Dynamics of a Distance-Based Population Diversity Measure†

Allen L. Barker
Datafilter Consulting
Charlottesville, Virginia
alb@datafilter.com

W. N. Martin
Computer Science Department
University of Virginia
martin@virginia.edu

Abstract- We study a class of steady-state genetic algorithms where, at each time step, two parents are selected to produce a child which then replaces one member of the population at the next time step. We consider the finite-population case. A general crossover and mutation operation are defined, as well as a genomic distance between individuals. Certain specific properties are required to hold for such operations and distance functions, and we present examples of crossover operations, mutation operations, and distance functions which meet the requirements. We then define the sum over all pairwise population distances as a measure of the diversity of a population and consider the time evolution of the expected diversity of a population. We show conditions where, under uniform, independent selection of parents and the individual to be replaced, the expected diversity monotonically approaches a fixed point. For this case we calculate an explicit formula for the expected diversity at each time step.

1 Introduction

In this paper we consider the diversity of finite populations of individuals under a genetic algorithm. The type of genetic algorithm considered selects two parents at each time step, as well as an individual to be replaced. The population at the next time step consists of the same population except that the individual selected for replacement is replaced by a child of the two parents. This is an example of a *steady-state* genetic algorithm [Mit96]. Similar techniques should also be applicable to other definitions of genetic algorithms.

The diversity measure we define is based on a general, additive genomic distance between individuals. We analyze the dynamics of this diversity measure under the specific class of genetic algorithms considered. In certain cases, like uniform selection without mutation (to be defined explicitly later), the expected diversity measure is strictly decreasing, causing a population to converge toward a state where all individuals have identical genome sequences. The intuitive idea is that when parents mate and produce a child, the child is "close" in some sense to the parents. If the child then replaces a different individual, the diversity of the population will tend to decrease on average. When mutation of individuals is included in the uniform selection case the expected diversity approaches a fixed point. We calculate explicit expressions for the expected diversity measure under uniform selection which show this trend. Because of the generalized crossover operation used in the analysis, the results hold for all forms of crossover.

Section 2 describes the class of genetic algorithms we consider. Section 3 defines the general form of the crossover operation and provides some examples. Section 4 defines a class of additive distance functions between the genomic representations of individuals and presents some examples. Section 5 defines the mutation operation and a basic assumption about its interaction with the distance function in terms of expected distances. We give some examples where this assumption holds and present a basic lemma used in later results. Section 6 defines the diversity measure as the sum of all pairwise distances within a population, and Section 7 offers a recursive expression for the expected diversity of a population at time step k. In Section 8 we consider the dynamics of the diversity measure in the case of independent, uniform selection. We present some theorems for this case which explicitly characterize the expected diversity of a population over time. Section 9 contains some conclusions.

2 Algorithm Definition

In this section we define the specific class of algorithms to be analyzed. In general terms, the algorithm works on a population of individuals. An individual is represented by some string or vector, often representing a genome sequence or a problem instance. Each individual is also associated with a positive, scalar fitness value.

The algorithm begins with some initial population of individuals. At each successive discrete time instance two individuals are chosen from the population to be mated, and one individual is chosen for replacement. This selection is random, with selection probabilities in the general case determined by the fitness values of the individuals. A crossover operator is applied to the two individuals to be mated, producing a new individual (the offspring or child). The population at the following time step is created by replacing the individual chosen for removal with the child.

We denote the population at time k by

$$S^{(k)} = (s_1^{(k)}, \ldots, s_n^{(k)}),$$

where the population size $n \geq 1$ is fixed for all k. The initial population is defined as $S^{(0)}$. The fitness values associated with individuals at time k are written as

†This paper appeared at the Congress on Evolutionary Computation, July 2000.

$$F^{(k)} = (f_1^{(k)}, \ldots, f_n^{(k)}).$$

We will often drop the superscripts (k) when there is no danger of confusion.

The algorithm moves from time k to time $k+1$ by first selecting three integer indices between 1 and n, representing the parents to be mated and an individual to be replaced. The indices of the parents chosen at time k will be written as $m^{(k)}$ and $w^{(k)}$, and the index of the individual to be replaced will be written $r^{(k)}$. These indices are chosen according to a known probability function $p(m^{(k)}, w^{(k)}, r^{(k)}|F^{(k)})$.

We assume a generalized crossover operator Φ which takes two individuals and produces two new individuals. This function is typically random, but we write it as a deterministic function by having it take an additional random vector as an argument (possibly a random vector with both discrete and continuous elements). We typically write this additional random vector as $\phi^{(k)}$. We discuss the specific properties required for the crossover operation in the next section. Only one child produced by the crossover operation is selected, and we assume the choice is random with equal probabilities for each child. A 0-1 random variable $\gamma^{(k)}$ will be used to denote this choice. After an individual is selected it is mutated and replaces individual s_r.

To represent the mutation operation we define a mutation operator Ψ which takes an individual and returns a modified individual. Mutation is also a random operation, and so takes a random vector $\psi^{(k)}$ as an additional argument. We discuss the specific properties required for the mutation operation in a later section.

For now we have

$$(c_1^{(k)}, c_2^{(k)}) = \Phi(s_m^{(k)}, s_w^{(k)}, \phi^{(k)}) \quad (1)$$

$$s_r^{(k+1)} = \begin{cases} \Psi(c_1^{(k)}, \psi_1^{(k)}) & \text{if } \gamma^{(k)} = 0 \\ \Psi(c_2^{(k)}, \psi_2^{(k)}) & \text{if } \gamma^{(k)} = 1 \end{cases}, \quad (2)$$

where $c_1^{(k)}$ and $c_2^{(k)}$ are the individuals produced as the children of $s_m^{(k)}$ and $s_w^{(k)}$. We assume the random vectors $\phi^{(k)}$ and $\psi^{(k)}$ are all independently generated. Defining $s_i^{(k+1)} = s_i^{(k)}$ for $i \neq r$ completes the definition of the algorithm. Note that there is one crossover at each time step.

3 The Generalized Crossover Operator

We require the crossover operator Φ to have a certain form. As mentioned earlier, we take individuals to be strings or vectors of some sort of element, having a fixed length h. Letting a and b be individuals, we write $a = (a_1, \ldots, a_h)$ and $b = (b_1, \ldots, b_h)$. Note that here we are using subscripts for elements of individuals rather than for individuals in a population.

The crossover operator on parents a and b, producing children x and y, is written as

$$\Phi(a, b, \phi) = (x, y), \quad (3)$$

where ϕ is a random vector as described earlier. We require that either $x_i = a_i$ and $y_i = b_i$ or else that $x_i = b_i$ and $y_i = a_i$, for each i. Thus the operator chooses elements of the parents and assigns them to the children, though mixing them up.

Example 1: One-point crossover. In this case the random vector ϕ passed in specifies some crossover point q between 1 and n. We then take

$$x = (a_1, \ldots, a_q, b_{q+1}, \ldots, b_h)$$

and

$$y = (b_1, \ldots, b_q, a_{q+1}, \ldots, a_h).$$

Example 2: Uniform crossover. [Sys89] In this case the random vector ϕ passed in contains h binary random variables, each with equal probabilities for 0 and 1. (More generally, as long as ϕ is independent, these 0-1 variables can have arbitrary distribution functions while still meeting the requirements for the generalized crossover operation.) We then choose $x_i = a_i$ and $y_i = b_i$ if $\phi_i = 0$ and $x_i = b_i$ and $y_i = a_i$ if $\phi_i = 1$.

We note that a "combining operator" other than crossover, taking two individuals and producing a third, could also be used provided it interacts correctly with the distance functions to be introduced later. That is, provided the distance relations to be defined in Section 5 still hold.

4 An Additive Distance Function Between Individuals

We assume there is a distance function $d(u, v)$ defined between any individuals u and v. This distance function is required to have the following properties:

- Symmetric: $d(u, v) = d(v, u)$,
- Non-negative: $d(u, v) \geq 0$,
- Zero only for identical individuals: $d(u, v) = 0$ if and only if $u = v$,
- Additive in the sense that $d(u, v) = \sum_{i=1}^{h} d_i(u_i, v_i)$, where the d_i are arbitrary functions (which, with some restrictions, may also depend on k) such that the previous properties hold for them (i.e. they are symmetric, non-negative, and zero only for identical elements).

Some examples of common distance functions with these properties follow.

Example 1: Hamming distance. In this case we treat the individuals as strings. For simplicity we restrict the strings to be binary strings in this example, i.e., strings over the alphabet $\{0, 1\}$. We define $d_i(u_i, v_i) = 0$ if $u_i = v_i$ and $d_i(u_i, v_i) = 1$ otherwise.

Example 2: Squared Euclidean distance. In this case we have individuals which can be considered to be vectors of integers, rationals, or real numbers. We define $d(u, v) = \sum_{i=1}^{h}(u_i - v_i)^2$. That is, $d_i(u_i, v_i) = (u_i - v_i)^2$. Note that the Euclidean distance (based on the Euclidean or l_2 norm) does not meet the additive requirement. Distance based on the the l_1 norm, i.e., the sum of the absolute differences, does meet the requirements and provides another example. Note that for populations of individuals with strictly 0-1 valued (binary) elements the squared Euclidean distance is identical to the Hamming distance.

Example 3: Combined forms. Since the function d_i can depend on i one can treat part of an individual as a string and part as made up of real numbers, as in the two earlier examples.

Combining a distance function of this form with the generalized crossover operation defined above we obtain the following lemma. In words, the lemma states that the sum of the distances from an arbitrary individual to a pair of parents equals the sum of the distances from that arbitrary individual to the two children produced by a crossover operation on those parents.

Lemma 1 *Let*

$$\Phi(a, b, \phi) = (x, y)$$

be a crossover operator of the type defined in Section 3, and let d be a distance function with the properties above. Then

$$d(s, a) + d(s, b) = d(s, x) + d(s, y)$$

for any individual s. Also, $d(a, b) = d(x, y)$.

Proof:

$$\begin{aligned}
d(s,a) + d(s,b) &= \sum_{i=1}^{h} d_i(s_i, a_i) + \sum_{i=1}^{h} d_i(s_i, b_i) \\
&= \sum_{i=1}^{h} [d_i(s_i, a_i) + d_i(s_i, b_i)] \\
&= \sum_{i=1}^{h} [d_i(s_i, x_i) + d_i(s_i, y_i)] \\
&= d(s, x) + d(s, y).
\end{aligned}$$

The result $d(a, b) = d(x, y)$ is clear from the additive property of d and the fact that $d_i(a_i, b_i) = d_i(x_i, y_i) = d_i(y_i, x_i)$ due to the symmetry of the d_i.

Note that this lemma holds regardless of the properties of the random variable ϕ and allows us to write the summed distances from any individual to a pair of children in terms of the summed distances to the parents. See [Eng97] for some similar distance properties derived for binary-string individuals under the Hamming distance. These properties include $d(a, b) = d(x, y)$ as above, as well as equalities such as $d(x, y) = d(a, x) + d(a, y)$. This latter equality follows from Lemma 1 by setting $s = a$ and applying $d(a, b) = d(x, y)$.

5 The Mutation Operator

As defined previously, the mutation operator takes an individual and a random vector and produces another, modified, individual. That is, $s = \Psi(x, \psi)$ applies the mutation operation to an individual x and produces a new individual s. The mutation operator is also required to meet certain requirements defined with respect to the expected value of the distance function d. We define those requirements later in this section.

Before proceeding, we discuss the notation we use for expected values. We use the symbol E with a superscript indicating the probability distribution with respect to which the expectation is taken. Thus, for example,

$$E^{p(u,v|z)}\left[g(u, v, z)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} du\ dv\ p(u, v|z)\left[g(u, v, z)\right]$$

for an arbitrary function g. The integrals are replaced by summation for discrete random variables (which we deal with in this paper). This allows us to manipulate expectation operators in some cases as if we are multiplying densities, e.g., $E^{p(u,v)} = E^{p(v)}E^{p(u|v)}$ and so forth.[1] We will write the operator E without the superscript to denote the full expectation with respect to all random variables of its argument (resulting in a number rather than a random quantity). When the brackets are left off, the multiplicative term to the right of the expectation operator is taken as its argument.

We assume that any random argument ψ to the mutation operator is independent from all other random variables. The condition we require for the mutation operator is that

$$E^{p(\psi)}d(\Psi(u, \psi), v) = \epsilon_1 d(u, v) + \epsilon_2, \qquad (4)$$

where here u and v are arbitrary individuals and we assume ϵ_1 is positive and ϵ_2 is non-negative. In words, the expected distance from a mutated individual $\Psi(u, \psi)$ to some other individual v is assumed to equal a constant times the distance from u to v, plus another constant. The constants ϵ_1 and ϵ_2 are fixed for a given mutation operation Ψ and probability function $p(\psi)$.

Applying (4) twice, we also have the relations

$$\begin{aligned}
E^{p(\psi_1,\psi_2)}d(\ \Psi(u, \psi_1),\ \Psi(v, \psi_2)\ ) &= \epsilon_1^2 d(u, v) + \epsilon_1\epsilon_2 + \epsilon_2 \\
E^{p(\psi_1,\psi_2)}d(\ \Psi(\Psi(u, \psi_1), \psi_2),\ v\ ) &= \epsilon_1^2 d(u, v) + \epsilon_1\epsilon_2 + \epsilon_2
\end{aligned}$$

and so forth, making use of the independence of the two (or more) mutation operations.

[1] Note that the expectation operators do not commute in general, though they do for expectations with respect to independent variables. Note also that for conditional distributions like $E^{p(u,v,z)} = E^{p(v,z)}E^{p(u|v,z)}$ the $p(v, z)$ expectation must be to the left. One way to view this is just as an ordering of the integration or summation operations for evaluating the total expectation (of a function of several random variables or vectors), making use of conditional probability rules. One can also view this as successive expectations with certain variables "given," though independent variables are usually dropped from the conditioning on a probability function or density.

Some examples of mutation operators which meet this condition follow, corresponding to the examples of distances given in Section 4.

Example 1: Random changes in a string. Suppose we are using the Hamming distance between binary strings as our d function, as in Example 1 of Section 4. A mutation is assumed to change exactly one string element, chosen uniformly. The index of the element to be chosen is passed in as random variable ψ. We write this index as q to indicate that, in general, other random variables might be passed in via ψ, though the expectation in this example is taken only with respect to the index variable q. That is, q might more generally be a name for an element of a random vector ψ which could contain other random variables as well. A change of a string is assumed to always flip the bit value u_q from its previous value to $1 - u_q$. Then we have

$$
\begin{aligned}
&E^{p(\psi)} d(\Psi(u,\psi),v) \\
&\quad= E^{p(q)}[\,(\sum_{i=1}^{h} d_i(u_i,v_i)\,) - d_q(u_q,v_q) + d_q(1-u_q,v_q)\,] \\
&\quad= d(u,v) + E^{p(q)}[\,d_q(1-u_q,v_q) - d_q(u_q,v_q)\,] \\
&\quad= d(u,v) + E^{p(q)}[\,(1-d_q(u_q,v_q)) - d_q(u_q,v_q)\,] \\
&\quad= d(u,v) + \frac{1}{h}\sum_{q=1}^{h}(1-2d_q(u_q,v_q)) \\
&\quad= (1-\frac{2}{h})d(u,v) + 1.
\end{aligned}
$$

The case where more than one element is mutated can be obtained by repeatedly applying the operator, or else derived similarly if some number of distinct elements are required to be mutated.

Example 2: Additive, zero-mean random noise. Suppose our d function is the squared Euclidean distance, as in Example 2 of Section 4. Recall that in this case we treat the elements as numbers. The mutation operator for this example is defined to add independent, zero-mean random noise to exactly one element of the individual, i.e., to one component of the vector. The additive noise w is assumed to have finite variance σ^2. Thus the random vector ψ contains two random variables: an integral index q into the elements of the individual and the real-valued, zero-mean random noise w. We assume q and w are independent. Then we have

$$
\begin{aligned}
&E^{p(\psi)} d(\Psi(u,\psi),v) \\
&\quad= E^{p(q,w)}[\,(\sum_{i=1}^{h} d_i(u_i,v_i)\,) \\
&\qquad -d_q(u_q,v_q) + d_q(u_q+w,v_q)\,] \\
&\quad= d(u,v) + E^{p(q,w)}[\,d_q(u_q+w,v_q) - d_q(u_q,v_q)\,] \\
&\quad= d(u,v) + E^{p(q,w)}[\,((u_q-v_q)+w)^2 - (u_q-v_q)^2\,] \\
&\quad= d(u,v) + E^{p(q)}E^{p(w)}[\,2w(u_q-v_q)+w^2\,] \\
&\quad= d(u,v) + \sigma^2.
\end{aligned}
$$

Similar expressions can be obtained for the cases where more than one element is perturbed.

Using property (4) of the mutation operator and Lemma 1 we obtain the following lemma.

Lemma 2 *Let*

$$\Phi(a,b,\phi) = (x,y)$$

be a crossover operator of the type defined in Section 3, and let d be a distance function with the properties defined in Section 4. Then, with Ψ a mutation operator of the type defined in Section 5,

$$
\begin{aligned}
&E^{p(\psi_1,\psi_2)}[\,d(s,\Psi(x,\psi_1)) + d(s,\Psi(y,\psi_2))\,] \\
&\quad= \epsilon_1(d(s,x)+d(s,y)) + 2\epsilon_2 \\
&\quad= \epsilon_1(d(s,a)+d(s,b)) + 2\epsilon_2
\end{aligned}
$$

for any individual s. Also,

$$
\begin{aligned}
&E^{p(\psi_1,\psi_2)} d(\Psi(x,\psi_1),\Psi(y,\psi_2)) \\
&\quad= \epsilon_1^2 d(x,y) + \epsilon_1\epsilon_2 + \epsilon_2 \\
&\quad= \epsilon_1^2 d(a,b) + \epsilon_1\epsilon_2 + \epsilon_2
\end{aligned}
$$

Proof: The expected value is linear, so the expected value of the sum is the sum of the expected values. Applying requirement (4) to both terms in the sum, along with Lemma 1, gives the first two equalities. For the third equality, set $c = \Psi(y,\psi_2)$ and take the expectation with respect to ψ_1 (using the independence of the ψ_i and requirement (4)). Then substitute back in for c and take the expectation with respect to ψ_2. Apply Lemma 1 to get the final line.

Note that, again, the random vector ϕ from the crossover operation does not enter into the result. In words, Lemma 2 states that the expected sum of the distances from an individual to the two mutated children equals a constant times the sum of the distances to the parents plus another constant.

Lemma 2 is the basic property that the crossover operator, mutation operator, and distance measure between individuals must satisfy for the results of this paper to hold. Alternate definitions of these operations are possible provided this condition still holds (and provided the distance properties of Section 4 still hold).

6 A Measure of Population Diversity

We now construct a measure of population diversity from the previously defined distance function between individuals. We assume the distance function d has all the properties previously specified.

We define a diversity measure $L^{(k)}$ of a population at time k as the sum over all pairwise distances between individuals in the population. That is,

$$L^{(k)} \equiv \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} d(s_i^{(k)}, s_j^{(k)}) \quad (5)$$
$$= \frac{n(n-1)}{2}\bar{d}^{(k)}, \quad (6)$$

where $s_i^{(k)}$ and $s_j^{(k)}$ are individuals in population $S^{(k)}$ and $\bar{d}^{(k)}$ is the average of all distances (not including zero self-distances) at time k. The following lemma follows directly from the properties of the distance function.

Lemma 3 $L^{(k)} = 0$ *if and only if all the individuals in population* $S^{(k)}$ *are identical, and* $L^{(k)} > 0$ *otherwise.*

Proof: From the requirements of the distance function d we know $d(s_i, s_j) = 0$ if and only if $s_i = s_j$, and is positive otherwise. The lemma then follows directly from the definition of $L^{(k)}$.

Actually, for some distance functions, a weaker condition will guarantee that all individuals in a population are identical. For example, using the discrete-valued Hamming distance we know the minimum difference between unequal individuals is 1, independent of h, so in this case $L^{(k)} < 1$ guarantees that all the individuals are identical. The following lemma is then clear:

Lemma 4 *If* d *is the Hamming distance then*

$$p(\text{all individuals are identical at time } k) = p(L^{(k)} < 1).$$

Better and more general bounds, involving n, are possible.

7 The Expected Diversity

We now write an expression for $L^{(k+1)}$ in terms of $L^{(k)}$. From the definition of the algorithm and of $L^{(k)}$ it is clear that to get $L^{(k+1)}$ we need to subtract off all distances to the individual that was removed and then add in all the distances to the new child. Exactly one child is produced at each time step. Recall that m and w are the indices for the individuals chosen as parents and that r is the index for the individual selected for removal. Where the superscript is left off it will be assumed to be k.

Throughout this section we define the new (potential) children as c_1 and c_2. That is, letting $(x, y) = \Phi(s_m, s_w, \phi)$ then $(c_1, c_2) = (\Psi(x, \psi_1), \Psi(y, \psi_2))$. One of the two children is chosen, randomly with equal probabilities, to go in slot r in the next generation. We write the 0-1 random variable representing this choice as γ. Writing the expression out we obtain

$$\begin{aligned} L^{(k+1)} &= L^{(k)} \\ &- \sum_{i\neq r} d(s_r, s_i) \\ &+ \gamma\sum_{i\neq r} d(c_1, s_i) + (1-\gamma)\sum_{i\neq r} d(c_2, s_i). \end{aligned}$$

Note that we can choose to sum over the zero self-distances or not, but leaving out other indices is significant. For example, leaving out the $i = r$ terms in the sums of the third line corresponds to not including the distance from the new child to the individual chosen to be removed.

We next take the expected value of $L^{(k+1)}$. Note that functionally, in terms of random variables,

$$L^{(k+1)} = g(S^{(k)}, F^{(k)}, m^{(k)}, w^{(k)}, r^{(k)}, \phi^{(k)}, \psi_1^{(k)}, \psi_2^{(k)}, \gamma^{(k)})$$

for some function g, since we assume that n and all other variables are known constants. Thus to find the full expectation $E[L^{(k+1)}]$ we take the expectation

$$\begin{aligned} &E^{p(S,F,m,w,r,\phi,\psi_1,\psi_2,\gamma)} L^{(k+1)} \\ &\quad= E^{p(S,F)} E^{p(m,w,r|S,F)} E^{p(\phi)} E^{p(\psi_1,\psi_2)} E^{p(\gamma)} L^{(k+1)} \\ &\quad= E^{p(S,F)} E^{p(m,w,r|F)} E^{p(\psi_1,\psi_2)} E^{p(\gamma)} L^{(k+1)}, \end{aligned}$$

where we have left off the superscripts k. Note we have used the independence assumptions and the assumption that m, w, and r are selected based only on F. The expectation $E^{p(\phi)}$ is dropped because, as will become clear, using Lemma 2 to evaluate $E^{p(\psi_1,\psi_2)}$ eliminates the variable ϕ. We will evaluate these expectations in right-to-left order, applying the operator sequence to $L^{(k+1)}$. We first take the expectation with respect to γ. Since γ is a uniform, 0-1 random variable we obtain

$$\begin{aligned} E^{p(\gamma)} L^{(k+1)} &= L^{(k)} \\ &- \sum_{i\neq r} d(s_r, s_i) \\ &+ \frac{1}{2}\sum_{i\neq r} d(c_1, s_i) + \frac{1}{2}\sum_{i\neq r} d(c_2, s_i). \end{aligned}$$

We now take the expected value with respect to the mutation noise by applying Lemma 2 to the distances to the children in the third line of the equation above. We obtain an expression entirely in terms of distances between population members at time k (and independent of $\phi^{(k)}$):

$$\begin{aligned} &E^{p(\psi_1,\psi_2)} E^{p(\gamma)} L^{(k+1)} \\ &\quad= L^{(k)} \\ &\quad- \sum_{i\neq r} d(s_r, s_i) \\ &\quad+ \frac{1}{2}\sum_{i\neq r}[\epsilon_1 d(s_m, s_i) + \epsilon_1 d(s_w, s_i) + 2\epsilon_2]. \end{aligned}$$

It follows after some simplification that

$$\begin{aligned} E^{p(\psi_1,\psi_2)} E^{p(\gamma)} L^{(k+1)} &= L^{(k)} + (n-1)\epsilon_2 \\ &- \sum_{i\neq r} d(s_r, s_i) \\ &+ \frac{\epsilon_1}{2}\sum_{i\neq r} d(s_m, s_i) \\ &+ \frac{\epsilon_1}{2}\sum_{i\neq r} d(s_w, s_i). \end{aligned}$$

The expected value with respect to the mutation operations has now been evaluated. We next take the expected values with respect to the selection of parents and individual to be replaced.

Taking the expectation and rearranging the summations we obtain,

$$
\begin{aligned}
& E^{p(m,w,r|F)} E^{p(\psi_1,\psi_2)} E^{p(\gamma)} L^{(k+1)} \\
&\quad = \quad L^{(k)} + (n-1)\epsilon_2 \\
&\quad + \quad E^{p(m,w,r|F)}[\\
&\quad - \quad \sum_{i \neq r} d(s_r, s_i) \\
&\quad - \quad \frac{\epsilon_1}{2} d(s_m, s_r) \\
&\quad + \quad \frac{\epsilon_1}{2} \sum_i d(s_m, s_i) \\
&\quad - \quad \frac{\epsilon_1}{2} d(s_w, s_r) \\
&\quad + \quad \frac{\epsilon_1}{2} \sum_i d(s_w, s_i)].
\end{aligned}
$$

We now use the linearity of the expected value operation and write

$$
\begin{aligned}
& E^{p(m,w,r|F)} E^{p(\psi_1,\psi_2)} E^{p(\gamma)} L^{(k+1)} \\
&\quad = \quad L^{(k)} + (n-1)\epsilon_2 \qquad (7) \\
&\quad - \quad E^{p(r|F)} \sum_{i \neq r} d(s_r, s_i) \\
&\quad - \quad \frac{\epsilon_1}{2} E^{p(m,r|F)} d(s_m, s_r) \\
&\quad + \quad \frac{\epsilon_1}{2} E^{p(m|F)} \sum_i d(s_m, s_i) \\
&\quad - \quad \frac{\epsilon_1}{2} E^{p(w,r|F)} d(s_w, s_r) \\
&\quad + \quad \frac{\epsilon_1}{2} E^{p(w|F)} \sum_i d(s_w, s_i).
\end{aligned}
$$

Notice that the expected value operations sum over any variables not present in a term, resulting in expectations with respect to the marginal probability functions. Further analysis will require making some assumptions about the distribution of m, w, and r.

Note that nowhere have we assumed m, w, and r to be distinct. Thus expression (7) holds for any of the cases where $m = w$, $m = r$, $w = r$, or $m = w = r$. This is important because it allows for — though does not require — various assumptions of independence between m, w, and r (conditional on F). For example, one might assume that $p(m,w,r|F) = p(m|F)p(w|F)p(r|F)$. We use this assumption in the following section.

8 Independent, Uniform Selection

In this section we assume that m, w, and r are independent and uniformly selected. That is, we take

$$p(m,w,r) = p(m)p(w)p(r) = \frac{1}{n^3}. \qquad (8)$$

The algorithm is independent of the fitness values (or they are assumed to all be equal at each time k) and so dependence on $F^{(k)}$ is removed from the expected value.

The marginal distributions, e.g. $p(m)$ and $p(m,r)$, are equal to $1/n$ for single arguments and $1/n^2$ for two arguments. We can therefore evaluate expression (7) with these distributions and obtain

$$
\begin{aligned}
& E^{p(m,w,r)} E^{p(\psi_1,\psi_2)} E^{p(\gamma)} L^{(k+1)} \\
&\quad = \quad L^{(k)} + (n-1)\epsilon_2 \\
&\quad - \quad \sum_r \frac{1}{n} \sum_i d(s_r, s_i) \\
&\quad - \quad \frac{\epsilon_1}{2} \sum_m \sum_r \frac{1}{n^2} d(s_m, s_r) \\
&\quad + \quad \frac{\epsilon_1}{2} \sum_m \frac{1}{n} \sum_i d(s_m, s_i) \\
&\quad - \quad \frac{\epsilon_1}{2} \sum_w \sum_r \frac{1}{n^2} d(s_w, s_r) \\
&\quad + \quad \frac{\epsilon_1}{2} \sum_w \frac{1}{n} \sum_i d(s_w, s_i)].
\end{aligned}
$$

Simplifying and writing in terms of $L^{(k)}$ we obtain

$$
\begin{aligned}
E^{p(m,w,r)} E^{p(\psi_1,\psi_2)} E^{p(\gamma)} L^{(k+1)} \quad &= \quad L^{(k)} + (n-1)\epsilon_2 \\
&- \quad \frac{2L^{(k)}}{n} \\
&- \quad \frac{L^{(k)}\epsilon_1}{n^2} \\
&+ \quad \frac{L^{(k)}\epsilon_1}{n} \\
&- \quad \frac{L^{(k)}\epsilon_1}{n^2} \\
&+ \quad \frac{L^{(k)}\epsilon_1}{n}.
\end{aligned}
$$

Simplifying again we have

$$
\begin{aligned}
& E^{p(m,w,r)} E^{p(\psi_1,\psi_2)} E^{p(\gamma)} L^{(k+1)} \\
&\quad = \quad (1 - \frac{2}{n} + \frac{2\epsilon_1}{n} - \frac{2\epsilon_1}{n^2}) L^{(k)} + (n-1)\epsilon_2 \qquad (9) \\
&\quad = \quad [1 - \frac{2}{n^2}(n - n\epsilon_1 + \epsilon_1)] L^{(k)} + (n-1)\epsilon_2.
\end{aligned}
$$

To further simplify the equations we define

$$\delta_2 = (n-1)\epsilon_2 \qquad (10)$$

$$\delta_1 = 1 - \frac{2}{n^2}(n - n\epsilon_1 + \epsilon_1). \qquad (11)$$

We now take the final expectation. Because (9) is a function only of random variable $L^{(k)}$, we only need to take the expectation $E^{p(L)}$ rather than $E^{p(S)}$. This follows from (5), because $L^{(k)} = g(S^{(k)})$ for some function g. Therefore $E[L^{(k)}] = E[g(S^{(k)})]$, and so $E^{p(L)}[L^{(k)}] = E^{p(S)}[g(S^{(k)})]$. We thus obtain

$$\begin{aligned} E[L^{(k+1)}] &= E^{p(L)} E^{p(m,w,r)} E^{p(\psi_1,\psi_2)} E^{p(\gamma)} L^{(k+1)} \\ &= E^{p(L)}[\delta_1 L^{(k)} + \delta_2] \qquad (12) \\ &= \delta_1 E[L^{(k)}] + \delta_2. \end{aligned}$$

Summarizing in a theorem, we have

Theorem 1 *Let m, w, and r be selected uniformly and independently (with replacement) from $\{1, \ldots, n\}$. Then*

$$E[L^{(k+1)}] = \delta_1 E[L^{(k)}] + \delta_2 \qquad (13)$$

$$E[\bar{d}^{(k+1)}] = \delta_1 E[\bar{d}^{(k)}] + \frac{2\delta_2}{n(n-1)} \qquad (14)$$

Theorem 1 explicitly describes the change in expected diversity after one time step in terms of the parameters δ_1 and δ_2 and thus also in terms of the direct parameters ϵ_1 and ϵ_2 of the model. The theorem might be used to aid in defining these mutation parameters, perhaps dynamically, to obtain a desired expected change in diversity.

Now, consider the parameters δ_1 and δ_2. Since we assumed earlier that ϵ_2 is non-negative we have $\delta_2 \geq 0$ from (10). Also, since ϵ_1 is positive, we know from (11) that $\delta_1 > 0$ if $n \geq 2$. Because the diversity is trivially always 0 when $n = 1$, we will assume henceforth that $n \geq 2$. It is clear from (13) that $E[L^{(k)}]$ diverges if $\delta_1 > 1$. If $\delta_1 = 1$ then the diversity also diverges (unless $\delta_2 = 0$ also, in which case the diversity is constant). Therefore, we consider the cases where $\delta_1 \neq 1$ and where $0 < \delta_1 < 1$.

The next theorem gives a formula for computing the expected diversity at any time step, based only on the initial average or expected population diversity. It can also be applied to calculate future expected values at any time step when the current expected diversity is known or can be estimated. Note that the next four theorems follow from Theorem 1, and so similar results apply whenever the expected value of the diversity measure obeys the formulas in Theorem 1 (and $\delta_1 > 0$ and $\delta_2 \geq 0$). That is, the same formulas may hold for different models with a redefinition of δ_1 and δ_2.

Theorem 2 *Let m, w, and r be selected uniformly and independently (with replacement) from $\{1, \ldots, n\}$. Let $k \geq 0$, and $\delta_1 \neq 1$. Then*

$$E[L^{(k)}] = \delta_1^k E[L^{(0)}] + \delta_2 \frac{1 - \delta_1^k}{1 - \delta_1}. \qquad (15)$$

Proof: Solve equation (13) for $E[L^{(k)}]$. A recurrence relation of the form $a_k = ca_{k-1} + f(k)$, has a general solution for $k \geq 1$ given by $a_k = c^k a_0 + \sum_{j=1}^{k} c^{k-j} f(j)$ [MKB83]. Also, $\sum_{k=0}^{n} a^k = \frac{a^{n+1}-1}{a-1}$ for all real $a \neq 0$ and $a \neq 1$.

In the following theorem we consider the difference values between the successive expected diversity values, i.e., $E[L^{(k+1)}] - E[L^{(k)}]$.

Theorem 3 *Let m, w, and r be selected uniformly and independently (with replacement) from $\{1, \ldots, n\}$, and let $k \geq 0$. Then*

$$E[L^{(k+1)}] - E[L^{(k)}] = \delta_1^k \left(\delta_2 - E[L^{(0)}](1 - \delta_1)\right). \qquad (16)$$

Proof: Rewrite $E[L^{(k+1)}] - E[L^{(k)}]$ by applying (15) to each term; simplify.

Theorem 3 shows that the difference values have the form $\delta_1^k \cdot const$, where the sign of the constant depends on the initial expected diversity value but is not a function of k. Thus the sign of the difference values is unchanging for a given initial expected diversity.

We next present a theorem about the convergence of the trajectories of expected diversity.

Theorem 4 *Let m, w, and r be selected uniformly and independently (with replacement) from $\{1, \ldots, n\}$. Let $0 < \delta_1 < 1$. Then all trajectories of $E[L^{(k)}]$ converge monotonically to the fixed point*

$$\xi \equiv \frac{\delta_2}{1 - \delta_1}.$$

Proof: Take the limit of (15) as $k \to \infty$. Monotonicity follows from (16), since the sign of the difference values is constant for a given initial condition and δ_1^k is positive and strictly decreasing.

Theorem 4 shows that if the expected diversity starts out larger than $\delta_2/(1 - \delta_1)$ it will decrease monotonically until it reaches that level. Since the delta values are based on the epsilon mutation constants, ξ will typically be small relative to the diversity of a random initial population. Considering the population of individuals as a cluster, and the diversity measure as a clustering measure, the expected "size" of the initial cluster decreases until it reaches a level where the decrease in diversity is countered by the increase due to the mutation operation. With no mutation (i.e., with $\epsilon_1 = 1$ and $\epsilon_2 = 0$) the expected diversity converges toward zero.

We now present another theorem which will allow us to use the expected value $E[L^{(k)}]$ to obtain a lower bound on the probability that $L^{(k)}$ is less than some given constant value.

Theorem 5 *Let m, w, and r be selected uniformly and independently (with replacement) from $\{1, \ldots, n\}$. Let $\alpha > 0$ be a given constant and let $\delta_1 \neq 1$. Then*

$$\begin{aligned} p(L^{(k)} < \alpha) &\geq 1 - \frac{E[L^{(k)}]}{\alpha} \\ &= 1 - \frac{\delta_1^k}{\alpha} E[L^{(0)}] - \frac{\delta_2(1-\delta_1^k)}{\alpha(1-\delta_1)}. \end{aligned}$$

Proof: From a generalization of the Chebyshev inequality, for a random variable $L^{(k)}$ taking only positive values [Pap65], we know

$$\begin{aligned} p(L^{(k)} \geq \alpha) &\leq \frac{E[L^{(k)}]}{\alpha} \\ 1 - p(L^{(k)} \geq \alpha) &\geq 1 - \frac{E[L^{(k)}]}{\alpha} \\ p(L^{(k)} < \alpha) &\geq 1 - \frac{E[L^{(k)}]}{\alpha}. \end{aligned}$$

Use Theorem 2 to write in terms of $L^{(0)}$.

This theorem can be used to show the convergence in probability of the zero mutation case to zero diversity. In a practical sense it allows for computing a bound on the the number of time steps until the diversity is below a certain arbitrary value α with some given probability. For example, one could use Lemma 4 and the Hamming distance along with this theorem to compute an upper bound on the number of time steps k until all individuals in a population are identical with probability, say, 0.99.

9 Discussion

We have defined the properties for a class of distance functions and defined a diversity measure in terms of it. Mathematically, of course, we have defined a term proportional to the average distance between individuals in a population and referred to it as the diversity of a population. There are many possible measures of a population which might reasonably be called the diversity, depending on the various senses in which the term is used.

The results we presented for the uniform selection case allow for explicit predictions of the expected diversity at a given time step. In order to illustrate Theorem 1 and its associated theorems concretely, we simulated the basic genetic algorithm in a computer program. The individuals are considered to be binary strings. The crossover operation used in the simulation is the uniform crossover, as described in Section 3, Example 2. The distance measure used is the Hamming distance, as described in Section 4, Example 1. The mutation operator consists of one random change to the binary string, as described in Section 5, Example 1. Note that here the choice of mutation operator defines the parameters ϵ_1, ϵ_2, δ_1, and δ_2 when the population size, n, and the number of elements in an individual, h, are given.

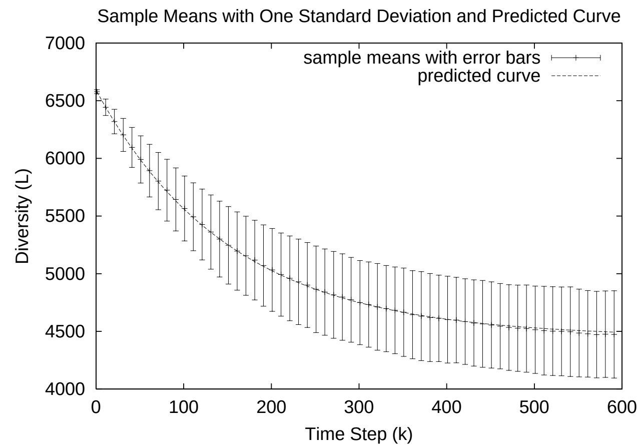

Figure 1: Simulation Results

The parameter values used in the simulation were $n = h = 30$. An initial population was selected uniformly from the space of populations (of size n) of binary strings and kept constant throughout the simulations. Then 1000 trials were conducted by running the genetic algorithm 1000 times on that initial population. Every 10 time steps the sample diversity was calculated, and the results were averaged over these 1000 trials. The simulation results are shown in Figure 1, along with the curve predicted by Theorem 1. At the scale shown, most of the mean values are indistinguishable from the predicted curve. The graph also gives an idea of the sample standard deviations resulting at these parameter and sampling values, which we have not discussed in this paper.

How these results relate to the behavior of typical genetic algorithms using fitness-based selection is still an open question for analytical or empirical research. The uniform selection case might provide useful heuristic values to apply in choosing mutation rates and time schedules for algorithms using fitness-based selection. See [BM99] for further discussion of the effects of fitness-based selection on expected diversity.

Bibliography

[BM99] Allen L. Barker and W. N. Martin. Population diversity and fitness measures based on genomic distances. Technical Report TR-99-24, University of Virginia Computer Science Department, 1999. ftp://ftp.cs.virginia.edu/pub/techreports/CS-99-24.ps.Z.

[Eng97] T. M. English. Some geometric and algebraic results on crossover. In P. Angeline, R. Reynolds, J. McDonnell, and R. Eberhart, editors, *Evolutionary Programming VI*, pages 285–295. Springer-Verlag, 1997.

[Mit96] Melanie Mitchell. *An Introduction to Genetic Algorithms*. MIT Press, 1996.

[MKB83] Joe Mott, Abraham Kandel, and Theodore Baker. *Discrete Mathematics for Computer Scientists*. Reston Publishing Co., 1983.

[Pap65] Athanasois Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1965.

[Sys89] G. Syswerda. Uniform crossover in genetic algorithms. In J. Schaffer, editor, *Proceedings of the Third International Conference on Genetic Algorithms*, pages 2–9. Morgan Kaufmann, 1989.