Abstract- Spatially-structured evolutionary algorithms (SSEAs) have allowed evolutionary search to be scaled up to increasingly larger and more difficult problems. While their use is becoming more widespread, the basic underlying theory behind them has some omissions. In particular, the effect of random genetic drift on spatial structures is largely unknown. This paper uses models derived from random walk theory to describe the behaviour of a specific class of SSEA. The match between the model and experimental findings is very good. It is the intention that this work serves as the basis for a more abstract model of drift that will encompass all spatially-structured population models.

1 Introduction

Evolutionary algorithms (EAs) are being applied to progressively larger and more difficult problems. In order to efficiently handle this scaling up to larger problems, EAs have incorporated new techniques into their frameworks. One such technique is the adoption of spatially structured populations. Spatially structured evolutionary algorithms (SSEAs) are useful because they are relatively easy to implement on parallel machines (Cantú-Paz, 2001; Marco Tomassini, 1999). In addition, SSEAs modify the behaviour of the algorithm in ways that often reduce the number of fitness evaluations required to find a suitable solution (Cantú-Paz, 2001).

While the behaviour of SSEAs has been researched extensively, one particular aspect that has received little focus is the impact incorporating space into a population has on random genetic drift. This paper attempts to take the first steps into addressing this issue. It uses random walk theory as a base to determine the resultant drift within spatially-structured populations. Initial results with one particular population structure, the ring topology, are encouraging.

The remainder of this paper is structured as follows: §2 briefly describes the concepts of random genetic drift that will be used throughout this paper. §3 provides a brief overview of spatially structured evolutionary algorithms. §4 describes the model used to determine fixation times in ring topologies and §5 concludes the paper with a brief summary of the work and directions that future research should take.

2 Genetic Drift

The evolution of populations is driven by two forces. The first, natural selection is a force that all evolutionary computationists are familiar with and has been explored in great depth by a number of researchers. The second force is random genetic drift, which is arguably just as important as selection in determining the direction of evolution for a population. Genetic drift is the force applied to a population as a result of errors in the stochastic sampling process.

Any evolutionary algorithm that performs non-deterministic selection is subject to random genetic drift. It is especially prominent in smaller populations like those typically found in many EAs. Like natural selection, genetic drift works directly to reduce the diversity of the population. However, genetic drift operates independently from fitness. Figure 1 shows the effect of random mating on the allele frequency of a given population. The same initial population was used for three separate runs. Each run resulted in a different outcome purely as a result of genetic drift. In one run, the final allele frequency remained relatively unchanged. In the other runs, the particular allele of interest either completely takes over ($p = 1$) or is completely eliminated from the population ($p = 0$).

The important observation to make is that genetic drift is unguided. Unlike natural selection, genetic drift does not focus on any individuals of a particular fitness. This means the best individuals in a population are equally subject to the effects of drift as the weakest. While mechanisms such as automatically copying the best individual into the next generation help to alleviate this situation, it is still important to have a proper understanding of genetic drift.

The nature of genetic drift in panmictic populations is well understood and mathematical models describing the dynamics of such systems have been presented in previous work (Crow and Kimura, 1970; Hideki Asoh and Heinz Mühlenbein, 1994). Any population that is undergoing evolution in the absence of selection will eventually become
homogeneous. The convergence of a drifting evolutionary algorithm using bitstrings converges at a rate given by the following equation:

$$\tau_p \approx C_0 N \left( a \log_e n + 1.0 \right)^b$$  \hspace{1cm} (1)

where \( N \) is the population size and \( n \) is the bitstring length. \( C_0 \) is a constant which depends on the ratio of alleles in the system, while \( a \) and \( b \) are experimentally determined parameters. For the case of a typical bitstring-based EA, \( C_0, a \) and \( b \) take on the values \( 1:4, 0:5 \) and \( 1:1 \) respectively.

The above equations describing random genetic drift no longer apply when spatially-structured populations are in use. The nature of fixation in spatially-structured populations has not been investigated. The following sections of this paper will attempt to address this issue.

3 Spatially-Structured Evolutionary Algorithms

The relative high computational costs of evolutionary algorithms prompted a need to implement traditional EAs on parallel machines. As a result of this work, techniques were discovered that not only allow for parallel execution of an EA, but also changed the behaviour of the algorithm (Cantú-Paz, 2001). These techniques can be grouped into two categories, the island model EAs and the spatially-structured EAs (SSEAs). Although SSEAs have had several prominent pieces of work written on them, it has traditionally been the island model which has received the greater attention in previous studies (Cantú-Paz, 2001).

Spatially-structured EAs contain a single population that is distributed over a specified topology (Robertson, 1987; Manderick and Spiessens, 1989). The shape of this topology is often a two-dimensional torus, however other shapes have been explored (Schwehm, 1992; Schwehm, 1996). Indeed, this paper will focus on the use of rings as a spatial structure. Selection in SSEAs is constrained by limiting mating to individuals that are within neighbourhoods. These neighbourhoods act as relatively isolated demes within the larger population. Genetic material of the individuals is spread through the population due to the fact that an individual can belong to more than one deme. This allows highly fit solutions to propagate through the population, which sometimes leads SSEAs to be known as diffusion EAs.

While SSEAs appear to be a simple extension to traditional, panmictic EAs, they introduce a number of new parameters into the algorithm:

1. **Topology.** As stated above, this is the definition of the shape of the space being used. Often it is dictated by the underlying hardware that the algorithm is running on.
2. **Neighbourhood Size.** This is the number of “steps” out from a locating individual that the algorithm can make when building a selection deme.
3. **Neighbourhood Shape.** This is the physical structure that the neighbourhood takes. There are a number of ways that this can be formed, but most commonly it takes the form of the individuals that are located in straight lines from the origin (the Von-Neumann neighbourhood), those that fall within a bounding box from the origin (the Moore neighbourhood) or a hybrid of the two that somewhat resembles a diamond. Obviously, the chosen topology also dictates to some degree the possible neighbourhood shapes.

3.1 Ring Topology

This paper makes use of one particular spatial structure; the ring topology. In this spatial structure, \( N \) individuals are placed at a equidistant locations in one dimensional space. The boundaries of this space, that is the ends of the line, are joined to form one continuous dimension, as is shown in Figure 2. Each dot in the ring represents a single individual. Demes are constructed via a parameter \( d \) which is the neighbourhood step size. The demes overlap, allowing for the slow exchange of genetic information throughout the population. Figure 2 shows demes around the individual at “12 o’clock” for several settings of \( d \). The members of the deme are shown in black. The deme size for a given value of \( d \) is \( 2d + 1 \). This gives a fine-grained tuning of deme sizes as necessary; a setting of \( d = 1 \) provides very small subpopulations while a value of \( d = \frac{N}{2} \) results in the ring converging to a panmictic population.
3.2 Selection Models in SSEAs

While this paper focuses on the behaviour of genetic drift in SSEAs, it is still important to briefly mention the concept of natural selection structured populations. The behaviour of selection in SSEAs has been researched intensively in the past (Sarma and Jong, 1997; Rudolph, 2000; Giacobini et al., 2003; Giacobini and Tomassini, 2003). The majority of this work has examined the takeover time of an SSEA, which is simply the number of generations required for a single individual to dominate the population (Goldberg and Deb, 1991). It is examined in the absence of mutation or recombination. More importantly, it is also examined in the absence of genetic drift, as the “best individual” has a probability of selection of 1. The omission of genetic drift from these selection models leaves an entire facet of SSEAs unexamined and is, in part at least, the justification for this paper.

4 Random Walks

Random walks are a commonly encountered phenomenon within the realms of physics and biology (Berg, 1983). In addition, the theory behind random walks have been used to describe the behaviour of certain aspects of evolutionary algorithms (Harik et al., 1999). Indeed, the fixation of an allele in a panmictic population via genetic drift can be viewed as a random walk (Crow and Kimura, 1970).

Three rules are present in the simple model of the random walk:

1. It is assumed that a particle has constant velocity, that is, it moves a constant amount at each time step.
2. A particle is never stationary, it must move at each time step.
3. The probabilities of a particle to move either left or right are constant. At each iteration, the particle will move left with a probability \( p \) and to the right with a probability of \( q = 1 - p \).
4. The movement of a particle in space is independent from any other particles present in the system.

When the above assumptions hold, information about the position of a particle in space after a certain number of steps can be easily established. This paper takes particular interest in one feature of the one-dimensional random walk; how long does it take on average for a particle to move from its starting position to an absorbing boundary.

4.1 Walks with Absorbing Boundaries

Imagine a particle in one-dimensional space as shown in Figure 3. The space has the dimension \( b \). The particle performs a random walk through the space, with an initial point \( x \). At each time step the particle moves a distance \( \delta \). The particle has an equal chance of moving left or right, and the probability of not moving at any time step is zero.

At each end of space there is a boundary. If the particle reaches a boundary, it is considered captured and will no longer move. A given property of such a random walk is that the particle is guaranteed to eventually reach one of these boundaries. Given this, the expected time to capture, \( W \), is established via the differential equation (Berg, 1983)

\[
\frac{d^2W}{dx^2} + \frac{1}{D} = 0.
\]

In this equation, \( D \) is a quantity known as the diffusion coefficient. \( D \) is a value that describes the rate of movement of a particle in space and is defined by the following equation (Berg, 1983)

\[
D = \frac{\delta^2}{2}. \tag{3}
\]

The equation for \( D \) also has a parameter \( \tau \) which is the interval over which distance samples are measured. In this paper, \( \tau = 1 \) and hence can be removed from the equation.
In order to determine $W$, some boundary conditions are required. The mean time to capture at a boundary is $W(0) = W(b) = 0$, therefore the equation to model $W$ becomes:

$$W(x) = \frac{1}{2D}(bx - x^2)$$

(4)

where $x$ is the initial point in space. If the particle starts at the furthest point from the boundaries, namely $\frac{b}{2}$, the equation collapses to

$$W\left(\frac{b}{2}\right) = \frac{b^2}{8D}$$

(5)

Substituting the equation for $D$ into 5 yields the following function for mean time to capture:

$$W\left(\frac{b}{2}\right) = \left(\frac{b}{2\delta}\right)^2$$

(6)

### 4.2 Spatially-Structured Genetic Drift as a Random Walk

An individual in a ring-structured population is only allowed to mate with individuals from within its deme. In order for the genes of two disparate individuals to interact, they must travel through space via the overlapping demes. Because the ring is simply a line in which the ends meet, the maximum distance between two individuals in a ring of size $N$ is $\frac{N}{2}$. When an individual mates with its neighbours, part of its genotype is placed in the neighbouring locations. The genetic information of that individual is now available to new, previously unreachable, demes. The probability of this mating occurring to the left or the right of the individual is equal. At each generation, the flow of genes may reach out to new demes, or may move back towards the deme from which they originated.

The description above for the behaviour of genes moving back and forth between demes is essentially a random walk. It is therefore possible to use the mean time to capture described in the previous section as a basis for modelling fixation in a ring structure. By setting $b = N$ and including the neighbourhood step size into Equation 6, the fixation time for a ring via genetic drift is:

$$\tau_r(N, d) = \tau_p(N) + \left(\frac{N}{2\delta(d)}\right)^2$$

(7)

where $\tau_p(N)$ is the fixation time for a panmictic population of size $N$ and $d$ is the neighbourhood step size. The distance $\delta$ is a function of the neighbourhood step size

$$\delta(d) = a + d$$

(8)

where $a$ is a parameter used to calibrate the fixation equation. A value for $a$ was found experimentally to be 0.3. Figure 4 shows a comparison between the model for ring fixation and values for ring fixation found experimentally. The model agrees with the experimental data to a high degree of accuracy.

#### 4.2.1 Arbitrary-length bitstrings

The paper so far has only discussed models of fixation using single locus, haploid individuals with two possible alleles (0,1). A single locus system is useful to describe the overall dynamics of a ring topology subject to genetic drift, but it does not encompass the behaviour of systems using more realistic representations. An extension to the model proposed is therefore required in order to understand the behaviour of genotypes with multiple loci within a ring structure.

One method to model arbitrary length genotypes is to take the view that each locus is a separate particle in space undergoing a random walk. Each random walk is running in parallel independently from the other loci in the system. The expected time therefore for all $n$ particles to be absorbed is then $\log_e n$ times the expected time for a single particle. The equation to model fixation in a ring topology thus becomes

$$\tau_r(N, n, d) = \tau_p(N) + (\log_e n + 1.0)\left(\frac{N}{2\delta(d)}\right)^2.$$  

(9)

The comparison between this model and experimental findings is shown in Figure 5. As before, there is a high agreement between the experimental data and the proposed model.

### 5 Conclusion

The use of spatially-structured populations is a useful technique to help improve the performance of evolutionary algorithms. While a number of areas in SSEAs are well understood, the impact of genetic drift in these systems has remained largely unknown. This paper has introduced a model for determining genetic drift in one specific SSEA; the ring topology. The model, based around the concept of a random walk in one-dimensional space, captures the essence of drift in a ring to a high degree of accuracy.

Many real-world problems are now being analysed through evolutionary search. Most of these problems have computationally expensive fitness functions, which severely limits the practical population sizes that can be used. Genetic drift is a significant problem for these smaller populations. This paper has shown that ring structures with small deme sizes are much less prone to diversity loss through genetic drift. One could argue here that, since the ring structure is simple and efficient to implement, there is little reason not to implement such a spatial model in most evolutionary algorithms. Doing so may limit the loss of good
Figure 4: Comparison of mathematical model of ring convergence and empirical data.

Figure 5: A model for ring convergence for arbitrary bitstring lengths.
individuals in these populations due to random sampling effects.

This paper is an initial investigation into the effects of drift in SSEAs and has focused on a single topology. There is some evidence, as brought forward by Rudolf (Rudolph, 2000), that the dynamics of SSEAs are determined by the connectivity of the space, rather than the space itself. The model given in this paper is specific to ring topologies. An interesting area for future work to follow is to investigate the possibility of extending the given model to encompass the notion of connectivity, not space, in determining the fixation rates of drifting SSEAs.

**Bibliography**


