The Dynamics of a Genetic Algorithm for Simple Random Ising Systems

Adam Prügel-Bennett* and Jonathan L. Shapiro
Department of Computer Science
University of Manchester
Manchester, M13 9PL, UK.

May 20, 1996

Abstract

A formalism is presented for analysing Genetic Algorithms. It is used to study a simple Genetic Algorithm consisting of selection, mutation and crossover which is searching for the ground states of simple random Ising-spin systems: a random-field ideal paramagnet and a spin-glass chain. The formalism can also be applied to other population based search techniques and to biological models of micro-evolution. To make the problem tractable, it is assumed that the population dynamics can be described by a few macroscopic order parameters and that the remaining microscopic degrees of freedom can be averaged out. The macroscopic quantities that are used are the cumulants of the distribution of fitnesses (or energies) in the population. A statistical mechanics model is presented which describes the population configuration in terms of the cumulants, this is used to derive equations of motion for the cumulants. Predictions of the theory are compared with experiments and are shown to predict the average time to convergence and the average fitness of the final population accurately. A simplified version of the equations is produced by keeping only leading nonlinear terms, and truncating the cumulant expansion. This is shown to give a novel description of the role of genetic operators in search, e.g. it is argued that an important role of crossover is to reduce the skewness of the population.

PACS 0550, 0250

*Present Address: NORDITA Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

Contents

1 Introduction

1.1 Genetic Algorithms

Genetic Algorithms (GAs) are a popular class of iterative search techniques used to find good solutions to hard problems. They use a search strategy, originally proposed by Holland [1], which is motivated by natural selection in evolution. A population of potential solutions undergoes a set of operations which mimic reproduction, mutation and natural selection so that the quality of solutions in the population increases over time. Genetic Algorithms have been successfully applied in a wide range of domains; and their use is ever increasing.

GAs differ from other search methods such as hill-climbing and simulated annealing in two respects. First, rather than making changes to a single candidate solution, a population of solutions is evolved. Second, in addition to undergoing local changes to the solutions (via mutation), there is the nonlocal crossover operation in which a new solution is produced from two parent solutions. These two properties give GAs their interest, they also make them hard to analyse. In this paper we use statistical mechanics techniques to derive a set of deterministic dynamical equations describing the average behaviour of simple GAs. Preliminary results have been reported elsewhere [2, 3]. We show how this technique can be used to predict the evolution of Genetic Algorithms on a simple problem.

In the simplest realization of a GA [4] each solution is represented by a genetic string or chromosome. A population of strings is generated at random to form the first generation. The "fitness" of each solution is measured using the appropriate cost-function or measure for the problem. A new generation is created by applying a set of "genetic operators" to the preceding generation. The selection operator copies each member of the population into the new population with a frequency (sometimes zero) depending on its fitness. Usually the size of the population is maintained at each generation. The two most commonly used genetic mixing operators are mutation and crossover; these are the only ones discussed in this paper. In mutation parts of the strings are randomly altered — this produces a "local" move in the search space. In crossover two solutions are merged together to form a new solution. The efficiency of a GA depends on the representation used and on how the various free parameters are chosen. These parameters include the population size, the selection rate, the mutation rate and the proportion of solutions that undergo crossover.

The dynamics of GAs are difficult to analyse – in addition to being high-dimensional and stochastic, the population-dependence and the crossover operator make the system nonlinear and strongly interacting. The result is a paucity of theoretical results on GA dynamics. Among methods which have been proposed and used to study GAs are Markov chain formulations and static fitness-function analysis. A number of researchers, notably Vose and co-workers [5, 6, 7], have applied Markov Chain analysis to this problem. However, this formulation is difficult to apply due to the high dimensionality and nonlinearity, and no results on realistic GAs have been produced. Static analysis is based on characterizing the fitness function rather than describing the dynamics. Walsh function analysis [8] and schemata theory [4] are examples of this approach. These approaches miss effects due to the fact that the region of fitness space that the population is in is changing in time. Consequently, these methods can make predictions which are wrong (e.g. see [9]). There is clearly room for a new formalism to describe GA dynamics, because methods which could be used to analyse GAs would be useful, both in applying them more effectively and in understanding more complicated population-based dynamics such as realistic population biological models.

In this paper, we apply statistical mechanics to Genetic Algorithms. We choose a small set of macroscopic quantities which we assume characterise the system. We then assume that all remaining degrees of freedom follow a maximum entropy distribution. The approach is analogous to how one computes properties of a magnetic system in statistical mechanics, where the macroscopic quantities are energy and magnetization and the assumed distribution is the Boltzmann distribution. Two further elements are essential in our calculation. First, Genetic Algorithms are dynamical systems. We use statistical mechanics to calculate statistical properties of the population which change in time. Thus, our approach converts a nonlinear dynamical system consisting of an enormous number of degrees of freedom into one consisting of only a few. Second, a Genetic Algorithm is intrinsically a finite-size and finite-population system, which means that more order parameters are required to get accurate results.

1.2 Overview of Our Approach

In order to study the dynamics of Genetic Algorithms, we consider two specific problems: finding the ground states for a random-field paramagnet and for a spin-glass chain. We derive deterministic equations which describe the average dynamics of the Genetic Algorithm. That is the main goal of this paper. The description is in terms of how the distribution of fitnesses in the population evolves, on average. We discuss how this helps in understanding how GAs work and suggests criteria for designing effective GAs. In the rest of this section we describe the problems tackled by the GA; the Genetic Algorithm we have used; and we summarise the approach taken and the results obtained.

The Problem

We have considered two test problems, which are finding the low lying states of the spin-glass chain and the random-field paramagnet. In both cases the fitness is taken as minus the energy of the magnet. For the spin-glass chain

$$\mathcal{F} = -E[\vec{S}] = \sum_{i=1}^{N} J_i S_i S_{i+1} \tag{1.1}$$

while for the random-field paramagnet

$$\mathcal{F} = -E[\vec{\tau}] = \sum_{i=1}^{N} J_i \tau_i \,. \tag{1.2}$$

where in both cases the J_i's are randomly chosen independent variables coming from a Gaussian distribution with mean 0 and width 1. The average maximum fitness is at $\mathcal{F}_{\max} = \sqrt{2/\pi}N$. Thermodynamically, the two problems are identical — the spin-glass chain can be transformed into a random-field paramagnet by making a change to "defect variables" $S_i S_{i+1} \to \tau_i$. The thermodynamics has been thoroughly investigated in other contexts (see for example [10, 11, 12, 13]). However, from the point of view of single spin flip dynamics the two problems are very different. In the random-field paramagnet the spins are all decoupled so there are no local minima. In contrast in the spin-glass problem there are very many (typically $2^{N/3}$) local minima. These correspond to islands of satisfied bonds separated by weak unsatisfied bonds. Removing an island involves flipping each spin in the island which breaks some strong bonds — these act as energy barriers.

The spin-glass chain is perhaps the simplest example of a hard search problem; one which is dominated by an exponentially growing number of minima. By making the chain sufficiently long we can make the problem hard enough to defeat any size of population. Even for this simple problem, however, the calculation of the GA is very complicated, and requires some approximations to be made. The random-field paramagnet is introduced as a related but simpler model for which the dynamics can be worked out completely.

The Genetic Algorithm

We will consider a Genetic Algorithm with three operators: selection, mutation and crossover. The solutions of the problem are represented by strings of spins (or defects in the case of the random-field paramagnet). The population consists of a set of P strings, $\vec{S}^\alpha$, (or $\vec{\tau}^\alpha$) where $\alpha = 1, \ldots, P$ denotes the different members of the population. The fitness of a string is denoted by $\mathcal{F}_\alpha = -E(\vec{S}^\alpha)$ (or $\mathcal{F}_\alpha = -E(\vec{\tau}^\alpha)$). The initial population is created by generating random strings. To perform selection we assign a "Boltzmann weight" (proportional to $\exp(\beta\mathcal{F}_\alpha)$) to each member of the population. The new population is selected with a probability proportional to its weight. The degree of selection is determined by the "inverse temperature", β, in the Boltzmann weight. This is scaled throughout the run so that the effect of selection remains roughly constant. We examine "Roulette wheel" selection where the members of the new population are drawn independently, although two alternative forms of selection are also discussed (see section 6.2: "Deterministic" selection where the members of the new population are chosen so that their proportion in the new population is as close as possible to their weight, and Truncation selection in which the best p members are duplicated).

Mutation is performed by flipping each spin (or defect) with a probability γ. For the random-field paramagnet we use "uniform" crossover where a child is produced by taking pairs of solution and choosing the spin for the child from either parent at random. For the spin-glass chain we use "single-point" crossover where the child is produced by crossing its parents at some randomly chosen bond. Each time we produce a child we also produce the complementary child by taking the spins that are left over from the two parents. This does not introduce any new correlations and is computationally efficient.

Deriving the Dynamical Equations

Our approach to studying this problem is to predict the change in the distribution of fitnesses in the population from one generation to the next. To do this the effect on the distribution of each of the GA operators; selection, mutation and crossover is calculated. Although the operators produce a random change to a single string, the statistical features of the population, such as the mean and standard deviation, change in a regular way (in finite populations there are small fluctuations). Denoting the distribution of fitnesses at generation t by $\rho_t(\mathcal{F})$, the evolution can be schematically represented by

$$\rho_t(\mathcal{F}) \xrightarrow{\text{selection}} \rho_t^s(\mathcal{F}) \xrightarrow{\text{mutation}} \rho_t^{sm}(\mathcal{F}) \xrightarrow{\text{crossover}} \rho_t^{smc}(\mathcal{F}) = \rho_{t+1}(\mathcal{F}). \tag{1.3}$$

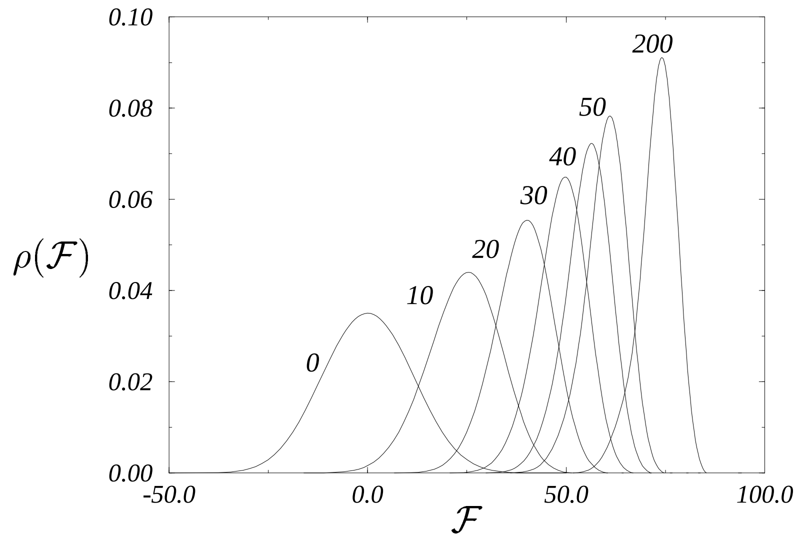


Figure 1: The distribution of fitness is shown after 0, 10, 20, 30, 40, 50 and 200 steps of the GA applied to the spin-glass chain. The curves were created by averaging over 10 000 samples for $N = 127$, $P = 50$, $\beta_s = 0.1$, $\gamma = 1/2N$ and $f = 1$. The average ground state energy is at $\mathcal{F}_{\max} \approx 101$.

The full dynamics is calculated by iterating the sequence in (1.3) starting from the initial distribution $\rho_0(\mathcal{F})$. The evolution in the distribution of fitnesses for the spin glass chain is illustrated in figure 1. The curves show the distribution $\rho_t(\mathcal{F})$ at generation 0, 10, 20, 30, 40, 50 and 200. The curves are produced by averaging over 10 000 simulations. The initial population is approximately Gaussian with zero mean. The simulations were performed using chains of $N = 127$ bonds; the population size was $P = 50$; roulette wheel selection was used with a selection strength of $\beta_s = 0.1$; the mutation rate was $\gamma = 1/(2N)$; single-point crossover was applied to *all* members of the population, i. e. $f = 1$ (the meanings of these parameters are described in the sections on selection, mutation and crossover). As the distribution evolves the average fitness of the population increases but the width of the distribution decreases until an equilibrium is reached.

Our goal, for each operator, is to calculate the distribution after that operator is applied given the distribution before it is applied. In principle, this is impossible for mutation and crossover, because these operators depend upon the configuration of the strings and that cannot be deduced from the distribution of fitnesses. (As an example, consider the case when the fitness distribution is a delta-function. That could mean that all of the strings are identical, or that they are all different but have the same fitness). Thus, we make an ansatz – that the configuration of the strings is maximally mixed given the distribution and a few other properties (the average inter-string correlation, see section 4). This is a maximum entropy ansatz, to apply it we use statistical mechanical techniques to compute the effect of each operator on the distribution.

In addition to this ansatz, there are two sources of approximation in our approach. In order to describe the distribution $\rho(\mathcal{F})$, we use a cumulant expansion (see, for example, [14, page 928]). The first two cumulants are the mean and variance of the distribution respectively. The higher order cumulants measure the distortion on the distribution away from a Gaussian (see Appendix A for details). To compute the full evolution of the GA we calculate the average mapping between the cumulants produced by each of the operators. These are then applied iteratively to the initial cumulants. The cumulant expansion is exact when an infinite number of cumulants are considered, but this leads to an infinite number of equations in an infinite number of variables. We truncate the cumulant expansion. The second source of error occurs due to sample to sample fluctuations that arise because of the stochastic nature of the operators. We calculate average behaviour. Cumulants self-average, but the dynamical equations are non-linear, and non-linear functions of cumulants need not self-average. For operators which have a non-linear dependence on the cumulants this approximation leads to systematic errors in the estimate of the average behaviour. We discuss these fluctuations in Appendix C.

Summary of Results and Discussion

The bulk of the paper contains the derivation of the dynamics. Here we give an indication of the effectiveness of the approach. Figure 2 shows the average evolution of the first four cumulants for the case $N = 127$, $P = 50$, $\beta_s = 0.2$, $\gamma = 1/N$ and $f = 1$ (the choice of parameters was motivated by the desire to make the graphs as readable as possible). The solid curves show the simulation results while the dashed curves show the theoretical predictions. The typical features of the simulations are that the theory is very accurate for short times; the prediction of the long term behaviour is qualitatively correct but often quantitatively wrong by a few percent for the first two cumulants.

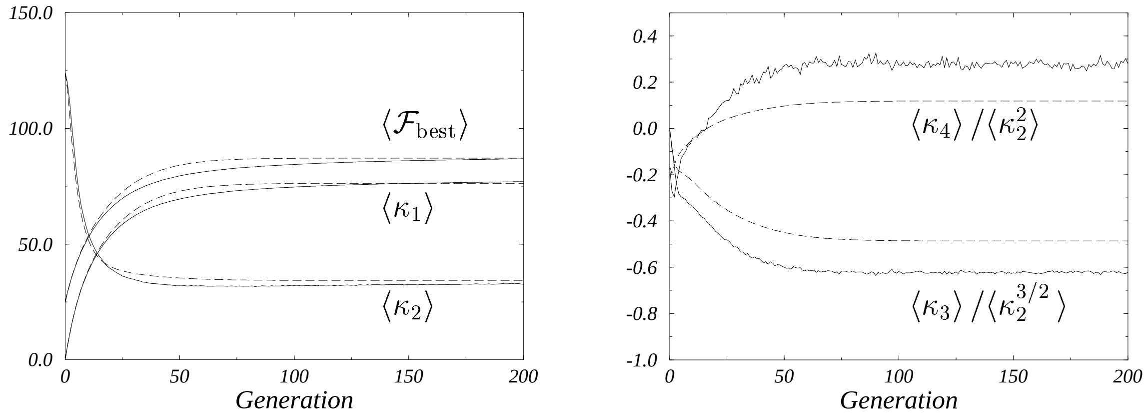


Figure 2: The solid curves show the evolution of the average of the first four cumulants and the fitness member of the best member of the population for the case $N = 127$, $P = 50$, $\beta_s = 0.2$, $\gamma = 1/N$ and $f = 1$. Again the results where averaged was over 10 000 samples. The dashed curve shows the theoretical prediction.

We know that one source of the discrepancy arises from ignoring the sample to sample fluctuations which are small, but accumulate throughout the run. Evidence that fluctuations introduces systematic error is discussed in Appendix C – when the nonlinear cumulants are treated more carefully, the discrepancy in the first two cumulants can be removed (in the special case we have considered). There is also clearly error introduced via the truncation of the cumulant expansion. This is most marked in the prediction of higher cumulants.

For a Genetic Algorithm to work efficiently the parameters, such as the selection and mutation rate, must be carefully chosen. One application of our approach is to help understand how these parameters govern the search. An example of what needs to be understood is shown in figure 3. This shows simulation results of best member of the population after 200 iterations versus the selection rate, for $N = 127$, $P = 50$, and at different mutation rates γ. For changing values of the mutation parameter, the optimal selection rate changes from small to large. In section 6.3 we study a model dynamics based on our calculation of the full dynamics. The model dynamics ignores some of the less important non-linear terms and so can be solve in closed form. This provides a semi-quantitative understanding of the role played by the various operators. For example it illustrates how the convergence rate will depend on the selection and mutation rate and how crossover improves performance by reducing the third cumulant.

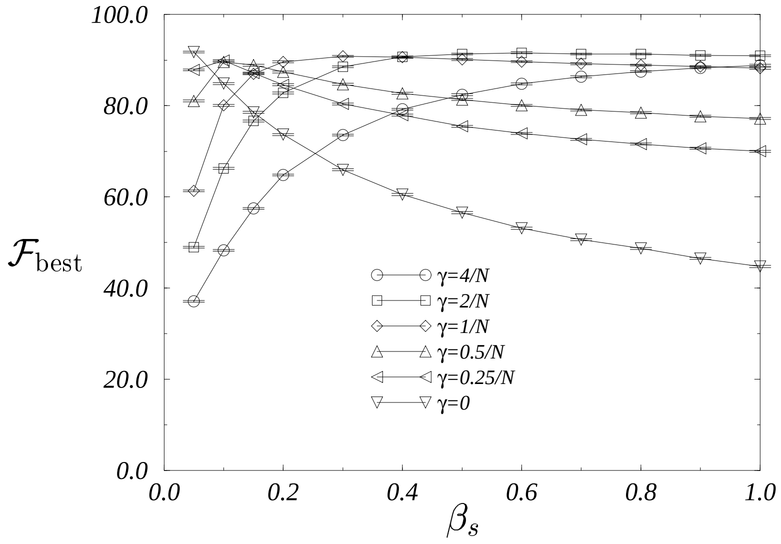


Figure 3: $\mathcal{F}_{\text{best}}$ versus β_s at different mutation rates after 200 generation. The chain consisted on $N = 127$ bonds; the population size was $P = 50$; Deterministic selection was used. The data points show the average over 1 000 simulations; the error bars show the estimated errors in the mean.

In the remainder of the paper, we give the technical details showing how to derive the dynamical equations, and give some approximate solutions in simple cases. This is done in terms of cumulants, κ_n denotes the nth cumulant. For each operator, we compute the cumulants after that operator is applied in terms of those before. The cumulant expansion is described in Appendix A. Herein, we consider the first four cumulants: the mean, the variance, and the third and fourth cumulant which are related to the skewness and kurtosis respectively.

The rest of the paper is organized is as follows:

Section 2: Selection. The effect of selection on the distribution of fitness is calculated for "roulette wheel" selection where the member of the population are weighted with their "Boltzmann" weight. This calculation is carried out for a finite size population using the technique used by Derrida for solving random energy model. The calculation depends only on the distribution of fitnesses and not on the particular string configurations, so the approach is problem-independent. Extension of this model to other types of selection are discussed in §6.2.

Section 3: Mutation. Mutation is calculated exactly for the first two cumulants. For the higher cumulants, the effect of mutation is found to depend on configurational properties of the strings. These configurational properties are estimated in the next section using a statistical mechanics model. The calculations, particularly for higher cumulants, are facilitated using diagram techniques; these are discussed in appendix D.

Section 4:The Statistical Mechanical Approach. The effects of mutation and crossover depend upon the configuration of strings as well as the distribution of fitnesses. These are found by modelling the population with a random population with the same first two cumulants. In this section, we present the statistical mechanics calculation to find properties of the population configuration, and present direct comparisons with measurements of these properties from simulations to see whether the assumption is appropriate.

Section 5: Crossover. We compute the effect of uniform crossover for the random-field paramagnet and single-point crossover on the spin-glass chain using similar techniques to those use for mutation. The effect of single-point crossover on the spin-glass chain is complicated because crossover can potentially break bonds and because single-point crossover introduces spatial correlations. We estimate the "interface energy" produced by crossover at the crossover site by using a statistical mechanics approach to infer the probability of breaking a bond. We estimate the mixing effect of crossover by assuming it is the same as for uniform crossover with a variable bias so that child will sometimes inherit more from one parent than the other.

Section 6: The Full Dynamics. In this section, we describe how the previous sections are put together to produce the formalism which results in the theoretical curve shown in figure 2. We also discuss possible extensions which we have not looked at in detail: deterministic selection, Truncation selection and performing crossover only on some of the population. Finally, we discuss a very simplified dynamics generated by keeping only a few cumulants and dropping most nonlinear terms and high order terms in the expansions. The remaining equations are not quantitatively correct, however; they illustrate how the final result depends upon the parameters of the model, may be useful in choosing optimum search parameters, and provide a picture of how GAs work.

Section 7: Conclusion. The approach to studying GAs reveals much about how they work and what are important features in understanding them. Suggestions for further work are presented.

Appendix A: Representing the Distribution of Fitnesses. To calculate the effect of selection we represent the distribution of fitnesses by a function which depends on the cumulants. In this appendix we show how this function is generated from the cumulants.

Appendix B: Finite Population Effects. Throughout the paper there are corrections due to the finite size of the population, which we discuss here.

Appendix C: Sample to Sample Fluctuations. One of the main assumptions made throughout the paper is that the statistics of the population do not fluctuate. Because the population is finite this assumption is incorrect and there are fluctuations of order $1/\sqrt{P}$. In this appendix we illustrate that these fluctuations can lead to systematic errors in the estimate of the average behaviour of the cumulants because selection depends non-linearly on some of the cumulants.

Appendix D: Diagrammatics for Mutation and Crossover. We show how graphs can be used to facilitate the calculation of mutation and crossover.

Appendix E: The Initial Distribution. To compute the full dynamics we need to know the initial distribution of fitnesses, which we calculate here.

2 Selection

2.1 The Selection Procedure

Selection is the operation whereby more fit strings are increased in the population at the expense of less fit ones. There are a number of methods of selecting the fitter strings. In most of this paper we consider "Roulette wheel" selection. A probability or weight, p_α, is assigned to each member of the population according to its fitness, and the members of the new population are drawn independently from the old population with a frequency proportional to their weight. This selection procedure has the drawback that by chance some good members of the population may be lost. An improved selection procedure is "deterministic" selection [15] (which also goes by other names) where each member is chosen with a frequency as close to Pp_α as possible. A third selection procedure is Truncation selection where the top p members of the population are chosen and reproduced to create the next generation. We have studied only roulette wheel in detail; in §6.2 we show that similar analysis could be applied to deterministic selection and Truncation selection.

As well as different selection procedures there are also many different methods for assigning weights to the members of the population. The most common is to weight each string in proportion to its fitness. The method we have used is to weight each member of the population with its "Boltzmann probability"

$$p_\alpha = \frac{\mathrm{e}^{\beta \mathcal{F}_\alpha}}{Z}, \qquad Z = \sum_{\alpha=1}^{P} \mathrm{e}^{\beta \mathcal{F}_\alpha}. \tag{2.1}$$

The parameter β controls the strength of selection; when $\beta = 0$ all members of the population have the same weight, while as $\beta \to \infty$ only the fittest solution is selected. This selection procedure has been used previously, for example, see [16]. This selection procedure has several advantages over fitness proportional weighting. There is an explicit parameter, β, controlling selection. The fitnesses need not be positive numbers; in fitness proportional selection they must be in order to be interpreted as probabilities. Boltzmann selection is unchanged by adding a constant to the fitness. Furthermore by setting β inversely proportional to the width of the fitness distribution ($\sqrt{\kappa_2}$) selection is unchanged by a rescaling of the fitnesses. This invariance has the advantage that the effect of selection is roughly constant throughout the entire run of the GA (it is not exactly constant as selection also depends on the higher cumulants which change during a run). This rescaling plays the same role as linear fitness rescaling [4] — a discussion of the connections between these two types of scaling is given in [3]. In the present paper we have used scaled selection with a selection parameter defined by $\beta_s = \beta\sqrt{\kappa_2/2\log(P)}$.

The cumulants after selection are given by

$$\kappa_n^s = \frac{\partial^n}{\partial t^n} \log\left(\sum_{\alpha=1}^{P} n_\alpha \mathrm{e}^{t\mathcal{F}_\alpha}\right)\Bigg|_{t=0} \tag{2.2}$$

where n_α is the number of times member α has been selected. On average $\langle n_\alpha \rangle = P p_\alpha$. For deterministic selection the n_α will differ from its average value by at most 1 while in roulette wheel selection the fluctuations in n_α will be of order $P p_\alpha(1 - p_\alpha)$.

2.2 Roulette Wheel Selection With Boltzmann Weights

In Roulette wheel selection each new member is independently chosen from the old population with a probability given by their Boltzmann weight. This is exactly equivalent to the problem of solving the thermodynamics of Derrida's Random Energy Model (REM) [17]. In the REM the system occupies a state with a probability equal to its Boltzmann weight where the states are drawn independently from a distribution (in the original REM this was taken to be Gaussian).

We model the effect of roulette wheel selection by choosing P solutions independently form a continuous distribution with the correct first few cumulants (see Appendix A) and weighting each by the probability, p_α, of selecting it. To calculate the average effect of roulette wheel selection we just average over all possible choices. Since we have chosen Boltzmann weight the cumulant generating function (2.2) simplifies to

$$\kappa_n^s = \frac{\partial^n}{\partial \beta^n} \left\langle \log(Z) \right\rangle_\rho \tag{2.3}$$

where Z is given in equation (2.1) and

$$\langle \log(Z) \rangle_\rho = \left[\prod_{\alpha=1}^{P} \int_{-\infty}^{\infty} \rho(\mathcal{F}_\alpha)\, \mathrm{d}\mathcal{F}_\alpha\right] \log(Z). \tag{2.4}$$

Provided the states are reasonably dense we would expect this to be a good approximation. It is likely to be most inaccurate when the population has almost converged so that all members of the population occupy only a few states (in this case a cumulant expansion may not be the best way to describe the distribution). This problem is similar to that found in the low temperature phase of the REM [18, 19]. However, for our problem the GA will spend almost all its time in reaching this state.

Selection depends critically on the size of the population — for an infinite population the optimal solution could be obtained by choosing a sufficiently large selection rate. To evaluate equation (2.4) for a finite population we use Derrida's trick of writing the logarithm in terms of the integral representation

$$\log(Z) = \int_0^\infty \frac{\mathrm{e}^{-t} - \mathrm{e}^{-tZ}}{t}\, \mathrm{d}t. \tag{2.5}$$

This decouples the sum over strings in the logarithm. Averaging over the fitness distribution we find

$$\langle \log(Z) \rangle_\rho = \int_0^\infty \frac{\mathrm{e}^{-t} - f^P(t,\beta)}{t} \, \mathrm{d}t \tag{2.6}$$

where

$$f(t,\beta) = \int_{-\infty}^{\infty} \rho(\mathcal{F}) \, \mathrm{e}^{-t\mathrm{e}^{\beta\mathcal{F}}} \, \mathrm{d}\mathcal{F}. \tag{2.7}$$

From equations (2.3) and (2.7) the (average) cumulants after selection are given by

$$\kappa_n^s = -P \int_0^\infty \frac{\partial^n}{\partial \beta^n} f^P(t,\beta) \, \frac{\mathrm{d}t}{t}. \tag{2.8}$$

The cumulants can therefore be calculated by performing the double integral over $\mathcal{F}$ and t numerically.

We can also obtain an analytical expression for the effect of selection in the weak selection or "high temperature" limit (β small). This is useful for understanding many of the properties of selection. For small β we can expand the term $\exp(-t\mathrm{e}^{\beta\mathcal{F}})$ in equation (2.7) (this is valid only when $\mathcal{F} \ll \log(t)/\beta$, however, the error coming from large $\mathcal{F}$ is suppressed by the $\rho(\mathcal{F})$ factor provided β is sufficiently small). Equation (2.7) becomes

$$f(t,\beta) = \int_{-\infty}^{\infty} \rho(\mathcal{F}) \left(1 - t\mathrm{e}^{\beta\mathcal{F}} + \frac{t^2}{2}\mathrm{e}^{2\beta\mathcal{F}} + \cdots\right) \, \mathrm{d}\mathcal{F} = 1 - t\hat{\rho}(\beta) + \frac{t^2}{2}\hat{\rho}(2\beta) + \cdots \tag{2.9}$$

where $\hat{\rho}(n\beta)$ is the moment generating function defined by

$$\hat{\rho}(n\beta) = \int_{-\infty}^{\infty} \rho(E) \, \mathrm{e}^{n\beta\mathcal{F}} \, \mathrm{d}\mathcal{F}. \tag{2.10}$$

Exponentiating the first term and raising to the power P, equation (2.9) becomes

$$f^P(t,\beta) = \mathrm{e}^{-tP\hat{\rho}(\beta)} \left(1 + \frac{Pt^2}{2}\left(\hat{\rho}(2\beta) - \hat{\rho}(\beta)^2\right) + \cdots\right) \tag{2.11}$$

substituting back into equation (2.6) and performing the integral over t we find

$$\langle \log(Z) \rangle_\rho = \log\left(P\hat{\rho}(\beta)\right) - \frac{1}{2P}\left(\frac{\hat{\rho}(2\beta)}{\hat{\rho}^2(\beta)} - 1\right) \tag{2.12}$$

where we have used equation (2.5) to evaluate the first term. The average cumulants after selection, κ_n^s, are, to order $1/P$ and for small β, related to those before selection by

$$\kappa_n^s = \frac{\partial^n}{\partial \beta^n}\left(\sum_{m=1}^{\infty} \kappa_m \frac{\beta^m}{m!} - \frac{1}{2P}\exp\left\{\sum_{l=2}^{\infty}(2^l - 2)\kappa_l \frac{\beta^l}{l!}\right\}\right). \tag{2.13}$$

(Setting $\beta = 0$, gives the leading finite population effects, equations (B.3–B.5)).

From equation (2.13) we find the first few cumulants are

$$\kappa_1^s = \kappa_1 + \beta\left(1-\frac{1}{P}\right)\kappa_2 + \frac{\beta^2}{2}\left(1-\frac{3}{P}\right)\kappa_3 + \frac{\beta^3}{3!}\left[\left(1-\frac{7}{P}\right)\kappa_4 - \frac{6}{P}\kappa_2^2\right] + \cdots, \quad (2.14)$$

$$\kappa_2^s = \left(1-\frac{1}{P}\right)\kappa_2 + \beta\left(1-\frac{3}{P}\right)\kappa_3 + \frac{\beta^2}{2}\left[\left(1-\frac{7}{P}\right)\kappa_4 - \frac{6}{P}\kappa_2^2\right] + \cdots, \quad (2.15)$$

$$\kappa_3^s = \left(1-\frac{3}{P}\right)\kappa_3 + \beta\left[\left(1-\frac{7}{P}\right)\kappa_4 - \frac{6}{P}\kappa_2^2\right] + \cdots. \quad (2.16)$$

If we truncate the cumulant expansion we obtain a set of closed expressions. To leading order in β and P the first cumulant is shifted by $\beta\kappa_2$. To the same order the second cumulant is reduced by a factor $1-(1+3\beta^2\kappa_2+\cdots)/P$ and shifted by $\beta\kappa_3$. Even when $\beta=0$, the variance is reduced by a factor $1-1/P$. This arises from random fluctuations — by chance some members of the population will be selected by more than their weight while other members of the population by less. The fractional change in the variance κ_2^s/κ_2 as a function of the scaled selection parameter $\beta_s=\beta\sqrt{\kappa_2/2\log(P)}$ for a Gaussian is shown in figure 4, for $P=2^5$, 2^{10} and 2^{20}. The dashed lines show the small β expansion, $\kappa_2^s/\kappa_2 = 1-(1-4\log(P)\beta_s^2)\exp(4\log(P)\beta_s^2)/P$. Notice that increasing the selection rate β_s initially has little effect on κ_2^s but increases the mean fitness (whose initial growth is proportional to β_s). However, we observe that the curves have a shoulder above which the population converges rapidly (this shoulder is a reflection of the first order phase transition that occurs in the infinite system at $\beta_s=1$). For high values of β the increase in the mean fitness saturates. This suggests an optimum in the selection rate around the shoulder of the curves — above that value selection removes diversity from the population, below that value selection increases the mean fitness less than it could. It is also important to note that in a finite population, the third cumulant, or skewness, is driven negative by selection. This is because the solutions around the best solution are amplified while those in the low energy tail are thinned out. This leads to a densely populated region at high fitness and a sparsely populated tail at low fitness, thus giving rise to a negative skewness. From equation (2.8) we observe that κ_3^s for the Gaussian is proportional to the derivative of the curves in figure 4. If one of the genetic operators could reduce the magnitude of the skewness without affecting the mean or variance, this would increase the fittest members of the population, thereby improving the search. We shall see shortly that this is a possible role of crossover.

Notice that in equations (2.14) – (2.16), κ_2 always occurs with the factor $(1-1/P)$; κ_3 with the factor $(1-3/P)$ and so forth. These are precisely the expected values of the cumulants when you draw P data points from a parent population with cumulants κ_1, κ_2, etc., to order $1/P$ (because equation (2.12) is valid only to order $1/P$). This makes perfect sense, because selection is indeed the act of drawing data from a parent population. These factors recur throughout this work, and are described in Appendix B.

3 Mutation

The mutation operator causes random changes to individual bits of the genetic strings. Thus, it instigates local exploration, and is like the move sets of many other search

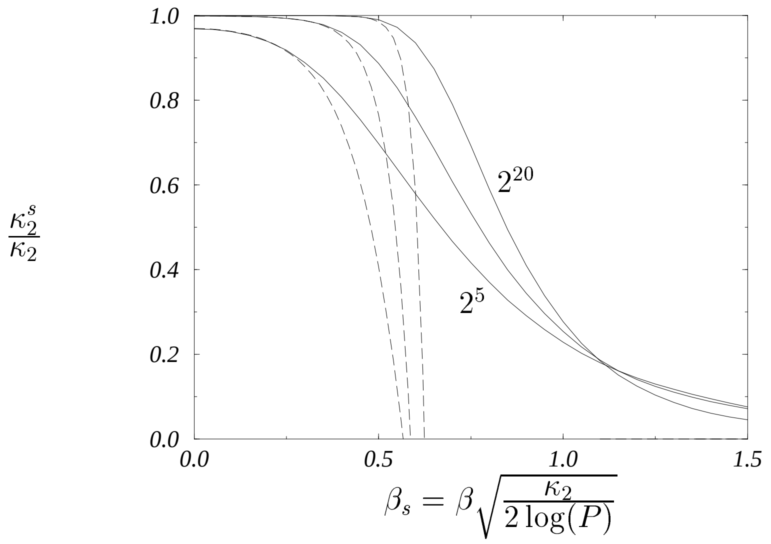


Figure 4: The change in width of fitness distribution after selection starting from a Gaussian distribution for $P = 2^5$, 2^{10} and 2^{20}. The solid lines are calculated by numerical integration. The dashed lines show the small β expansion.

algorithms. Mutation acts on each member of the population independently. We consider both mutation on the random-field paramagnet and on the spin-glass chain.

Unlike selection, the effect of mutation (and crossover) on a string will not depend only on the distribution of fitnesses, it will also depend on the configuration of strings. For example, mutation will depend upon whether the strings in the population get their fitnesses by satisfying a few strong bonds (J_i's) or many small bonds. (In fact these configurational effects do not alter the change in the mean fitness caused by mutation but show up in the higher order statistics.) This effect will manifest itself by a dependence of higher cumulants on the average bond satisfaction at a site

$$\langle \tau_i^\alpha \rangle = \frac{1}{P}\sum_{\alpha=1}^{P} \tau_i^\alpha,$$

as a function of the bond strength J_i. Other configurational averages such as the average inter-string correlation will also be important.

To calculate the effect of mutation we must be able to estimate the configurational properties from the few macroscopic quantities we have kept. We do this using a maximum entropy calculation — we assume that we have chosen an appropriate set of macroscopic quantities in order to be able to infer the configurational information we need. We first calculate the effect of mutation in terms of cumulants and the configurational statistics; we defer the calculation of these configurational quantities to §4.

3.1 Mutation for the Random-Field Paramagnet

Here we consider mutation of the variables of the random-field problem, τ_i^α (the superscript α denotes which member of the population it is).

To calculate the average effect of mutation, we introduce a spin-flip operator σ_i^α which flips a spin on string α at site i. The spins are flipped with a probability γ, so that $\tau_i^\alpha \to \tau_i^{m\alpha} = \sigma_i^\alpha \tau_i^\alpha$ where

$$\sigma_i^\alpha = \begin{cases} 1 & \text{with probability } 1-\gamma \\ -1 & \text{with probability } \gamma. \end{cases} \tag{3.1}$$

The σ_i^α are independent of one another; they average to

$$\langle \sigma_i^\alpha \rangle_{\text{mut}} = 1 - 2\gamma \equiv \Gamma. \tag{3.2}$$

The fitness of a string after mutation is

$$\mathcal{F}_\alpha^m = \sum_i J_i \, \sigma_i^\alpha \, \tau_i^\alpha. \tag{3.3}$$

To calculate the average effect of mutation we need to average moments of $\mathcal{F}_\alpha^m$ over the population and all mutations, i.e.

$$\kappa_1^m = \left\langle \, \langle \mathcal{F}_\alpha^m \rangle_\alpha \, \right\rangle_{\text{mut}} \tag{3.4}$$

$$\kappa_2^m = P_2 \left\langle \left\langle (\mathcal{F}_\alpha^m)^2 \right\rangle_\alpha - \left\langle \mathcal{F}_\alpha^m \mathcal{F}_\beta^m \right\rangle_{\alpha\neq\beta} \right\rangle_{\text{mut}} \tag{3.5}$$

etc. ($P_2 = 1 - 1/P$ is the correction factor for a finite population, see Appendix B). We can commute the order in which we take the averages since the mutation on each string is independent. To calculate the average cumulants after mutation we calculate the moment for a single string and average over all possible mutations: we then average over all members of the population. So averaging equation (3.3) over all possible mutations, and using equation (3.2),

$$\langle \mathcal{F}_\alpha^m \rangle_{\text{mut}} = \Gamma \sum_i J_i \, \tau_i^\alpha = \Gamma \, \mathcal{F}_\alpha. \tag{3.6}$$

Averaging over all members of the population we find the average fitness after mutation is

$$\kappa_1^m = \left\langle \, \langle \mathcal{F}_\alpha^m \rangle_{\text{mut}} \, \right\rangle_\alpha = \Gamma \, \langle \mathcal{F}_\alpha \rangle_\alpha = \Gamma \kappa_1 \tag{3.7}$$

This averaging is quite straight-forward, although a graphical representation is useful in getting the combinatorial factors correct, as described in Appendix D. The results are as follows:

$$\kappa_1^m = \Gamma \kappa_1, \tag{3.8}$$

$$\kappa_2^m = \Gamma^2 \kappa_2 + P_2(1-\Gamma^2) \sum_i J_i^2, \tag{3.9}$$

$$\kappa_3^m = \Gamma^3 \kappa_3 - 2\Gamma(1-\Gamma^2) P_3 \sum_i J_i^3 \, \langle \tau_i^\alpha \rangle_\alpha \, , \tag{3.10}$$

$$\begin{aligned} \kappa_4^m = {} & \Gamma^4 \kappa_4 - 8 P_4 \Gamma^2 (1-\Gamma^2) \sum_i J_i^4 \Big(1 - \left\langle \tau_i^\alpha \tau_i^\beta \right\rangle_{\alpha\neq\beta}\Big) - 2P_4(1 - 4\Gamma^2 + 3\Gamma^4) \sum_i J_i^4 \\ & - \frac{6}{P} \left[P_2 (1-\Gamma^2)^2 (\sum_i J_i^2)^2 - 2\Gamma^2(1-\Gamma^2) \kappa_2 \sum_i J_i^2 \right]. \end{aligned} \tag{3.11}$$

Here $P_3 = 1 - 3/P + 2/P^2$ and $P_4 = 1 - 7/P + 12/P^2 - 6/P^3$.

Observe that the first two cumulants after mutation can be expressed in terms of the cumulants before mutation. The higher cumulants cannot. They depend on the configuration through $\langle \tau_i^\alpha \rangle_\alpha$ and $\langle \tau_i^\alpha \tau_i^\beta \rangle_{\alpha \neq \beta}$. In §4 we show how to estimate this using statistical mechanics.

As a check on our calculation we observe that the cumulants after mutation, equations (3.8), – (3.11), satisfy the conditions that $\kappa_n^m = \kappa_n$ when $\gamma = 0$ ($\Gamma = 1$) and $\kappa_n^m = (-1)^n \kappa_n$ when $\gamma = 1$ ($\Gamma = -1$). Also when $\gamma = 1/2$ ($\Gamma = 0$), so that the spins are flipped with a probability 1/2, then $\kappa_n^m = \kappa_n^0$, the cumulants for a random set of strings. Finally when the cumulants are initially random, $\kappa_n = \kappa_n^0$, then they remain unchanged for all Γ.

We note in passing that the fluctuations for a single string is given by

$$\left\langle (\mathcal{F}_\alpha^m)^2 \right\rangle_{\text{mut}} - \left\langle \mathcal{F}_\alpha^m \right\rangle_{\text{mut}}^2 = (1 - \Gamma^2) \sum_i J_i^2 = 4\gamma(1-\gamma) \sum_i J_i^2. \tag{3.12}$$

The fluctuation in κ_1^m will therefore be of the order of $\sqrt{4\gamma(1-\gamma)N/P}$, which are smaller than the change in κ_1 by a factor of $\sqrt{N(1-\gamma)/(P\gamma\kappa_1^2)}$.

3.2 Mutation for the Spin-Glass Chain

Flipping a spin S_i^α in the spin-glass chain causes two defects, τ_{i-1}^α and τ_i^α to change. Mutation of the spin variables is therefore approximately equivalent to defect mutation but with twice the mutation rate. (We will see there are a few corrections to this approximation.)

We can use the same techniques to calculate the effect of mutation in the spin-glass chain as we used in the previous section. Each spin is flipped with a probability γ and spin-flip variables σ_i^α are used to express this: $S_i^\alpha \to S_i^{m\alpha} = \sigma_i^\alpha S_i^\alpha$. Since the spins of the random-field problem are defects variables in the current problem, and flipping a spin in the spin-glass chain causes two defects to change, mutation in the spin-glass chain is approximately equivalent to mutation in the random-field problem with twice the mutation rate (but with corrections due to effects of flips at neighbouring sites). The fitness after mutation is

$$\mathcal{F}_\alpha^m = \sum_i J_i S_i^{m\alpha} S_{i+1}^{m\alpha} = \sum_i J_i \sigma_i^\alpha \sigma_{i+1}^\alpha S_i^\alpha S_{i+1}^\alpha \tag{3.13}$$

averaging over all mutations

$$\langle \mathcal{F}_\alpha^m \rangle_{\text{mut}} = \Gamma^2 \sum_i J_i S_i^\alpha S_{i+1}^\alpha = \Gamma^2 \mathcal{F}_\alpha \tag{3.14}$$

Averaging over all strings

$$\kappa_1^m = \Gamma^2 \kappa_1. \tag{3.15}$$

The calculation of the higher cumulants include an additional set of terms from those for defect mutation arising from flipping neighbouring spins. The second moment is

$$\left\langle (\mathcal{F}_\alpha^m)^2 \right\rangle_{\text{mut}} = \Gamma^4 \mathcal{F}_\alpha^2 + (1 - \Gamma^4) \sum_i J_i^2 + 2\Gamma^2(1 - \Gamma^2) \sum_i J_i J_{i+1} \tau_i^\alpha \tau_{i+1}^\alpha \tag{3.16}$$

Averaging over all strings in the population

$$\kappa_2^m = \Gamma^4 \kappa_2 + P_2(1-\Gamma^4) \sum_i J_i^2 + 2P_2 \Gamma^2 (1-\Gamma^2) \sum_i J_i J_{i+1} \left\langle \tau_i^\alpha \tau_{i+1}^\alpha \right\rangle_\alpha \tag{3.17}$$

The first two terms are identical to the defect variable result except with γ is replaced by $2\gamma(1-\gamma)$, or $\Gamma \to \Gamma^2$. Ignoring spatial correlations (and ignoring boundary effects) the last term is approximately proportional to

$$\sum_i J_i J_{i+1} \left\langle \tau_i^\alpha \tau_{i+1}^\alpha \right\rangle_\alpha \approx \frac{\kappa_1^2}{N} \tag{3.18}$$

For small γ, the last term is of the order of $8\gamma(\kappa_1)^2/N$. Initially this is small compared with the change in κ_2 due to the first two terms (which is order $8\gamma N$) and only becomes significant when the members of the population are very good. This term can be seen as a consequence of the local minima; similar terms appear in the higher cumulants. When the population is close to the ground state these terms become important. In this limit spatial correlations are likely to be very important.

The third cumulant is

$$\kappa_3^m = \Gamma^6 \kappa_3 - 2\Gamma^2(1-\Gamma^4) P_3 \sum_i J_i^3 \left\langle \tau_i^\alpha \right\rangle_\alpha + 6\Gamma^2 (1-\Gamma^2)^2 P_3 \sum_i J_i J_{i+1} J_{i+2} \left\langle \tau_i^\alpha \tau_{i+1}^\alpha \tau_{i+2}^\alpha \right\rangle_\alpha \tag{3.19}$$

The last terms is approximately proportional to κ_1^3/N^2, respectively.

The equations we have obtained for the cumulants after spin mutation satisfy all the checks described at the end of the last section except now flipping all the spins ($\gamma = 1$, $\Gamma = -1$) leaves all the cumulants unchanged as we would expect. (In contrast flipping all the defects changes all the odd cumulants $\kappa_n^m = -\kappa_n$.)

4 The Statistical Mechanics Approach

4.1 The Statistical Mechanics Ansatz

In the previous section we showed that the effect of mutation on the cumulants depends upon the fitness distribution (i.e. on the cumulants) *and* upon properties of the configuration of defects, specifically on the average defect per site and on the average correlation per site. As we have argued, one cannot, in principle, determine the configuration from the fitness distribution. The configuration depends upon which bonds are satisfied in each string, while the fitness only depends on the average bond satisfaction per string. In other words, one cannot tell which strings are in the population from which fitnesses are in the population.

We infer the constituents of the population by ansatz. We make the assumption often made in statistical mechanics, that the string variables are free to fluctuate subject to the constraint that a few "macroscopic" quantities are known, namely the fitness cumulants. In other words, our assumption is that of all the configurations which have the same distribution of fitnesses, the more numerous ones are more likely to describe the actual one. This can be seen as a maximum entropy assumption, the distribution of strings in the population is assumed to be the one which maximises the entropy given the fitness distribution. (For a good treatment of the relationship between entropy and statistical mechanics, see section II of [20].)

It should be emphasized that this is an assumption; it need not be true about GAs. It is likely to be true if the genetic operators are sufficiently effective at mixing so that all populations with given fitness properties are equally likely. It is also a requirement of the effectiveness of the approach that we chose the correct macroscopics; that the fitness properties chosen are appropriate for characterizing the system. In the next section, we describe a simple statistical mechanical model which implements the statistical mechanics assumption for the simple spin system we have looked at. We present direct evidence that the assumption is correct for these simple systems.

4.2 The Statistical Mechanics Model

We present a simple statistical mechanics model from which we can compute the configuration from two cumulants of the fitness distribution. The fluctuating quantity in this model is "average bond satisfaction" at each site

$$t_i = \langle \tau_i^\alpha \rangle_\alpha = \frac{1}{P} \sum_{\alpha=1}^{P} \tau_i^\alpha . \tag{4.1}$$

We calculate the distribution of satisfied bonds $P(t_i)$ for a particular coupling J_i via statistical mechanics. The average bond satisfaction and the higher order statistics can then be found by averaging over this distribution.

To find the distribution $P(t_i)$ we average over all possible values of t_i subject to the constraints that the average fitness is κ_1 and the average correlation is q. To impose these constraints, we introduce two Lagrange multipliers, z which couples to the average energy per site and x^2 which couples to the average correlation per site. In other words, the "effective Hamiltonian" of this statistical mechanics model consists of two terms,

$$z \sum_\alpha \sum_i J_i \tau_i^\alpha = zP \sum_i J_i t_i \tag{4.2}$$

and

$$\frac{x^2}{2} N \sum_\alpha \sum_\beta q^{\alpha\beta} = \frac{x^2}{2} \sum_\alpha \sum_\beta \sum_i \tau_i^\alpha \tau_i^\beta = \frac{x^2}{2} P^2 \sum_i t_i^2 . \tag{4.3}$$

The partition function for this model is

$$Z = \prod_i Z_i = \prod_i \operatorname*{Tr}_{\{t_i\}} \mathcal{N}(t_i) \mathrm{e}^{zPJ_i t_i + P^2 x^2 t_i^2/2} . \tag{4.4}$$

The number of ways of satisfying $P(1+t_i)/2$ of the bonds is given by the binomial coefficient

$$\mathcal{N}(t_i) = \frac{1}{2^P} \binom{P}{P(1+t_i)/2} . \tag{4.5}$$

Using Stirling's approximation the entropy is given by

$$S(t_i) = \log\left(\mathcal{N}(t_i)\right) = -\frac{P}{2} \log\left(1 - t_i^2\right) + \frac{Pt_i}{2} \log\left(\frac{1-t_i}{1+t_i}\right) \tag{4.6}$$

Using this to replace t_i with continuous variables and using the familiar properties of Gaussians to decouple the quadratic term yields for the partition function,

$$Z_i = \int dt_i \int \frac{\mathrm{d}\eta_i}{\sqrt{2\pi}} \mathrm{e}^{-\eta_i^2/2 + P\,G(t_i, J_i, \eta_i)} \tag{4.7}$$

where

$$G(t_i, J_i, \eta_i) = s(t_i) + zJ_it_i + x\eta_it_i \tag{4.8}$$

and $s(t_i) = S(t_i)/P$. The t_i integral is done by steepest descents, the value of t_i which dominates is given by the saddle point equation

$$\frac{\partial G}{\partial t_i} = 0 = \frac{\partial s(t_i)}{\partial t_i} + zJ_i + x\eta_i = -\mathrm{atanh}(t_i) + zJ_i + x\eta_i. \tag{4.9}$$

The full free-energy is found by summing on i. Derivatives in terms of z yield the cumulants,

$$\kappa_1 = \sum_i J_i \overline{\tanh(zJ_i + x\eta)} \tag{4.10}$$

$$\kappa_2 = P_2 \sum_i J_i^2 \Big(1 - \overline{\tanh^2(zJ_i + x\eta)}\Big), \tag{4.11}$$

where the bar denotes the average over the Gaussian noise η. The Lagrange multiplier x and z are chosen so that the first two cumulants take their measured values. (The factor P_2 in equation (4.11) is the usual finite-population correction; otherwise the calculation assumes infinite P.)

The required configurational averages are computed from the free energy of this model. The average correlation q, is found by differentiating in terms of x^2; this yields

$$q = \sum_i \overline{\tanh^2(zJ_i + x\eta)}. \tag{4.12}$$

The average bond satisfaction as a function of the bond strength can be found by averaging over all sites with bond strength between J_i and $J_i + dJ_i$. This gives

$$t_i = \overline{\tanh(zJ_i + x\eta_i)}. \tag{4.13}$$

All quantities required to compute the effects of mutation and crossover can be calculated in a similar manner.

The assumption we have made in deriving the distribution are that there are no spatial correlations and that the operators do not introduce any systematic bias into the distribution other than determining the average fitness and the average correlation. One test of how well these assumptions hold for the spin-glass chain is to directly measure the average value of the defect variable as a function of the bond strength. An example of such an experiment is shown in figure 5. This shows the average of $\langle \tau_i^\alpha \rangle_\alpha$ and its square over 10 000 simulations for all bonds whose couplings lie in a bin of size 0.1 after 25 generations. These values are plotted against J_i (solid lines) together with the theoretical prediction (dashed lines) where x and z are chosen to satisfy equations (4.10) and (4.11). There is a small discrepancy between the theoretical curves and the simulations which are probably due to neglecting spatial correlations. Nevertheless, at least in this example, the theoretical predictions seem to be in reasonable agreement with simulations. Since a maximum entropy assumption is a mixing assumption, it is significant that this assumption holds in the transient regime. We will show later that using this assumption, we get good predictions over a wide range of generations (c.f. figure 2).

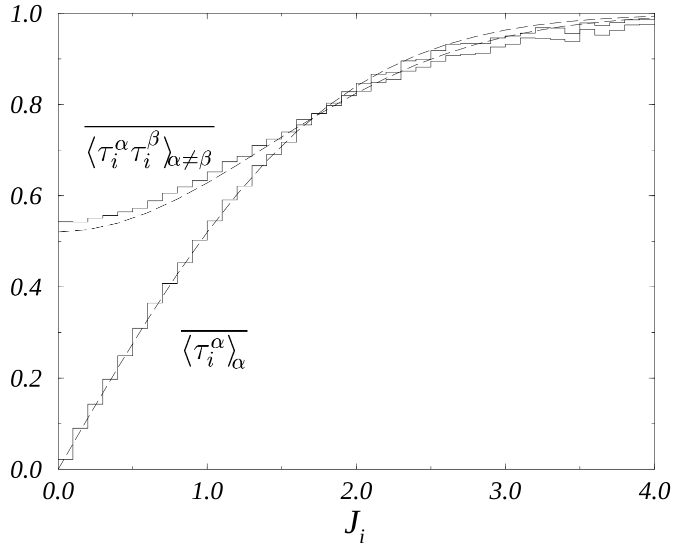


Figure 5: The curves show the average of $\langle\tau_i^\alpha\rangle_\alpha$ and $\langle\tau_i^\alpha\tau_i^\beta\rangle_{\alpha\neq\beta}$ measured from simulations (solid lines) and the theoretical prediction (dashed curve) plotted against the coupling strength J_i. The parameters are the same as those used in figure 2.

To calculate the effect of crossover for average couplings we replace the sum over J_i with an integral weighted by $P(J_i)$. The terms $\sum_i J_i^n \langle\tau_i^{\alpha_1}\cdots\tau_i^{\alpha_m}\rangle_{\alpha_1\neq\cdots\neq\alpha_m}$ are replaced by their average value

$$f_{n,m}(x,z) = N\int\frac{\mathrm{d}J}{\sqrt{2\pi}}\mathrm{e}^{-J^2/2}\int\frac{\mathrm{d}\eta}{\sqrt{2\pi}}\mathrm{e}^{-\eta^2/2}J^n\tanh^m(zJ+x\eta). \tag{4.14}$$

(Rather than evaluate these integrals directly it is useful to transform to polar coordinates r and θ define by $J=\sqrt{2r}\cos\theta$ and $\eta=\sqrt{2r}\sin\theta$. After some rearrangement the integrals then decouple.)

5 Crossover

Crossover is the operator in which two strings combine to produce new solutions. Crossover is a very interesting operator as it gives Genetic Algorithms their distinctive flavour; it is also the most complicated to analyse. Not only does it perform highly non-local moves in the space of solutions, it also depends on a pair of solutions. This means the effect of crossover will depend on pairwise statistics. Like mutation the effects of crossover will depend on the configurational properties as well as the fitness distribution.

There are several ways of performing crossover. The simplest to calculate is "uniform crossover" where a child is constructed from its two parents by choosing each spin at random from either of the parents. This form of crossover is most effective when there is little interaction between the spins. For the random-field paramagnet this is a very effective form of crossover — we calculate the effect of uniform crossover on the random-field paramagnet in the next section. However, for the spin-glass chain uniform crossover will typically result in a large number of broken bonds so that the child would be very unlikely to have the same fitness as its parents. For this problem a much superior crossover is "single-point crossover" where a bond is selected at random and the child is produced by taking all the spins on the left of the bond from the first parent and all the spins on the right from the second parent. Thus at most only one bond will be broken. There will be on average a small cost or "interface energy" due to breaking this bond. Interface energies will arise in a large number of problems. For a Genetic Algorithm to be efficient the benefits of mixing must be sufficient to compensate for this interface energy. In calculating the effect of crossover on the spin-glass chain we separate the effect of mixing and the interface energy. The effect of mixing we estimate by assuming that it is similar to that for biased uniform crossover. In doing this we ignore spatial correlations which causes single-point crossover to explore the solution space slower than uniform crossover. The mixing effect of crossover is discussed in §5.2 and the interface energy is calculated in §5.3.

5.1 Crossover for the Random-Field Paramagnet

For the random-field paramagnet we use a uniform crossover where the strings are paired and a new child is produced by randomly choosing spins from either of the parent strings. The crossover can be biased by choosing spins from one parent with a probability a and from the other parent with a probability $1-a$. For this problem there is no loss in fitness due to crossover and the optimum value for a is $1/2$. Nevertheless we will calculate the effect of mixing for arbitrary a as it will allow us to apply these results to the spin-glass chain.

We consider crossing two strings $\vec{\tau}^\alpha$ and $\vec{\tau}^\beta$ to form a child $\vec{\tau}^c$. As with mutation, we introduce crossover variables χ_i^c such that the spins of the child will be $\tau_i^c = \chi_i^c \tau_i^\alpha + (1-\chi_i^c)\tau_i^\beta$, where

$$\chi_i^c = \begin{cases} 1 & \text{with probability } a \\ 0 & \text{with probability } 1-a \end{cases} \tag{5.1}$$

Thus on average $A = aN$ spins will come from $\vec{\tau}^\alpha$ and the rest from $\vec{\tau}^\beta$. There is a different and independent set of χ_i^c's for each α β pair. After crossover the child will have fitness

$$\mathcal{F}_c = \sum_i J_i \Big(\chi_i^c \tau_i^\alpha + (1-\chi_i^c)\tau_i^\beta \Big) \tag{5.2}$$

averaging over all possible crossovers, and using $\langle \chi_i^c \rangle_\chi = a$, we find

$$\langle \mathcal{F}_c \rangle_\chi = a\mathcal{F}_\alpha + (1-a)\mathcal{F}_\beta. \tag{5.3}$$

Averaging over all pairs of parents α and β

$$\kappa_1^c = \langle a\mathcal{F}_\alpha + (1-a)\mathcal{F}_\beta \rangle_{\alpha\neq\beta} = \kappa_1 \tag{5.4}$$

Thus the first cumulant is unchanged on average. We have assumed that all the children are independently chosen. If we had generated the complementary child ($\tau_i^{\bar{c}} = (1-\chi_i^c)\tau_i^\alpha + \chi_i^c \tau_i^\beta$) then κ_1 would be conserved exactly by crossover.

To calculate the higher order moments we use the property that $(\chi_i^c)^n = \chi_i^c$, $(1-\chi_i^c)^n = 1-\chi_i^c$ and $\chi_i^c(1-\chi_i^c) = 0$. The terms contributing to the higher moments can again be represented in terms of graphs (see appendix D). After averaging over all possible crossovers the second moment is

$$\Big\langle \mathcal{F}_c^2 \Big\rangle_\chi = a^2 \sum_{i\neq j} J_i J_j \tau_i^\alpha \tau_j^\alpha + (1-a)^2 \sum_{i\neq j} J_i J_j \tau_i^\beta \tau_j^\beta + 2a(1-a) \sum_{i\neq j} J_i J_j \tau_i^\alpha \tau_j^\beta + \sum_i J_i^2. \tag{5.5}$$

Adding and subtracting diagonal terms this becomes

$$\left\langle \mathcal{F}_c^2 \right\rangle_\chi = a^2 \mathcal{F}_\alpha^2 + (1-a)^2 \mathcal{F}_\beta^2 + 2a(1-a)\mathcal{F}_\alpha \mathcal{F}_\beta + 2a(1-a) \sum_i J_i^2 (1 - \tau_i^\alpha \tau_i^\beta). \tag{5.6}$$

Averaging over all pairs of parents and subtracting off $\langle \mathcal{F}_c \rangle_\chi^2$ we find

$$\kappa_2^c = \left(a^2 + (1-a)^2 \right) \kappa_2 + 2a(1-a) P_2 \sum_i J_i^2 \left(1 - \left\langle \tau_i^\alpha \tau_i^\beta \right\rangle_{\alpha \neq \beta} \right). \tag{5.7}$$

However assuming no spatial correlations so that $\left\langle \tau_i^\alpha \tau_j^\alpha \right\rangle_\alpha = \left\langle \tau_i^\alpha \tau_j^\beta \right\rangle_{\alpha \neq \beta}$ for $i \neq j$, then

$$\left\langle \mathcal{F}_\alpha^2 \right\rangle_\alpha - \left\langle \mathcal{F}_\alpha \mathcal{F}_\beta \right\rangle_{\alpha \neq \beta} = \sum_i J_i^2 \left(1 - \left\langle \tau_i^\alpha \tau_i^\beta \right\rangle_{\alpha \neq \beta} \right). \tag{5.8}$$

substituting this into equation (5.7) we find $\kappa_2^c = \kappa_2$. Thus, on average, crossover does not change the second cumulant either. This is a consequence of the assumption about spatial correlations which is implicit in the statistical mechanics model which we use to calculate configuration averages. For the models we have studied this assumption is found to be valid at least as a good approximation. This may not be true in general, in which case a more sophisticated approach would be required (see [21]).

Crossover also leaves the average correlation in the population unchanged since

$$q = \left\langle q^{\alpha\beta} \right\rangle_{\alpha \neq \beta} = \frac{1}{N} \sum_i \left\langle \tau_i^\alpha \tau_i^\beta \right\rangle_{\alpha \neq \beta} \tag{5.9}$$

and, although crossover swaps defects between the chains, it leaves them at the same site so $\langle \tau_i^\alpha \tau_i^\beta \rangle_{\alpha \neq \beta}$ is unchanged.

The third cumulant is computed in a similar manner. Of the terms contributing to $\mathcal{F}_c^3$ the only complicated one is

$$\left\langle \mathcal{F}_\alpha \sum_i J_i^2 \tau_i^\alpha \tau_i^\beta \right\rangle_{\alpha \neq \beta} = \left\langle \mathcal{F}_\gamma \sum_i J_i^2 \tau_i^\alpha \tau_i^\beta \right\rangle_{\alpha \neq \beta \neq \gamma} + \sum_i J_i^3 \left(\langle \tau_i^\alpha \rangle_\alpha - \left\langle \tau_i^\alpha \tau_i^\beta \tau_i^\gamma \right\rangle_{\alpha \neq \beta \neq \gamma} \right) \tag{5.10}$$

Averaging over all pairs of parents and using the definition of the third cumulant,

$$\kappa_3^c = \left(a^3 + (1-a)^3 \right) \kappa_3 - 6a(1-a) P_3 \sum_i J_i^3 \left(\langle \tau_i^\alpha \rangle_\alpha - \left\langle \tau_i^\alpha \tau_i^\beta \tau_i^\gamma \right\rangle_{\alpha \neq \beta \neq \gamma} \right) \tag{5.11}$$

The second term on the right hand side of equation (3.10) is estimated from a maximum entropy argument. The fourth cumulant after crossover is

$$\begin{aligned} \kappa_4^c \;=\; & (a^4 + (1-a)^4)\kappa_4 \\ & -4\mathcal{A}(2-\mathcal{A}) \left[P_4 \sum_i J_i^4 \left(1 - 4 \left\langle \tau_i^\alpha \tau_i^\beta \right\rangle_{\alpha \neq \beta} + 3 \left\langle \tau_i^\alpha \tau_i^\beta \tau_i^\gamma \tau_i^\delta \right\rangle_{\alpha \neq \beta \neq \gamma \neq \delta} \right) + \frac{3}{P} P_2 \kappa_2^2 \right] \end{aligned} \tag{5.12}$$

where $\mathcal{A} = a(1-a)$.

(Note that in calculating the effect of crossover we did not need to know the distribution of string correlations $q^{\alpha\beta} = \vec{\tau}^\alpha \cdot \vec{\tau}^\beta / N$. This was avoided by averaging over all pairs of parents which then gives the effect of crossover in terms of $\langle \tau_i^\alpha \rangle_\alpha$, etc. These can be estimated from a knowledge of κ_1 and κ_2 using the results of §4.)

We see that crossover does not change the first two cumulants. The third and fourth cumulants are reduced in magnitude and evolved towards their "natural values". The equations for the evolution of the cumulants can be written,

$$\kappa_3^c = \kappa_3 - 3\mathcal{A}\left(\kappa_3 - \kappa_3^{nat}\right) \tag{5.13}$$

$$\kappa_4^c = \kappa_4 - 2\mathcal{A}(2-\mathcal{A})\left(\kappa_4 - \kappa_4^{nat}\right). \tag{5.14}$$

The fixed point values are given by

$$\kappa_3^{nat} = -2P_3 \sum_i J_i^3 \left(\langle \tau_i^\alpha \rangle_\alpha - \left\langle \tau_i^\alpha \tau_i^\beta \tau_i^\gamma \right\rangle_{\alpha\neq\beta\neq\gamma} \right) \tag{5.15}$$

$$\kappa_4^{nat} = -2P_4 \sum_i J_i^4 \left(1 - 4\left\langle \tau_i^\alpha \tau_i^\beta \right\rangle_{\alpha\neq\beta} + 3\left\langle \tau_i^\alpha \tau_i^\beta \tau_i^\gamma \tau_i^\delta \right\rangle_{\alpha\neq\beta\neq\gamma\neq\delta} \right) - \frac{6}{P} P_2 \kappa_2^2. \tag{5.16}$$

These are the maximum entropy values for the cumulants given the constraints on the first two cumulants. The rate at which these higher cumulants approach their natural values is a measure of the efficiency of the mixing produced by crossover. This rate depends on the bias a. The most efficient mixing occurs when $a = 1/2$. When a takes this value, the higher cumulants return to their "natural values" most quickly. However, for many problems the interface energy may grow with the bias so that their may be some optimal bias less than 1/2. (For problems like the random-field paramagnet where there is no interface energy an alternative crossover strategy would be to select the spins for each child from any of the parents — this procedure, called Bit-Based Simulated Crossover, has been previously studied [22]. This procedure would cause maximal mixing and would reduce the higher cumulants immediately to their "natural values". We have not, however, studied this crossover procedure since it is only useful in a very special class of problems.)

5.2 Crossover for the Spin-Glass Chain

In the spin-glass chain there are interactions between spins, thus uniform crossover will tend to break bonds at many sites and result in children which are of much lower fitness than the parent strings. A better form of crossover to use for this problem is single-point crossover. To implement single-point crossover, a bond A is chosen at random; all the spins on the left of site A are taken from the first parent and the remaining spins are taken from the second parent. Thus the child produced by crossing two chains $\vec{S}^\alpha$ and $\vec{S}^\beta$ at a bond J_A will have fitness

$$\mathcal{F}_c = \sum_{i=1}^{A-1} J_i S_i^\alpha S_{i+1}^\alpha + J_A S_A^\alpha S_{A+1}^\beta + \sum_{i=A+1}^{N} J_i S_i^\beta S_{i+1}^\beta. \tag{5.17}$$

The only bond which can be changed by crossover is at the interface. We can treat the "interface energy" separately by writing equation (5.17) as

$$\mathcal{F}_c = \sum_{i=1}^{A} J_i S_i^\alpha S_{i+1}^\alpha + \sum_{i=A+1}^{N} J_i S_i^\beta S_{i+1}^\beta + \Delta_I \tag{5.18}$$

where Δ_I, the interface energy, is

$$\Delta_I = J_A S_A^\alpha (S_{A+1}^\beta - S_{A+1}^\alpha). \tag{5.19}$$

The interface energy can be either positive or negative or zero. As the fitness of the population improves the interface energy is more likely to be negative than positive, although as the population correlates it is also more likely to be zero. For crossover to be effective we would like to have a small interface energy and a large degree of mixing.

To make the interface energy small it is important to use an effective representation of the problem. If we had (perversely) represented the spin-glass chain by a randomly ordered string then single-point crossover would be likely to break many bonds ($2A(1 - A/N)$ bonds on average), and the child would be unlikely to have a similar fitness to its parents. We could also reduce the interface energy by choosing the crossing point so that no fitness is lost. However, this trick is not always feasible. For example in higher dimensional spin-glasses the interface energy comes from a hyper-surface. The "cutting plane" could be relaxed to reduce the interface energy but this would introduce an additional computational cost.

To estimate the effect of mixing produced by single-point crossover we assume that it is the same as the mixing produced by uniform crossover where the bias a is chosen at random (since, on average, the same number of spins are swapped between the parent chains). Using the results from the previous section we find that again the first two cumulants are unchanged while the third and fourth cumulants are changed by

$$\begin{aligned}
\kappa_3^c &= \frac{\kappa_3}{2} - P_3 \sum_i J_i^3 \Big(\langle \tau_i^\alpha \rangle_\alpha - \left\langle \tau_i^\alpha \tau_i^\beta \tau_i^\gamma \right\rangle_{\alpha\neq\beta\neq\gamma} \Big) && (5.20)\\
\kappa_4^c &= \frac{2\kappa_4}{5} - \frac{6}{5}\left[P_4 \sum_i J_i^4 \Big(1 - 4\left\langle \tau_i^\alpha \tau_i^\beta \right\rangle_{\alpha\neq\beta} + 3\left\langle \tau_i^\alpha \tau_i^\beta \tau_i^\gamma \tau_i^\delta \right\rangle_{\alpha\neq\beta\neq\gamma\neq\delta} \Big) + \frac{3}{P} P_2 \kappa_2^2 \right] && (5.21)
\end{aligned}$$

In making this approximation we have ignored spatial correlations which build up and reduces the mixing rate. Thus, this approximation is likely to over estimate the effectiveness of single-point crossover in reducing the higher cumulants.

5.3 The Interface Energy

For the spin-glass chain there is a contribution to the fitness at the site of crossover, an interface energy, (5.19). Although it is relatively small, it decreases the average fitness at every generation so its effect accumulates. To calculate the effect of the interface energy we write the fitness after crossover, equation (5.18), as $\mathcal{F}_c = \mathcal{F}_d + \Delta_I$, where $\mathcal{F}_d$ is equal to the sum of the two subchains not including the interface. For the spin-glass chain, the interface makes a order $1/N$ contribution to the fitness. This is sufficiently small that a simple approximate method will be sufficient. For more realistic problems, the interface contribution may be significant, and a more rigorous treatment may be required.

In essence, there is a contribution to the interface when the two spins on either side of the crossover site on one parent are the opposite of those on the other parent. If the bond at the crossover site is a satisfied bond, fitness is decreased by $2J_i$, if it is unsatisfied, fitness is increased by the same amount, otherwise fitness is unchanged at the interface (assuming that both complementary children are produced). Thus, in earlier work [2], we wrote this effect as follows,

$$\langle \Delta_I \rangle \approx -\frac{(1-c)\kappa_1}{N}, \tag{5.22}$$

where

$$\left\langle \frac{1}{N} \sum_i S_i^\alpha S_i^\beta \right\rangle_{\alpha \neq \beta} = c. \tag{5.23}$$

The factor involving the correlation gives the probability that the spins at the interface are opposite. Thermodynamically it would be zero; it is not because selection tends to correlate the population. We estimated this from the second cumulant by assuming that the defects are independent of the couplings (this is a crude approximation, but sufficient accurate for estimating the interface energy), then

$$\kappa_2 = \sum_i J_i^2 (1 - \left\langle S_i^\alpha S_{i+1}^\alpha \right\rangle_\alpha^2) = N(1-q) \tag{5.24}$$

where q is the average defect correlation. If we assume that the defect and spin correlation arise from the same process (so that $c = q$) then the first cumulant is approximately[1] $\kappa_1 \kappa_2 / N^2$.

We continue to use this simple approximation. The nature of the approximation is that it treats the spins as spatially independent. If we write the first cumulant after crossover as $\kappa_1^c = \kappa_1^d + \langle \Delta_I \rangle$, where we have averaged over all possible crossover sites and all pairs of parents, the interface contribution is,

$$\langle \Delta_I \rangle = \frac{1}{N} \sum_{A=1}^{N} J_A \left\langle S_A^\alpha (S_{A+1}^\beta - S_{A+1}^\alpha) \right\rangle_{\alpha \neq \beta}. \tag{5.25}$$

Treating the spin correlation as independent of the coupling J_i,

$$\left\langle S_A^\alpha S_A^\beta \right\rangle_\beta \approx c, \tag{5.26}$$

and $\left\langle S_A^\beta \right\rangle_\beta \approx c S_A^\alpha$. This gives the result (equation 5.22).

The second moment is given by

$$\kappa_2^c = \left\langle \mathcal{F}_c^2 \right\rangle - \left\langle \mathcal{F}_c \right\rangle^2 = \kappa_2^d + 2\Big(\langle \mathcal{F}_d \Delta_I \rangle - \langle \mathcal{F}_d \rangle \langle \Delta_I \rangle \Big) + \left\langle \Delta_I^2 \right\rangle - \left\langle \Delta_I \right\rangle^2 \tag{5.27}$$

assuming no spatial correlations and using the same approximation as that used above

$$\left\langle \Delta_I^2 \right\rangle - \left\langle \Delta_I \right\rangle^2 \approx \frac{2(1-c)}{N} \sum_i J_i^2 \left\langle \tau_i^\alpha \right\rangle_\alpha^2 - \frac{(1-c)^2 \kappa_1^2}{N^2}. \tag{5.28}$$

The sum in the first term is approximately equal to $N - \kappa_2 / P_2$; this is zero for a random distribution but grows as the population correlates. The change in the higher cumulants are all of order 1 and therefore are less important for the higher cumulants. In this section we have made assumptions which are not strictly correct and which will give rise to errors. However, because of the smallness of the interface energy these errors are in practice not very significant. To obtain a more accurate approximation we would need to use a maximum entropy argument similar to that developed in section 4; this would be necessary in models where the interface energy was larger.

[1] Note this disagrees with the result given in reference [2] by a factor of 2

6 The Full Dynamical Equations

6.1 Overview

In the previous four sections, we have shown how to derive the equations which describe simple genetic operators. Here we show how to put these equations together. The result is a mathematical system which accurately predicts the dynamics of the simple Genetic Algorithm operating on the random-field paramagnet and the spin-glass chain.

The idea is to produce the distribution of fitnesses at time $t+1$ from the distribution of fitnesses at time t by computing the changes to this distribution caused by selection, mutation, and crossover. In our simulations, the fitness distribution is characterized by the first four cumulants. This procedure is then iterated an arbitrary number of times to get from the initial distribution to the final one.

First we must calculate the initial conditions. In the limit of infinite strings, the initial distribution will be Gaussian. It is straight-forward to compute the first four cumulants in the finite system; this is done in appendix E. The effect of selection on these four cumulants is found by numerically integrating the integrals produced by applying (2.8) to equations (2.7) and (2.6). Note that in equation (2.7) knowledge of the full distribution $\rho(\mathcal{F})$ is required. However, $\rho(\mathcal{F})$ is not known; only the first four cumulants of $\rho(\mathcal{F})$ are known. To deal with this, we use a model form of the distribution which has the correct four cumulants; how this is done is shown in appendix A.

Mutation operates next on the cumulants. For the random-field paramagnet the equations are (3.8)–(3.11); there are comparable equations for the spin-glass chain. These equations involve configurational averages which are computed from the cumulants by invoking the statistical mechanics ansatz and computing these averages from the statistical mechanics (using equations (4.10) and (4.11) to set the Lagrange multipliers).

The effect of crossover on the cumulants is computed via equations (5.11) and (5.13) for uniform crossover acting on the random-field paramagnet and equations (5.20) and (5.21) together with the interface energy equations (5.22) – (5.28) for the spin-glass chain. As in mutation, the configurational averages are computed from the statistical mechanics model.

The result of iterating this procedure 200 times starting from from initial state is shown in figure 2 at the beginning of the paper for a typical run of the GA. The calculation takes a few seconds. In that figure, the four cumulants are shown computed as described above and as measured in runs of a real GA. Also shown is a comparison of the best member of the population as predicted by our theory and as measured in GA simulations. (How to compute the fitness of the best string is shown in appendix A.) Although there are clearly systematic errors in the predictions of the theory, it does predict the asymptotic best fitness and the search time extremely well. These two quantities are very important characteristics of a search algorithm: what is the quality of the solutions found and how long it takes to find them. However, this does not only represent a fairly accurate method of predicting the behaviour of a GA, it is, as far as we know, the only method of predicting GA dynamics on a system of realistic size and with a non-trivial fitness function.

There are three potential sources of error in our calculation. The assumption that the relationship between the fitness cumulants and configurational average is predicted by the statistical mechanical model may be wrong. Some systematic errors are apparent in figure 5. The truncation of the cumulant expansion is a second source of errors. Obviously, one could go to higher order, although the calculation of the mutation and crossover effects would be quite complicated. The third source of error is in sample-to-sample fluctuations. The transformation which predicts the cumulants at a given time in terms of those in the previous time is nonlinear. Although cumulants are self-averaging quantities of which sample-to-sample fluctuations should be small, nonlinear functions of cumulants are not necessarily self-averaging. The way to rectify this problem would be to introduce an additional set of variables — the sample variance for each cumulant. These would then be evolved along with the cumulants which would now depend upon these new variables as well as the other cumulants. We have not studied this approach, however[2].

6.2 Possible Extensions

In the GA literature there are many forms of Genetic Algorithms used; what we have described is the simplest version. It is possible to extend the calculations to include many of the common extensions. We show some of these below.

6.2.1 Deterministic Selection

In Roulette wheel selection, the strings are chosen probabilistically, with a probability of string α given by p_α. In Deterministic selection [15], the strings are chosen deterministically. The number of times string α is chosen, n_α, is as close as possible to Pp_α found by rounding up or down. Thus, n_α fluctuates by at most one between different trials. These fluctuations will slightly reduce the cumulants, but much less than in Roulette wheel selection. If we ignore these fluctuation then we can approximate the effect of selection using the parent population values for the cumulants rather than the expected values of a finite sample. For example, if we assume κ_1 does not fluctuate, then the expected value of κ_2 is not reduced by the factor $(1 - 1/P)$. Thus, we can estimate the effect by assuming each member of the new population is selected independently from the "infinite" distribution. That is the distribution whose cumulants are the unbiased estimators for the infinite population (obtained by inverting equations (B.3–B.5) in Appendix B). The qualitative shape of the selection curves are the same in both Deterministic selection and Roulette wheel selection: consequently much of the discussion given in the previous section also applies to Deterministic selection. In §6.3, we show how the choice of selection procedure can lead to different final populations.

6.2.2 Truncation Selection

In both Roulette wheel selection and Deterministic selection, a string's presence in the population is relative to its fitness. A very fit string could completely dominate a population. An alternative approach is simply to chose the fittest r (with $1 \leq r \leq P$) members of the population and reproduce these. These can then be reproduced in proportion to fitness, or with equal probability to create the next generation. This method is called

[2]This suggestion has subsequently been carried out for the case of a biological model with just mutation and selection [23]. This provides a much more accurate model of the dynamics; more information about the evolution, namely the expected sample to sample fluctuations; and a more coherent framework, as fluctuations are taken properly into account.

Truncation selection. To illustrate the flexibility of our approach, we discuss how Truncation selection can be study in this formalism. However, we have not studied this selection procedure in any detail.

For simplicity, assume that the fittest r members of the population are reproduced with equal probability. It is simplest to treat this by replacing it with a selection procedure in which all members of the population with a fitness greater than $\mathcal{F}_0$ are selected, where $\mathcal{F}_0$ is chosen so that on average r members of the population will be chosen. Clearly $\mathcal{F}_0$ must satisfy

$$\int_{\mathcal{F}_0}^{nat} \rho(\mathcal{F}) \, \mathrm{d}\mathcal{F} = \frac{r}{P}. \tag{6.1}$$

This allows r to fluctuate, with fluctuations of order $1/\sqrt{P}$ which arise due to the finite population. (Whereas, our treatment of Roulette wheel selection used Boltzmann statistics, Truncation selection uses Fermi-Dirac statistics. In allowing r to fluctuate, we are in essence treating it in a grand canonical ensemble). Within this approximation we can calculate selection as described in §2 where now equation (2.2) becomes

$$\kappa_n^s = \frac{\partial^n}{\partial t^n} \log \left(\sum_\alpha \Theta(\mathcal{F}_\alpha - \mathcal{F}_0) \mathrm{e}^{t\mathcal{F}_\alpha} \right) \Bigg|_{t=0} \tag{6.2}$$

where $\Theta(x)$ is the usual step function. Using the same trick as for Boltzmann selection we find that on average the moments are given by

$$\mu_n^s = \frac{M_n(\mathcal{F}_0)}{M_0(\mathcal{F}_0)} \tag{6.3}$$

where

$$M_n(\mathcal{F}_0) = \int_{\mathcal{F}_0}^{nat} \mathcal{F}^n \, \rho(\mathcal{F}) \, \mathrm{d}\mathcal{F} \tag{6.4}$$

(We have ignored corrections to order $1/P$ as these are less important than for Boltzmann selection.) The cumulants being constructed from the moments in the usual way. These results appear intuitively obvious. Their simplicity is probably a result of using the weak constraint of fixing $\mathcal{F}_0$ so that on average r members of the population are chosen, rather than the imposing the stronger constraint that exactly r members of the population are chosen.

6.2.3 Sub-unity Crossover Rate

We have allowed all strings to participate in crossover. In applications of GAs, most researchers allow a percentage of the old population to be copied into the new population without crossing over. This is important in problems where the probability of a child having a similar or better fitness than its parents is low. In our problems it was found in simulations that there was no advantage in using sub-unity crossover — this was presumably because the interface energy for these problems was so low. We show here that it is straightforward to calculate the effect of combining populations with different distributions. Assuming that we take a fraction f of our new population from a population with distribution $\rho^a(\mathcal{F})$ and the rest from a population with a distribution $\rho^b(\mathcal{F})$ then

$$\rho(\mathcal{F}) = f\rho^a(\mathcal{F}) + (1-f)\rho^b(\mathcal{F}). \tag{6.5}$$

multiplying both sides by $e^{\omega\mathcal{F}}$ and integrating over $\mathcal{F}$

$$\log\Big(\hat{\rho}(\omega)\Big) = \sum_n \frac{\kappa_n \omega^n}{n!} = \log\Big(f\hat{\rho}^a(\omega) + (1-f)\hat{\rho}^b(\omega)\Big) \tag{6.6}$$

expanding $\hat{\rho}^a(\omega)$ and $\hat{\rho}^b(\omega)$ in terms of cumulants and equating terms of the same order in ω

$$\kappa_1 = f\kappa_1^a + \bar{f}\kappa_1^b \tag{6.7}$$

$$\kappa_2 = f\kappa_2^a + \bar{f}\kappa_2^b + f\bar{f}\Big(\kappa_1^a - \kappa_1^b\Big)^2 \tag{6.8}$$

$$\kappa_3 = f\kappa_3^a + \bar{f}\kappa_3^b + 3f\bar{f}\Big(\kappa_1^a - \kappa_1^b\Big)\Big(\kappa_2^a - \kappa_2^b\Big) + f\bar{f}(1-2f)\Big(\kappa_1^a - \kappa_1^b\Big)^3 \tag{6.9}$$

etc, where $\bar{f} = 1 - f$. For the spin-glass chain crossover involves only a very small cost on average and there is no improvement in performance by not performing crossover on all members of the population.

6.3 Model Dynamics

In the previous sections, we showed how to develop a mathematical model of Genetic Algorithms. This model was shown to be fairly accurate. However, it is quite complicated, even for the simple systems we have studied, involving multiple integrals and solution of implicit function as well as nonlinear iterative functions. In order to use the formalism to help understand how Genetic Algorithms work, we present here some highly simplified model dynamics. These models are generated by linearizing the equations where possible and truncating the expansions. We treat three types of Genetic Algorithm: one using selection and mutation only; one using selection, mutation and crossover; and finally a GA using selection and crossover only. Roulette wheel and deterministic selection are given a unified treatment. Even though these model dynamics are greatly simplified, they reproduce the qualitative features of the Genetic Algorithm for this problem and help to reveal much about how Genetic Algorithms work.

Selection and Mutation Only

The simplest model which reveals some properties of genetic search consists of mutation and selection only (no crossover) as they effect the first two cumulants. By considering two cumulants only, we are making a Gaussian approximation. This will break down as the optimal state is reached (since there is no high fitness tail there). We shall see later that the third cumulant plays a crucial role in the GA. Nonetheless, the two cumulant model is a useful first approximation.

To two cumulants, the effect of selection can be expressed,

$$\kappa_1^s = \kappa_1 + s_{12}(\tilde{\beta}_s)\sqrt{\kappa_2} \tag{6.10}$$

$$\kappa_2^s = s_{22}(\tilde{\beta}_s)\kappa_2 \tag{6.11}$$

where $\tilde{\beta}_s = \beta\sqrt{\kappa_2} = \sqrt{2\log(P)}\beta_s$. Here we are rescaling the selection rate at each generation so that $\tilde{\beta}_s$ remains constant, as suggested in §2. The coefficients s_{ij} (which couple the i^{th} cumulant after selection to a function of the j^{th} cumulant prior to selection) are functions of the selection rate and the population size, and are constants otherwise. The function $s_{22}(\tilde{\beta}_s)$ is shown in figure 4 for different population sizes, it is initially constant at $(1 - 1/P)$ then has a shoulder and decays to zero. The other coefficient $s_{12}(\tilde{\beta}_s)$ is given by the integral of these curves — it grows initially at a rate $(1-1/P)\tilde{\beta}_s$ but then tails off to an asymptote. Using deterministic selection the function $s_{12}(\tilde{\beta}_s)$ and $s_{22}(\tilde{\beta}_s)$ are rescaled by a factor $1/(1-1/P)$. For small values of the selection parameter

$$s_{12}(\tilde{\beta}_s) \approx \tilde{\beta}_s \left[1 - \frac{1}{P}\left(e^{\tilde{\beta}_s^2} - \zeta\right)\right] \tag{6.12}$$

$$s_{22}(\tilde{\beta}_s) \approx 1 - \frac{1}{P}\left(e^{\tilde{\beta}_s^2}(1+2\tilde{\beta}_s^2) - \zeta\right) \tag{6.13}$$

where $\zeta = 0$ for roulette wheel selection and 1 for deterministic selection. We see that selection increases the mean in proportion to the standard deviation but decreases the variance. Increasing the amount of selection initially increases the average fitness without effecting the variance, but for larger selection rates causes a big loss in genetic diversity. These equations suggest an optimal value for the selection parameter $\tilde{\beta}_s$. It should be chosen around the shoulder of the curve $s_{22}(\tilde{\beta}_s)$, where one get the biggest improvement in the mean fitness for the smallest loss of variance in the population. To simplify the expressions, the explicit dependents of the coefficients on $\tilde{\beta}_s$ will sometimes be dropped.

The effect of mutation on the first two cumulants is to change them by an amount, (see (3.15–3.17)),

$$\begin{aligned} \kappa_1^m &= \Gamma\kappa_1 \\ \kappa_2^m &= \Gamma^2\kappa_2 + (1-\Gamma^2)N. \end{aligned} \tag{6.14}$$

Here $\Gamma = (1-2\gamma)$ and γ is the mutation probability as before. This equation is for the random-field paramagnet. For the spin-glass chain the same equations can be used by replacing Γ with Γ^2 if corrections due to flipping neighbouring spins are ignored. This equation shows that mutation makes the average fitness worse and drives the variance towards that of a random population.

Applying the selection and mutation operators together produces a change in the cumulants of

$$\kappa_1(t) = \Gamma\kappa_1(t-1) + \Gamma s_{12}\sqrt{\kappa_2(t-1)} \tag{6.15}$$

$$\kappa_2(t) = \Gamma^2 s_{22}\kappa_2(t-1) + (1-\Gamma^2)N. \tag{6.16}$$

Using the initial conditions $\kappa_1(0) = 0$ and $\kappa_2(0) = N$ we find that the first two cumulants evolve as

$$\kappa_1(t) = s_{12}\sum_{t'=1}^{t}\Gamma^{t-t'+1}\sqrt{\kappa_2(t'-1)}. \tag{6.17}$$

$$\kappa_2(t) = \frac{(1-\Gamma^2)N}{1-\Gamma^2 s_{22}} + \frac{N\Gamma^2(1-s_{22})}{1-\Gamma^2 s_{22}}\left(\Gamma^2 s_{22}\right)^t. \tag{6.18}$$

The population converge towards a dynamic equilibrium distribution where the improvement due to selection is balanced by a loss of fitness caused by mutation. The variance at equilibrium is

$$\kappa_2(\infty) = \frac{(1-\Gamma^2)N}{1-\Gamma^2 s_{22}} \tag{6.19}$$

and the equilibrium average fitness is

$$\kappa_1(\infty) = \frac{s_{12}\Gamma\sqrt{\kappa_2(\infty)}}{1-\Gamma}. \tag{6.20}$$

From these equations, we can see how the selection rate and the mutation rate affect both the time for the algorithm to converge and the quality of the solutions it converges to. The rate at which $\kappa_2(t)$ converges to its equilibrium value is $-\log(\Gamma^2 s_{22})$. This convergence rate is thus a sum of two factors: the convergence due to mutation and the convergence due to selection. For large selection rates and/or large mutation rate κ_2 converges rapidly towards its asymptotic limit. The convergence of κ_1 is slower. It converges first through κ_2 and then converges at a rate $-\log(\Gamma)$. Thus the final convergence is controlled by the mutation rate alone.

The fitness of the equilibrium population depends upon the amount of selection and the mutation rate. The selection optimum suggested above is born out in the effect of selection on the final fitness. As the selection parameter $\tilde{\beta}_s$ increases from zero, initially $\kappa_1(\infty)$ increases through $s_{12}(\tilde{\beta}_s)$. However, as the selection rate reaches and exceeds an optimal value $s_{22}(\tilde{\beta}_s)$ will decrease rapidly, making $\kappa_2(\infty)$ and thus $\kappa_1(\infty)$ smaller.

The mutation rate also has a dramatic effect on the mean fitness of the equilibrium population. The final fitness improves by decreasing the mutation rate. Note that the fitness of the optimal string is of order N. However, it is apparent from the equations above that the final mean fitness will be of order $\sqrt{N}$ unless the mutation rate is very small, of order $1/N$. If the mutation rate is taken as $\gamma = g/N$, then equations (6.20) and (6.19) become

$$\kappa_2(\infty) \approx \frac{4g}{1-s_{22}} \tag{6.21}$$

$$\kappa_1(\infty) \approx \sqrt{\frac{N^2 s_{12}^2}{g\,(1-s_{22})}}. \tag{6.22}$$

Thus, for large mutation rate the system reaches a point where there is high diversity in the population, but little improvement has occurred; if the mutation rate is very small (order $1/N$) the system can get much closer to the optimum whence the variance will be small (order 1). For a uniform-field paramagnet, the $1/N$ scaling of the optimal mutation rate has been long known (see, for example, reference [24] and references therein). Remember, however, that the convergence rate increases with mutation rate. If the mutation rate is sufficiently small to find a good solution, the convergence will be very slow.

There is a trade-off; small mutation rate gives slow convergence to a good solution while a large mutation rate results in rapid convergence to a poor solution. However, the final values for the cumulants depend solely on the mutation when equilibrium is reached. It is possible that convergence times could be decreased by using a large mutation for early times and then annealing the mutation rate down to a small value. Time-dependent mutation rate could be put into the equations for the first two cumulants in order to study the efficacy of this idea, and to learn about appropriate annealing schedules. We leave that work for another day.

There is a major problem with the model dynamics described above. It appears that one can make the fitness arbitrarily large by a sufficiently small mutation rate. This is nonsense, obviously the fitness cannot increase more than that of the fittest string. This is a consequence of the two-cumulant approximation. The Gaussian approximation is also essentially an infinite population and infinite string length approximation, since a finite population can not be exactly Gaussian and the density of states for a finite population is non-Gaussian. In an infinite population, there are always strings in the tails of the distribution, so you can always increase the fitness[3]. To understand the system for very small mutation rates, for larger selection rates, and to get important finite-population effects, we must go beyond the Gaussian approximation.

We get a more complete picture by including the third cumulant. To do this, we use the small β expansion for selection, equations (2.14)–(2.16), and drop all non-linear terms coupling κ_2 and $\tilde{\kappa}_3 = \kappa_3/\sqrt{\kappa_2}$. This yields a set of coupled equations

$$\kappa_1^s = \kappa_1 + s_{12}(\tilde{\beta}_s)\sqrt{\kappa_2} + s_{13}(\tilde{\beta}_s)\frac{\tilde{\kappa}_3}{\sqrt{\kappa_2}} \tag{6.23}$$

$$\kappa_2^s = s_{22}(\tilde{\beta}_s)\kappa_2 + s_{23}(\tilde{\beta}_s)\tilde{\kappa}_3 \tag{6.24}$$

$$\tilde{\kappa}_3^s = -s_{32}(\tilde{\beta}_s)\kappa_2 + s_{33}(\tilde{\beta}_s)\tilde{\kappa}_3 \tag{6.25}$$

where, for small $\tilde{\beta}_s$,

$$s_{32}(\tilde{\beta}_s) = \frac{\tilde{\beta}_s e^{\tilde{\beta}_s^2}}{P}(6 + 4\tilde{\beta}_s^2) \tag{6.26}$$

$$s_{13}(\tilde{\beta}_s) = \frac{\tilde{\beta}_s^2}{2}\left[1 - \frac{3}{P}(e^{\tilde{\beta}_s^2} - \zeta)\right] \tag{6.27}$$

$$s_{23}(\tilde{\beta}_s) = \tilde{\beta}_s\left[1 - \frac{3}{P}\left(e^{\tilde{\beta}_s^2}(1 + 2\tilde{\beta}_s^2) - \zeta\right)\right] \tag{6.28}$$

$$s_{33}(\tilde{\beta}_s) = 1 - \frac{3}{P}\left(e^{\tilde{\beta}_s^2}(1 + 9\tilde{\beta}_s^2 + 6\tilde{\beta}_s^4) - \zeta\right) \tag{6.29}$$

(where again $\zeta = 0$ for roulette wheel selection and 1 for deterministic selection). It is important that selection introduces a negative third cumulant (unless the population size is infinite). The tail is long on the low fitness side and short on the high fitness side. This is because the fit solutions are more rare.

We assume the effect of mutation on the third cumulant is to reduce it by

$$\tilde{\kappa}_3^m \approx \Gamma^3\tilde{\kappa}_3 \tag{6.30}$$

where we have ignored the correction terms in (3.10). (These are initially zero but change with κ_1 and κ_2; since they are non-linear they have been ignored in this model dynamics.)

[3]There is a long history of approximating population dynamics by taking the infinite population limit. It is interesting that these results indicate that an approach which treats only the mean and variance of a population is implicitly working in this unrealistic limit.

Within this approximation, the change in the cumulants at each generation is

$$\begin{aligned}
\kappa_1(t) &= \Gamma\kappa_1(t-1) + \Gamma s_{12}\sqrt{\kappa_2(t-1)}\left(1 + \frac{s_{13}}{s_{12}}\frac{\tilde{\kappa}_3(t-1)}{\kappa_2(t-1)}\right) && (6.31)\\
\kappa_2(t) &= \Gamma^2 s_{22}\kappa_2(t-1) + \Gamma^2 s_{23}\tilde{\kappa}_3(t-1) + (1-\Gamma^2)N && (6.32)\\
\kappa_3(t) &= -\Gamma^3 s_{32}\kappa_2(t-1) + \Gamma^3 s_{33}\tilde{\kappa}_3(t-1). && (6.33)
\end{aligned}$$

The second and third cumulants evolve to equilibrium values

$$\begin{aligned}
\kappa_2(\infty) &= \frac{(1-\Gamma^3 s_{33})(1-\Gamma^2)N}{(1-\Gamma^2 s_{22})(1-\Gamma^3 s_{33}) + \Gamma^5 s_{23}s_{32}} && (6.34)\\
\tilde{\kappa}_3(\infty) &= \frac{-\Gamma^3 s_{32}(1-\Gamma^2)N}{(1-\Gamma^2 s_{22})(1-\Gamma^3 s_{33}) + \Gamma^5 s_{23}s_{32}} && (6.35)
\end{aligned}$$

while the equilibrium average energy is

$$\kappa_1(\infty) = \frac{s_{12}\sqrt{\kappa_2(\infty)}}{1-\Gamma}\left(1 + \frac{s_{13}}{s_{12}}\frac{\tilde{\kappa}_3(\infty)}{\kappa_2(\infty)}\right). \qquad (6.36)$$

Comparing these with equations (6.19) and (6.20) we see that the Genetic Algorithm is less effective than the two cumulant model predicts. The third cumulant reduces $\kappa_1(\infty)$ in two ways: it reduces it directly because the equilibrium distribution has a negative skewness, $\tilde{\kappa}_3(\infty)/\kappa_2(\infty)$; it also reduces $\kappa_1(\infty)$ indirectly by reducing the equilibrium variance $\kappa_2(\infty)$. The first of these effects is quite small

$$\frac{s_{13}\tilde{\kappa}_3(\infty)}{s_{12}\kappa_2(\infty)} \overset{\beta,\gamma\ small}{\approx} \frac{\tilde{\beta}_s^2}{1-\zeta+2\alpha g+\mathcal{O}(\tilde{\beta}_s^2)}. \qquad (6.37)$$

where $\alpha = P/N$, and again $\gamma = g/N$. The more significant effect is the reduction in the equilibrium variance caused by the skewness

$$\kappa_2(\infty) = \frac{N(1-\Gamma^2)}{1-\Gamma^2 s_{22} + \Upsilon} \qquad (6.38)$$

where

$$\Upsilon = \frac{\Gamma^5 s_{23}s_{32}}{1-\Gamma^3 s_{33}} \overset{\beta,\gamma\ small}{\approx} \frac{2\tilde{\beta}_s^2}{1-\zeta+2\alpha g+\mathcal{O}(\tilde{\beta}_s^2)}. \qquad (6.39)$$

Note that in the two cumulant approximation $\Upsilon = 0$. Additionally a large negative third cumulant will reduce the position of the best member of the population relative to the average member.

We see that including the third cumulant regulates the final mean fitness. As the mutation rate goes to zero $\kappa_1(\infty)$ remains finite. At the same time, the third cumulant reduces the effectiveness of the Genetic Algorithm. The reason is clear, the distribution is not Gaussian, the high-fitness tail is underpopulated. We shall see that one important role of crossover is to reduce the third cumulant and make the distribution more Gaussian. This reduces the deleterious effect of the skewness seen here.

Selection, Mutation and Crossover

We now include crossover to our three cumulant model. We have shown that crossover reduces higher cumulants. The third cumulant is reduced towards it "natural value" by a factor of $1/r$ (where $r = 2$ for the spin-glass chain with single-point crossover, see equation (5.20), and $r = 4$ for the random-field paramagnet with uniform crossover, see equation (5.11) with $a = 1/2$). The natural value is initially zero but grows with κ_1 and κ_2. Neglecting this natural value, the third cumulant after crossover is $\tilde{\kappa}_3^c = \tilde{\kappa}_3/2$. For the random-field paramagnet, this is the only effect of crossover. For the spin-glass chain, crossover also reduces the mean fitness via the "interface energy".

For a GA using selection, mutation and crossover, κ_2 and $\tilde{\kappa}_3$ evolve to equilibrium values

$$\kappa_2(\infty) = \frac{(r-\Gamma^3 s_{33})(1-\Gamma^2)N}{(1-\Gamma^2 s_{22})(r-\Gamma^3 s_{33})+\Gamma^5 s_{23}s_{32}} \tag{6.40}$$

$$\tilde{\kappa}_3(\infty) = \frac{-r\Gamma^3 s_{32}(1-\Gamma^2)N}{(1-\Gamma^2 s_{22})(r-\Gamma^3 s_{33})+\Gamma^5 s_{23}s_{32}} \tag{6.41}$$

In order to see the effect of the interface term, we will make its contribution explicitly separate. Assuming an interface energy proportional to the average energy $\langle \Delta_I \rangle = -\kappa_1 \mathcal{I}$ then the equilibrium average energy is

$$\kappa_1(\infty) = \frac{s_{12}\sqrt{\kappa_2(\infty)}}{1-\Gamma+\mathcal{I}}\left(1+\frac{s_{13}}{s_{12}}\frac{\tilde{\kappa}_3(\infty)}{\kappa_2(\infty)}\right) \tag{6.42}$$

For the random-field paramagnet $\mathcal{I} = 0$ while for the spin-glass chain $\mathcal{I} = \kappa_2/N^2$ (remember also for the spin-glass chain we replace Γ with Γ^2).

In order to understand the effect of crossover on the equilibrium fitness, we must understand the effects on the skewness and on the second cumulant. Crossover decreases the final skewness and increases the final variance. This may be difficult to see by comparing the equations for the equilibrium fitness with and without crossover, but note that $(1-\Gamma^3 s_{33})$ can be very small, while $(r-\Gamma^3 s_{33})$ is of order 1. The other terms in the denominators are also small.

Thus, the equilibrium fitness is changed due to three effects. First, the skewness is decreased in magnitude (remember that it is negative) which increases the final fitness. Second, the final fitness is proportional to the standard deviation, and this is increased by crossover. Finally, there is the $\mathcal{I}$ in the denominator of equation (6.42). This is the sole effect of the interface term and is a small correction unless the selection rates and mutation rates are very small.

The magnitudes of these effects can be estimated in the small selection limits. The larger effect is on the equilibrium variance. Again the equilibrium variance has the form of equation (6.38), but now Υ is given by

$$\Upsilon = \frac{\Gamma^5 s_{23}s_{32}}{r-\Gamma^3 s_{33}} \overset{\beta,\ \gamma\ small}{\approx} \frac{6\tilde{\beta}_s^2}{P(r-1)+\mathcal{O}(\tilde{\beta}_s^2)}. \tag{6.43}$$

For small $\tilde{\beta}_s$ this is substantially smaller than without mutation. The qualitative behaviour is in fact close to the two cumulant model where the skewness is completely neglected. For small $\tilde{\beta}_s$ and γ, the equilibrium variance is

$$\kappa_2(\infty) \approx \frac{4gP}{4g\alpha + 1 - \zeta + \mathcal{O}(\tilde{\beta}_s^2)} \tag{6.44}$$

where $\zeta = 0$ for Roulette wheel selection and 1 for Deterministic selection. The equilibrium variance initially decreases only slowly with $\tilde{\beta}_s$ so that $\kappa_1(\infty)$ improves initially as the selection rate increases. However, when $\tilde{\beta}_s$ is of order $\sqrt{\log(P)}$ the equilibrium variance begins to decrease substantially. This defines an optimal selection rate above which the average fitness declines.

Crossover also has an effect on the first cumulant coming from the interface energy. For small $\tilde{\beta}_s$ and γ this reduces $\kappa_1(\infty)$ by a factor

$$\frac{1}{1 + (\mathcal{I}/2\gamma)}. \tag{6.45}$$

This is only significant for small $\tilde{\beta}_s$ and when γ is small (so that the disruption due to the interface energy is of the same order as that due to mutation).

Selection and Crossover Only

When there is no mutation the population converges completely. Considering only the first two cumulants and ignoring the interface energy

$$\kappa_2(t) = s_{22}\kappa_2(t-1) = \left(s_{22}\right)^t N \tag{6.46}$$

and

$$\kappa_1(t) = \kappa_1(t-1) + s_{12}\sqrt{\kappa_2(t-1)} = s_{12}\sqrt{N}\left(\frac{1 - \left(\sqrt{s_{22}}\right)^t}{1 - \sqrt{s_{22}}}\right). \tag{6.47}$$

The population therefore converges to

$$\kappa_1(\infty) = \frac{s_{12}\sqrt{N}}{1 - \sqrt{s_{22}}} \overset{\beta_s \ small}{\approx} \frac{2P\tilde{\beta}_s\sqrt{N}}{1 - \zeta + 3\tilde{\beta}_s^2}. \tag{6.48}$$

This calculation shows a very strong advantage for deterministic selection in the large N limit. For Roulette wheel selection ($\zeta = 0$) convergence always occurs at a value far from the optimum. The final equilibrium value is of order $\sqrt{N}$, although choosing the optimum value of $\tilde{\beta}_s$ similar to that we found for selection, mutation and crossover will improve things somewhat. For Deterministic selection ($\zeta = 1$) the best results are obtained by taking β_s as small as possible. In this case, the optimum string can be approached. However, the convergence rate is $-\log(s_{22})/2 \approx (1 - \zeta + 3\tilde{\beta}_s^2)/(2P)$ so the convergence time increases as we decrease the selection rate. In contrast to using selection and mutation alone, keeping diversity in the population is very important when we include crossover.

7 Conclusion

In this paper, we have shown how to derive a set of equations describing dynamics of a Genetic Algorithm. These equations describe the evolution of the first four cumulants of the distributions of fitnesses in the population. The effect of selection depends only on the distribution of fitnesses and is therefore problem independent. The effect of mutation and crossover depends on the problem being tackled and how it is represented; to compute their effects required an assumption about the configurations in the population. The derivation is quite complicated. However, the approach is advantageous in that it predicts the equilibrium and the time to reach it; comparisons with simulations suggest that they do so fairly accurately.

The development of this formalism raises several issues about Genetic Algorithms. First, we have made an assumption about the configurations in the population, namely the statistical mechanics ansatz. This assumption was essential, because without it we could not infer the configuration of strings from the fitness cumulants. That this assumption works suggests that there is "microscopic disorder", that is that the Genetic Algorithm is sufficiently mixing so that all strings with the same fitness statistics are equally likely. Whether it is a property of the simple problem which we looked at or of classes of problems is an important question.

We would not have recognized the need for this assumption had we studied a spatially homogeneous problem. Many theorists in the GA community have used as a test problem a paramagnet in a uniform field. In the GA literature, the problem is called *ones-counting* or *one-max*, the fitness is equal to the number of ones in a binary string. For this problem, string configuration can be inferred directly from the fitness, so our ansatz is not required. Thus, we feel that theoretical methods developed by studying *ones-counting* may not extend to more complex problems.

Another approach to describing Genetic Algorithms has been proposed by Vose and his co-workers [5, 6, 7]. In this framework the Genetic Algorithm is viewed as producing a mapping between all the possible "micro-states" of the system. Thus, for each operator one associates a transition matrix describing the probability of each string being mapped to all the other strings. For a population of P binary strings, each of which can be in any of 2^N possible states, the size of the state space will be $2^{PN}/P!$. The practical problem this approach faces is in determining the structure of these matrices. In our approach we calculate the transition probabilities between the "macro-states" of the system as defined by the cumulants. This vastly reduces the size of the space needed to describe the system. Of course, this mapping is highly non-linear so that we need to iterate the mapping between cumulants numerically. However, our approach allows us to calculate the dynamics for a real problem, albeit rather trivial.

We argue that our view of how the GA works would not be seen in the infinite population limit. In particular, the role of the third cumulant would be missed. Selection introduces a third cumulant, and a role of crossover seems to be to reduce it, effectively moving low-fitness individuals to the high-fitness tail. In the infinite population limit, selection does not introduce a third cumulant to Gaussian distributions. Therefore, theoretical approaches based on infinite population formulations may not see these effects.

In this paper we have done little more than derive and present the formalism; there is much scope for further work. The role and effect of genetic operators could be studied more fully. In particular, we have treated single-point crossover via approximations which prevent a full understanding of this operator and when it is likely to be effective. This must be studied more fully. The effect of the search parameters, whether there are quantities which could help determine their values, and whether they should change over time are questions of practical interest. They could be studied using this formalism. We have considered only the average behaviour, and have argued that sample-to-sample fluctuations lead to systematic errors. One could evolve the sample variances as well as the sample average for the population cumulants to try to get more accurate results[4]. The extensions, such as deterministic selection and Truncation selection should be looked at in detail to test whether the formulation is accurate and to learn when these operators are likely to be effective. Most importantly, other test problems must be considered. The problems we have considered may have atypical features, for example, the interface energy is less than the average cost of mutation, thus mutation in this problem may play a less important role than it would in a typical application. Whether the approach presented here is of general applicability can only be determined by applying it to other problems [5]. Only if we can analyse a range of test problem can we learn generic characteristics of how Genetic Algorithms work.

Appendices

A Representing the Distribution of Fitnesses

We represent the distribution of fitnesses $\rho(\mathcal{F})$ using a truncated cumulant expansion. The first cumulant, κ_1, is the average fitness for the population while the second cumulant, κ_2, measures the variance in the population. The higher cumulants measure the distortion of the distribution away from a Gaussian. The third and fourth cumulants are defined by

$$\kappa_3 = \frac{1}{P}\sum_\alpha (\mathcal{F}_\alpha - \kappa_1)^3 \tag{A.1}$$

$$\kappa_4 = \frac{1}{P}\sum_\alpha (\mathcal{F}_\alpha - \kappa_1)^4 - 3\kappa_2^2 \tag{A.2}$$

To obtain a dimensionless measure of the distortion of the distribution away from a Gaussian the higher cumulants are divided by $\kappa_2^{n/2}$. The normalized third and fourth cumulants are known as the skewness and kurtosis. Unfortunately estimating the average skewness and kurtosis from the average cumulants gives rise to some systematic errors due to the sample to sample fluctuations (see Appendix C).

To calculate the moments, μ_n, analytically we can use the moment generating function $\hat{\rho}(t)$ defined by

$$\hat{\rho}(t) = \int_{-\infty}^{\infty} \mathrm{e}^{t\mathcal{F}}\, \rho(\mathcal{F})\, \mathrm{d}\mathcal{F} = \tilde{\rho}(-\mathrm{i}t) \tag{A.3}$$

where $\tilde{\rho}(\mathrm{i}t)$ is the Fourier transform of the distribution (and where the integral is rotated

[4]This has now been carried out; see footnote 2

[5]Since submitting this paper several works in this direction have either been completed or are close to completion [21, 25, 26]

to be along the real axis). The n^{th} moment is given by

$$\mu_n = \frac{\partial^n}{\partial t^n}\hat{\rho}(t)\bigg|_{t=0} = \frac{\partial^n}{\partial t^n}\int_{-\infty}^{\infty}\rho(\mathcal{F})\,\mathrm{e}^{t\mathcal{F}}\,\mathrm{d}E\bigg|_{t=0} = \int_{-\infty}^{\infty}\mathcal{F}^n\,\rho(E)\,\mathrm{d}\mathcal{F}. \tag{A.4}$$

The cumulants, κ_n, are generated from the logarithm of the moment generating function

$$\log\Big(\hat{\rho}(t)\Big) = \sum_{n=1}\frac{\kappa_n t^n}{n!}. \tag{A.5}$$

To represent the distribution $\rho(\mathcal{F})$ with cumulants κ_n, $n = 1, \ldots, n_c$, numerically we approximate it by the function

$$\rho(\mathcal{F}) = \frac{\mathrm{e}^{-(\mathcal{F}-\kappa_1)^2/(2\kappa_2)}}{\sqrt{2\pi\kappa_2}}\left(1+\sum_{n=3}^{n_c}\frac{\kappa_n}{\kappa_2^{n/2}}\,u_n\!\left(\frac{\mathcal{F}-\kappa_1}{\sqrt{\kappa_2}}\right)\right) \tag{A.6}$$

where $u_n(x)$ are related to the Hermite polynomials and are defined by $u_0(x) = 1$, $u_1(x) = x$ and

$$u_n(x) = \frac{1}{n}\Big(x\,u_{n-1}(x) - u_{n-2}(x)\Big). \tag{A.7}$$

The first few polynomials are $u_2(x) = (x^2-1)/2$, $u_3(x) = (x^3-3x)/3!$, $u_4(x) = (x^4-6x^2+3)/4!$. Equation (A.6) does not define a probability distribution as it is not necessarily positive, however, it does have the correct moments and provides a good approximation given our lack of knowledge to the higher cumulants. (It would be possible to obtain a better approximation by using a maximum entropy argument to fix the higher cumulants. We have not pursued this approach as the current approximation is sufficiently accurate for our purposes.)

To compute the expected best member of the population we use

$$\mathcal{F}_{\text{best}} = P\int_{-\infty}^{\infty}\mathcal{F}_1\,\rho(\mathcal{F}_1)\left[\prod_{\alpha=2}^{P}\int_{-\infty}^{\mathcal{F}_1}\rho(\mathcal{F}_\alpha)\;\mathrm{d}\mathcal{F}_\alpha\right]\mathrm{d}\mathcal{F}_1. \tag{A.8}$$

The product of integrals gives the probability that all the other members of the population have fitness less than $\mathcal{F}_1$. From equation (A.6) and using the properties of the Hermite polynomials we find

$$\int_{-\infty}^{\mathcal{F}_1}\rho(\mathcal{F}_\alpha)\;\mathrm{d}\mathcal{F}_\alpha = \Phi\!\left(\frac{\mathcal{F}_1-\kappa_1}{\sqrt{\kappa_2}}\right) - \frac{\mathrm{e}^{-(\mathcal{F}_1-\kappa_1)^2/(2\kappa_2)}}{\sqrt{2\pi}}\sum_{n=3}^{n_c}\frac{\kappa_n}{n\kappa_2^{n/2}}\,u_{n-1}\!\left(\frac{\mathcal{F}_1-\kappa_1}{\sqrt{\kappa_2}}\right) \tag{A.9}$$

where $\Phi(x)$ is the normal probability function,

$$\Phi(x) = \int_{-\infty}^{x}\mathrm{e}^{-z^2/2}\frac{\mathrm{d}z}{\sqrt{2\pi}}. \tag{A.10}$$

To evaluated equation (A.8) we perform the integral over $\mathcal{F}_1$ numerically.

B Finite Population Effects

There are corrections to the cumulants arising from the finite size of the population. These leave the first cumulant unchanged but reduces all the higher cumulants. The variance for a finite population, κ_2^f, is given by

$$\begin{aligned}\kappa_2^f &= \frac{1}{P}\sum_\alpha \mathcal{F}_\alpha^2 - \left(\frac{1}{P}\sum_\alpha \mathcal{F}_\alpha\right)^2 = \left(\frac{1}{P} - \frac{1}{P^2}\right)\sum_\alpha \mathcal{F}_\alpha^2 - \frac{1}{P^2}\sum_\alpha\sum_{\beta\neq\alpha}\mathcal{F}_\alpha\mathcal{F}_\beta \\ &= \left(1-\frac{1}{P}\right)\left(\left\langle \mathcal{F}_\alpha^2\right\rangle_\alpha - \left\langle \mathcal{F}_\alpha\mathcal{F}_\beta\right\rangle_{\alpha\neq\beta}\right).\end{aligned} \tag{B.1}$$

where

$$\left\langle \mathcal{F}_\alpha\mathcal{F}_\beta\right\rangle_{\alpha\neq\beta} = \frac{1}{P(P-1)}\sum_\alpha\sum_{\beta\neq\alpha}\mathcal{F}_\alpha\mathcal{F}_\beta \tag{B.2}$$

But $\langle\mathcal{F}_\alpha^2\rangle - \left\langle \mathcal{F}_\alpha\mathcal{F}_\beta\right\rangle_{\alpha\neq\beta}$ is the variance for the infinite population so

$$\kappa_2^f = \left(1-\frac{1}{P}\right)\kappa_2^\infty = P_2\kappa_2^\infty, \qquad P_2 = 1 - \frac{1}{P}. \tag{B.3}$$

Similarly

$$\kappa_3^f = P_3\kappa_3^\infty, \qquad P_3 = 1 - \frac{3}{P} + \frac{2}{P^2} \tag{B.4}$$

and

$$\kappa_4^f = P_4\kappa_4^\infty - \frac{6}{P}P_2(\kappa_2^\infty)^2, \qquad P_4 = 1 - \frac{7}{P} + \frac{12}{P^2} - \frac{6}{P^3} \tag{B.5}$$

etc. We can invert these equations to obtain the unbiased estimators for the cumulants of the infinite population.

The initial variance for the spin-glass chain, κ_2^0, is on average $P_2\sum_i J_i^2$, while the fourth cumulant, κ_4^0, is $-2P_4\sum_i J_i^4 - 6P_2(\sum_i J_i^2)^2/P$.

C Sample to Sample Fluctuations

The operators of a GA produce a stochastic change in the cumulants which causes sample to sample fluctuations to grow during the evolution of the GAs. These fluctuation cause a systematic error in the estimating the average behaviour of operators which depend non-linearly on the cumulants. In the GA we have considered this error arises mainly in our prediction of selection. It also causes small error in our estimate of $\mathcal{F}_{\text{best}}$ and in our estimate of the effect of the interface energy to the second cumulant.

To understand these errors we consider a population which experiences Roulette wheel selection only with $\beta = 0$. Thus the members of the new population are selected at random from the old population. For this system the average dynamics can be calculated exactly. On average, the odd cumulants all remain zero throughout the run. The average second cumulant evolves according to (see Appendix B)

$$\langle\kappa_2(t)\rangle = \left(1-\frac{1}{P}\right)\langle\kappa_2(t-1)\rangle = \left(1-\frac{1}{P}\right)^t\langle\kappa_2(0)\rangle \tag{C.1}$$

while on average the fourth cumulant changes according to

$$\langle \kappa_4(t) \rangle = \left(1 - \frac{7}{P} + \frac{12}{P^2} - \frac{6}{P^3}\right) \langle \kappa_4(t-1) \rangle - \frac{6}{P}\left(1 - \frac{1}{P}\right) \left\langle \kappa_2^2(t-1) \right\rangle . \tag{C.2}$$

The average change in the fourth cumulant thus depends on the $\langle \kappa_2^2(t) \rangle$. For this system it is easy to show that

$$\left\langle \kappa_2^2(t) \right\rangle = \left(1 - \frac{1}{P^2}\right) \left\langle \kappa_2^2(t-1) \right\rangle + \frac{1}{P}\left(1 - \frac{2}{P} + \frac{1}{P^2}\right) \langle \kappa_4(t-1) \rangle \tag{C.3}$$

which evolves differently to $\langle \kappa_2(t) \rangle^2$. To illustrate these differences we assume that the initial members of the population are drawn from a Gaussian distribution with mean fitness zero and width 1. The fourth cumulant then evolves according to

$$\langle \kappa_4(t) \rangle = \frac{6(P-1)}{5P-6} \left[\left(1 - \frac{6}{P} + \frac{11}{P^2} - \frac{6}{P^3}\right)^{t+1} - \left(1 - \frac{1}{P}\right)^{t+1} \right]. \tag{C.4}$$

Had we instead estimated $\langle \kappa_2^2 \rangle$ by $\langle \kappa_2 \rangle^2$ (as we have done in this paper) we would have found

$$\left\langle \kappa_4^{est}(t) \right\rangle = \frac{6(P-1)}{5P-6} \left[\left(1 - \frac{7}{P} + \frac{12}{P^2} - \frac{6}{P^3}\right)^{t+1} - \left(1 - \frac{2}{P} + \frac{1}{P^2}\right)^{t+1} \right]. \tag{C.5}$$

In both cases the average fourth cumulant grows negative before reaching a minimum and decaying to zero, however, the estimated behaviour predicts a smaller absolute magnitude for the fourth cumulant. We get an even more dramatic error if we try to estimate the average kurtosis $\langle \kappa_4/\kappa_2^2 \rangle$ by $\langle \kappa_4 \rangle / \langle \kappa_2 \rangle^2$. For the system discussed above the kurtosis after initially going negative turns around and becomes positive. The estimated kurtosis always remains negative going to an asymptote of $-6(P-1)/(5P-6)$.

In the Genetic Algorithm the sample to sample fluctuation gives rise to smaller errors because the effect of mutation and crossover is to reduce the higher cumulants (although they also produce a stochastic change in the cumulants which adds to the fluctuations). These fluctuations will also be much smaller in Deterministic selection.

D Diagrammatics for Mutation and Crossover

The calculation of the effect of mutation and crossover on the cumulants is facilitated by using graphs to represent each term in the expansion. In this appendix we briefly describe the graphs used in the calculation of mutation for the defect variables (§3.1), the calculation of mutation for the spin variables (§3.2) and for crossover (§5.1).

Graphs for Mutation in the Random-Field Paramagnet

The calculation for mutation of the defect variables is the simplest example of the use of graphs. Mutation of the spin variables and crossover require additional sets of graphs, although the calculations are quite similar. For defect variables each graph corresponds to a term in the cumulant expansion.

The defect variables τ_i^α are represented by the symbol $\bullet$ while the couplings (which in this case act as random-fields) are represented by the symbol $\times$. The term $\sum_i J_i \tau_i^{m\alpha}$

can then be represented by the "windmill graph" shown in figure 6(a). After averaging over all possible mutation this term becomes $\Gamma \mathcal{F}_\alpha$. Averaging over all members of the population gives $\Gamma \kappa_1$. As this is the only graph that contributes to the first cumulant $\kappa_1^m = \Gamma \kappa_1$.

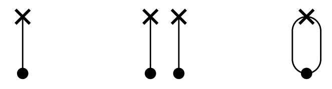


Figure 6: The graphs represent terms contributing to κ_1^m and κ_2^m for mutation of the defect variables. The terms represented by each graph are (a) $\Gamma \sum_i J_i \tau_i^\alpha$, (b) $\Gamma^2 (\sum_i J_i \tau_i^\alpha)^2$ and (c) $(1-\Gamma^2)\sum_i J_i^2$.

To calculate $(\mathcal{F}_\alpha^m)^2$ we pull out the diagonal term

$$(\mathcal{F}_\alpha^m)^2 = \left(\sum_i J_i F_i^\alpha \tau_i^\alpha \right)^2 = \sum_{i \neq j} J_i J_j \sigma_i^\alpha \sigma_j^\alpha \tau_i^\alpha \tau_j^\alpha + \sum_i J_i^2. \tag{D.1}$$

Averaging over all possible mutations and add $\Gamma^2 \sum_i J_i^2$ to the first term and subtract it from the second term gives

$$(\mathcal{F}_\alpha^m)^2 = \Gamma^2 \left(\sum_i J_i \tau_i^\alpha \right)^2 + (1-\Gamma^2) \sum_i J_i^2 = \Gamma^2 \mathcal{F}_\alpha^2 + (1-\Gamma^2) \sum_i J_i^2. \tag{D.2}$$

We can represent these two terms by the graphs shown in figure 6(b) and (c). The graph in figure 6(b) consists of two disconnected windmill graphs. The graph contributes a term $\Gamma^2 \mathcal{F}_\alpha^2$, which is just the product of the single term contribution times a combinatoric factor (in this case 1). The second graph represents the diagonal term $(1-\Gamma^2)\sum_i J_i^2$. The weighting factor $1-\Gamma^2$ can be calculated by subtracting the weighting factor for the graph (b) from 1, the factor we got from averaging the diagonal term over all mutations. To calculate the second cumulant we average over all members of the population and subtract $(\kappa_1^m)^2$,

$$\kappa_2^m = P_2 \left(\left\langle (\mathcal{F}_\alpha^m)^2 \right\rangle_\alpha - \left\langle \mathcal{F}_\alpha^m \mathcal{F}_\beta^m \right\rangle_{\alpha \neq \beta} \right) = \Gamma^2 \kappa_2 + P_2 (1-\Gamma^2) \sum_i J_i^2. \tag{D.3}$$

Note that on subtracting out $\langle \mathcal{F}_\alpha^m \mathcal{F}_\beta^m \rangle_{\alpha \neq \beta}$ and averaging over the members of the population the graph in figure 6(b) contributes a term $\Gamma^2 \kappa_2$. In general graphs consisting of n disconnected windmill graphs contribute a term $\Gamma^n \kappa_n$.

The graphs contributing to $(\mathcal{F}_\alpha^m)^3$ are shown in figure 7. On averaging over all mutation they give

$$(\mathcal{F}_\alpha^m)^3 = \Gamma^3 \mathcal{F}_\alpha^3 + 3\Gamma(1-\Gamma^2) \mathcal{F}_\alpha \sum_i J_i^2 - 2\Gamma(1-\Gamma^2) \sum_i J_i^3 \tau_i^\alpha. \tag{D.4}$$

The first two graphs consist of disconnected terms so their weight factor is straightforward to compute; they are (a) Γ^3 and (b) $3\Gamma(1-\Gamma^2)$, where the 3 is the combinatoric factor for this graph. The third graph is a new connected graph whose weight factor can be calculated by subtracting the weighting factors for the other two graphs from Γ, the contribution obtained by averaging the diagonal term over all mutations. To find the third cumulant we average over all members of the population and subtract out a combination of the first and second moments. The first graph contributes a term $\Gamma^3\kappa_3$. The second graph is completely cancelled — in fact any graph made up from disconnected graphs in which one of the constituent parts consists of a connected graph with an even number of spins will cancel. Thus

$$\kappa_3^m = \Gamma^3\kappa_3 - 2P_3\Gamma(1-\Gamma^2)\sum_i J_i^3 \left\langle \tau_i^\alpha \right\rangle_\alpha \tag{D.5}$$

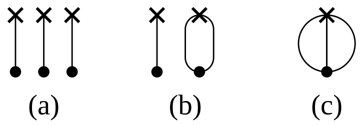


Figure 7: The graphs represent terms contributing to $(\mathcal{F}_\alpha^m)^3$ for defect spin mutation. The weighting factor associated with each of the graphs are (a) Γ^3, (b) $3\Gamma(1-\Gamma^2)$ and (c) $-2\Gamma(1-\Gamma^2)$.

With these rules we can calculate the contribution to higher orders. The only graphs that contribute to κ_4^m are shown in figure 8. To calculate the weighting factor for figure 8(c) we need to subtract the weighting factors from *all* the graphs that contribute to $(\mathcal{F}_\alpha^m)^4$ (there are an additional four graphs to the ones shown in figure 8, although their contribution to κ_4^m all cancel). The graph in figure 8(b) does not cancel because each of the constituent graphs consists of an odd numbers of spins. On subtracting off the appropriate lower moments and averaging over all members of the population it contributes a term proportional to

$$\left\langle \mathcal{F}_\alpha \sum_i J_i^3 \tau_i^\alpha - \mathcal{F}_\alpha \sum_i J_i^3 \tau_i^\beta \right\rangle_{\alpha\neq\beta} = \sum_i J_i^4 \Big(1 - \left\langle \tau_i^\alpha \tau_i^\beta \right\rangle_{\alpha\neq\beta}\Big). \tag{D.6}$$

There is in addition a correction to κ_4^m of $-6P_2(\kappa_2)^2/P$ due to the finite size of the population. Adding in this additional term to the three terms coming from figure 8 we arrive at equation (3.11).

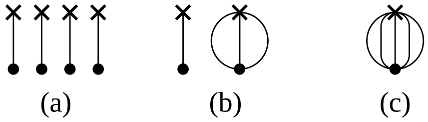


Figure 8: The graphs represent terms contributing to κ_4^m for defect spin mutation. The weighting factor associated with each of the graphs are (a) Γ^4, (b) $-8\Gamma^2(1-\Gamma^2)$ and (c) $-2(1-\Gamma^2)(1-3\Gamma^2)$.

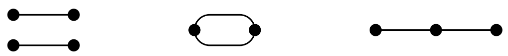
Graphs for Mutation in the Spin-Glass Chain

We can calculate the mutation of the spin variables, S_i^α, by using (one dimensional) lattice graphs. We again represent the spins by $\bullet$. The term $J_i S_i^{m\alpha} S_{i+1}^{m\alpha}$ can be represented by a dumb-bell t—t. On averaging over all mutations each unpaired spin contributes a term Γ. Thus $\mathcal{F}_\alpha^m = \Gamma^2 \mathcal{F}_\alpha$, and on averaging over all members of the population $\kappa_1^m = \Gamma^2 \kappa_1$.

The graphs contributing to $(\mathcal{F}_\alpha^m)^2$ are shown in figure 9. The first two graphs are topologically identical to those in figure 6 for the defect variable mutation. The only difference is that Γ is now replaced by Γ^2. The graph shown in figure 9(c) is a new type of graph arising from flipping neighbouring spins. For these "chain graphs" the paired spins contribute a weighting factor of $1-\Gamma^2$. The totally contribution of this graph is

$$2\Gamma^2(1-\Gamma^2)\sum_i J_i J_{i+1} \tau_i^\alpha \tau_{i+1}^\alpha \tag{D.7}$$

where the 2 is a combinatoric factor which arises because the ordering of the bonds is significant. The contribution of the three graphs give for the second cumulant equation (3.17). The higher order cumulants are found in an analogous way to the defect variable mutation.



(a) (b) (c)

Figure 9: The graphs represent terms contributing to $(\mathcal{F}_\alpha^m)^2$ for mutation of the spin variables. The weighting factor for the three graphs are (a) Γ^4, (b) $1-\Gamma^4$ and (c) $2\Gamma^2(1-\Gamma^2)$.

Graphs for Crossover

In calculating the effect of crossover we again use defect variables but we have to take into account bonds coming from either of the parents. We can represent these two types of bonds, $J_i\tau_i^\alpha$ and $J_i\tau_i^\beta$ by using thin and thick lines joining the defects to the couplings. For each bond $J_i\tau_i^\alpha$ we associate a weight a while for each bond $J_i\tau_i^\beta$ we associate a weight $1-a$. Disconnected graphs contribute a term equal to the product of their constituent graphs times a combinatoric factor. To calculate the weight of connected graphs we subtract the weight factors for all the disconnected graphs that contain the same elements as the connected graph from the contribution from the connected graph alone. This contribution is either a if the connected graph contains just τ_i^α bonds, $1-a$ if it contains just τ_i^β bonds or zero if it contains a mixture of bonds. The possible graphs are topologically identical to those used in the defect mutation calculation but now each graph appears several times with different proportions of τ_i^α bonds and τ_i^β bonds.

There are six graphs contributing to $\mathcal{F}_c^2$, these are shown in figure 10. On averaging over all pairs of parents the first three graphs contribute the terms

$$a^2\left\langle\mathcal{F}_\alpha^2\right\rangle_\alpha, \qquad (1-a)^2\left\langle\mathcal{F}_\alpha^2\right\rangle_\alpha, \qquad 2a(1-a)\left\langle\mathcal{F}_\alpha\mathcal{F}_\beta\right\rangle_{\alpha\neq\beta}$$

respectively. Subtracting $\langle \mathcal{F}_\alpha \mathcal{F}_\beta \rangle_{\alpha\neq\beta}$ the third term cancels while the first two terms give a contribution $(a^2 + (1-a)^2)\kappa_2$. The last three graphs contribute the terms

$$a(1-a)\sum_i J_i^2, \qquad a(1-a)\sum_i J_i^2, \qquad -2a(1-a)\sum_i J_i^2 \left\langle \tau_i^\alpha \tau_i^\beta \right\rangle_{\alpha\neq\beta}$$

respectively. These three terms add up to give a contribution $\sum_i J_i^2 \left(1 - \left\langle \tau_i^\alpha \tau_i^\beta \right\rangle_{\alpha\neq\beta}\right)$ to κ_2^c.

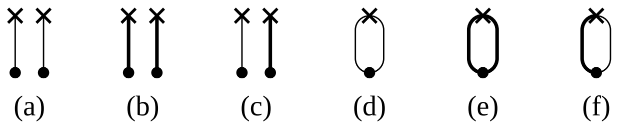


Figure 10: The graphs represent terms contributing to $(\mathcal{F}_\alpha^c)^2$. The weighting factor for the six terms are (a) a^2, (b) $(1-a)^2$, (c) $2a(1-a)$, (d) $a(1-a)$,(e) $a(1-a)$ and (f) $2a(1-a)$.

There a fourteen graphs contributing to $\mathcal{F}_c^3$, although we can obtain half of them by replacing τ_i^α by τ_i^β (the weighting factors are obtained by replacing a by $1-a$ and visa versa). In figure 11 we show half the graphs contributing to this term. Adding the symmetric partners graphs (c) and (g) cancel all together. As with mutation, the graphs with disconnected bubbles (in this case (b), (e)) cancel when we subtract off the lower moments to obtain the cumulant. Also the graphs containing disconnected graphs with both unpaired τ_i^α and τ_i^β bonds (here graphs (d)) also cancel. Thus the only surviving graphs are (a) and (f) and their symmetric partners. Graph (a) and its partner graph contributes a term $(a^3 + (1-a)^3)\kappa_3$. After subtracting off the lower moments. Graph (f) contributes a term to κ_3^c proportional to

$$\left\langle \mathcal{F}_\alpha \sum_i J_i^2 \tau_i^\alpha \tau_i^\beta - \mathcal{F}_\gamma \sum_i J_i^2 \tau_i^\alpha \tau_i^\beta \right\rangle_{\alpha\neq\beta\neq\gamma} = \sum_i J_i^3 \left(\langle \tau_i^\alpha \rangle_\alpha - \left\langle \tau_i^\alpha \tau_i^\beta \tau_i^\gamma \right\rangle_{\alpha\neq\beta\neq\gamma} \right). \tag{D.8}$$

Adding together these terms κ_3^c is given by equation (5.11).

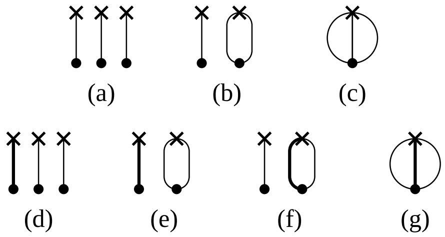


Figure 11: The graphs represent terms contributing to $(\mathcal{F}_\alpha^c)^3$. The weighting factor for the different graphs are (a) a^3, (b) $3a^2(1-a)$, (c) $a(1-a)(1-2a)$, (d) $3a^2(1-a)$ (e) $3a(1-a)^2$ and (f) $-6a^2(1-a)$ and (g) $-3a(1-a)(1-2a)$.

The fourth cumulant involves 33 different graphs, although most of these occur with a symmetric partner and many cancel to give (5.13).

E The Initial Distribution

The initial strings are generated by randomly setting each spin to ± 1. Since each bond contributes a factor of either $\pm J_i$ to the fitness we would expect from the central limit theorem that the distribution of fitnesses to be approximately Gaussian with zero mean and width $\sqrt{N}$. There are, however, correction coming from both the finite length of the chain and from the finite size of the population. We calculate the corrections due to the finite length of the string here. In Appendix B we discuss the corrections due to the finite population size.

The density of states for a spin-glass chain is given by

$$\Omega(\mathcal{F}) = \frac{1}{2^{N+1}} \operatorname*{Tr}_{\{\vec{S}\}} \delta\left(\mathcal{F} - \sum_{i=1}^{N} J_i S_i S_{i+1}\right) = \frac{1}{2^N} \operatorname*{Tr}_{\{\vec{\tau}\}} \delta\left(\mathcal{F} - \sum_{i=1}^{N} J_i \tau_i\right) \tag{E.1}$$

where the trace implies a sum over all possible configurations of the spins. Using the integral representation for the delta function this becomes

$$\Omega(\mathcal{F}) = \frac{1}{2^N} \operatorname*{Tr}_{\{\tau_i\}} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{\mathrm{d}t}{2\pi}\, \mathrm{e}^{-t\mathcal{F} + t\sum_{i=1}^{N} J_i \tau_i} = \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{\mathrm{d}t}{2\pi}\, \mathrm{e}^{-t\mathcal{F} + \sum_{i=1}^{N} \log\left(\cosh(J_i t)\right)}. \tag{E.2}$$

The cumulant generating function for the density of states is given by (see Appendix A)

$$\log\left(\hat{\Omega}(t)\right) = \sum_i \log\left(\cosh(tJ_i)\right). \tag{E.3}$$

Thus the cumulants for the density of states of the spin-glass chain are

$$\kappa_n = \sum_{i=1}^{N} \frac{\partial^n}{\partial t^n} \log\left(\cosh(tJ_i)\right)\bigg|_{t=0} = a_n \sum_{i=1}^{N} (J_i)^n \tag{E.4}$$

where the odd cumulants are all identically zero while for n even $a_n = 2^n(2^n - 1)B_n/n$ and B_n is the n^{th} Bernoulli number ($a_2 = 1$, $a_4 = -2$, etc.). On average the even cumulants are

$$\langle \kappa_n \rangle = N a_n (n-1)!! \tag{E.5}$$

The fluctuations (for samples with different J_i's) in the even cumulant are of the order of

$$\sqrt{\langle \kappa_n^2 \rangle - \langle \kappa_n \rangle^2} = \sqrt{\frac{(2n-1)!! - ((n-1)!!)^2}{N}}\, \langle \kappa_n \rangle\,. \tag{E.6}$$

The fluctuations decrease as the length of the chains, N, increases, however, the fluctuations grow rapidly with the order of the cumulant, n. From equation (E.5) we find the average variance is $\kappa_2 = N$ and the average fourth cumulant is $\kappa_4 = -6N$, etc. The kurtosis, κ_4/κ_2^2, for the initial distribution is $-6/N$ — as N grows the initial distribution becomes more nearly Gaussian.

Acknowledgement. We would like to thank Magnus Rattray for helpful comments and suggestions in producing this paper.

References

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

[2] Prügel-Bennett A. and J. L. Shapiro. An Analysis of Genetic Algorithms Using Statistical Mechanics. *Phys. Rev. Lett.*, 72(9):1305–1309, 1994.

[3] Shapiro J. L., A. Prügel-Bennett, and M. Rattray. A Statistical Mechanical Formulation of the Dynamics of Genetic Algorithms. In T. C. Fogarty, editor, *Lecture notes in Computer Science* **865**, pages 17–27, Berlin, 1994. Springer-Verlag.

[4] David E. Goldberg. *Genetic Algorithms in Search, Optimization & Machine Learning.* Addison-Wesley (Reading, Mass), 1989.

[5] A Nix and Michael D. Vose. Modeling genetic algorithms with markov chains. *Annals of Mathematics and Artificial Intelligence*, 5:79–88, 1991.

[6] Michael D. Vose and Gunar E. Liepins. Punctuated equilibria in genetic search. *Complex Systems*, 5:31–44, 1991.

[7] Michael D. Vose. Modelling simple genetic algorithms. In Darrell Whitley, editor, *Foundations of Genetic Algorithms 2.* Morgan Kaufmann (San Mateo, CA), 1992.

[8] David E. Goldberg. Genetic algorithms and walsh functions: Part I, a gentle introduction. *Complex Systems*, 3:129–152, 1990.

[9] John J. Grefenstette. Deception considered harmful. In L. Darrel Whitley, editor, *Foundations of Genetic Algorithms 2*, San Mateo, 1993. Morgan Kaufmann.

[10] T Li. Structure of metastable states in random Ising chain. *Phys. Rev. B*, 24(11):6579–6587, 1981.

[11] B. Derrida and E. J. Gardner. Metastable states of a spin-glass chain. *J. Physique*, 47:959, 1986.

[12] G. Grinstein, A. N. Berker, J. Chalupa, and Wortis M. Exact renormalization group with griffiths singularities and spin-glass behaviour: The random ising chain. *Phys. Rev. Lett.*, 36:1508–1511, 1976.

[13] H. H. Chen and S. K. Ma. Low-temperature behaviour of a one-dimensional random ising model. *J. Stat. Phys.*, 29(4):717–746, 1982.

[14] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*, volume 55 of *Applied Mathematics Series.* Dover, Washington, 9th. edition, 1964.

[15] James Edward Baker. Reducing bias and inefficiency in the selection algorithm. In *Proceedings of the Second International Conference on Genetic Algorithms.* Lawrence Erlbaum Associates (Hillsdale), 1987.

[16] M. De La Maza and B. Tidor. Increased flexibitity in genetic algorithms: The use of variable boltzman selective pressure to control propagation. In *Proceedings of the ORSA CSTS Conference: Computer Science and Operations Research: New Developments in Their Interfaces*, 1991.

[17] B. Derrida. Random-energy model: An exactly solvable model of disordered systems. *Phys. Rev.*, B24:2613–2626, 1984.

[18] J. Cook and B. Derrida. Finite size effects in random energy models and in the problem of polymers in a random medium. *J. Stat. Phys.*, 63:505, 1991.

[19] Derrida B., R. B. Griffiths, and A. Prügel-Bennett. Finite-Size Effects and Bounds for Perceptron Models. *J. Phys. A: Math. Gen.*, 24:4907–4936, 1991.

[20] Shang-Keng Ma. *Statistical Mechanics*. Singapore:World Scientific, 1985.

[21] M. Rattray. The dynamics of a genetic algorithm under stabilizing selection. *Complex Systems*, 1996. Accepted for publication.

[22] G. Syswerda. Simulated crossover in genetic algorithms. In L. Darrel Whitley, editor, *Foundations of Genetic Algorithms 2*, San Mateo, 1993. Morgan Kaufmann.

[23] Prügel-Bennett A. Modelling Evolving Populations. Submitted to the *Journal of Theoretical Biology*, 1996.

[24] H. Mühlenbein and D. Schlierkamp-Voosen. The science of breeding and its application to the breeder genetic algorithm. *Evolutionary Computation*, 2(3):335–360, 1994.

[25] M. Rattray and J. Shapiro. The dynamics of genetic algorithms for a simple learning problem problem. *J. Phys. A.*, 1996. Submitted for publication.

[26] J. L. Shapiro and A. Prügel-Bennett. Genetic algorithms can beat stochastic hill climbing: the basin with a barrier. In preparation., 1996.