The Role of Selection in Evolutionary Algorithms

Dr. Krasimir Kolarov
Interval Research Corporation
1801-C Page Mill Road, Palo Alto, CA 94304
E-mail: kolarov@interval.com

Abstract

The primary goal of this paper is to understand the dynamics of interaction in a population of individuals using a simple but representative evolutionary approach. We study the effect of different parameters, like population size, strength of selection and different selection regimes, on the performance of an evolutionary algorithm in terms of speed of fixation and fixation to suboptimal individuals. We look at evolution as a process of adaptation rather than an optimization, thus our evolutionary model is closer to the Population Genetics approach than to the most commonly used Genetic Algorithm model.

1. Introduction

Genetic Algorithm (GA) is a heuristic computation procedure that is inspired by evolutionary theory. The stochastic character of the GA operators is one of the reasons why it is difficult to explain that it works well for certain problems, or to predict whether the same approach will work in some other problem. There have been very few attempts to theoretically analyze the dynamics of interaction between the different operators and the performance of GA (e.g. Holland's *building blocks* and *schema* [5]). Our major goal in this paper is to understand this dynamics for a simple, but representative system.

An interesting question is to analyze the importance of different **fitness regimes** for the viability of a GA system and to include the role of dynamics in this analysis. We chose a simple type of fitness landscape and performed experiments with it. Criterion for termination in our model is **fixation** of the population (i.e. the state when all individuals in the population have exactly the same genotype). In a large number of GA systems, the simulation runs until the first individual with an optimal fitness value appears. In nature populations of organisms do not stop their evolution after "the best" individual has appeared. In an effort from each individual to survive and reproduce, the next generation is formed. The best individual that is created at some generation might disappear later as a result of the interaction within the system. Thus we need to study the dynamics of interaction and look at the evolution process as **adaptation**, rather than **optimization**.

We focused on the role of the *population size* and the *type of selection* in the performance of this simple GA system. We analyze the mechanism and rate of fixation on sub-optimal fitness values . Since we use fixation as a criterion for termination, the mutation rate for all reported experiments was set to zero. We use the number of cases that fix on a genotype with a fitness different from the optimal one as our sub-optimal fixation level X. Fixation on sub-optimal fitness can result from what is known in population genetics as "genetic drift". In general in finite population models there is always drift due to the statistical nature of the process of sampling that produces offspring from the parental types. When genetic drift dominates the selection pressure, the population may fix on a genotype with fitness different from the optimal. In our case, fixation may occur not only due to random sampling error but also due to the existence of local optima in the fitness landscape.

The problem of finding the appropriate (and in some cases the optimal) population size have been discussed for certain applications in the GA literature (see [3], [9]). There has also been research of the role of different selection schemes on GA performance [7]. A theoretical study of the role of genetic drift was presented in [4].

We will give a detailed description of our model in Section 2. Section 3 considers a simple Gaussian landscape in which we varied the strength and the type of selection. Section 4 analyzes the results from the experiments and draws some conclusions.

2. The model

In an effort to address some of the questions raised so far, we performed experiments with a simple GA type system. Our model includes a finite number of diallelic haploid organisms. The chromosome consists of one string of bits and we chose each genotype or bit string to be of fixed length 20. Each allele can have two values: 0 or 1. The GA operators are crossover and mutation and there is random mating within non-overlapping generations. The initial population is selected at random with equal probability of 0's and 1's. Our model has fixed rates of recombination and mutation. We use single (one-point) recombination with the break point selected at random, uniformly across the chromosome. The offspring are subjected to mutation and proportional selection. New offspring are accumulated until the fixed population size is reached. At that time we have formed the new generation and the current one becomes its parental generation. Thus we do not have explicit elitism - the new generation is entirely formed by applying the genetic operators to the previous one. For each case (i.e. random initial population) this process continues until all the individuals within the same generation have the same genotype - the condition for fixation explained above. For statistical test we run each set of input parameter settings for 100 cases.

This method of constructing the offspring generation where every member of the new population is added as a result of individual "selection" after application of the genetic operators to the previous generation, is a characteristic of population genetic analysis. In a typical GA analysis the individuals in the next population are generated as a result of "group" competition, i.e. their fitness is compared against each other in order to select the offspring. In our case an individual is selected with respect to fixed exogenous fitness function rather than relative to other individuals in the present population. In other words we implement fitness based proportional selection, independent of the fitness values for the rest of the population. This is a significant difference which deserves further analysis. For the purposes of this paper we will mention that as a result of our model, in a typical case, the first generation (after the initial random one) is the one that takes the most amount of computing time. Typically in that generation a significant number of good individuals are created, after which it takes fewer generations overall for the whole population to fixate. Thus, when comparing the results of our simulation with similar ones done with a typical GA setup, we need to pay close attention not only to the number of generations to fixation but to the running time to fixation of the algorithm as well.

We can derive iterative relationships that describe the dynamics of interaction in the system.

Let us denote with $F(s)$ the value of the fitness function for an individual with s number of 1's. With $\varpi_k(s)$ we denote the expected number of individuals with fitness $F(s)$ in the population at generation k. As we have mentioned before N is the population size and l is the length of the genotype (in this paper $l = 20$). The population size at every generation is constant. The average expected fitness of the population at generation k is:

$$\Phi_k = \frac{1}{N}\sum_{s=0}^{l} F(s)\varpi_k(s) \quad \text{for every } k \qquad (1)$$

We have shown in [6] that when there is no recombination (i.e. selection alone is the driving process):

$$\varpi_{k+1}(s) = \frac{F(s)}{\Phi_k}\varpi_k(s) \qquad (2)$$

This relation is similar to Fisher's Fundamental Theorem of Natural Selection.

For the case when we have recombination with fixed rate r, the relationship is somewhat more complicated, i.e.:

$$\varpi_{k+1}(s) = N[r\frac{1}{(l+1)N^2}(\sum_{x=0}^{s}\sum_{z=x}^{l}\sum_{p=s-x}^{l} B(x,z,p,s).$$

$$\omega_k(z)\omega_k(p)) + (1-r)\frac{1}{N}\varpi_k(s)]\frac{F(s)}{\Phi_k} \qquad (3)$$

As one can see in [6] the coefficients $B(x,z,p,s)$ are numerical constants for given x, z, p and s. The formulas above are valid with slight modifications for any single event recombination (not necessarily simple one-point ones).

3. Experiments with Different Selection Models

In this section we will consider a simple Gaussian fitness function. We would like to analyze the dynamics of interaction in the system when we vary the strength and the type of selection. We computed 100 cases for each run, which allowed us within reasonable time constraints to vary all the parameters of the model. In addition we performed a few simulations with a larger number of cases to verify our conclusions.

The fitness function in the experiment is the following Gaussian:

$$F(i) = e^{-\frac{(i-\mu)^2}{2\sigma^2}} \qquad (4)$$

Here i represents the number of 1's in the genotype, thus there are 21 discrete values for the fitness function. This fitness has been discussed for the diploid case in [1] and [2].

We experimented with three different values for the mean, μ , of the fitness function. The case when $\mu = 10$ corresponds to a **stabilizing selection**. In that case the mean of the fitness function coincides with the mean of the initial population (because of the equal probability for 0 and 1 in the initial population). When $\mu = 6$ the selection is **directional** and the peak of the fitness function is somewhat displaced to the left of the mean of the initial population. Thus there is selection pressure toward the mean value. This selection pressure is even stronger in the third case of $\mu = 0$ where we have again **directional selection** and the fitness is a monotonically decreasing function as a function of the number of 1's in the genotype.

To vary the selection strength we gave σ the following values: 0.1; 1.0; 2.5; $\sqrt{10}$; 5.0; $2\sqrt{10}$; 7.5 and 10.0. As the value of σ decreases the fitness function gets a sharper peak. The population size takes the values 25, 50, 100, 150, 200 and 240. The recombination rate is 0.4 and there is no mutation.

To analyze the influence of the different parameters, the data from the simulations were organized in 3 dimensional plots of the fixation on suboptimum, X, with respect to every two of the parameters (σ, μ and *Population Size*) for all possible values of the third parameter. In these plots X is shown in increasing order along the Z-axis and the input parameters are in decreasing order along the other two axes (See Figures 1, 2, 3).

	σ								μ			PopSize					
	10	**7.5**	**6.3**	**5**	**3.2**	**2.5**	**1**	**0.1**	**10**	**6**	**0**	**240**	**200**	**150**	**100**	**50**	**25**
Mean	350	323	303	272	198	168	127	114	289	287	136	395	360	297	221	120	58
Min	18	19	18	13	7	4	2	2	8	2	2	3	10	4	5	2	2
Max	1855	1417	1363	1154	808	979	890	804	1225	1855	1393	1855	1276	1417	877	502	254
St.Dev.	140	121	115	97	65	67	68	69	103	109	81	142	128	110	85	49	27

Table 1. Number of Generations to Fixation for the different settings.

The comparison for the different speeds of fixation (number of generations to fixation, whether or not the population fixates to the optimum genotype), is given in Table 1. The numbers in each column correspond to a fixed value for one of the parameters and all possible values for the other two parameters. The standard deviation follows the trend of the mean in all cases. The corresponding standard error is on the average 0.4 for the different values of σ, m and PopSize.

In order to better analyze the distribution of the number of generations to fixation, we experimented with a larger number of cases (500), just for the sets of parameter values $\sigma = 10$, $\mu = 10$ and all different population sizes. In Figure 4 the horizontal axis shows the number of generations to fixation, with each unit representing 30 generations. Thus if a case has fixed at generation 440, it gets a bar with height 1 at horizontal position 15. The heights of all the cases with the same horizontal position are summed and the result is the plot in Figure 4. We can derive several statistical characteristics of the distribution from this plot. Analyzing these distributions for the different values of PopSize, we can see the shift to the right with the increase of the Population Size.

4. Analysis and Conclusions

The experiments described above reinforced some of our expectations and known theoretical results from the population genetics field about the dynamics of a simple GA system and introduced some new interesting features.

With respect to fixation on suboptimal fitness in the simulations we can conclude that:

- The increase of the Population Size decreases the rate of "wrong" fixation due to less sampling error.

- The increase of σ for a given μ (i.e. as the selection strength is weakened) leads to an increased suboptimal fixation. "Wrong" fixation is relatively high when the fitness landscape has a relatively flat top around the optimum and the selection pressure is not enough to drive it to the true optimum.

These two results are in accord with the results in population genetics that are due to sampling error. As described in [8] selection strength has to be above certain value for given population size, in order for the system to overcome the "genetic drift". Genetic drift in that context refers to the sampling error by the gametes - an internal source of chance effects in evolution.

As we will see, another way of analyzing these results is using equations $(1)-(3)$ from Section 2.

Continuing with the conclusions from our experiments we can state that:

- Beyond a certain selection pressure point (Population Size ³ 100), the selection parameter μ does not influence the likelihood of fixation for a given population size. For small population size, directional selection causes more "wrong" (suboptimal) fixations than stabilizing selection. The increase of σ leads to an increasing rate of fixation on suboptimum values.

- For a given value of σ , directional selection with small populations causes a larger rate of suboptimum fixation than:

a) stabilizing selection under the same conditions;

b) directional selection with large populations.

A reasonable explanation for the above two conclusions is as follows. With stabilizing selection and equal chance for 0's and 1's in the initial population its Gaussian distribution is centered above the genotype with ten 1's. That is also where the optimal individuals are located, thus it takes less time and less "wrong" fixation for the population to settle at the optimum. On the other side, with directional selection the optimum of the population is displaced toward one end of the fitness function thus requiring more time for the population to fixate. The longer time for small populations allows for more incorrect fixation due to sampling error, on the way to the global optimum.

The results for rate of fixation reflect the speed for all fixations, correct or incorrect. We can summarize them as follows:

- The time to fixation increases with the increase of σ- the inverse steepness of the fitness function (strength of selection).

- As expected, larger populations take longer to fixate.

These results also follow from equations $(1)-(3)$ in Section 2. Since the rate of selection $\frac{F(s)}{\Phi_k}$ does not depend on the population size, clearly larger populations will take longer to bring all except one of the $\varpi_k(s)$ to zero (i.e. to fixate). In addition when F is relatively flat on the top (larger σ), for Gaussian initial distribution, there are several suboptimal values of s around the optimum for which both $F(s)$ and $\varpi_k(s)$ (and thus their product) are close to the optimal ones, i.e. selection pressure is weak. From (2) and (3) follows that in future generations the number of these suboptimal individuals will be close to the number of the optimal. Thus on average it will take longer to fixate on the optimal individual and the probability of wrong fixation will be higher.

Another conclusion from the results in Section 3 is the following:

- Stabilizing selection takes longer to fixate than strong directional selection. However it also produces less "wrong" fixations for large enough population. This is a new insight that is a direct result from this work. Here is our explanation:

With stabilizing selection when we start with a random population with equal chance for each allele to be 0 or 1, we have a large number of individuals that have different genotypes but have close fitness values. This is due to the fact that the mean of the fitness function and the mean of the initial population coincide. Thus it is rather difficult to choose among them for the next generation and in general it takes longer for the population to fix. At the same time there is less wrong fixation for large enough populations because the probability is higher of creating and maintaining the best individual .

In the case of directional selection, in the very first generation we have only very few individuals that are better than the rest because the mean of the initial population is displaced from the mean of the fitness function. Thus, through successive generations we slowly build good individuals but they do not leave much chance for other good ones to appear because they dominate the population. Consequently in this case fixation usually occurs faster than in the stabilizing case, but we also have more occurrences of dominating drift.

The discussion above reflects on the question about maintaining variability throughout the generations. A good measure of diversity in population genetics is the **heterozygosity** of the population at a given time. For our haploid case we use the quantity H below as a measure of the heterozygosity in terms of p_i - the frequency of 1's per site i at the given time.

$$H = \sum_{i=0}^{l} 2\, p_i(1 - p_i) \qquad (5)$$

Figure 5 shows the results of a simulation calculating the heterozygosity for three simple type of selection functions:

- Additive (with fitness function $F(s) = s$, where s again is the number of ones in the genotype);

- Multiplicative ($F(s) = k^s$, where k =1.2);

- 'Weak" Multiplicative ($F(s) = 1.1^s$ for multiplicative function with smaller k than above).

All the simulations were done for six different recombination rates (r = 0.5; 0.4; 0.3; 0.2; 0.1 and 0.0). The comparison for the speed of fixation confirms our previous observation that steeper landscape fixate faster.

We can also see that for all three selections, the diversity is maintained longer in the population with smaller recombination values. We can conclude that the steeper the fitness function the faster is diversity lost. The case when there is no recombination shows a singularity in the weak multiplicative vs. multiplicative fitness that will be analyzed further in the future.

Acknowledgments

I thank Dr. Aviv Bergman from Interval Research Corp. for all the joint work, discussions and ideas he shared. I thank Dr. Bergman and Professor Marc Feldman from Stanford University for the detailed and useful comments on the paper.

References

[1] A.Bergman and M.Feldman, *Recombination Dynamics and the Fitness Landscape*, Physica D 56, 1992, p. 57-67.

[2] A.Bergman and M.Feldman, *More on Selection for and against Recombination,* Journal of Theoretical Population Biology, Vol. 38, No.1, August 1990.

[3] D.Goldberg, *Sizing Populations for Serial and Parallel Genetic Algorithms,* Proceedings of the Third International Conference on Genetic Algorithms, 1989, p. 70-79.

[4] I.Harvey, *The puzzle of the Persistent Question Marks: A Case Study of Genetic Drift*, Proceedings of the Fifth International Conference on Genetic Algorithms, 1993, p. 15-22.

[5] J.Holland, *Adaptation in Natural and Artificial System*, MIT Press, 1992.

[6] K.Kolarov, *The Dynamics of Populations in Evolutionary Algorithms: A Parametric Analysis*, submitted to the journal Evolutionary Computation, July 1994.

[7] B.Manderick, M. de Weger and P.Spiessens, *The Genetic Algorithm and the Structure of the Fitness Landscape*, Proceedings of the Fourth International Conference on Genetic Algorithms, 1991, p. 143-150.

[8] J.Roughgarden, *Theory of Population Genetics and Evolutionary Ecology: An Introduction*, Macmillan Publishing Co., 1979.

[9] R.Smith, *Adaptively Resizing Populations: An Algorithm and Analysis*, Proceedings of the Fifth International Conference on Genetic Algorithms, 1993, p. 653.

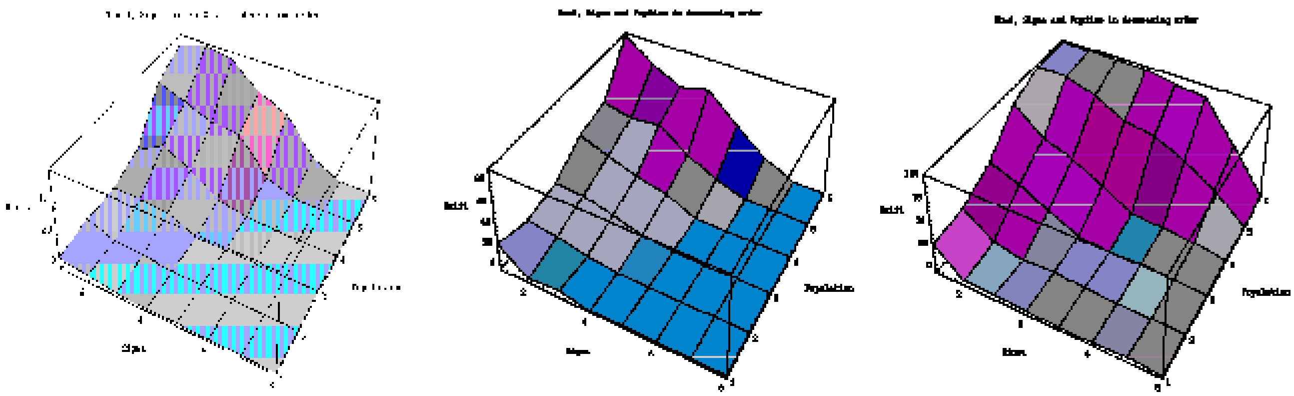

Figure 1. Suboptimal Fixation for given selection type μ

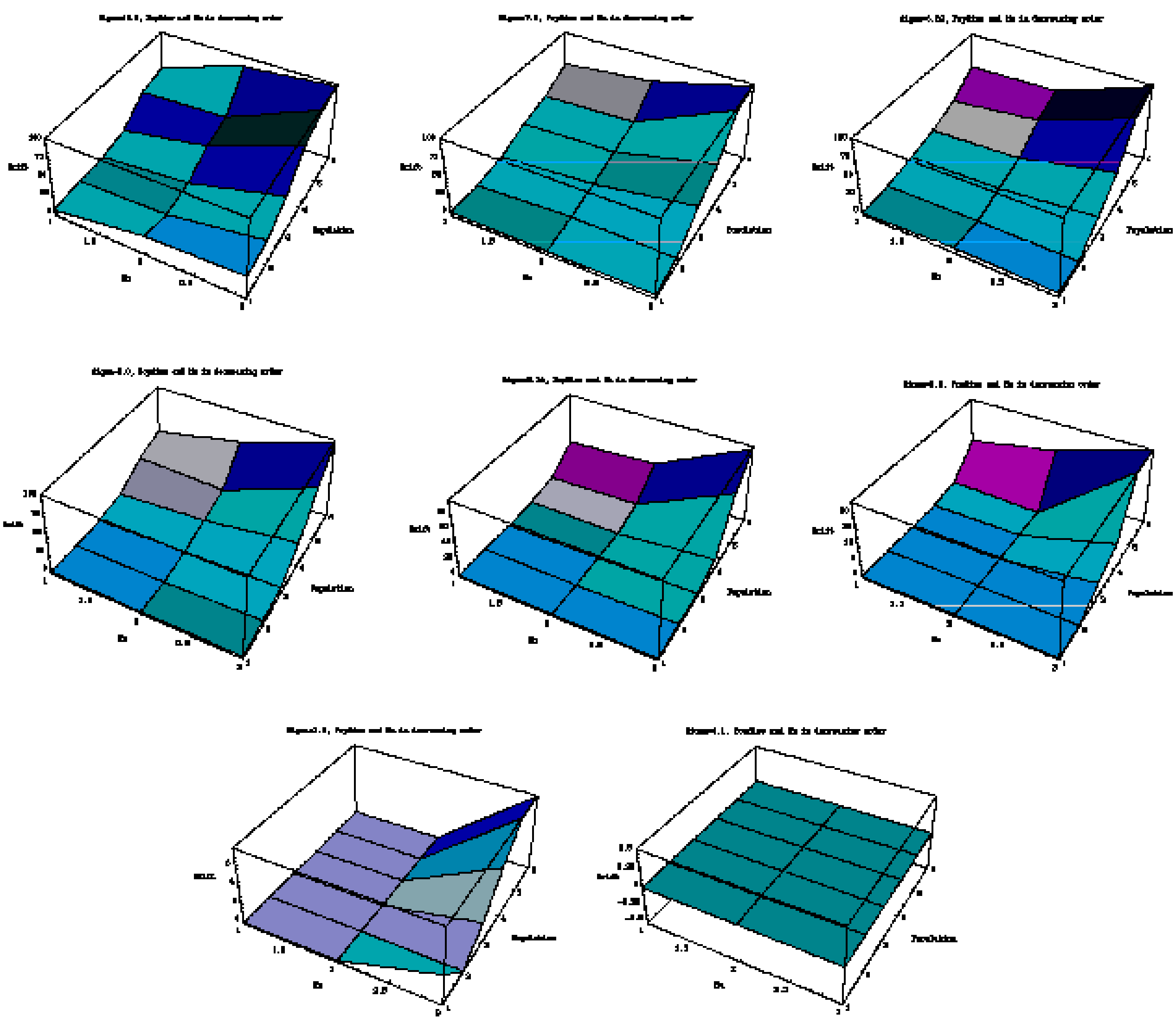

Figure 2. Suboptimal Fixation for given strength of selection σ

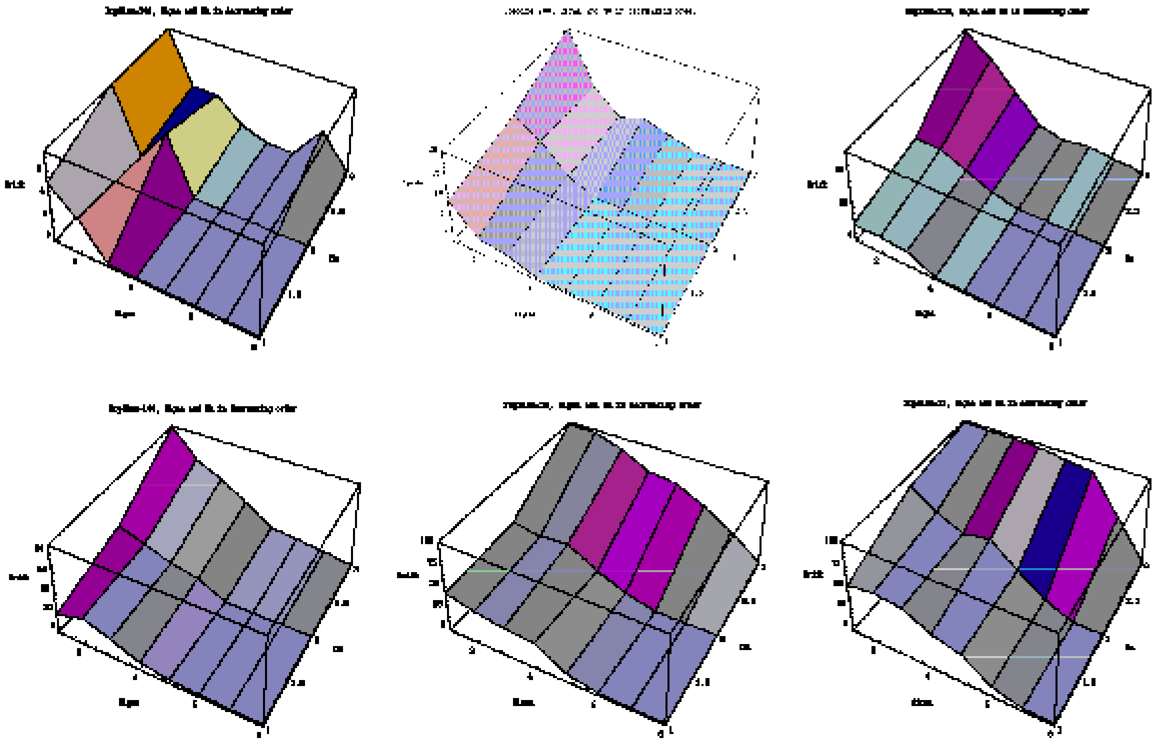

Figure 3. Suboptimal Fixation for given PopSize

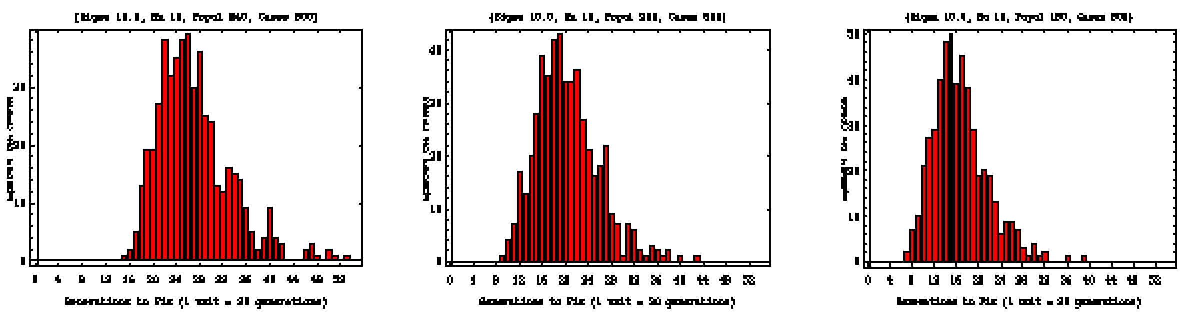

Figure 4. Distribution of the generations to fixation

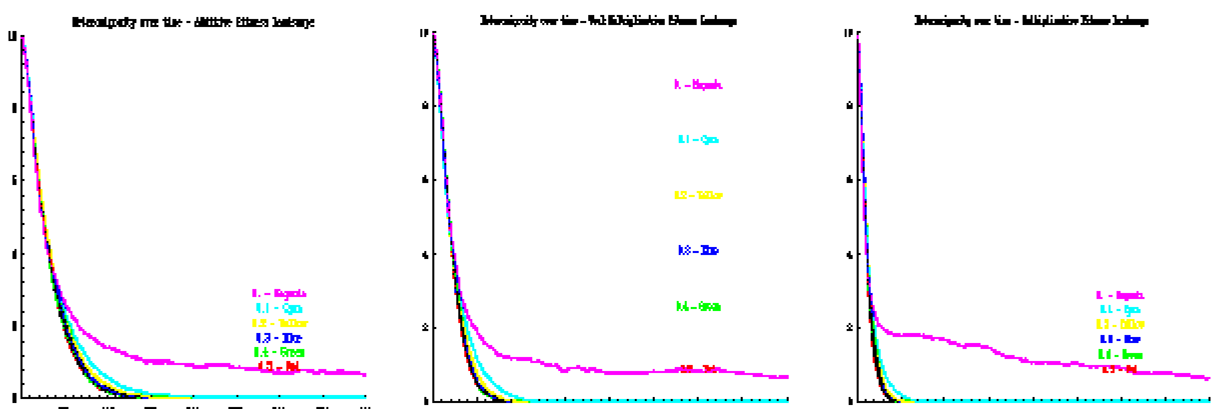

Figure 5. Heterozigocity as a measure of Diversity