A Statistical Mechanical Formulation of the Dynamics of Genetic Algorithms

Jonathan Shapiro, Adam Prügel-Bennett, and Magnus Rattray
Department of Computer Science, The University, Manchester
Oxford Road
Manchester M13 9PL

Abstract: A statistical mechanical formulation of the dyamics of genetic algorithms is described. This formulation allows the derivation of equations which predict the distributions of fitness with the population at one generation in terms of the distribution at the previous generation. The effects of selection are problem independent, and the formulation predicts an optimal value of selection. Crossover and mutation are discussed in terms of a test problem – search for the low energy states of a random spin chain. The theory is compared with simulations and the agreement is good.

1 Introduction

The effectiveness of genetic algorithms can depend crucially on how they are carried out. In order to apply a genetic algorithm effectively, two types of decisions must be made. The first is the choice of representation – solutions to the problem must be represented as strings and genetic search operators (e.g. crossover, selection) must be chosen appropriately to this. The second is the choice of parameters of the algorithm, these include the population size and parameters which govern the relative importance of the search operators. Good choices of problem representation and parameter values can lead to genetic algorithms which are extremely effective; poor choices can render the method little better than random search.

In order to optimize the performance of a GA, it is useful to understand its dynamics. A very general description of GA dynamics is given by the Schema picture of Holland [1]. Principles of GA design have been put forward by Goldberg [2] and made precise and extended to general representations by Radcliffe [3] based on this idea. However, this does not predict the dynamics of a particular genetic algorithm in detail. Alternatively, there are detailed mathematical formulations of genetic algorithms which do give a detailed description. The two most developed are the Markov Chain formulation of Vose et. al. [4], and the Walsh-function analysis introduced by Bethke [5] and much used by Goldberg and collaborators [6,7,8]. The former is a beautiful and exact theory of the evolution of a GA. It has been useful in revealing the types of dynamics which can occur (e.g. punctuated equilibria). However, it is of little practical utility, as in order to make predictions, knowledge of transition probabilities between all strings is required. These are never known in real applications. A similar type of objection can be made for Walsh-function analysis – it has been very useful in revealing why certain problems are easy or difficult for GAs, and has been used to cook up difficult problems to test GAs on. In particular, the notion of *deceptive problems* has been very usefully identified and understood as the local minima of GA's. However, the Walsh expansion for a fitness function will not be known in general, so it is not a formalism which can be used to learn about specific problems.

In this paper, we present a new formalism for studying the dynamics of genetic algorithms. We present a statistical mechanics formulation of GA's which can predict the evolution of the algorithm in some cases. For very simple cases, the dynamics can be solved exactly. For more complex problems, not all properties of the dynamics can be predicted, but by making some assumptions, it may be possible to approximate certain features. This formulation is based upon properties of the distribution of fitnesses within a population, which can be easily measured. The utility of this approach is two-fold. First, since the dynamics of the GA can be solved in some cases, it can serve as a test-bed for ideas about how GA's perform, and how to improve GA's performance. Second, since it can approximate or reveal some features of the algorithm in general, it might be useful in helping a user of GA's to find good values of search parameters or good representations.

2 Statistical Mechanics Formulation of Genetic Algorithms

Statistical mechanics is a branch of physics which deals with systems consisting of a huge number of interacting entities. It provides methods for calculating the gross or average properties of these systems by treating the small-scaled motions (which are probably unknow in detail) as random. It is usually applied to the computation of thermodynamic properties of particles (e.g. molecules, electrons) interacting with a heat bath, but has also been applied recently to a host of other problems in coding theory, neural networks and optimization (for examples, see [9]). As an example, statistical mechanics was found to be useful in setting the parameters for simulated annealing [10]. Statistical Mechanics has been previously applied to the study of genetic dynamics (see for example [11]). However, these previous studies there was no notion of *fitness*; these were studies of genetic variability and converge through genetic drift.

There may seem to be no relationship between the motions of atoms in a gas and the dynamics of genetic search, so one might wonder what the relevance of statistical mechanics might be. The connection is that both involve motion which is both directed and random. Atoms in a gas would relax into a state of lowest energy, except that thermal motion causes states of high entropy to be more likely. The result is a trade-off between energy minimization and entropy maximization; the relative importance of each is controlled by the temperature. Likewise, in genetic search, selection tends to move the population towards states of high fitness, while the randomizing operators mutation and crossover explore states of high entropy. However, it is complicated to make this connection precise, because there is no notion of equilibrium in the GA dynamics, and because selection and particularly crossover depend upon the population configuration.

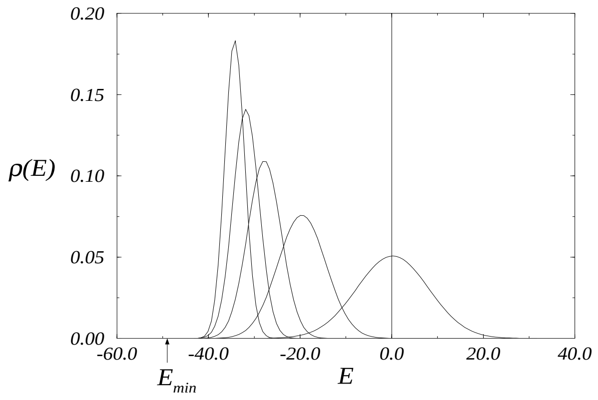

Figure 1: A GA is used to find low energy states of a spin-glass. The distribution of energy in a population of 50 is shown after 0, 10, 20, 30 and 40 steps of the GA. The search occurs in 63 dimensions. The curves show 1 000 samples averaged. Figure is from reference 12

It is also worth pointing out that statistical mechanics predicts the average properties of a system in the limit of large system size. This formalism will predict the average properties of the dynamics. Fluctuations which will occur in individual runs will not be considered. The formalism will be appropriate for studying very high-dimensional search spaces. It will allow the study of test problems which are complementary to the usual test-bed problems, namely one and two dimensional function optimization problems.

We consider the evolution of the distribution of fitnesses within a population [12]. The goal is to predict the distribution of fitnesses at one timestep in terms of this distribution at the previous timestep. For example, suppose a GA is being used to find the minima of an objective function E. A typical evolution is shown in figure 1. Here the distribution of objective functions $\rho_t(E)$ of each individual in the population is shown at different times t in the genetic search. Initially, the distribution is nearly Gaussian; during the evolution of the GA it the average fitness of the population increases (decreasing E), the variation in the population decreases, and the distribution loses the Gaussian shape (although this is not apparent from the figure). Throughout this paper, we shall consider the task of the GA to minimize an objective function E, and we will refer to the distribution of E's rather than the distribution of fitnesses (so the fitness is $-E$ plus some constant to insure non-negativity). We will refer to E as the "energy" in analogy with statistical mechanics.

To study the evolution of the energy distribution, we calculate the effect of selection, crossover, and mutation on an arbitrary distribution. The full evolution of the GA is easily calculated by iterating

$$\rho_t(E) \stackrel{\text{selection}}{\longrightarrow} \rho_t^s(E) \stackrel{\text{crossover}}{\longrightarrow} \rho_t^c(E) \stackrel{\text{mutation}}{\longrightarrow} \rho_t^m(E) = \rho_{t+1}(E), \tag{1}$$

starting from the initial population $\rho_0(E)$.

Of course, since the population size is finite, the actual distribution will consist of delta-functions. We will concentrate on the *statistical* properties of this distribution, the mean, variance, and higher moments averaged over samples. In particular, we consider a cumulant expansion for the distribution function. Let. E^α denote the energy of a member of the population, $\alpha = 1, \dots P$ where P is the size of the population. The cumulant expansion is defined by the generating function which is given by the log of the Fourier transform of the distribution function [13]

$$G(t) = \log(\left\langle \mathrm{e}^{itE^\alpha} \right\rangle) \tag{2}$$

through the following equation for the n^{th} cumulant κ_n

$$\kappa_n = (-i)^n \frac{\partial^n}{\partial t^n} G(t). \tag{3}$$

Here $< \cdots >$ denotes average over the distribution $\rho(E)$. (In statistical mechanics terminology, $G(t)$ is the log of the partition function, and the inverse temperature $\beta = -it$). The cumulants contain the same information as the distribution function. The first two cumulants are simply the mean and the variance. All cumulants except the first two are zero for a Gaussian distribution, so the higher cumulants measure the non-Gaussian nature of the function. The third cumulant is called the skewness – this measures the degree of asymmetry (about the mean) of the distribution. The fourth cumulant – called the excess – gives the first indication whether the function falls off more quickly or more slowly than a Gaussian.

Our description of the evolution of the GA will be in terms of the change in time of the cumulants due the the genetic operators. The cumulant expansion is convenient for a number of reasons. First, cumulants can readily be measured. Second, cumulants are a more stable measure than moments, because they tend to self-average. This means average values are the same as typical values, and the averages predicted from statistical mechanics will be a good approximation. Finally, the cumulants provide a interesting picture concerning how GA's work. For example, as we shall see, selection increases the skewness of the population. The low-fitness tail is long, whereas the high fitness tail is sharp. Thus, it is useful to study the effect on the skewness of crossover and mutation. The extend to which these operators decrease the skewness relative to the amount that they decrease the mean fitness is a measure of the effectiveness of these operators.

3 Selection

Selection is the mechanism which increases the mean fitness of the population by chosing more fit individuals with a higher probability than unfit ones. Many ways of performing selection have been proposed, the best-known being "roulette-wheel" selection. In this method, one choses each individual with independent probabilities which determined by the fitness. We use a Boltzmann weighting,

$$p^\alpha = \frac{\mathrm{e}^{-\beta E}}{Z}, \qquad Z = \sum_{\alpha=1}^{P} \mathrm{e}^{-\beta E}. \tag{4}$$

where β controls the amount of selection. For $\beta = 0$ each member of the population has the same weight, and the only change is due to genetic drift; when $\beta \to \infty$ only the most fit individual is selected; β small corresponds to proportional selection.

In order to compute the effect on the cumulants of selection, the generating function must be computed. This is given by

$$\left\langle \log(Z) \right\rangle_\rho = \left[\prod_{\alpha=1}^{P} \int_{-\infty}^{\infty} \rho(E_\alpha)\ \mathrm{d}E_\alpha \right] \log(Z) \tag{5}$$

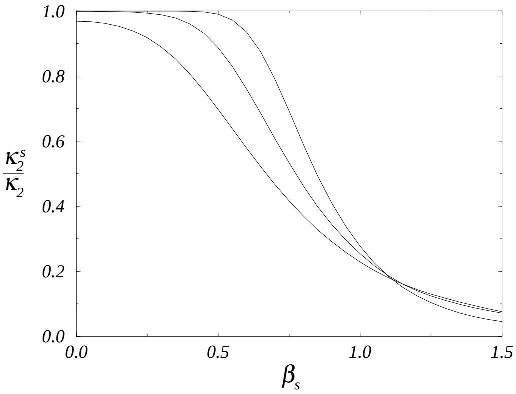

Figure 2: The curves show the change in the variance of the energy distribution after selection starting from a Gaussian distribution for $P = 2^5$, 2^{10} and 2^{20} versus the scaled selection parameter $\beta_s = \beta(\kappa_2/2\log(P))^{1/2}$. The solid lines are calculated by numerical integration.

This is equivalent to a statistical mechanical model, the Random Energy Model, proposed by Derrida [14] as a model of spin-glasses, where fitness plays the role of energy. Using the methods which developed for this model, the cumulants after selection can be computed from those before. Details will be given elsewhere [12,14].

To leading order the first cumulant after selection becomes $\kappa_1^s = \kappa_1 - \beta\kappa_2 + \cdots$. Thus the mean energy is shifted by an amount proportional to the selection parameter β times the variance, κ_2. The variance is changed by an amount $\kappa_2^s = (1 - 1/P)\kappa_2 - \beta\kappa_3 + \cdots$. In Fig. 2 we show the rate of convergence, κ_2^s/κ_2, versus the (scaled) selection parameter $\beta\sqrt{\kappa_2/2\log(P)}$, for populations of size 2^5, 2^{10} and 2^{20}, starting from a Gaussian distribution (i.e. $\kappa_n = 0$, $n > 2$). The curves have been calculated by numerical integration using Gaussian quadrature. The ordinate, to first order in β, is proportional to the shift in the mean energy of the distribution. We can see from Fig. 2 that even for $\beta = 0$ (arbitrary selection) there is an intrinsic convergence rate, which reduces the variance in the population by a factor $1 - 1/P$. This arises because, by chance, some members of the population will not be selected while other members will be selected more than once.

The effects of selection on higher cumulants can also be computed [12]. Selection introduces a skewness into the distribution which slows down the shift in the average energy and increases the rate of convergence. The skewness is in favor of the higher energy states, so a role of crossover and mutation will be to replace some of the higher energy tail with lower than average states. Of course, these two operators will also increase the mean energy, this show the trade-off which it will be important to optimize.

3.1 Optimal Selection Parameters

If we assume that the goal of selection is to provide the optimal improvement in the mean fitness of the population while having the least deletarious effect on the diversity, an optimal amount of selection can be chosen. As can be seen from the curve in figure 2, the curve is initially flat for small β. The improvement in the mean energy is proportional to β so it pays to increase β, since there is no extra loss of diversity in the flat region.

However, passed an optimal value around the shoulder of the curves in the figure, the diversity in the population decreases very rapidly with selection and the GA will be unlikely to find a good solution. Thus we see for Boltzmann selection there is an optimal choice of the selection parameter, β, and we can predict how it scales. Since the variance decreases as the population converges the degree of selection should be increased, as has already been observed [2].

4 Crossover and Mutation

4.1 A Test Problem

Crossover and mutation are problem specific. In order to study these, we must consider specific problems. Here we report studies of a GA which searches for low lying states of a one dimensional spin chain with random nearest-neighbor couplings. The energy of a configuration of spins, $\vec{S} = (S_1, S_2, \ldots, S_N)$ is

$$E(\vec{S}) = -\sum_{i=1}^{N-1} J_i S_i S_{i+1}, \tag{6}$$

where each spin takes the value $+1$ or -1 and the couplings J_i are drawn from a Gaussian distribution. The average ground state energy is $E_{min} = -\sum_i |J_i|$. Although this problem is trivial to solve it is nevertheless interesting in that is has an exponential number of local minima under single spin flip dynamics (typically $2^{N/3}$), and it is an example of search on a high-dimensional (N dimensions) space. Thus, it would be very difficult for a hill-climbing algorithm to solve. On the other hand, it is perhaps the easy type of problem for a GA, because the fitness function is very local. In the language of GA's, it would have zero epistatis, except for the small interface energy. The thermodynamics of this model have been extensively studied [15,16].

4.2 Mutation

Mutation is the operator which changes genes with a small random probability. It is a straight-forward calculation to find the distribution of energies after mutation. We will only describe the results here, refering the reading to the more detailed paper [14] for the calculation.

Mutation changes the mean energy by a small amount proportional to κ_1/N. The $1/N$ is because mutation can only change two of the N bonds of the chain; the proportionality to the mean energy is because the more likely it is that a bond is statisfied, the more likely it is that mutation will break that bond. Mutation increases the variance if it is more narrow than the natural variance for the spin chain and narrows it if it is wider than that variance. The natural variance is that of a random population. This is obvious, because repeated application of mutation should produce a randomized population. The effects on higher cumulants are small.

It is possible to produce a closed form expression for these cumulants, but this is beyond the scope of this qualitative description. Mutation is not a particularly effective search operator for this problem, because of the large number of local minima.

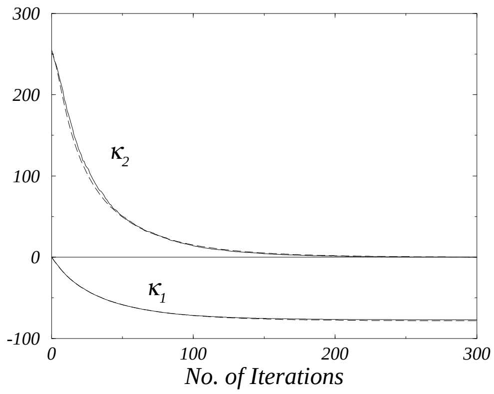

Figure 3: Comparison of theory and simulations for selection and crossover using $\beta = 0.01$ with $P = 50$ and $N = 255$. The solid curves show the simulations averaged over 500 samples. The theory is shown by dashed curves, which are nearly obscured by the simulations. From reference 12

4.3 Crossover

Crosssover is the most important operator, because it is what distinguishes GA's from stochastic gradient search. It is also the most complicated to understand, because the effect depends upon the correlations between the strings within a population. This information is not contained in the distribution function. For example, two strings with very similar energies could be very similar or very different genetically. The energies produced by crossing these strings will obviously depend on how correlated they are. However, the degree of correlation between the strings will not be determined by their energies, but by their common ancestry. Thus, we must model the evolution of the correlation between the spins as well as the dynamics of the energy distribution.

Like mutation, crossover is likely to break a bond at the crossover point. This will be proportional to κ_1/N. The constant of proportionality is $(1-q)$ where q is the average correlation (for example, if the two strings were identical, there would be no energy change). This can be estimated as κ_2/N. Beyond the contribution of the interface energy change, crossover has no effect on the first two cumulants. It does have the effect of decreasing the higher cumulants.

For the spin-chain, crossover is a more effective search operator than mutation. This can be seen in the cumulants and correlations. Mutation introduces an average fitness cost which increases as the GA evolves. The cost of crossover is less, since it is mediated by the correlations which increase during evolution. Thus, the best choice of operators would be to use no mutation and crossover with an optimal amount of mixing. Figure 3 shows a comparison of simulation (solid line) with the theory derived from iterating the equations for the first six cumulants(dashed curves). The agreement with simulations is very good provided β and P are not too large. For larger values of these parameters the approximations used begin to break down; a more accurate treatment is being developed.

5 Future Work

The strong localness of the spin-chain is a very unrealistic aspect of this test problem. This meant that mutation was clearly less effective than crossover. In more realistic problems, crossover could have a much larger effect on the mean fitness, and mutation a much smaller one.

The spin-chain was chosen as a first test problem because it is a trivial model to solve. It will be useful to extend this work to higher dimensional spin-glass, where mutation will have a much more important role to play in reducing the amount of converge caused by selection.

References

1. J. H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press (Ann Arbor), 1975.

2. D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley (Reading, Mass), 1989.

3. N. J. Radcliffe, Equivalent Class Analysis of Genetic Algorithms, Complex Systems 5, 183 (1991).

4. M. D. Vose and G. E. Liepins, Punctuated Equilibria in Genetic Search, Complex Systems 5, 31-44, 1991.

5. A. D. Bethke, *Genetic Algorithms as Function Optimizers,* Doctoral Dissertation, University of Michigan, 1981.

6. D. E. Goldberg, Genetic Algorithms and Walsh Functions: Part I, a Gentle Introduction, Complex Systems, 3, 129-152, 1989.

7. D. E. Goldberg, Genetic Algorithms and Walsh Functions: Part II, Deception and its Analysis, Complex Systems, 3, 153-171, 1989.

8. D. E. Goldberg and M. Rudnick, Genetic Algorithms the Variance of Fitness, Complex Systems, 5, 265-178, 1991.

9. M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond,* World Scientific (Singapore) 1987.

10. P. J. M. van Laarhoven and E. H. L. Aarts, *Simulated Annealing: Theory and Applications*, Kluwer Academic Press (Dordrecht) 1987.

11. C. Tsallis, Exactly Solvable Model for a Genetically Induced Geographical Distributions of a Population, Physica A 194, 502-518, 1993.

12. A. Prügel-Bennett and J.L. Shapiro, An Analysis of Genetic Algorithms Using Statistic Mechanics, Phys. Rev. Lett., 72(9) p1305, 1994.

13. M. Abramowitz and I.A. Segun, *Handbook of Mathematical Functions*, Dover Press, 1964.

14. B. Derrida, Random-energy Model: An exactly Solvable Model of Disordered Systems, Phys. Rev. B24, 2613 (1984).

15. A. Prügel-Bennett and J.L. Shapiro, The Dynamics of Genetic Algorithms for the Ising Spin-Glass Chain, in preparation.

16. T. Li, Phys. Rev. B **24**, 6579 (1981).

17. H. H. Chen and S. K. Ma, J. Stat. Phys. **29**, 717 (1982).