{x}$ and $\mathbf{y}$ are arbitrary, it follows that $\tilde{p}^n_{\mathcal{U},\,\mathcal{V}}>0\ \forall\mathcal{U},\,\mathcal{V}\in\mathcal{X}/\sim$. This shows that the Markov chain $M/\sim$ 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 $M/\sim$ can now be verified by direct computation. Indeed, we obtain

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

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

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

Although theorem 11 is rather elementary it allows us to make useful observations of the following type:

Corollary 12 *Suppose we are given an irreducible Markov chain $\mathcal{M}$ over the state space $\mathcal{X}$, and let $\mathcal{X}=A\cup B$ with $A\cap B=\emptyset$. Suppose for every $a\in A$ we have $p_{a,\,B}<\epsilon$ while for every $b\in B$ $p_{b,\,A}>\kappa$. Then, if π denotes the unique stationary distribution of $\mathcal{M}$, we have $\pi(B)<\frac{\epsilon}{\kappa}$ and $\pi(A)>1-\frac{\epsilon}{\kappa}$.*

Proof. Let $\sim$ denote the equivalence relation corresponding to the partition $\{A,\,B\}$ of $\mathcal{X}$. Now consider the Markov chain $\mathcal{M}/\sim$. This is a Markov chain determined by the 2×2 transition matrix

$$\begin{pmatrix}p_{A,A} & p_{B,\,A}\\ p_{A,B} & p_{B,\,B}\end{pmatrix}$$

where $p_{A,B} = \frac{1}{\pi(A)} \sum_{a \in A} \pi(a) \cdot p_{a,\,B} < \frac{1}{\pi(A)} \sum_{a \in A} \pi(a) \cdot \epsilon = \epsilon$ and, likewise, $p_{B,A} = \frac{1}{\pi(B)} \sum_{b \in B} \pi(b) \cdot p_{b,\,A} > \frac{1}{\pi(B)} \sum_{b \in B} \pi(b) \cdot \kappa = \kappa$. According to theorem 11, Markov chain $\mathcal{M}/\sim$ is irreducible and its unique stationary distribution $\tilde{\pi}$ satisfies $\tilde{\pi}(\{A\}) = \pi(A)$ and $\tilde{\pi}(\{B\}) = \pi(B)$. Moreover, by direct computation it is easy to see that the stationary distribution of the Markov chain determined by the 2×2 transition matrix above (i.e. of the Markov chain $\mathcal{M}/\sim$) is $\tilde{\pi}(\{A\}) = \frac{p_{B,A}}{p_{A,B}+p_{B,A}}$ and $\tilde{\pi}(\{B\}) = \frac{p_{A,B}}{p_{A,B}+p_{B,A}}$. We finally obtain the desired inequalities $\pi(B) = \tilde{\pi}(\{B\}) = \frac{p_{A,B}}{p_{A,B}+p_{B,A}} < \frac{\epsilon}{\kappa}$ and $\pi(A) = 1 - \pi(B) > 1 - \frac{\epsilon}{\kappa}$

Corollary 12 can be somewhat strengthened by observing that any power of a Markov transition matrix determining an irreducible Markov chain also determines an irreducible Markov chain having the same stationary distribution as the original one. Applying corollary 12 to every power of a Markov transition matrix then gives us the following fact:

Corollary 13 *Suppose we are given an irreducible Markov chain $\mathcal{M}$ over the state space $\mathcal{X}$, and let $\mathcal{X} = A \cup B$ with $A \cap B = \emptyset$. Suppose for every $a \in A$ we have $p^n_{a,\,B} < \epsilon_n$ while for every $b \in B$ $p^n_{b,\,A} > \kappa_n$. Then, if π denotes the unique stationary distribution of $\mathcal{M}$, we have $\pi(B) < \inf\{\frac{\epsilon_n}{\kappa_n} \mid n \in \mathbb{N}\}$ and $\pi(A) > 1 - \inf\{\frac{\epsilon_n}{\kappa_n} \mid n \in \mathbb{N}\}$.*

6 Main Results:

Corollary 13 readily implies some basic observations about the rate of concentration of the stationary distribution of Markov chains modelling EAs. Our bound applies to a rather wide class of EAs described in the definition below:

Definition 14 Let $\mathcal{A}$ denote an evolutionary algorithm determined by a cycle $C = \{s_n\}_{n=1}^{j}$. We say that $\mathcal{A}$ is ergodic if there exists an i such that s_i is the elementary step of ergodic mutation and s_k is non-annihilating for every k with $1 \leq k \leq j$ (see definitions 6 and 7). Moreover, if, in addition, all but one of the elementary steps in C are pure (see definition 5), we say that $\mathcal{A}$ is standard.

First of all, we need to establish the irreducibility of the Markov chain modelling EAs in the framework of the current paper:

Proposition 15 *Suppose we are given an ergodic algorithm $\mathcal{A}$ (see definition 14 above). Then the Markov chain modelling the algorithm $\mathcal{A}$ (see section 4) is irreducible.*

Proof. Let $C = \{s_n\}_{n=1}^{j}$ denote the cycle determining the algorithm $\mathcal{A}$. Suppose, s_i is the elementary step defining mutation. We shell prove that the Markov transition matrix associated to the algorithm $\mathcal{A}$ has all positive entries (which suffices for irreducibility). Indeed, fix populations $\mathbf{x}$ and $\mathbf{y} \in \Omega^m$. Construct a sequence of states (populations) $\mathbf{x} = \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_i$ as follows: given that $\mathbf{x}_l$ is already constructed, let $\mathbf{x}_{l+1}$ be any population such that $\mathbf{x}_{l+1}$ is obtained with positive probability as an output of the elementary step s_l on the input $\mathbf{x}_l$ (such a population exists since the total probability of getting at least somewhere is $1 > 0$). Now choose a sequence $\mathbf{y}_{i+1}, \mathbf{y}_{i+2}, \ldots, \mathbf{y}_n = \mathbf{y}$

recursively as follows: $\mathbf{y}_{i-1}$ is chosen so that the elementary step s_{i-1} produces $\mathbf{y}$ on the input $\mathbf{y}_{i-1}$ with positive probability. Such a $\mathbf{y}_{i-1}$ can be found thanks to the assumption that the EA $\mathcal{A}$ is standard (see definitions 14 and 6). Since the elementary step s_i is an ergodic mutation, it follows that the probability of getting from $\mathbf{x}_i$ to $\mathbf{y}_{i+1}$ upon the application of s_i is positive. It follows now that the probability of obtaining $\mathbf{y}$ from $\mathbf{x}$ upon the completion of the cycle C is at least as big as the product of positive numbers which is again positive and the desired conclusion now follows.

The main idea of what follows is to apply corollaries 12 and 13 to the subset $H \subseteq \Omega^m$ of the homogenous populations and its complement.

Corollary 16 *Suppose we are given a standard algorithm $\mathcal{A}$ with mutation rate ϵ. Let π denote the unique stationary distribution of the Markov chain associated to the algorithm $\mathcal{A}$. Then we have $\pi(H) \geq 1 - \frac{m^{m+1}}{(1-\epsilon)^m}\epsilon$.*

Proof. This is an immediate application of corollary 12. Indeed, the event of destroying a given homogenous population is equivalent to the event of applying the nonidentity transformation to either one of the elements of that population and so is a union of events happening with probability ϵ each. Hence the probability of destroying a given homogenous population is bounded above by $m\epsilon$. The probability of passing from a non-homogenous population to a homogenous one is at least as large as the probability of consecutive m independent drawings of the most fit individual. The probability of picking the most fit individual in a population is bounded below by the probability of picking a given individual form a population where all the individuals have the same fitness, which is $\frac{1}{m}$. Doing so consecutively and independently m times is $\frac{1}{m^m}$. Afterwards, with probability $(1-\epsilon)^m$ everyone stays the same. The desired equation now follows immediately from corollary 12

The bound in corollary 16 is a rather weak one. This is not too surprising the more so that it applies to a wide class of algorithms. One should be able to improve the bound in corollary 16 for specific types of algorithms using corollary 13 instead of corollary 12.

Corollary 17 *Suppose we are given a standard algorithm $\mathcal{A}$ with mutation rate ϵ. Let $T(\mathbf{x})$ denote the random variable measuring the number of steps it takes for an EA to reach a homogenous population starting with the population $\mathbf{x}$.Let π denote the unique stationary distribution of the Markov chain associated to the algorithm $\mathcal{A}$. Then we have $\pi(H) \geq 1 - \frac{2m \cdot \max_{\mathbf{x}\in\Omega^m - H} E(T(\mathbf{x}))}{(1-\epsilon)^{2m \max_{\mathbf{x}\in\Omega^m - H} E(T(\mathbf{x}))}}\epsilon$.*

Proof. First note that

$$\forall\, \mathbf{x} \in \Omega^m - H \; p^L_{\mathbf{x},H} \geq P(T(\mathbf{x}) < L) \cdot (1-\epsilon)^{Lm}.$$

By Markov inequality we have

$$P(T(\mathbf{x}) < L) \geq 1 - \frac{E(T(\mathbf{x}))}{L} \geq \frac{1}{2} \quad \text{for } L \geq 2E(T(\mathbf{x})).$$

We then deduce that

$$p^{2\max_{\mathbf{x}\in\Omega^m - H} E(T(\mathbf{x}))}_{\mathbf{y},H} \geq \frac{1}{2}(1-\epsilon)^{2m \max_{\mathbf{x}\in\Omega^m - H} E(T(\mathbf{x}))} \quad \forall\, \mathbf{y} \in \Omega^m - H.$$

Just like in the proof of corollary 16, we have

$$p_{\mathbf{y},\Omega^m-H}^{2\max_{\mathbf{x}\in\Omega^m-H}E(T(\mathbf{x}))} \leq m \cdot \max_{\mathbf{x}\in\Omega^m-H} E(T(\mathbf{x}))\epsilon$$

which finally gives $\pi(H) \geq 1 - \frac{2m\cdot\max_{\mathbf{x}\in\Omega^m-H}E(T(\mathbf{x}))}{(1-\epsilon)^{2m\max_{\mathbf{x}\in\Omega^m-H}E(T(\mathbf{x}))}}\epsilon$

7 Conclusions and Future Work:

In the current paper we constructed a quotient (or, in the language used by the evolutionary computation community, a "coarse graining") of an irreducible Markov chain with respect to an arbitrary equivalence relation. As an illustration of how this construction can be useful we established some inequalities which tell us haw fast the stationary distribution of a Markov chain modelling an EA concentrates on the homogenous populations (populations consisting of the repeated copies of a single individual only). It was shown (see corollary 12 and 13 that the stationary distribution value of the set of homogenous populations is bounded below by $1 - k\epsilon$ where ϵ is the mutation rate and k is a multiplicative constant depending on the population size. It should be noted that a similar result was obtained in [7] by completely different methods. Our method seems to show a little more though. Corollary 13 demonstrates that this bound is closely related to the maximum over all populations $\mathbf{x}$ of the expected waiting times until we reach a homogenous population starting with the population $\mathbf{x}$.) We hope that the quotient of Markov chains construction can be useful for other purposes as well.

References

1. 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*.
2. Fisher, R. A. 1930. The Genetical Theory of Natural Selection. *Oxford: Clarendon Press.*
3. Mitavskiy, B., Rowe, J. (to appear). An Extension of Geiringer Theorem for a wide class of evolutionary search algorithms. *Evolutionary Computation*.
4. 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.
5. Mitavskiy, B., Rowe, J. (accepted) Some Results about the Markov Chains Associated to GPs and General EAs. *Theoretical Computer Science*
6. Schmitt, L. (2001) "Theory of Genetic Algorithms." *Theoretical Computer Science*, 259: 1-61.
7. 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.
8. Radcliffe, N. (1994). The algebra of genetic algorithms. *Annals of Mathematics and Artificial Intelligence*, 10:339-384. http://users.breathemail.net/njr/papers/amai94.pdf
9. Vose, M. (1999). The simple genetic algorithm: foundations and theory. *MIT Press, Cambridge, Massachusetts*.
10. Wright, S. 1931. Evolution in Mendelian Populatinos. *Genetics* 16:97-159.