On the Limit of Long Strings

Adam Prügel-Bennett
ISIS, ECS
University of Southampton
UK

Abstract

To gain a clearer understanding of the dynamics of evolving populations the limit of infinitely long strings is examined. The limit clarifies some of the behaviour of evolving systems. For example, this limit removes any complications due to transient behaviour. Care is required in taking the limit to ensure that the quantities considered are meaningful. For concreteness a simple Genetic Algorithm acting on the problem of ones-counting is considered. A population evolving under selection and mutation alone is first examined. In this limit a remarkably simple picture of the evolutionary flow emerges. The discussion is widened to include crossover. This gives a new insight into how diversity induced by crossover improves search. Finally the approach to the limit is examined. The limit provides a way of separating the initial transient dynamics from the long time (asymptotic) dynamics.

1 Introduction

Over the last five years a formalism has been developed for calculating the dynamics of Genetic Algorithms [1–5]. The formalism provides an accurate quantitative description of a GA's dynamics, which affords many useful insights into how GAs work (see, for example, [6–8]). However, the formalism gives a solution in terms of a set of coupled non-linear equations describing the evolution of a number of macroscopic variables. Unfortunately, how the best solution found by the GA depends upon its parameters often remains obscure. In this paper we examine the limit of infinitely long strings, $L \to \infty$. To use an analogy we can think of understanding the dynamics of an evolving populations as similar to the task of completing a jig-saw puzzle. Understanding the limits is like starting at the edges of the puzzle. They are usually easier to put together and allow a secure basis from which to build. We will see that the long string limit eliminates some of the details which hide the generic behaviour of evolving populations. However, care is necessary in taking the limit to retain the relevant behaviour. It is also important to understand the approach to the limit (i. e. the finite string corrections) and in what circumstances the behaviour being described is relevant to a real GA. Taking this limit is not new, it has been used before, most notably in Muller's ratchet (see [5, 9] for a discussion). However, the limit has not been used to investigate GA operators.

The limit is illustrated by considering a simple GA working on the ones-counting problem. The limiting procedure is first discussed with regards to selection and mutation. In this case the infinite string limit can be simulated and compared with finite string simulations. The limit is shown to provide a good quantitative description of a GA for moderate size strings ($L \approx 1000$), but also the correct qualitative behaviour for shorter strings. We next explore the problem of crossover. The situation is more complicated, but reveals much about the behaviour of this operator. Finally we discuss the approach to the limit and how the limit separates the asymptotic dynamics from the transient dynamics. In practical situations the transient dynamics are very important.

Before discussing the limit of long strings we review the formalism for solving the dynamics of GAs. This formalism provides many of the ideas and the language for discussing the limit. This review forms the content of the next section. In section 3 we describe a classic GA and the ones-counting problem which acts as an example for the rest of this paper. Having cleared the ground in the previous two sections we start on the major contribution of this paper in section 4 when we describe the limit of long strings. Crossover is described in section 5. Section 6 contains a discussion of the initial dynamics and the approach to the limit. Finally in section 7 we draw some conclusions arising from this work. Derivations of the main formulae are given in the appendix.

2 Solving the Dynamics of GA

This paper builds on the formalism developed by Prügel-Bennett, Shapiro and Rattray to predict dynamics of evolving populations. In this section I outline some of developments in creating the formalism. The underlying idea behind this approach is that the GA operators, selection, mutation and crossover, are stochastic so that a detailed knowledge of the population (such as the genotype of each member of the population) is not required to describe the evolution. Rather we can use a number of "macroscopic variables" which describe statistical properties of the whole population. An obvious set of statistical properties to use are those describing the distribution of efficacies[1] in the population. Since the distribution tends to be bell shaped the cumulants, κ_n, are an appropriate set of variables. These include the average fitness, κ_1, the variance in the fitness, κ_2, and higher cumulants describing in more detail the shape of the distribution [1, 2] (see appendix A.1 for a discussion of cumulants). To calculate the effect of selection the cumulants contain all the necessary information. However for the calculation to be exact all the cumulants would be necessary. To make the problem tractable we use only a few cumulants, throwing away some information about the shape of the distribution. The errors produced through this truncation are however very small. To calculate the effect of crossover we also need to know the average correlation between the strings in the population. Initially these were estimated from the variance in the population, but later Rattray showed how to include the average correlation as a macroscopic variable, giving a more accurate and cogent scheme for calculating the dynamics of GAs with crossover [3, 4]. For some problems additional information is required to predict how the macroscopic variables are changed by a genetic operator. This information can often be estimated from a maximum entropy calculation [2].

In the original description of the formalism the evolution was described in terms of the average macroscopic variables alone. In a sense it was a mean field description of the dynamics. However, this approximation gave rise to discrepancies; selection depends non-linearly on the cumulants, so that fluctuations due to the stochastic nature of the evolution operators produce systematic errors. To incorporate fluctuations into the formalism an ensemble of populations was considered [5]. Rather than use the raw macroscopic variables, we can use "ensemble variables" describing the average value of the macroscopic variables and the fluctuations (for example, by using the covariances

[1] We use the word efficacy to denote the goodness of a solution and reserve the word fitness for its strict biological meaning, that is, the fitness is proportional to the number of its offspring in the next generation.

between the macroscopic variables). The ensemble description gives a very accurate prediction of the evolutionary dynamics as well as providing additional information about the fluctuations.

In calculating the effect of mutation and crossover it is necessary to study a particular problem. To date we have considered a number of "toy problems", including ones-counting, a random ideal paramagnet (similar to ones-counting except each gene has a different strength), a spin-glass chain [2], the subset-sum problem [3], a perceptron learning problem [4], and some simple multi-well potential problems [6]. In all cases the objective function was simple, although some of the problems have features in common with hard optimization problems. For example, the spin-glass chain has an exponential number of local minima, while the subset-sum problem is weakly NP-hard.

A quantitatively accurate theory of GA dynamics provide a number of benefits. We can be sure that we have included all the important features of the GA. It is possible to compare different types of operators, for example, we have looked at different types of selection (scaled and unscaled) Boltzmann selection, truncation selection [2], tournament and ranking selection [10]. Recently the dynamics of a steady-state GA have been studied to understand how it differs from a generational GA [11]. Having an analytical solution helped to understand the effect of noisy fitness evaluation [7] and the related problem of diploidy [10]. These effects are subtle and would be difficult to pick out from simulations alone. A further example of the use of this formalism has been in understanding the dynamics of GAs with a multi-well fitness functions [6]. Rather than solving the model analytically the formalism was used to derive the asymptotic behaviour. The problem illustrated how Genetic Algorithms could sometimes outperform simulated annealing. As a final example Bornholdt examined evolution on different fitness landscapes by parameterizing the effect of crossover [8]. Although not exact this approach opens up the possibility to fit the parameters from empirical data allowing much more complicated problems to be studied.

Despite the successes, the formalism fails to provide a transparent view of how the performance of a GA depends on its parameters. In part this is because the method reduces the problem to a set of coupled non-linear equations. It is difficult to extract from these the long time behaviour as a function of the GA's parameters. An attempt to do this was made in an earlier paper by semi-linearizing the non-linear equations which could then be solved in closed form [2]. Although, this provided a useful caricature of how GAs worked, it suffered from lack of accuracy. The aim of this paper is to initiate a programme to obtain a more satisfactory description of a GA's dynamics.

3 A GA for Ones-counting

Before starting our main discussion we define a model system around which to base our later discussion. We take the problem of ones-counting where the efficacy of a string is

$$E^\mu = \sum_{i=1}^{L} X_i^\mu \tag{1}$$

where the string is given by $\vec{X}^\mu = (X_1^\mu, \ldots, X_L^\mu)$, and X_i^μ are binary variables. The superscript, $\mu = 1, 2, \ldots, P$, label the different members of the population. The efficacy is equal to the number of 1's in the string. We select the strings with a probability proportional to their Boltzmann weight

$$p^\mu = \frac{\mathrm{e}^{\beta E^\mu}}{Z}, \qquad\qquad Z = \sum_{\nu=1}^{P} \mathrm{e}^{\beta E^\nu}. \tag{2}$$

This model is known in the biological literature as a multiplicative fitness landscape since each good allele ($X_i^\mu = 1$) increases the fitness by a multiplicative factor $\exp(\beta)$. We use roulette wheel selection on the whole population to form a mating pool.

A new generation is created by applying genetic operators to the mating pool. We consider two genetic operators, mutation and crossover. In mutation we decide whether to mutate each variable, $X_i^\mu \to 1 - X_i^\mu$, with a small probability u; each variable being considered independently. Crossover is achieved by pairing all the members of the mating pool and performing uniform crossover to create two children (one the complement of the other). The full GA dynamics consists of repeatedly applying selection, followed by the genetic operators.

We are well aware that this strategy is far from optimal and that the problem being tackled is pathologically easy. However, these are not the concerns of this paper, more efficient operators could also be investigated as could different sets of problems. The aim of this paper is to understand what it means to take the long string limit and what are its implications.

4 The Long String Limit

To understand the behaviour of GAs we consider the limit $L \to \infty$. This limit can be obtained by considering finite strings but make them longer and longer. We focus on those quantities that tend to a finite value in this limit. Examples of these are the efficacy per variable, E^μ/L, and the rate of change of efficacy. The limiting value of these quantities provide an accurate approximation for finite strings provided L is sufficiently large. One immediate benefit of taking this limit is that we have eliminated the parameter L from the analysis.

For most problems the infinite string limit is unobtainable and we must approximate it by considering longer and longer strings. However, for ones-counting we can simulate a GA with infinite length strings, at least, if the GA consists of selection and mutation alone. This is useful as it allows us to compare our approximation to the limit found by repeatedly increasing the length of the strings to the exact value (i. e. it provides a useful sanity check). To simulate a GA with infinite length strings we consider a population with efficacy

$$E^\mu = \kappa_1 + \Delta^\mu \tag{3}$$

where κ_1 is the average efficacy (of order L) and Δ^μ is a random component of order 1. For Boltzmann selection the probability of selecting a member μ is given by

$$p^\mu = \frac{\mathrm{e}^{\beta E^\mu}}{\sum_\nu \mathrm{e}^{\beta E^\nu}} = \frac{\mathrm{e}^{\beta \Delta^\mu}}{\sum_\nu \mathrm{e}^{\beta \Delta^\nu}} \tag{4}$$

where we have cancelled $\exp(\beta\kappa_1)$ in the numerator and denominator. Boltzmann selection (in common with scaled selection, tournament selection, ranking selection and many other forms of selection) depends only on the difference in the efficacy of the members of the population. We therefore concentrate our attention on the Δ^μ's and consider κ_1 as fixed. To simulate mutation for the ones-counting problem it is not necessary to know the string variables as long as we know the string efficacy since this tell us the number of variables that are 0 and 1. The change in efficacy caused by mutation is given by

$$E^{\mu\prime} = E^\mu + n_B - n_D \tag{5}$$

where n_B is the number of beneficial mutations and n_D is the number of detrimental mutations. The probability of a particular number of beneficial and detrimental mutations is given by the Poisson distributions

$$\begin{aligned} P(n_B|u, E^\mu) &= \frac{(u(L-E^\mu))^{n_B}\,\mathrm{e}^{-u(L-E^\mu)}}{n_B!} \\ P(n_D|u, E^\mu) &= \frac{(uE^\mu)^{n_D}\,\mathrm{e}^{-uE^\mu}}{n_D!}. \end{aligned} \tag{6}$$

To simulate mutation we choose n_B and n_D to be the appropriate Poisson deviates. The average change in fitness due to mutation is $u(L - 2E^\mu)$, while fluctuations are of order $\sqrt{uL}$. For the number of beneficial mutations to have a chance of out numbering the detrimental mutations we require $u \sim \mathcal{O}(1/L)$. That is, there should be a finite number of mutations per string.

We can use equations (4) and (5) to simulate the dynamics of a GA performing selection and mutation. Since both selection and mutation cause changes of efficacy of order 1 the "macroscopic efficacy", κ_1, does not change in finite time. Consequently we need only consider what happens to the Δ^μ's. This leads to a considerable simplification because the initial dynamics (transient behaviour) is irrelevant. The population settles down to a fixed distribution (or more accurately an ensemble of distributions) with a well defined rate of change of average efficacy $\mathrm{d}\kappa_1/\mathrm{dt}$. (This may appear contradictory, we observe that κ_1 changes in time, but assume that it remains fixed. However, this is justified as the change in κ_1 is of order 1, while κ_1 itself is of order L, thus it takes of order L time steps for κ_1 to change appreciably. In the limit $L \to \infty$ the average efficacies remain unchanged for finite times.)

In figure 1 we show a graph of the rate of change of efficacy as a function of the average efficacy per variable (solid line). Also shown in the graph are simulation results for populations with finite length strings (dashed lines). The velocities for the finite size systems were obtained by starting the population off in both the best and worst cases and measuring the rate of change of efficacy as the population evolves.

We notice that the velocities for the finite strings converge rapidly to those for the infinite string, indicating that the infinite string limit provides a quantitatively accurate description of the dynamics for moderate size strings ($L \approx 1000$). As the string size decreases to around 100 variables there are significant finite size corrections, although interestingly the curves show similar qualitative behaviour. For small β the change in κ_1 due to selection and mutation is given by

$$\frac{\mathrm{d}\kappa_1}{\mathrm{dt}} \approx \beta\kappa_2 + u(L - 2\kappa_1) \tag{7}$$

where κ_2 is the variance in the population. See appendices A.2 and A.3 for a derivation of this formula. The value of κ_2 is very nearly independent of κ_1. Consequently, the velocity is very nearly a linear function of κ_1. Assuming this is the case then we can solve the dynamics in closed form

$$\kappa_1(t) = \kappa_1^* + (\kappa_1(0) - \kappa_1^*)\mathrm{e}^{-t/\tau} \tag{8}$$

where $\kappa_1(t)$ is the average efficacy at time t; $\kappa_1^* \approx L/2 + \beta\kappa_2/(2u)$ is the

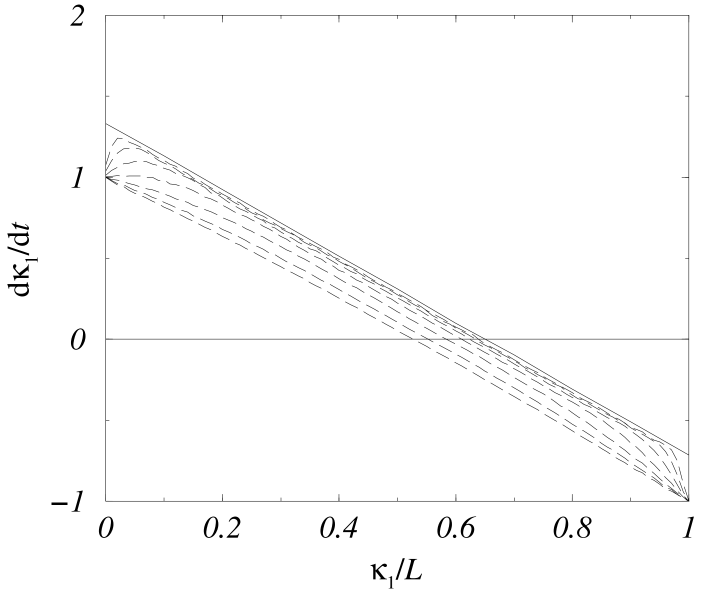

Figure 1: Plot of the evolutionary velocity, $\mathrm{d}\kappa_1/\mathrm{dt}$, versus the average efficacy per variable, κ_1/L, for a population size $P = 100$, with a selection strength $\beta = 0.01$ and mutation rate $u = 1/L$. The data points joined by the solid line show the simulation results for infinitely long strings, while the dashed curves show the same quantity for strings of length $L = 2^n$ where $n = 5$, 6, ..., 11. The results for finite strings converge to the infinitely long string as L increases. The data was obtained from simulations averaged over at least 1400 runs. The size of the errors are indicated by the wiggles in the curves.

equilibrium efficacy where $\mathrm{d}\kappa_1/\mathrm{dt} = 0$; and $-1/\tau \approx -2u$ is the slope in figure 1. Calculating the variance in the distribution of efficacies, κ_2, is more complicated. Mutation increases the variance by an amount uL, however, the effect of selection depends on the full distribution of efficacies [5]. We could investigate the role of the GA parameters such as population size, selection or mutation rate within this framework. For example, increasing the population size increases κ_2 and thus increases the equilibrium efficacy. However, a detailed study of the role of the parameters deserves its own paper. We will rather discuss the thornier problem of crossover.

5 Crossover

Crossover is much more complicated as it depends on the correlations between members of the population. It would still be possible to simulate a model with infinite strings provided we kept the matrix of pairwise correlations. We could alternatively perform bit simulated crossover in which all members of the population were crossed with each other [12]. In this case we would only need to keep the average correlation for the whole population. However, a simpler approach is to study the limit by considering longer and longer strings. In figure 2 we show the rate of change of the average efficacy for a GA which includes crossover. The curves are calculated for finite strings where we have doubled the string length many times. The curves converge to a limit curve which again appears to be almost linear. To reach the limit requires considerably longer strings ($L \approx 10\,000$) than for selection and mutation alone.

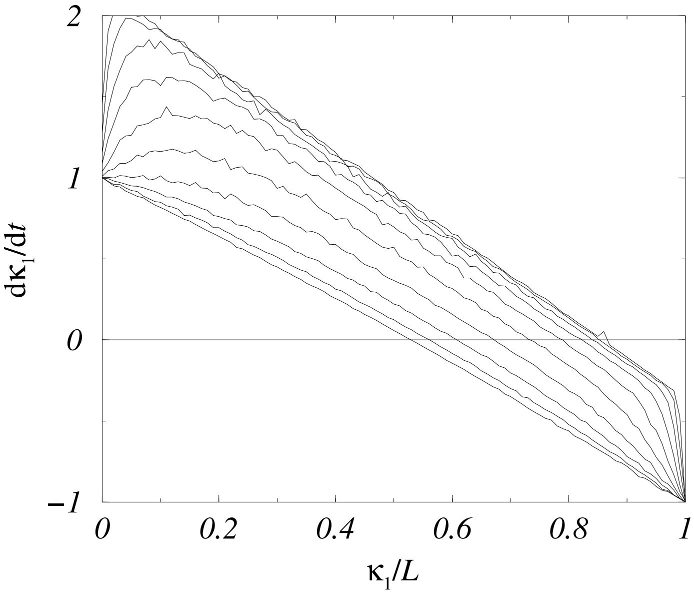

Figure 2: Plot of the evolutionary velocity, $\mathrm{d}\kappa_1/\mathrm{dt}$, versus the average efficacy per variable κ_1/L. The same parameters are used as in figure 1 but uniform pairwise crossover is applied at each generation. The curves are for strings of length $L = 2^n$ where $n = 5, 6, \ldots, 15$.

Comparing figures 1 and 2 we see that crossover increases the equilibrium efficacy. The reason for this is that it increases the variance in the efficacies of the members of the population. A simple calculation reveals that crossover does not change the average efficacy, but causes a change in the variance of

$$\kappa_2' = \frac{1}{2}\left(\kappa_2 + \frac{1}{4}L(1-q)\right) \tag{9}$$

where q is the average correlation defined by

$$q = \frac{1}{P(P-1)}\sum_{\mu}\sum_{\nu\neq\mu} q^{\mu\nu} \tag{10}$$

and $q^{\mu\nu}$ is the pair correlation

$$q^{\mu\nu} = \frac{1}{L}\sum_i (2X_i^{\mu}-1)(2X_i^{\nu}-1). \tag{11}$$

See appendix A.4 for a derivation of this result. The main effect of crossover is to drive κ_2 towards $L(1-q)/4$. The effect of selection and mutation on the variance in rather minor. Thus the variance and consequently the rate of change of fitness is governed by the correlation. This is in contrast to the situation of selection and mutation alone where the variance is determined by a balance between selection and mutation acting on the variance. Crossover does not change the correlation which is determined by selection and mutation. Mutation decreases the correlation by a factor $1-4u$, while selection increases mutation. In the limit of long strings $L(1-q)$ converges to a finite value depending only on the average efficacy.

6 Transience

For finite strings, the initial transient behaviour will be very important. This is particularly true for crossover. Supposing we crossed two strings both with efficacy κ_1 and with a pair correlation of q. We create two children one with efficacy $\kappa_1 + C$, the other with efficacy $\kappa_1 - C$ (since the children are complementary). It is easy to show that C has a binomial distribution with zero mean and variance

$$(1-q)L\frac{\kappa_1}{L}\left(1-\frac{\kappa_1}{L}\right). \tag{12}$$

The typical change in efficacy is of the order of the square root of the variance or of order $\sqrt{(1-q)L}$. For a random population $q=0$ and so that in the initial or transient dynamics the rate of change of efficacy will be of order $\sqrt{L}$. The population will correlate rapidly due to selection so that in the long time or asymptotic regime the rate of change of efficacy was of order 1. It is useful to distinguish between these two regimes, initially a GA will be evolving very quickly in its transient state until the population has converged and the GA slows down to its asymptotic velocity. This transient time remains finite even for very long strings, while the asymptotic behaviour takes of order L generation to reach the vicinity of the equilibrium point. Thus as L increases the transient behaviour becomes less and less relevant.

One of the intriguing questions is whether it was possible to exceed the equilibrium fitness for finite size strings starting from a random population? However, we found no sets of parameters where we could exceed the equilibrium fitness this way. The population would always converge to the asymptotic behaviour by the time that the average fitness reached its equilibrium value. Why this should be remains an open question.

7 Discussion

The limit of long strings provides a transparent framework for understanding the dynamics of GAs. It allows us to separate the long time, asymptotic dynamics from the initial, transient dynamics. These can be studied separately and provide a more profound understanding of how GAs work. In this paper we focussed on the problem of ones-counting which is a particularly simple problem. However, the long string limit demonstrates clearly the way that crossover changes the dynamics. This programme is at a very early stage, with many open questions. The analysis of more complicated fitness landscapes with many local minima would provide much greater insight into the dynamics of real GAs, but this may be a highly non-trivial exercise. Although it is often possible to simulated different size problems it may not be possible to get close to a limit. It is also possible that with many local minima the asymptotic behaviour depends on the initial conditions. Nevertheless the first step in this direction indicates that this may be a fertile territory for investigating GA dynamics.

Acknowledgements

I would like to thank the anonymous referees for their comments and Bill Spears and Nikolay Nikolaev for their constructive criticism of the manuscript.

A Mathematical Details

In the text we left out many of the mathematical details. In this appendix we sketch how the key equations can be derived. Full derivations would, however, be too long and have been given elsewhere.

A.1 Cumulants

Cumulants provide a convenient measure for the shape of a distribution. Given some distribution $\rho(x)$ we can compute the cumulants, κ_n through the cumulant generating function $G(t)$ defined as

$$G(t) = \log\left(\int_{-\infty}^{\infty} \mathrm{e}^{t\,x}\,\rho(x)\,\mathrm{d}x\right) = \sum_{n\geq 1} \frac{\kappa_n t^n}{n!}.$$

From this definition we see

$$\kappa_n = \left.\frac{\partial^n G(t)}{\partial t^n}\right|_{t=0} = G^{(n)}(0)$$

and, consequently, the first two cumulants are

$$\begin{aligned}\kappa_1 &= \int_{-\infty}^{\infty} x\,\rho(x)\,\mathrm{d}x \\ \kappa_2 &= \int_{-\infty}^{\infty} x^2\,\rho(x)\,\mathrm{d}x - \left(\int_{-\infty}^{\infty} x\,\rho(x)\,\mathrm{d}x\right)^2.\end{aligned}$$

In other words the first two cumulants are the average value of the distribution and the variance of the distribution. The third cumulant is equal to the third moment around the average, the forth cumulant is a function of the fourth moment and the second moment. From the definition it is apparent that for a Gaussian all cumulants of third order and higher will be zero. Thus the higher cumulants can be regarded as a measure of the distortion of a distribution away from a Gaussian.

A.2 Selection

We consider Boltzmann selection, as its effect on the cumulants is simple to calculate. For simplicity we will consider the effect on an infinite population. Assuming that the distribution of efficacies before selection is given by $\rho(E)$, then the distribution after selection is given by

$$\frac{\mathrm{e}^{\beta E}\rho(E)}{\int_{-\infty}^{\infty} \mathrm{e}^{\beta E}\,\rho(x)\,\mathrm{d}x}.$$

We can calculate the cumulants using a cumulant generating function

$$G_s(t) = \log\left(\int_{-\infty}^{\infty} \mathrm{e}^{(t+\beta)x}\, \rho(x)\,\mathrm{d}x\right).$$

Note that the normalization factor can be ignored. The cumulants after selection are given by the derivative, $\kappa_n^s = G_s^{(n)}(0)$, however, we can simplify this expression as t and β play a similar role. Thus we can use the generating function

$$G_s(\beta) = \log\left(\int_{-\infty}^{\infty} \mathrm{e}^{\beta x}\, \rho(x)\mathrm{d}x\right).$$

The cumulants after selection are given by $\kappa_n^s = G_s^{(n)}(\beta)$, but since $G_s(\beta)$ is the generating function for the cumulants of the original distribution we have

$$\kappa_n^s = \frac{\partial^n}{\partial \beta^n} \sum_{m\geq 1} \frac{\kappa_m \beta^m}{m!} = \sum_{m\geq 0} \frac{\kappa_{n+m}\beta^m}{m!}.$$

Consequently, the average efficacy after selection is given by

$$\kappa_1^s = \kappa_1 + \beta\kappa_2 + \frac{\beta^2}{2}\kappa_3 + \cdots \tag{13}$$

For small β the increase in the average efficacy is proportional to κ_2 — this is the origins of the first term in equation (7). This derivation ignores finite population effects. These are treated in detail in reference [5].

A.3 Mutations

The probability of mutating any bit is u. Since the string consists of E^μ ones the average number of detrimental mutations is uE^μ. Similarly the average number of beneficial mutations is $u(L - E^\mu)$. The average change in efficacy of an individual is thus $u(L - 2E^\mu)$. Averaging over the whole population the change in κ_1 is equal to $u(L - 2\kappa_1)$. This is the origin of the second term in equation (7).

A.4 Crossover

Uniform crossover takes two strings $\vec{X}^\mu$ and $\vec{X}^\nu$ and produces a child $\vec{X}^\alpha$ such that

$$X_i^\alpha = c_i X_i^\mu + (1 - c_i) X_i^\nu$$

where

$$c_i = \begin{cases} 1 & \text{with probability } 1/2 \\ 0 & \text{with probability } 1/2. \end{cases}$$

If we average over all ways of doing mutation

$$\langle X_i^\alpha \rangle_{mut} = \frac{1}{2} X_i^\mu + \frac{1}{2} X_i^\nu.$$

We can use this to compute what average effect of crossover on the efficacies.

$$\begin{aligned} \langle E^\alpha \rangle_{mut} &= \left\langle \sum_i X_i^\alpha \right\rangle_{mut} \\ &= \frac{1}{2} \sum_i X_i^\mu + \frac{1}{2} \sum_i X_i^\nu = \frac{1}{2} E^\mu + \frac{1}{2} E^\nu. \end{aligned}$$

Thus on average the child has the mean efficacy of its parents. Averaging over all members of the population we find the average efficacy, κ_1, is unchanged by crossover.

We can use the same procedure to calculate the effect of crossover on the second cumulant. We first consider

$$\begin{aligned} \left\langle (E^\alpha)^2 \right\rangle_{mut} &= \left\langle (\sum_i X_i^\alpha)^2 \right\rangle_{mut} \\ &= \sum_{i \neq j} \left\langle X_i^\alpha X_j^\alpha \right\rangle + \sum_i \left\langle (X_i^\alpha)^2 \right\rangle_{mut}. \end{aligned}$$

But, $(X_i^\alpha)^2 = X_i^\alpha$, since X_i^α is zero or one. Furthermore, since crossover acts independently at each loci $\left\langle X_i^\alpha X_j^\alpha \right\rangle = \langle X_i^\alpha \rangle \left\langle X_j^\alpha \right\rangle$. Thus

$$\begin{aligned} \left\langle (E^\alpha)^2 \right\rangle_{mut} &= \sum_{i \neq j} \langle X_i^\alpha \rangle \left\langle X_j^\alpha \right\rangle + \sum_i \langle X_i^\alpha \rangle_{mut} \\ &= \sum_{i \neq j} \left(\frac{1}{2} X_i^\mu + \frac{1}{2} X_i^\nu \right) \left(\frac{1}{2} X_j^\mu + \frac{1}{2} X_j^\nu \right) + \sum_i \left(\frac{1}{2} X_i^\mu + \frac{1}{2} X_i^\nu \right) \\ &= \sum_{i = j} \left(\frac{1}{2} X_i^\mu + \frac{1}{2} X_i^\nu \right) \left(\frac{1}{2} X_j^\mu + \frac{1}{2} X_j^\nu \right) \\ &\quad + \sum_i \left(\left(\frac{1}{2} X_i^\mu + \frac{1}{2} X_i^\nu \right) - \left(\frac{1}{2} X_i^\mu + \frac{1}{2} X_i^\nu \right)^2 \right) \\ &= \left(\frac{1}{2} E^\mu + \frac{1}{2} E^\nu \right)^2 + \frac{1}{8} \left(1 - \sum_i (2X_i^\mu - 1)(2X_i^\nu - 1) \right). \end{aligned}$$

The second cumulant after selection is given by

$$\kappa_2^s = \left\langle \frac{1}{P}\sum_\alpha (E^\alpha)^2 - \frac{1}{P(P-1)}\sum_{\alpha\neq\beta} E^\alpha E^\beta \right\rangle_{mut} .$$

Substituting in the expressions for $\langle (E^\alpha)^2\rangle_{mut}$ and $\langle E^\alpha\rangle_{mut}$ we obtain equation (9). A more comprehensive treatment of crossover is given in reference [2].

References

[1] A. Prügel-Bennett and J. L. Shapiro. An analysis of genetic algorithms using statistical mechanics. *Phys. Rev. Lett.*, 72(9):1305–1309, 1994.

[2] A. Prügel-Bennett and J. L. Shapiro. The dynamics of a genetic algorithm for simple random ising systems. *Physica D*, 104:75–114, 1997.

[3] M. Rattray. The dynamics of a genetic algorithm under stabilizing selection. *Complex Systems*, 9(3):213–234, 1995.

[4] M. Rattray and J. Shapiro. The dynamics of genetic algorithms for a simple learning problem problem. *J. Phys. A.*, 29:7451–7473, 1996.

[5] A. Prügel-Bennett. Modelling evolving populations. *J. Theor. Biol.*, 185:81–95, 1997.

[6] J. L. Shapiro and A. Prügel-Bennett. Genetic algorithms dynamics in two-well potentials with basins and barriers. In R. K. Belew and M. D. Vose, editors, *Foundations of Genetic Algorithms 4*, pages 101–139, California, 1997. Morgan Kaufmann.

[7] M. Rattray. Noisy fitness evaluations in genetic algorithms and the dynamics of learning. In R. K. Belew and M. D. Vose, editors, *Foundations of Genetic Algorithms 4*, pages 117–139, San Francisco, 1996. Morgan Kaufmann.

[8] S. Bornholdt. Probing genetic algorithm performance of fitness landscapes. In R. K. Belew and M. D. Vose, editors, *Foundation of Genetic Algorithms 4*, pages 141–154, San Francisco, 1996. Morgan Kaufmann.

[9] G. Woodcock and P. G. Higgs. Population evolution in a single peak fitness landscape. *J. Theor. Biol.*, 176:61–73, 1996.

[10] M. Rattray. *Modelling the Dynamics of Genetic Algorithms using Statistical Mechanics.* PhD thesis, Manchester University, Manchester, UK, 1996.

[11] A. Rogers and A. Prügel-Bennett. Modelling the dynamics of steady-state genetic algorithms. In W. Banzhaf and C. Reeves, editors, *Foundations of Genetic Algorithms 5*, San Francisco, 1998. Morgan Kaufmann. In press.

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