MODELLING GENETIC ALGORITHMS AND EVOLVING POPULATIONS

By
Alexander Rogers
B.Sc.(Hons)

A thesis submitted for the degree of
Doctor of Philosophy

Department of Electronics and Computer Science
University of Southampton
United Kingdom

September 2000

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING

ELECTRONICS AND COMPUTER SCIENCE DEPARTMENT

Doctor of Philosophy

Modelling Genetic Algorithms and Evolving Populations

by Alexander Rogers

A formalism for modelling the dynamics of genetic algorithms using methods from statistical physics, originally due to Prügel-Bennett and Shapiro, is extended to ranking selection, a form of selection commonly used in the genetic algorithm community. The extension allows a reduction in the number of macroscopic variables required to model the mean behaviour of the genetic algorithm. This reduction allows a more qualitative understanding of the dynamics to be developed without sacrificing quantitative accuracy.

The work is extended beyond modelling the dynamics of the genetic algorithm. A caricature of an optimisation problem with many local minima is considered — the basin with a barrier problem. The first passage time — the time required to escape the local minima to the global minimum — is calculated and insights gained as to how the genetic algorithm is searching the landscape. The interaction of the various genetic algorithm operators and how these interactions give rise to optimal parameters values is studied.

Contents

List of Figures

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institution of learning.

Acknowledgements

I would like to thank Adam Prügel-Bennett for providing direction and supervision throughout the course of this research and for not solving all the problems himself. I would also like to thank Neil Lawrence and Bart Naudts for many useful discussions and Margaret Ann and Eleanor Mary for providing diversion.

I gratefully acknowledge a research studentship awarded by the Engineering and Physical Sciences Research Council (ref. 97307210).

Chapter 1

Introduction

1.1 The Genetic Algorithm

The genetic algorithm (GA) came to popularity through the work of John Holland [11] in 1975. It is now commonly seen as a generic stochastic search algorithm and as such is often applied to combinatorial optimisation problems. These problems generally exhibit un-characterised problem spaces which are highly dimensional and have many local minima; features which prevent the use of traditional optimisation techniques.

At its simplest the genetic algorithm functions as a simple model of an evolving population. Potential solutions to the problem under study are mapped onto a binary string which represents the genetic material of the individual. A population of such individuals is maintained and allowed to evolve in a caricature of natural evolution. A cost function maps the encoded solution to a fitness value on which selection acts: replicating fitter individuals and culling the less fit. Mutation acts to generate small changes in the genetic code of each individual and recombination or crossover allows individuals to exchange genetic material. By repeating this process, it is hoped that good solutions to the problem are evolved.

Genetic algorithms operate in the realm of stochastic search operators and compete with more established techniques such as the Metropolis algorithm [17] or simulated annealing [14]. In these algorithms, the problem is again represented in a form which allows local moves to be generated. If only moves which improve fitness are accepted, the algorithms rapidly become trapped in local minima. To avoid this, the Metropolis algorithm allows moves which are detrimental to fitness with some

small probability. The probability of these moves is controlled by a temperature parameter. If the criteria for accepting moves is chosen correctly, the algorithm asymptotically samples according to the Gibbs distribution.

If the temperature is low and the system is in equilibrium, there is a high probability of sampling the global minimum. One difficulty of this approach is that at low temperature, reaching equilibrium can be very slow. Simulated annealing remedies this by allowing a gradual lowering of the temperature – annealing. If done slowly, so that the system remains in equilibrium, the algorithm is shown to converge to the global minimum. In practice, such time scales are not available and the algorithm eventually freezes into a local minimum. Choosing the annealing schedule is critical to the success of any simulated annealing algorithm.

The genetic algorithm is differentiated from these stochastic search methods by the maintenance of a population and the possibility of performing recombination. In the common understanding of the genetic algorithm, the population of individuals, performing local search through mutation, allows the algorithm to escape local minima. Crossover allows moves on the landscape which are more global in nature, potentially finding new fitter areas for further search. Whilst there is no single parameter such as temperature, it is clear that the dynamics of the genetic algorithm are parameterised indirectly through the choice of selection scheme, mutation rate and crossover.

Understanding the interplay of these various parameters is a very complex problem and little progress has been made. Yet without some theoretical basis, practitioners are forced to work by a set of 'rules of thumb' which have been shown to work on previous problems. There is a clear need for a comprehensive theory which would give some insight into the sort of problems where genetic algorithms should excel and gives some guidance as to how to set the various parameters.

1.2 Genetic Algorithm Theory and Modelling

The original work of Holland [11] introduced the idea of schema. The schema is a subset of the binary string which contributes to the fitness of the individual. Holland argued that whilst explicitly calculating the fitness of only P individuals, the genetic algorithm implicitly calculates and processes P^3 schema. He called this *implicit parallelism* and argued that this was the basis of the search power of the genetic algorithm.

Much of the genetic algorithm theory literature is still based around the schema theory and it has received renewed interest recently through the work of Stephens [44]. The original problems remain despite being reformulated. It results in an inequality for the probability of occurrence of each schema in the next generation, if the average fitness of each schema in the current generation is known. It is thus not able to predict the dynamics of the evolving population and it is unclear how the schema theorem will help understanding of genetic algorithms.

A less expansive approach is to try to develop a simple model of the genetic algorithm. Rather than attempting to make general comments on a broad range of algorithms and problems, the details of a number of simple cases may be studied to try to develop an insight into how the genetic algorithm is working. To be of use, such a model must capture the essential elements of the genetic algorithm but not become so complex that the underlying process is obscured by the details. Creating such a model is difficult for a number of reasons:

- The problem size is extremely large. For a binary string of typical length 100, there are 2^{100} possible genotypes or binary combinations.
- The small population sizes of typically 100 individuals, sample a very small fraction of the total search space. Thus theoretical results based on infinite population models are often misleading. Indeed much of the interesting phenomena observed in the evolution of genetic algorithms are features of a finite population.
- The translation from the encoding of the binary string to a fitness value for any particular individual is often highly non-linear and simple cases need to be studied to make any real progress.
- The interaction between population members in terms of the selection of fitter members and the transfer of genetic material between them in crossover is fundamental to the power of the genetic algorithm. It is thus not possible to model an average population member and the entire population of interacting individuals must be modelled.

Due to these complications, most attempts at modelling genetic algorithms concentrate on simple models and problem spaces. Although this means that the models are caricatures of the real world, it is hoped that the insights gained can be used in the study of real problems and real algorithms.

1.2.1 Microscopic models

The evolution of a genetic algorithm from one generation to another is simply subject to the influence of selection, mutation and crossover. If at each generation, the population can be exactly described, the resulting evolution may be described by a Markov chain. This approach has been pioneered by Vose and collaborators [50, 51, 27, 53] and is detailed in Vose's recent book [52].

Such models consider all the microscopic detail of the population and construct transition matrices which describe the change in the population due to the various genetic operators. To describe problems of reasonable size, very large matrices are generated and in general, these may only be solved in the infinite population limit.

Relating infinite population results back to the finite population case must be done carefully and it is here that the microscopic models encounter difficulties. The state space of the genetic algorithm is vast compared to the typical size of population used. Thus the response of an infinite population can be very different from that of the finite population under analysis and many of the interesting features of the genetic algorithm relate to the existence of a finite population.

1.2.2 Macroscopic models

Whilst analysis of the population at the microscopic level can be exact, it is often not what is of interest. Macroscopic descriptions such as the average fitness of the population or the fittest member of the population tend to be more useful in gaining a qualitative understanding of what is happening.

This situation is similar to that in physics when modelling the properties of a material. The behaviour and state of every individual atom contributes to the bulk properties of the material such its magnetism or temperature. However it is these bulk properties which are of interest and much can be said about the behaviour of the material without having to worry about the microscopic details. For example, thermodynamics allows the behaviour of gases to be predicted without resorting to a calculation of the velocity of every molecule.

When describing such systems in macroscopic terms, some information is being discarded. In general, properties which can not be calculated from the macroscopic descriptors can be inferred by a maximum entropy argument based on the large numbers present. Macroscopic models of genetic algorithms are often referred to as statistical physics or statistical mechanics models for this reason.

Crutchfield and Van Nimvegen at the Santa Fe Institute have taken this approach [49, 48]. By starting at the level of the transition matrices and extracting the macroscopic descriptors which are of interest, they have been able to model the dynamics of a simple genetic algorithm on the royal road functions [18]. Unfortunately including crossover into their formalism has proved to be very difficult. Whilst this does not prevent the analysis of problems such as the royal road function, where crossover is shown to be of little benefit, it limits the applicability to other problem spaces.

Another approach to the macroscopic modelling of the genetic algorithm is to start with a macroscopic description of the population, or more precisely its fitness distribution, and model the effect of selection. Theile and Blickle [3] compared selection schemes by considering the mean and variance of fitness in an infinite population and Mühlenbein [26, 23, 24] modeled a special class of genetic algorithm using a similar technique. In general, the accuracy of these models precluded them from considering more that one generation and they gave qualitative results rather then quantitative predictions of the dynamics of a genetic algorithm.

The formalism developed by Prügel-Bennett and Shapiro [29, 43, 41, 30, 42] and later extended by Rattray [32] represents the most sophisticated of these approaches. The population fitness distribution is described by its cumulants and finite population effects are calculated accurately. The accuracy of the model allows the calculations for one generation to be iterated and the dynamics of the genetic algorithm followed over many generations. Significantly, not only is selection considered but the formalism is extended to mutation and crossover. This allows the dynamics of a simple genetic algorithm to be completely modelled on a number of simple problems. It is this formalism which provides the basis for the work in this thesis.

1.2.3 Biological models

For obvious reasons many of the simple models of genetic algorithms are very similar to biological models of evolving populations. The field of quantitative genetics is much more established than that of genetic algorithms with the work of Fisher [10] and Wright [54] dating back to the nineteen twenties and thirties.

Despite borrowing some terminology from population geneticists, for example genetic drift, there is little crossover from this work into mainstream genetic algorithm literature. This is probably because of a difference of focus between the two groups. Biologists are interested in the frequency of a particular trait within a population

and thus the models concentrate on the particular allele frequency at each loci; equivalent to considering the individual probabilities that each bit is either a one or zero.

Despite this difference in focus, models of evolving populations which now appear to be very similar to simple models of genetic algorithms have been proposed by many researchers, amongst them Moran [21, 20, 19] and Kimura [13]. Typically these are solved by making a diffusion equation approximation to the Markov chain analysis; a technique which is of direct use in the microscopic modeling of genetic algorithms.

1.3 Thesis Goal

The aim of modelling the genetic algorithm is to gain insight into how the algorithm works. Genetic algorithms are complex systems and to model them a number of simple cases must be considered. The aim is always to reduce the complexity of the model without removing the fundamental features which are of interest. By studying these simple cases, techniques and insights are gained which will hopefully be of use to other more complex cases.

The formalism developed by Prügel-Bennett and Shapiro and later extended by Rattray has been applied to a range of simple cases by these researchers. Due to the complexities involved, much of the focus of the work so far has been in deriving the formalism. Whilst being quantitatively accurate, the model developed is not particularly amenable to qualitative analysis.

In this thesis, the formalism is extended to a more common form of selection scheme and in the process, significantly simplified. This simplification reduces the number of macroscopic variables required to describe the genetic algorithm and allows a more qualitative understanding of the dynamics to be developed without sacrificing quantitative accuracy.

The first steps are taken in extending the work beyond simply modeling the dynamics of the genetic algorithm. A caricature of an optimisation problem with many local minima is considered – the basin with a barrier problem. The first passage time – the time required to escape the local minima to the global minimum – is calculated and insights gained into how the genetic algorithm is searching the landscape. The interaction of the various genetic algorithm operators and how these interactions give rise to optimal parameters values is studied.

The work presented in this thesis has previously been published in a number of sources [35, 34, 37, 39, 36, 38].

1.4 Thesis Outline

The thesis is presented as detailed below:

- Chapter 2 - Statistical Physics Formalism
  The formalism developed by Prügel-Bennett and Shapiro is presented and extended to the case of steady state genetic algorithms.
- Chapter 3 - Genetic Drift in Selection Schemes
  The comparison of genetic drift in selection schemes is generalised by developing a simple method of calculating the rate of genetic drift. The technique is applied to a range of genetic algorithm selection schemes and those used in evolutionary strategies.
- Chapter 4 - Ranking Selection
  The formalism originally developed by Prügel-Bennett and Shapiro is extended to the case of ranking selection. Finite population effects for both roulette wheel and stochastic universal sampling are calculated and compared.
- Chapter 5 - Crossover and the Onemax Problem
  The dynamics of a full genetic algorithm under selection, mutation and crossover is modelled on the onemax problem. Closed form expressions are derived for the equilibrium point.
- Chapter 6 - Stabilising Selection
  A model of a hard optimisation problem – the basin with a barrier problem – is introduced and the analysis of ranking selection is extended to the case of stabilising selection in order to model the dynamics of the genetic algorithm on this problem.
- Chapter 7 - Solving the Basin with a Barrier
  The analysis of stabilising selection is used to calculate the time to solve the basin with a barrier problem. The influence of the genetic algorithm parameters – population size, mutation rate and selection scheme – on this time are explored.
- Chapter 8 - Biological Models
  The models of genetic algorithms developed in the thesis are compared to biological models of evolving populations and some biologically interesting results developed.

- Chapter 9 - Conclusions and Future Directions
  The limitations of the model developed so far is discussed and the direction of future work highlighted.

Chapter 2

Statistical Physics Formalism

2.1 Introduction

The formalism developed by Prügel-Bennett and Shapiro [29, 43, 41, 30, 42] and later extended by Rattray [32] allows the dynamics of a simple genetic algorithm to be modelled. It is based on a macroscopic description of the population fitness distribution using the cumulants of the distribution. Selection is modelled by calculating the effect on these cumulants.

The formalism was initially developed considering generational Boltzmann selection and is presented first in this form. The value of the formalism for exploring questions of interest in the genetic algorithm community is demonstrated by extending it to the case of steady state selection. This work represents the first formal comparison of the two schemes – previous analysis being based on empirical observations.

2.2 Generational Selection

In the generational or canonical genetic algorithm, selection is applied once to an initial population, generating a new population of P individuals. As selection operates solely on the fitness of individuals within the population, a population with discrete fitnesses can be considered without initially having to consider the details of any particular problem space or encoding.

2.2.1 The Model Genetic Algorithm

The model genetic algorithm considered consists of a population of P individuals each with some assigned fitness value, F. At each time generation, Boltzmann

roulette wheel selection is performed whereby P population members are drawn independently with replacement from the original population using the weighting

$$w_\alpha = \frac{\mathrm{e}^{\beta F_\alpha}}{Z} \tag{2.1}$$

where β is a parameter allowing control of the selection strength and Z is the normalisation factor over the population

$$Z = \sum_{\alpha=1}^{P} \mathrm{e}^{\beta F_\alpha} \tag{2.2}$$

commonly referred to as the partition function.

2.2.2 Selection

The macroscopic variables chosen to model the fitness distribution are the distribution cumulants, denoted by K_n. These are natural variables to describe distributions close to a Gaussian. The first two cumulants are familiar as the mean and variance. Higher order cumulants describe the deviation away from a Gaussian - the third and fourth being related to the skewness and kurtosis. Unlike distribution moments, cumulants are invariant under a change of mean.

The cumulants of any continuous distribution, $\rho(F)$, may be calculated from the cumulant generating function

$$K_n = \frac{\mathrm{d}^n}{\mathrm{d}z^n} G(z) \,|_{z=0} \tag{2.3}$$

where

$$G(z) = \log\left(\int \rho(F)\, \mathrm{e}^{zF} \mathrm{d}F\right). \tag{2.4}$$

In an infinite population, selection is deterministic. However, genetic algorithms typically consider small populations where the stochastic nature of the selection scheme is important. The ensemble average over many such finite populations is considered. In doing so, the ensemble is described as a continuous probability distribution. Any particular finite population is considered as a sample of this ensemble probability distribution.

The generating function for the ensemble distribution is thus given by the average over the weighted finite population

$$G(z) = \left\langle \log\left(\sum_{\alpha=1}^{P} w_\alpha \mathrm{e}^{zF_\alpha}\right)\right\rangle \tag{2.5}$$

where the angle brackets denote averaging over all ways of sampling a finite population from the ensemble probability distribution and all ways of performing selection on that finite population. Using the Boltzmann weighting from equation (2.1) allows the logarithm to be expanded

$$\log\left(\sum_{\alpha=1}^{P} w_\alpha \mathrm{e}^{zF_\alpha}\right) = \log\left(\sum_{\alpha=1}^{P} \mathrm{e}^{\beta F_\alpha}\mathrm{e}^{zF_\alpha}\right) - \log\left(\sum_{\alpha=1}^{P} \mathrm{e}^{\beta F_\alpha}\right). \tag{2.6}$$

The second term does not depend on z and can thus be neglected. Changing variables gives

$$G(\beta) = \left\langle \log\left(\sum_{\alpha=1}^{P} \mathrm{e}^{\beta F_\alpha}\right)\right\rangle \tag{2.7}$$

where the cumulants after selection are now given by

$$K_n = \frac{\mathrm{d}^n}{\mathrm{d}\beta^n} G(\beta). \tag{2.8}$$

To evaluate equation (2.7), Prügel-Bennett and Shapiro took a technique from statistical physics used by Derrida to solve the random energy model [7]. The logarithm is represented as

$$\log(Z) = \int_0^\infty \frac{\mathrm{e}^{-t} - \mathrm{e}^{-tZ}}{t}\mathrm{d}t. \tag{2.9}$$

The generating function thus becomes

$$G(\beta) = \int_0^\infty \frac{\mathrm{e}^{-t} - g^P(t,\beta)}{t}\mathrm{d}t \tag{2.10}$$

where

$$g(t,\beta) = \int \rho(F)\,\mathrm{e}^{-t\mathrm{e}^{\beta F}}\mathrm{d}F \tag{2.11}$$

and $\rho(F)$ describes the ensemble fitness distribution as a continuous function.

2.2.3 Integrating around a Gaussian

The generating function derived above may be evaluated numerically assuming the ensemble fitness distribution, $\rho(F)$, can be described in terms of its initial cumulants. A convenient form is an expansion around a Gaussian using the Gram-Charlier expansion

$$\rho(F) = \frac{1}{\sqrt{2\pi K_2}} \exp\left(\frac{-(F-K_1)^2}{2K_2}\right)\left[1+\sum_{i=3}^{n}\frac{K_i}{i!K_2^{i/2}}H_i\left(\frac{F-K_1}{\sqrt{K_2}}\right)\right]$$

where $H_i(x)$ are Hermite polynomials

$$H_i(x) = (-1)^i e^{\frac{x^2}{2}} \frac{\mathrm{d}^i}{\mathrm{d}x^i}\left(\mathrm{e}^{-\frac{x^2}{2}}\right) \tag{2.12}$$

and n is the number of cumulants used. Whilst the resulting distribution is close to a Gaussian and has the correct cumulants, it can be negative in places and is thus an approximation to the correct probability distribution.

When n is two, the expansion has exactly the form of a Gaussian. Adding in further terms distorts the Gaussian to give the desired distribution. The first two terms expanded out are

$$\begin{aligned} H_3(x) &= x^3 - 3x \\ H_4(x) &= x^4 - 6x^2 + 3. \end{aligned} \tag{2.13}$$

To fully describe the distribution, an infinite number of cumulants are required. In general, a truncated set may be used to give quantitatively good results. Prügel-Bennett and Shapiro found that four cumulants were normally sufficient but calculating up to eight was sometimes necessary to generate quantitative agreement with simulation results.

2.2.4 Finite Sample Effects

Evaluating equation (2.10) and (2.11) using the Gram-Charlier expansion gives the cumulants of the ensemble, K_n, after selection. This includes the stochastic effects which arise through selection operating on a finite population. However the cumulants of any finite population drawn from the ensemble, κ_n, will differ slightly due to well known sampling effects.

The cumulants of a finite population drawn from the ensemble may be found from the identities

$$\begin{aligned}\kappa_1 &= K_1 \\ \kappa_2 &= P_2 K_2 \\ \kappa_3 &= P_3 K_3 \\ \kappa_4 &= P_4 K_4 - 6P_2 \left(K_2\right)^2 / P \end{aligned} \tag{2.14}$$

where $P_2 = \left(1 - \frac{1}{P}\right)$, $P_3 = P_2\left(1 - \frac{2}{P}\right)$ and $P_4 = P_2\left(1 - \frac{6}{P} + \frac{6}{P^2}\right)$. Here K_n represents the cumulants of the continuous probability distribution and κ_n are the cumulants of a finite population sampled from it. These results were originally shown by Fisher [10].

2.2.5 Weak Selection Expansion

Calculating the integrals in equations (2.10) and (2.11) numerically gives very little intuitive insight into the processes at work. Under weak selection — small β — the integrals may be performed analytically by expressing the weighting as a series expansion in terms of β. The result is a set of truncated cumulant expansions.

$$\begin{aligned}\langle K_1\rangle_s &= K_1 + \beta\left(1 - \frac{1}{P}\right)K_2 + \frac{\beta^2}{2}\left(1 - \frac{3}{P}\right) + \frac{\beta^3}{3!}\left[\left(1 - \frac{7}{P}\right)K_4 - \frac{6}{P}K_2^2\right] + \ldots \\ \langle K_2\rangle_s &= \left(1 - \frac{1}{P}\right)K_2 + \beta\left(1 - \frac{3}{P}\right)K_3 + \frac{\beta^2}{2}\left[\left(1 - \frac{7}{P}\right) - \frac{6}{P}K_2^2\right] + \ldots \\ \langle K_3\rangle_s &= \left(1 - \frac{3}{P}\right)K_3 + \beta\left[\left(1 - \frac{7}{P}\right)K_4 - \frac{6}{P}K_2^2\right] + \ldots \\ \langle K_4\rangle_s &= \left(1 - \frac{7}{P}\right)K_4 - \frac{6}{P}K_2^2 + \ldots \end{aligned} \tag{2.15}$$

where $\langle \ldots \rangle_s$ represents the expected value after selection.

The weak selection expansions gives a great deal more insight as the effect on the cumulants of selection is quite clear.

An initial infinite population described by a Gaussian will exhibit an increase in mean proportional to the selection strength. The variance and higher cumulants of the distribution will be unchanged by selection. The result is a Gaussian with steadily increasing mean.

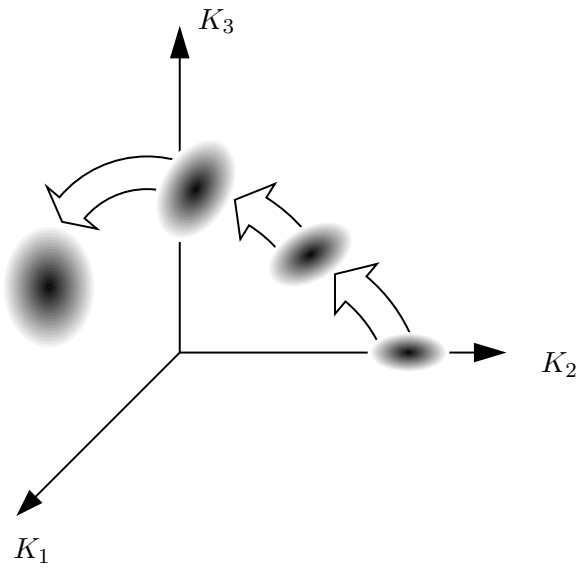


Figure 2.1: Representation of an ensemble of populations evolving in the phase space of macroscopic variables.

The finite population effects calculated from selection lead to a significant deviation from this infinite population case. The variance of the population is decreased by a term proportional to the population size. This effect is well known and referred to as genetic drift.

The higher order cumulants also become significant. The population becomes skewed as the third cumulant goes negative. The skewness causes a further reduction in variance which ultimately causes a slower increase in the mean.

Significantly, the infinite population approximation does not capture the interesting dynamics of the system. Correct calculation of the finite population effects is essential to fully understand the dynamics.

2.2.6 Mean Behaviour

In the analysis so far, only the mean values of the cumulants have been considered. Whilst this is acceptable when only one generation is considered, there are fluctuations about these mean values which become relevant when the entire trajectory of the population is considered. As the cumulant expansions are clearly non-linear, the variance around the mean value of each cumulant will be significant. For example, $\langle K_2^2 \rangle \neq \langle K_2 \rangle^2$. Figure 2.1 shows the ensemble of populations evolving in the phase space of the macroscopic cumulants.

Prügel-Bennett extended the formalism to include the variance of the cumulants and covariances between cumulants [28]. Their effect was generally shown to be small. Inspection of the cumulant expansion in equation (2.15) would lead one to expect this, as non-linear terms have small coefficients in the expansions. In some circumstances, such as calculating the equilibrium point in a system subject to Boltzmann selection and mutation however, the fluctuations in the extremes of the distribution are most significant and can not be ignored.

2.3 Steady State Selection

In the generational genetic algorithm, the entire new population is selected from the past generation at one go. A popular alternative to this is the steady state genetic algorithm where one or more fit population members are selected at a time and used to replace unfit population members.

Understanding the advantages or disadvantages of replacing only a fraction of the population was a goal of some of the earliest work in genetic algorithms. De Jong [5] introduced the term *generation gap* to describe the size of the generation overlap. Many empirical comparisons of generational and steady state genetic algorithms exist in the literature, however it is often the case that other significant changes are made to the algorithm masking any influence of the selection scheme. Whilst some careful comparisons have been performed [6, 46], there is still little understanding of the differences and a theoretical comparison is of value.

2.3.1 Calculating Selection

Under generational selection, the cumulants were calculated directly from the generating function. Calculating steady state selection takes a different approach whereby the cumulants are calculated by considering the change under a finite population when one individual is reproduced and another is deleted from the population.

The cumulants of a finite population are given by the standard definitions

$$\begin{aligned} \kappa_1 &= \langle F \rangle \\ \kappa_2 &= \langle F^2 \rangle - \langle F \rangle^2 \\ \kappa_3 &= \langle (F - \langle F \rangle)^3 \rangle \\ \kappa_4 &= \langle (F - \langle F \rangle)^4 \rangle - 3\,(\kappa_2)^2 \end{aligned} \tag{2.16}$$

where $\langle \ldots \rangle$ represents the expected values over the population.

Under steady state selection, one individual, μ, is selected from the population and reproduced. The population size is kept constant by deleting another individual, ν. The cumulants after selection are thus given by

$$\begin{aligned} \kappa_1 &= \left(\langle F \rangle + \frac{F_\mu}{P} - \frac{F_\nu}{P} \right) \\ \kappa_2 &= \left(\langle F^2 \rangle + \frac{F_\mu^2}{P} - \frac{F_\nu^2}{P} \right) - \left(\langle F \rangle + \frac{F_\mu}{P} - \frac{F_\nu}{P} \right)^2 . \end{aligned} \tag{2.17}$$

Expanding these terms and then averaging over all ways of selecting population member μ and ν gives

$$\begin{aligned} \langle \kappa_1 \rangle_s &= \kappa_1 + \frac{\langle F_\mu \rangle_s}{P} - \frac{\langle F_\nu \rangle_s}{P} \\ \langle \kappa_2 \rangle_s &= \kappa_2 + \frac{\langle F_\mu^2 \rangle_s}{P} - \frac{\langle F_\nu^2 \rangle_s}{P} - 2\kappa_1 \frac{\langle F_\mu \rangle_s}{P} + 2\kappa_1 \frac{\langle F_\nu \rangle}{P} \\ &\qquad - \frac{\langle F_\mu^2 \rangle_s}{P^2} + 2\frac{\langle F_\mu \rangle_s \langle F_\nu \rangle_s}{P^2} - \frac{\langle F_\nu^2 \rangle_s}{P^2} \end{aligned} \tag{2.18}$$

where $\langle \ldots \rangle_s$ represents the average over all ways of performing selection. The higher cumulants are calculated in the same fashion but involve rather more algebra.

Provided that the expected fitnesses of the individuals which are reproduced and deleted can be found, the cumulants after selection can be calculated directly. Whilst the individual to be reproduced is drawn from the population based on its Boltzmann weighting, a number of strategies exist for selecting the individual to be deleted. The two considered here are Boltzmann deletion where it is drawn with an inverse weighting calculated with $-\beta$ rather than β. The simpler alternative is to simply delete at random.

2.3.2 Strong Selection

For a finite population, the expected fitness of the individual selected for reproduction, μ, is simply given by summing over the Boltzmann weightings

$$F_\mu^n = \sum_{\alpha=1}^{P} w_\alpha F_\alpha^n \tag{2.19}$$

where F_μ^n is the nth power of F_μ. To deal with the summation in the partition function, the following identity is used

$$\frac{1}{A} = \int_0^\infty \mathrm{e}^{-tA}\mathrm{d}t. \tag{2.20}$$

Equation (2.19) is transformed to

$$F_\mu^n = \sum_{\alpha=1}^{P} \mathrm{e}^{\beta F_\alpha} F_\alpha^n \int_0^\infty \prod_{\beta=1}^{P} \mathrm{e}^{-t\mathrm{e}^{\beta F_\beta}}\mathrm{d}t. \tag{2.21}$$

Averaging gives

$$\left\langle F_\mu^n \right\rangle_s = P \int_0^\infty \left\langle F_\alpha^n \mathrm{e}^{\beta F_\alpha - t\mathrm{e}^{\beta F_\alpha}} \right\rangle \left\langle \mathrm{e}^{-t\mathrm{e}^{\beta F_\beta}} \right\rangle^{P-1} \mathrm{d}t. \tag{2.22}$$

Describing the ensemble cumulant distribution as a continuous function gives the final result

$$\left\langle F_\mu^n \right\rangle_s = P \int_0^\infty \int_{-\infty}^\infty F^n \mathrm{e}^{\beta F - t\mathrm{e}^{\beta F}} \rho\left(F\right) \mathrm{d}F \left[\int_{-\infty}^\infty \mathrm{e}^{-t\mathrm{e}^{\beta F}} \rho\left(F\right) \mathrm{d}F\right]^{P-1} \mathrm{d}t. \tag{2.23}$$

The expected fitness of the selected individual, μ, may thus be calculated by integrating this result numerically using the Gram-Charlier expansion to describe $\rho\left(F\right)$ in terms of the cumulants of the ensemble fitness distribution.

2.3.3 Weak Selection

The numerical integrals calculated above may again be performed analytically in the case of weak selection. When selection is weak, the extremes of the population become less significant and equation (2.19) may be approximated as

$$\left\langle F_\mu^n \right\rangle_s \approx \frac{\displaystyle\int_{-\infty}^\infty \rho\left(F\right) \mathrm{e}^{\beta F} F^n \mathrm{d}F}{\displaystyle\int_{-\infty}^\infty \rho\left(F\right) \mathrm{e}^{\beta F} \mathrm{d}F}. \tag{2.24}$$

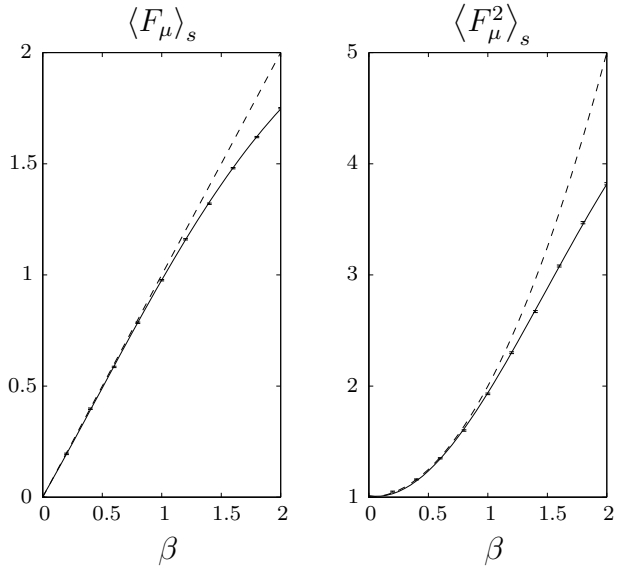


Figure 2.2: Comparison of simulation results, numerical integration (*solid line*) and weak selection approximation (*dashed line*) when selecting with varying selection strength. Simulations are for a population of 100 whose fitnesses are drawn from a unit Gaussian and are averaged over 10 000 selections.

Using the Gram-Charlier expansion, these integrals may be calculated analytically for a truncated set of cumulants. The result is a series expansion in β

$$\begin{aligned}
\langle F_\mu \rangle_s &= K_1 + K_2\beta + \dots \\
\langle F_\mu^2 \rangle_s &= K_2 + K_1^2 + (2K_1K_2 + K_3)\,\beta + \dots \\
\langle F_\mu^3 \rangle_s &= K_3 + 3K_1K_2 + K_1^3 + \left(3K_1K_3 + 3K_2^2 + 3K_1^2K_2 + K_4\right)\beta + \dots \\
\langle F_\mu^4 \rangle_s &= K_4 + 3K_2^2 + K_1^4 + 6K_1^2K_2 + 4K_3K_1 \\
&\quad + \left(4K_2K_1^3 + 4K_1K_4 + 12K_2^2K_1 + 10K_3K_2 + 6K_3K_1^2\right)\beta + \dots \quad .
\end{aligned} \tag{2.25}$$

Figure 2.2 shows a comparison of this weak selection case against the strong selection numerical integration for changing selection strength. The original population consists of 100 members with fitnesses drawn from a unit Gaussian. The error bars are simulation results averaged over 10 000 selections. As selection strength increases, the fitness of the selected member gradually tends to a limit — that of the fittest member of the finite population. The weak selection approximation is clearly valid for selection strengths less than unity.

2.3.4 Small Beta Expansion

Using the terms derived above and an equivalent set for $\langle F_\nu^n \rangle_s$, found by substituting $-\beta$ for β, the cumulants after selection when using Boltzmann deletion may be found

$$\begin{aligned}
\langle K_1 \rangle_s &= K_1 + \frac{2K_2}{P}\beta + \dots \\
\langle K_2 \rangle_s &= K_2 - \frac{2K_2}{P^2} + \frac{2K_3}{P}\beta \ + \dots \\
\langle K_3 \rangle_s &= K_3 - \frac{6K_3}{P^2} + \frac{2K_4}{P}\beta + \dots \\
\langle K_4 \rangle_s &= K_4 - \frac{14K_4}{P^2} - \frac{12K_2^2}{P^2} + \dots \quad .
\end{aligned} \tag{2.26}$$

The case of random deletion can be calculated simply by using

$$\left\langle F_\mu^n \right\rangle_s = \langle F^n \rangle \tag{2.27}$$

where $\langle \dots \rangle$ represents the average over the population. The resulting expressions are

$$\begin{aligned}
\langle K_1 \rangle_s &= K_1 + \frac{K_2}{P}\beta + \dots \\
\langle K_2 \rangle_s &= K_2 - \frac{2K_2}{P^2} + \frac{K_3}{P}\beta + \dots \\
\langle K_3 \rangle_s &= K_3 - \frac{6K_3}{P^2} + \frac{K_4}{P}\beta + \dots \\
\langle K_4 \rangle_s &= K_4 - \frac{14K_4}{P^2} - \frac{12K_2^2}{P^2} + \dots \quad .
\end{aligned} \tag{2.28}$$

The two sets of expressions are clearly very similar. The most obvious difference is simply the factor of two in each term containing the selection strength β. The strategy of deleting members using the Boltzmann weighting leads to a doubling of the effective selection strength. In the weak selection limit, doubling the selection strength and deleting at random is equivalent to using Boltzmann deletion.

2.3.5 Comparison with Simulation Data

Figure 2.3 shows a comparison of simulation results and the theoretical expressions. The simulations were performed with a population of 100 whose initial fitness was drawn from a unit Gaussian. The selection strength is 0.05 and the simulations are averaged over 10 000 runs.

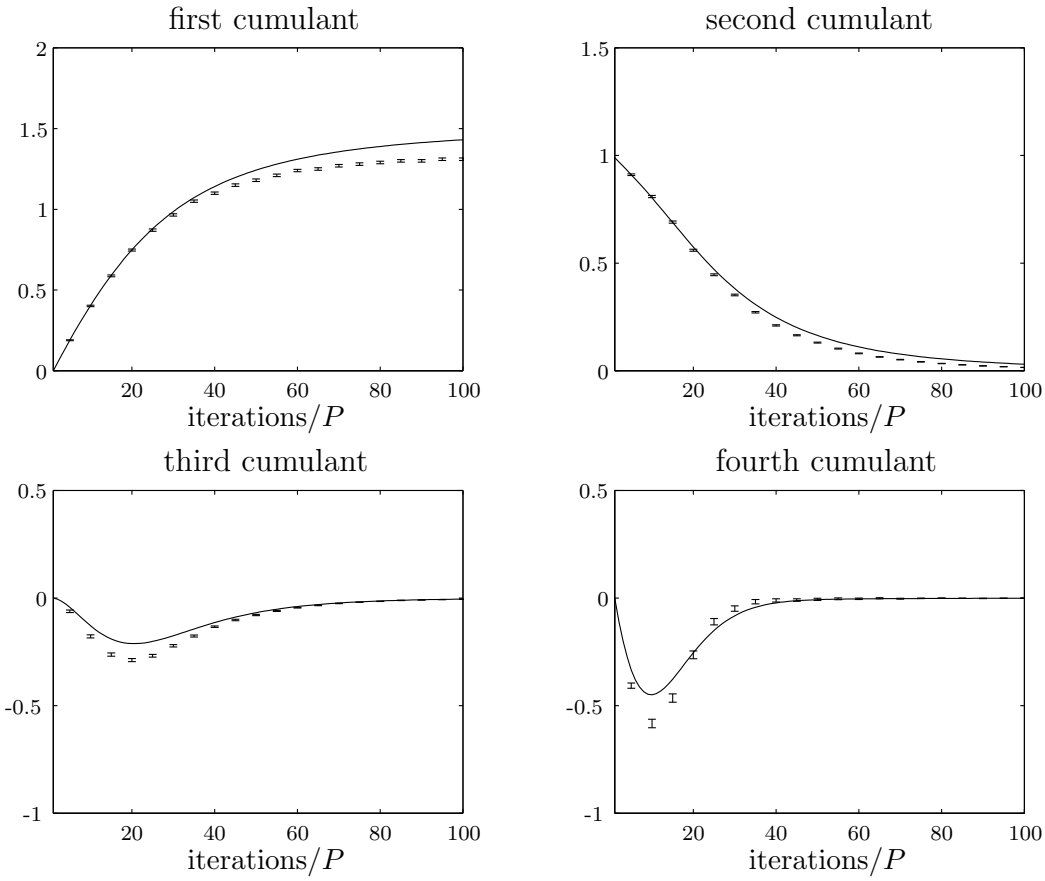


Figure 2.3: Theory predictions plotted against experimental data averaged over 10 000 runs for a simple steady state genetic algorithm using weak selection. The population size is 100 and the selection strength, β, is 0.05.

The theoretical results are calculated using the weak selection expansions for the first six cumulants. The agreement between theory and simulation is qualitatively good. For better quantitative agreement, more cumulants may be calculated.

2.4 Comparison of Generational and Steady State GA

The weak selection expansions for all three selection schemes — Boltzmann deletion, random deletion and generational — allow an easy comparison. The two steady state expressions have terms to $1/P^2$ rather than $1/P$ as they describe the change after each selection/deletion process. In this weak selection case, the cumulants do not change significantly with each selection and we can consider a quasi-static case and compare the two expressions directly despite this factor of P.

The terms independent of selection strength, β, describe the change in the population due to the stochastic nature of the selection scheme — genetic drift. Both the steady state selection schemes exhibit double the rate of genetic drift seen in the generational case. This doubling is due to the extra randomness introduced in the choice of which population member is deleted.

When $\beta = 0$, selection is independent of fitness and the new population is simply randomly sampled from the original. In this case the expressions decouple and become exact

$$\begin{aligned} \langle K_2 \rangle_s &= \left(1 - \frac{1}{P}\right) K_2 \qquad \text{generational selection} \\ \langle K_2 \rangle_s &= \left(1 - \frac{2}{P^2}\right) K_2 \qquad \text{steady state selection.} \end{aligned} \tag{2.29}$$

In an empirical comparison of generational and steady state GA, De Jong [6] noted an increase in a measure he called *allele loss*. This relates approximately to the increased rate of genetic drift inherent in steady state cases. Interestingly, in population genetics this comparison was performed in the nineteen fifties by Moran [21]. Although the analysis was quite different, the same doubling in genetic drift was observed as discussed in chapter eight.

Figure 2.4 shows a comparison of simulation results for generational selection and steady state selection with random deletion. Both use a population size of 100 and a selection strength of 0.05. The increased rate of genetic drift in the steady state genetic algorithm causes the variance to decrease more rapidly. Ultimately this results in a final lower mean fitness.

Interestingly, the three cases can be shown to give approximately the same dynamics by rescaling the parameters. Since Boltzmann deletion exhibits twice the selection strength and twice the rate of genetic drift it gives rise to the same dynamics but in half the time — $P/2$ selections being equivalent to one generation. The same is the case for the steady state genetic algorithm with random deletion if the selection strength is doubled. Figure 2.5 shows the strong selection results overlaid with this rescaling of parameters. The time scales of the steady state algorithms plotted at $P/2$ iterations equal to one generation. The three cases clearly exhibit the same dynamics with the first two cumulants almost perfectly overlaying.

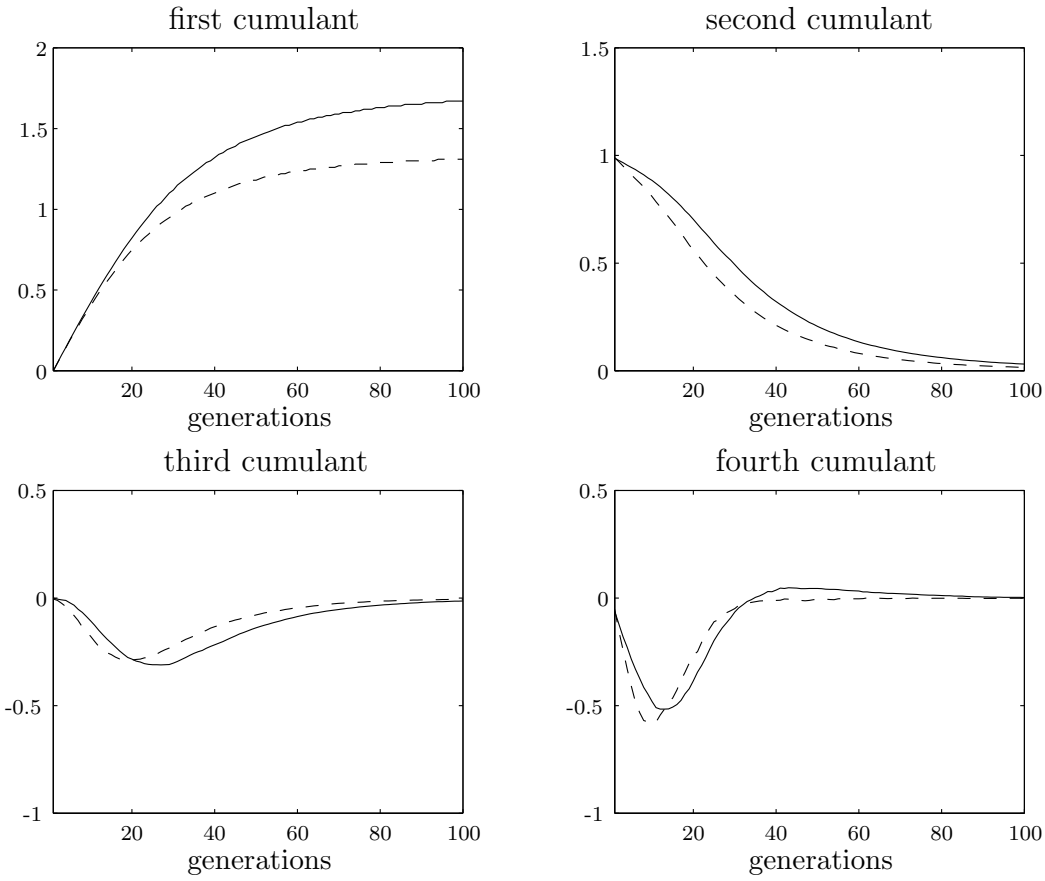


Figure 2.4: Comparison of generational (*solid line*) and steady state (*dashed line*) selection using random deletion. Population size is 100 and selection strength, β, is 0.05.

2.5 Discussion

The formalism developed by Prügel-Bennett and Shapiro gives a theoretical grounding on which to base investigations. The analysis of steady state selection shows that by careful theoretical comparison, results can be elucidated which have defied empirical comparisons. The results of this analysis however were dependent on the weak selection approximation and the derivation of the truncated expressions.

Whilst under the formalism as presented, the use of Boltzmann selection appears to ease the calculations, it also leads to a number of disadvantages. The large number of coupled equations required to describe the dynamics of the system mean that whilst quantitative analysis may be performed accurately, qualitative insights are still difficult in all but the simplest cases.

Under Boltzmann selection, an infinite population behaves qualitatively differently

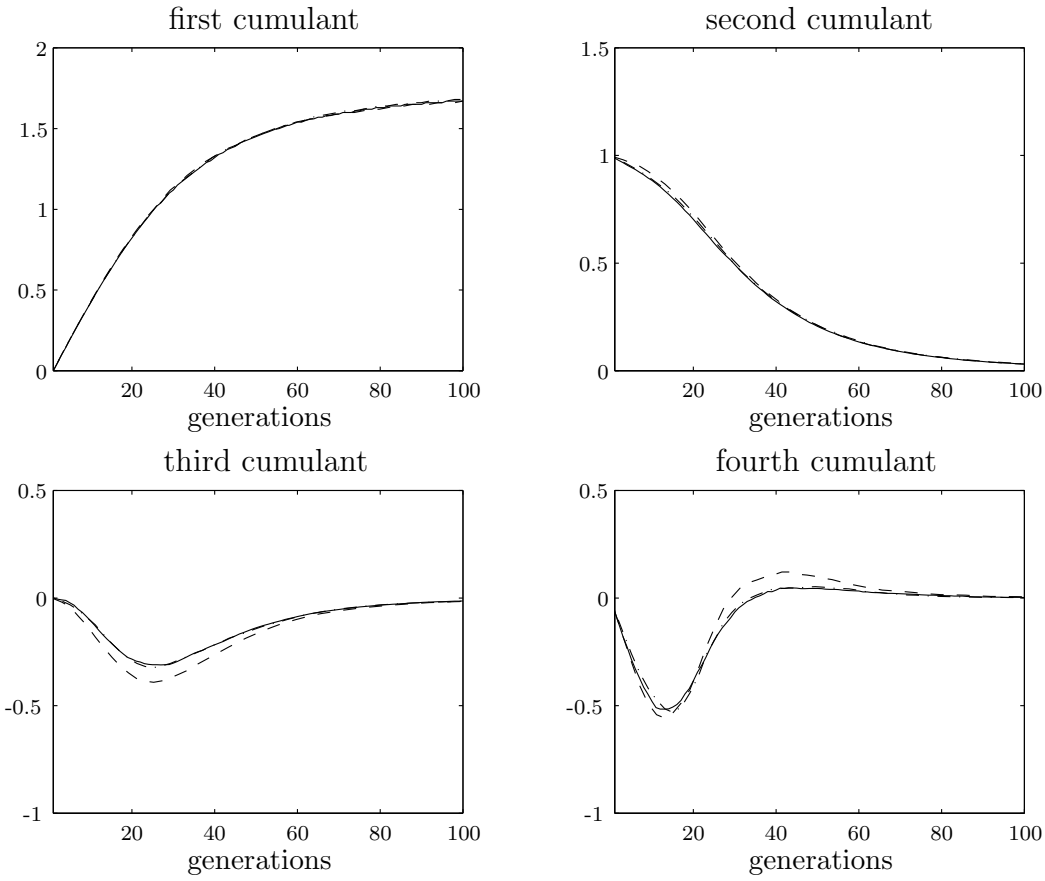


Figure 2.5: Comparison of steady state with random deletion (*dashed line*), steady state with Boltzmann deletion (*dot-dashed line*) and generational selection (*solid line*) when parameters are rescaled. Population size is 100 and for the generational genetic algorithm selection strength, β, is 0.05. Steady state genetic algorithms are rescaled as $P/2$ iterations equal to one generation.

from a finite one. In an infinite population, the mean of the distribution increases unceasingly. Clearly a finite population can only evolve as far as the fittest member of the initial population and thus the finite population effects must be calculated very accurately to capture this behaviour. An infinite population approximation is of no benefit.

Perhaps the strongest objection of the genetic algorithm community is that Boltzmann selection is not commonly used in practice and the weak selection required for the expansions to hold is an unrealistic restriction.

These problems are addressed in later chapters of the thesis. Initially a comparison of selection schemes is developed which allows the rate of genetic drift inherent in the scheme to be compared in a much simpler manner. The formalism is then

applied to the case of ranking or tournament selection — a much more common form of selection. In the process, the major problems of Boltzmann selection are addressed.

Chapter 3

Genetic Drift in Selection Schemes

3.1 Introduction

Genetic drift is a term borrowed from population genetics where it is used to describe changes in gene frequencies through neutral sampling of the population. It is a phenomenon observed in genetic algorithms due to the stochastic nature of the selection operator, and is one of the mechanisms by which an initially diverse population can converge to a population of P identical members.

In chapter two, the formalism developed by Prügel-Bennett and Shapiro was used to analytically compare the dynamics of the generational and steady state genetic algorithm. In the weak selection limit, it was seen that the significant difference between the two is a doubling in the rate of genetic drift. Such a calculation is complex and some other technique is sought to calculate genetic drift in general selection schemes.

Analysis of genetic drift is often performed by calculating the Markov chain transition matrices and hence finding the time for the system to reach an absorption state where all population members are identical. This measure is commonly known as the convergence time. Comparisons in the genetic algorithm literature are often performed numerically in this fashion [5, 40]. In population genetics some work has been to done to solve this analytically however the results are approximations and are difficult to generalise to other cases [21, 13, 9].

Chapter two showed that the change in mean fitness at each generation is a function of the population fitness variance. At each generation this variance is reduced by two factors. One factor is selection pressure producing multiple copies of fitter

population members. The other factor is independent of fitness and is due to the stochastic nature of the selection operator — genetic drift. By considering neutral selection, the effect of selection pressure is decoupled and genetic drift seen directly.

This chapter presents a method of calculating the rate of genetic drift in terms of this change in population fitness variance. Unlike calculations in terms of convergence time, this approach lends itself to an exact analytical solution. A general expression for the change in population fitness variance due to genetic drift is derived and applied to a range of selection schemes.

Generational and steady state selection is compared. Using the concept of *generation gap*, G, introduced by De Jong [5, 6] to describe the percentage of the population selected from the initial population at each time step, the rate of genetic drift is calculated between the two extremes.

The formalism is also extended to other non-traditional selection schemes such as that used in Eshelman's CHC algorithm [8]. Schaffer *et al.* [40] recently used a numerical Markov chain comparison to show that a simple model of CHC style selection exhibits half the rate of genetic drift of the traditional genetic algorithm. This is shown to be the case analytically.

The simple model of the CHC algorithm is equivalent to selection schemes in evolution strategies and the approach is generalised for these selection schemes.

3.2 Population Fitness Variance

For an initial population of P discrete members, each with fitness F, the variance, κ_2, of the population fitness distribution is simply

$$\begin{aligned} \kappa_2 &= \langle F^2 \rangle - \langle F \rangle^2 \\ &= \frac{1}{P} \sum_{\alpha=1}^{P} F_\alpha^2 - \left(\frac{1}{P} \sum_{\alpha=1}^{P} F_\alpha \right)^2 . \end{aligned} \tag{3.1}$$

Separating terms that are not independent gives

$$\kappa_2 = \left(\frac{1}{P} - \frac{1}{P^2} \right) \sum_{\alpha=1}^{P} F_\alpha^2 - \frac{1}{P^2} \sum_{\alpha \neq \beta} F_\alpha F_\beta . \tag{3.2}$$

A selection scheme is then applied to this population and a new population of P individuals drawn from it. In this new population there are now n_α copies of

population member α and the variance of the new population fitness distribution is

$$\kappa_2 = \frac{1}{P}\sum_{\alpha=1}^{P} n_\alpha F_\alpha^2 - \left(\frac{1}{P}\sum_{\alpha=1}^{P} n_\alpha F_\alpha\right)^2. \tag{3.3}$$

Again separating terms that are not independent gives

$$\kappa_2 = \sum_{\alpha=1}^{P}\left(\frac{n_\alpha}{P} - \frac{n_\alpha^2}{P^2}\right) F_\alpha^2 - \sum_{\alpha\neq\beta}\frac{n_\alpha n_\beta}{P^2} F_\alpha F_\beta. \tag{3.4}$$

To consider the average case, the average over all ways of performing selection is taken. In the case of neutral selection, n_α is independent of F_α and these terms may be taken outside the summation. The expected population fitness variance is

$$\langle\kappa_2\rangle_s = \left(\frac{\langle n\rangle}{P} - \frac{\langle n^2\rangle}{P^2}\right)\sum_{\alpha=1}^{P} F_\alpha^2 - \frac{\langle n_\alpha n_\beta\rangle}{P^2}\sum_{\alpha\neq\beta} F_\alpha F_\beta. \tag{3.5}$$

As the selection scheme must maintain a constant population size, $\langle n\rangle = 1$. This gives the identity

$$\left(\sum_{\alpha=1}^{P} n_\alpha\right)^2 = P^2 = \sum_{\alpha=1}^{P} n_\alpha^2 + \sum_{\alpha\neq\beta} n_\alpha n_\beta. \tag{3.6}$$

Averaging over all possible selections gives

$$P^2 = P\langle n^2\rangle + P\left(P-1\right)\langle n_\alpha n_\beta\rangle \tag{3.7}$$

and thus

$$\langle n_\alpha n_\beta\rangle = \frac{P - \langle n^2\rangle}{P-1}. \tag{3.8}$$

Substituting this expression into equation (3.5) gives

$$\langle\kappa_2\rangle_s = \frac{P-\langle n^2\rangle}{P-1}\left[\left(\frac{1}{P} - \frac{1}{P^2}\right)\sum_{\alpha=1}^{P} F_\alpha^2 - \frac{1}{P^2}\sum_{\alpha\neq\beta} F_\alpha F_\beta\right]. \tag{3.9}$$

The term within the square brackets is simply the fitness variance of the initial population given in equation (3.2) and thus

$$\langle \kappa_2 \rangle_s = \frac{P - \langle n^2 \rangle}{P - 1} \kappa_2. \tag{3.10}$$

The change in population fitness variance for any selection scheme is simply found by calculating $\langle n^2 \rangle$ — the expected square of the number of times any population member is selected. This is related to the variance in the number of times any member is selected — $\mathsf{V}[n]$. As $\mathsf{V}[n] = \langle n^2 \rangle - \langle n \rangle^2$, equation (3.10) may be rewritten in these terms

$$\langle \kappa_2 \rangle_s = \left(1 - \frac{\mathsf{V}[n]}{P - 1} \right) \kappa_2. \tag{3.11}$$

This expression is the basis for the impending results. It describes the change in population fitness variance due to the stochastic nature of selection — genetic drift — in terms of the variance in the number of times any individual is selected.

3.3 Results

To compare each selection scheme, it is only necessary to calculate $\mathsf{V}[n]$. To allow direct comparison between traditional generational selection, the results are normalised to one generation — steady state selection is performed P times and selection with generation gap G, $1/G$ times. The ratio r is defined as the change in variance after one generation

$$r = \frac{\langle \kappa_2 \rangle_s}{\kappa_2}. \tag{3.12}$$

This gives a very simple picture of the change in genetic drift for differing selection schemes. Whilst the first expression for generational selection is exact, the other expressions are approximations that are accurate to terms in $1/P$.

$$\begin{aligned}
\text{Generational:} \quad & r = 1 - \frac{1}{P} \\
\text{Steady State:} \quad & r \approx 1 - \frac{2}{P} \\
\text{Generation Gap G:} \quad & r \approx 1 - \frac{2 - G}{P} \\
\text{CHC Algorithm:} \quad & r \approx 1 - \frac{1}{2P}
\end{aligned}$$

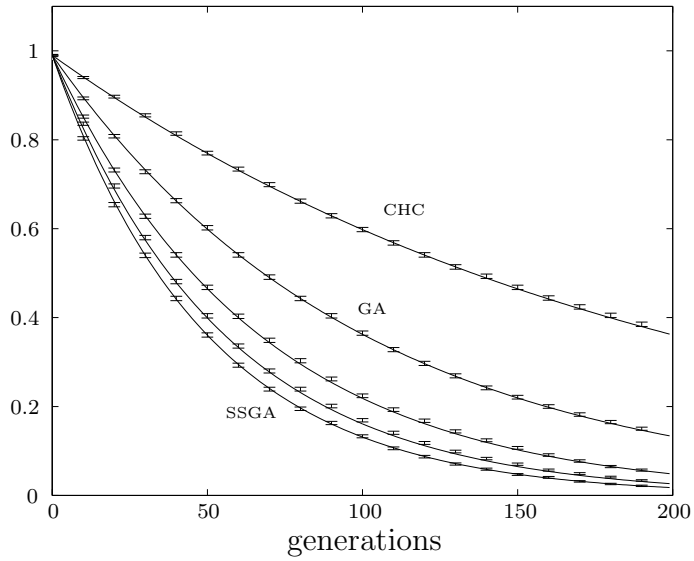


Figure 3.1: Population fitness variance for five different selection schemes. Solid lines are analytical results and error bars are simulation results averaged over 10 000 runs. Curves presented are steady state (*SSGA*), generation gap G=0.2, generation gap G=0.5, generational (*GA*), and a simple model of the CHC algorithm (*CHC*). Population size is 100.

The rate of genetic drift in generational selection is well known as the result of sampling P times with replacement from a finite population.

The rate of genetic drift in steady state selection is twice that of generational selection as was shown in chapter two. Varying the generation gap produces a smooth progression between these two extremes.

The simple model of the CHC algorithm shows half the genetic drift of the generational selection scheme, in agreement with the empirical observations by Schaffer *et al.* [40].

Figure 3.1 shows a comparison of these analytical results with simulation data. A population of 100 was initially drawn from a normal distribution ($K_2 = 1$) and selection repeatedly performed. The plot shows the decreasing population fitness variance for five different selection schemes — steady state selection (SSGA), generation gap $G = 0.2$, generation gap $G = 0.5$, traditional generational selection (GA), and CHC style selection (CHC). Simulation data were averaged over 10 000 runs.

3.4 Performing the Calculations

To calculate $\mathsf{V}[n]$ for each selection scheme is an exercise in probability. Two results from standard probability theory regarding binomial and hypergeometric distributions are used [22].

Selecting from a population with replacement gives rise to a binomial distribution $B(N,p)$ where selection occurs N times with probability of success p. In this case, the variance is the number of times any individual occurs is given by

$$\mathsf{V}[n] = Np(1-p).$$

When selecting without replacement, the result is the hypergeometric distribution $H(M,m,N)$. Here M is the size of the population, N is the number of times selection is applied and m is the number of copies of each individual in the initial population. This gives the result

$$\mathsf{V}[n] = \frac{Nm(M-N)(M-m)}{M^3 - M^2}.$$

In each case $\mathsf{V}[n]$ is calculated and used in equation (3.11) to give the expected change in population fitness variance and thus the rate of genetic drift.

3.4.1 Generational Selection

In a generational selection scheme under random sampling, P members are drawn from a population with replacement. This gives rise to a binomial distribution, $B(P,1/P)$ and thus

$$\mathsf{V}[n] \;=\; 1 - 1/P.$$

From equation (3.11), this gives

$$\langle \kappa_2 \rangle_s = \left(1 - \frac{1}{P}\right)\kappa_2. \tag{3.13}$$

Using the definition of r in equation (3.12) gives

$$r = 1 - \frac{1}{P}. \tag{3.14}$$

3.4.2 Steady State Selection

In the steady state genetic algorithm one member is selected at random, replicated and replaces another random member.

To calculate this, the population is divided into two. One member is drawn with replacement into subpopulation A and then $P-1$ members are drawn without replacement into subpopulation B. These two subpopulations are then combined to form the next population. Subpopulation A uses the binomial distribution $B(1, 1/P)$ and hence

$$\mathsf{V}[n_A] = (P-1)/P^2.$$

Subpopulation B uses a hypergeometric distribution $H(P, 1, P-1)$ and hence

$$\mathsf{V}[n_B] = (P-1)/P^2.$$

Since the two populations are independent, summing gives the final population result

$$\mathsf{V}[n] = \mathsf{V}[n_A] + \mathsf{V}[n_B] = 2(P-1)/P^2.$$

From equation (3.11), this gives

$$\langle \kappa_2 \rangle = \left(1 - \frac{2}{P^2}\right) \kappa_2. \tag{3.15}$$

It is often more convenient to compare P of these selections to one generational selection so using the definition of r as the change after one generation

$$\begin{aligned} r &= \left(1 - \frac{2}{P^2}\right)^P \\ &\approx 1 - \frac{2}{P}. \end{aligned} \tag{3.16}$$

It is clear that the rate of genetic drift is twice that of the generational case.

3.4.3 Varying Generation Gap

To generalise between these two cases the concept of generation gap (G) introduced by De Jong [5] is used. GP members are selected with replacement from the original

population.

Again two subpopulations are considered. For subpopulation A the binomial distribution $B(GP, 1/P)$ is used and hence

$$\mathsf{V}[n_A] = G(1 - 1/P).$$

For subpopulation B the hypergeometric distribution $H(P, 1, P - GP)$ is used and hence

$$\mathsf{V}[n_B] = G - G^2.$$

Summing for the final population gives

$$\mathsf{V}[n] = 2G - G^2 - G/P.$$

From equation (3.11), this gives

$$\langle \kappa_2 \rangle = \left(1 - \frac{2G - G^2 - G/P}{P - 1}\right) \kappa_2. \tag{3.17}$$

To compare this to one generation, the selection operator is applied $1/G$ times. Thus approximating to first-order terms in $1/P$ gives

$$\begin{aligned} r &= \left(1 - \frac{2G - G^2 - G/P}{P - 1}\right)^{\frac{1}{G}} \\ &\approx 1 - \frac{2 - G}{P}. \end{aligned} \tag{3.18}$$

There is a gradual transition between the two rates of genetic drift as generation gap changes.

3.4.4 CHC Algorithm

Eshelman's CHC algorithm uses another non-traditional form of selection whereby crossover is performed amongst the initial population and then selection is performed without replacement from the combined population of parents and offspring.

A simple model of this proposed by Schaffer is to duplicate each member of the population and then draw P members from the population of $2P$ without replacement. In terms of evolution strategies this is $(\mu + \lambda)$ selection with $\lambda = \mu$.

This selection gives rise to a hypergeometric distribution $H(2P, 2, P)$ where selection is performed P times from an initial population of $2P$ which consists of two copies of each individual.

$$\mathsf{V}[n] = (P-1)/(2P-1).$$

From equation (3.11), this gives

$$\langle \kappa_2 \rangle = \left(1 - \frac{1}{2P-1}\right) \kappa_2. \tag{3.19}$$

As we draw P members from the population, this can be compared directly to the generational case simply by making a first-order approximation

$$r \approx 1 - \frac{1}{2P}. \tag{3.20}$$

Thus genetic drift in this model of CHC selection is at half the rate of that of the traditional generational algorithm.

3.5 Evolutionary Strategies

The model of CHC selection considered is similar to many evolutionary strategy selection schemes. The formalism presented can easily be extended to these strategies. In general these selection schemes are described as $(\mu + \lambda)$ strategies. From an initial population of size μ, λ offsprings are produced and then selection acts on both the parents and the offsprings to produce the next population of size μ.

Consider a $(\mu + \lambda)$ evolution strategy where $\mu = P$ and $\lambda = sP$ where s is some fraction, $0 \leq s \leq 1$. Selection occurs from two subpopulations, one consisting of $P(1-s)$ individuals and the other of size $2sP$ containing sP pairs. If n_1 is the number of individuals and n_2 the number of pairs in the final population, the variance in the number of times any population member is selected can be shown to be simply

$$\mathsf{V}[n] = \frac{2n_2}{P} \tag{3.21}$$

as $P\langle n \rangle = n_1 + 2n_2$, $P\langle n^2 \rangle = n_1 + 4n_2$ and $\langle n \rangle = 1$. The number of pairs in the final population is simply found by considering the number of pairs produced when X individuals are drawn without replacement from the subpopulation of pairs whose

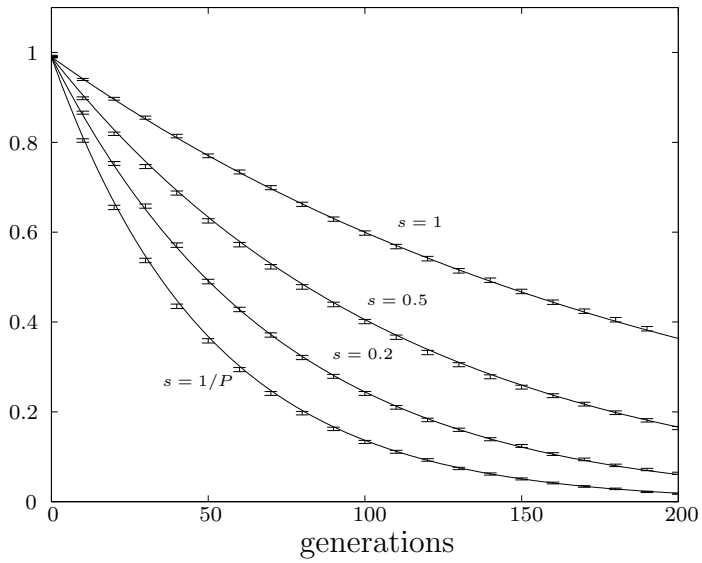


Figure 3.2: Population fitness variance for $(\mu + s\mu)$ selection for varying s. Solid lines are analytical results and error bars are simulation results averaged over 10,000 runs. Population size is 100.

size is $2sP$ and is given by

$$n_2 = \frac{X}{2}\frac{X-1}{2sP-1}. \tag{3.22}$$

Substituting equation (3.22) into equation (3.21) and averaging over X gives

$$\mathsf{V}[n] = \frac{\langle X^2\rangle - \langle X\rangle}{P(2sP-1)}. \tag{3.23}$$

The expectations of X — $\langle X^2\rangle$ and $\langle X\rangle$ — are described by a hypergeometric distribution given by $H(P(1+s), 2sP, P)$, as P individuals are drawn without replacement from a population of $P(1+s)$. Using the standard results for the hypergeometric distribution given earlier and substituting these results into equation (3.23), gives the result

$$\mathsf{V}[n] = \frac{2s(P-1)}{(1+s)[P(1+s)-1]}. \tag{3.24}$$

As before, substituting $\mathsf{V}[n]$ directly into equation (3.11) and normalising the expression by applying selection $1/s$ times, gives the final rate of genetic drift

$$\begin{aligned} r &= \left(1 - \frac{2s}{(1+s)\left[P(1+s)-1\right]}\right)^{1/s} \\ &\approx 1 - \frac{2}{(1+s)^2 P}. \end{aligned} \tag{3.25}$$

The rate of genetic drift covers the same range as that seen for the genetic algorithm selection schemes. Figure 3.2 shows a plot of these analytical result against simulation data. Four different values of s are considered and the population size is again 100.

3.6 Discussion

Analysing genetic drift in terms of the change in population fitness variance allows exact analytical expressions to be derived for any selection scheme. From these expressions comparisons of the effect that genetic drift has on the convergence of a genetic algorithm under varying generation gap can by made.

Figure 3.3 shows the population fitness mean and variance for steady state, generational, and varying generation gap ($G = 0.2$ and 0.5) implementations of an actual genetic algorithm on the one-max problem where the fitness is proportional to the number of ones in a binary string of 96 bits. Probabilistic tournament selection is used where two individuals are drawn from the population and the fitter of the two selected with probability s. In this case $s = 0.1$. All use a population size of 50 and the rate of mutation at each bit is 1/96. Finally uniform crossover is perform whereby the bits of the offspring are drawn at random from two parents. CHC is not included in the comparison as the other features of the algorithm lead to more significant differences than genetic drift alone.

Selection pressure is the same in each case as evidenced by the identical initial gradients of the mean fitness curves. As variance decreases through selection, the change in mean fitness decreases. For the steady state genetic algorithm, variance decreases fastest due to the higher rate of genetic drift and thus the mean fitness evolves to a lower final value.

These results illustrate how genetic drift can influence the convergence of a genetic algorithm. Definitive statements about the performance of different selection

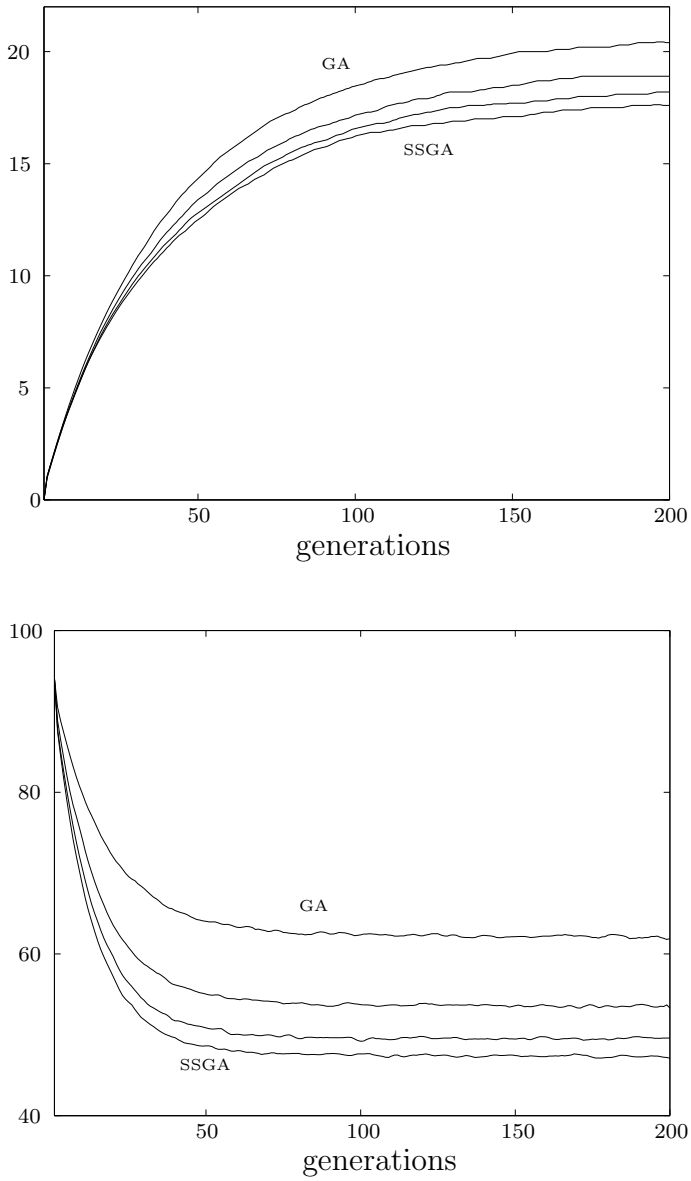


Figure 3.3: Population mean fitness (*upper plot*) and variance (*lower plot*) for four different selection schemes. Simulation results are averaged over 10 000 runs and the error bars are the thickness of the lines. Curves presented are (in order) steady state (*SSGA*), generation gap G=0.2, generation gap G=0.5, and generational (*GA*).

schemes are difficult to make. However it is clear that genetic drift is another factor, alongside more commonly understood factors such as selection pressure, which affects the convergence of the genetic algorithm and can be controlled by the choice of selection scheme.

Chapter 4

Ranking Selection

4.1 Introduction

The original formalism of Prügel-Bennett and Shapiro considered Boltzmann selection. This has a number of features which make it attractive to the formalism, namely the easily parameterised selection strength and the exponential relationship enabling weak selection expansions to be derived.

It also has some significant disadvantages; the most commonly raised one being that it is not generally used in the genetic algorithm community. Of more significance is the weighting which is applied to the extremes of the population through the exponential relationship. These extremes are ill defined under a cumulant expansion and thus a large number of cumulants are required to achieve quantitatively good results. The large number of macroscopic variables required makes qualitative understanding difficult.

The extremes of the distribution are also those areas where the difference between a finite and an infinite population are most pronounced. Under a finite population, these areas are sparsely populated. This leads to finite population effects being critical to the correct prediction of the dynamics of the genetic algorithm. A simplifying infinite population model is of no use as it behaves qualitatively differently.

One of the most common forms of selection in the genetic algorithm community is ranking selection or binary tournament selection. These are commonly observed to give similar results and in fact can be shown to be mathematically equivalent. For selection schemes where the weighting of each individual is a simple function of its fitness the effect of selection may be calculated exactly as previously done in

chapter two for Boltzmann selection. However in ranking selection, the additional relationship to the fitness of other members of the population makes this impossible and a new method of calculating the effect of selection is introduced.

In the analysis so far, roulette wheel selection has been considered. That is, population members are drawn with replacement from the population with a probability based on their weighting. This is equivalent to spinning a roulette wheel with P unequal size bins, P times.

Whilst on average the population members are drawn with probabilities given by their weighting, the process is stochastic and there is some variance in this number. An alternative scheme suggested by Baker [2] and commonly referred to as Baker selection or stochastic universal sampling, is used to address this issue. Instead of spinning a single ball P times, a P armed pointer is spun once.

Whilst it is commonly held that Baker selection is superior, there are no theoretical or empirical comparisons beyond Baker's original work. Under the formalism presented, the difference between these schemes can be compared.

4.2 Ranking Selection

In any selection scheme dependent on the absolute fitness value of the population members, there is a risk that an extremely fit individual will monopolise the population. Ranking selection was suggested by Baker [1] as a means of minimising this chance and has become a standard form of selection in the genetic algorithm literature.

Rather than using the absolute values, the population is ranked in order of fitness. The expected number of times that the population member of rank i will be represented in the next generation is controlled by the parameter MAX and is given by

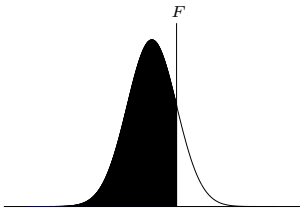
$$n_i = \mathrm{MAX} - 2\,(\mathrm{MAX} - 1)\,\frac{i-1}{P-1}. \tag{4.1}$$

The fittest population member is expected to be represented MAX times and the least fit $(2 - \mathrm{MAX})$ times. MAX may take any value between one and two.

Calculating the effect of selection is first done by considering an infinite population.

4.2.1 Infinite Population Model

In the infinite population limit, the ranking of any individual is proportional to its position within the population.



Thus the expected number of occurrences for an individual of fitness F is given by

$$n_F = (2 - \mathrm{MAX}) + 2\,(\mathrm{MAX} - 1) \int_{-\infty}^{F} \rho\,(F')\,\mathrm{d}F' \tag{4.2}$$

where $\rho\,(F')$ describes the continuous fitness distribution.

The first and second moments of the population distribution after selection are found by integrating the weighting over the distribution

$$\begin{aligned} \langle F \rangle &= \int_{-\infty}^{\infty} F n_F \rho\,(F)\,\mathrm{d}F \\ &= K_1 + (\mathrm{MAX} - 1)\sqrt{\frac{K_2}{\pi}} \\ \langle F^2 \rangle &= \int_{-\infty}^{\infty} F^2 n_F \rho\,(F)\,\mathrm{d}F \\ &= K_2 + K_1^2 + 2\,(\mathrm{MAX} - 1)\,K_1\sqrt{\frac{K_2}{\pi}}. \end{aligned} \tag{4.3}$$

Thus the first two cumulants after selection are given by

$$\begin{aligned} \langle K_1 \rangle_s &= K_1 + (\mathrm{MAX} - 1)\sqrt{\frac{K_2}{\pi}} \\ \langle K_2 \rangle_s &= \left[1 - \frac{(\mathrm{MAX} - 1)^2}{\pi}\right] K_2 \end{aligned} \tag{4.4}$$

where $\langle \ldots \rangle_s$ represents the average overall ways of performing selection.

Unlike Boltzmann selection, the variance is decreased by a factor determined by the selection pressure, even in the infinite population limit. The mean increases by a factor dependent on $\sqrt{K_2}$ — a measure of the width of the distribution.

4.2.2 Tournament Selection

Under binary tournament selection, two individuals are drawn independently from the population, compared and the fitter of the two is selected. The probability that one member of fitness F is fitter than another drawn from the population is given by

$$P_{\text{fitter}}(F) = \int_{-\infty}^{F} \rho(F')\,\mathrm{d}F'. \tag{4.5}$$

When integrated over the population distribution, the result is identical to that of ranking selection when MAX = 2. Indeed the two strategies are equivalent. Changing the parameter MAX is equivalent to introducing a probabilistic element into tournament selection.

The infinite population analysis of tournament selection for the binary case and larger tournament sizes has been performed by Blickle and Thiele [3].

4.2.3 Finite Population Effects

For Boltzmann selection, the calculation of finite population effects was integral to the formalism. Integrating ranking selection into this method has proved to be too difficult due to the non-linear relationship between the weighting of an individual and the weighting of the other members of the population.

Instead of performing an exact calculation, an approximation is developed which captures the underlying principle without the extraneous complications. This approximation however, is uncontrolled and the justification can only be presented as an *a posteriori* comparison of simulation and theory predictions.

Using the approximation presented here as the starting point, Prügel-Bennett has subsequently calculated the exact finite population effects. Whilst it has proved possible to do so, the calculation is complex and the resulting expressions of little value in developing a qualitative understanding.

The first cumulant is unaffected by finite populations. However, the second cumulant, the variance will exhibit an additional loss due to the stochastic nature of the selection scheme operating on a finite population. Selection is considered to be a two part process.

- The change in the cumulants of the ensemble distribution are calculated by considering an infinite population. The results of this have been calculated in section 4.2.1.

- The additional effect of a finite population is modelled by calculating the loss in variance when a finite population is sampled with the ranking assigned independently of fitness. This has been calculated in chapter two for various selection schemes and must be extended to ranking selection.

These two factors are exact when considered independently — an infinite population subject to selection or a finite population subject to neutral sampling. To combine these two terms and maintain this independence, the final result is assumed to be the product of these two factors. There is some justification for this. Examination of the weak selection expansion of Boltzmann selection shows that to first order terms, the finite population effects are a multiplicative factor.

4.2.4 Roulette Wheel and Stochastic Universal Sampling

The first form of selection proposed for the genetic algorithm was roulette wheel selection where each population member is simply drawn with replacement from the population. Baker [2] noted that whilst any individual with rank i is expected to occur n_i times after selection, the stochastic nature of roulette wheel selection allows anywhere between 0 and P copies to be selected. This is the source of convergence of a finite population due to stochastic effects — genetic drift.

Baker proposed stochastic universal sampling (SUS) as a selection scheme which limits the range of possible occurrences to either $\lfloor n_i \rfloor$ (n_i rounded down to the nearest integer) or $\lceil n_i \rceil$ (n_i rounded up to the nearest integer). Whilst no arguments were made as to the virtue of doing this in the original paper, it is generally understood that the use of stochastic universal sampling reduces the effects of genetic drift. Intuitively it can be seen that limiting the range of possible occurrences will reduce $\langle n^2 \rangle$ and hence reduce the loss in variance through stochastic effects.

Calculating $\langle n^2 \rangle$ for both selection schemes allows them to be compared.

Roulette Wheel Selection

Selecting using roulette wheel selection gives rise to a binomial distribution in which m trials are made with a probability of success p. The standard result for a binomial distribution is

$$\langle n^2 \rangle = m\,(m-1)\,p^2 + mp. \tag{4.6}$$

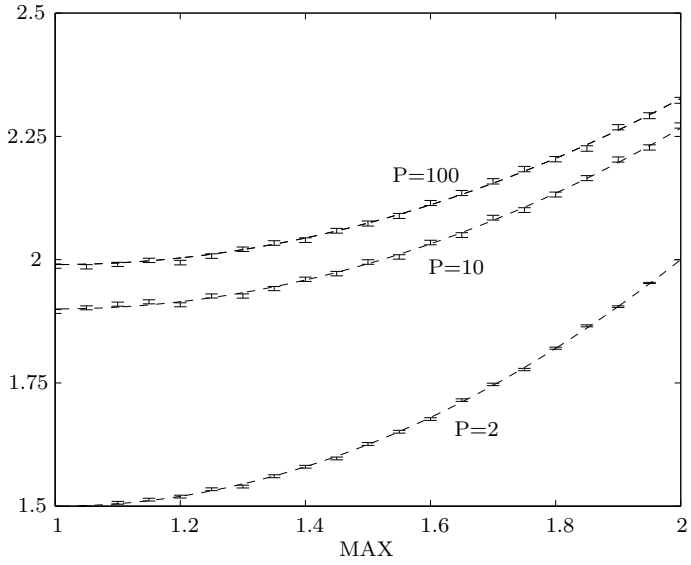


Figure 4.1: Comparison of simulation and theory calculation of $\langle n^2 \rangle$ for roulette wheel selection. Three population sizes are shown and simulation results are averaged over 10 000 runs.

Assuming independence between each population member, $\langle n^2 \rangle$ can be found by averaging over the weighting for each rank. If the probability of success is n_i/P and P trials are made, the resulting expression is

$$\langle n^2 \rangle = \frac{1}{P} \sum_{i=1}^{P} \left[P\left(P-1\right) \frac{n_i^2}{P^2} + P \frac{n_i}{P} \right]. \tag{4.7}$$

Using the expression for n_i given in equation (4.1) and performing the summation gives

$$\langle n^2 \rangle = 3 + \frac{(P+1)\left(\mathrm{MAX}^2 - 2\mathrm{MAX} - 2\right)}{3P}. \tag{4.8}$$

Fig 4.1 shows a comparison of simulation results and this theory prediction.

Stochastic Universal Sampling

In stochastic universal sampling, for any value of n_i either $\lfloor n_i \rfloor$ or $\lceil n_i \rceil$ copies exist after selection. The probabilities of either is given by

$$n_i = \begin{cases} \lceil n_i \rceil & \text{with probability } n_i - \lfloor n_i \rfloor \\ \lfloor n_i \rfloor & \text{with probability } \lceil n_i \rceil - n_i. \end{cases} \tag{4.9}$$

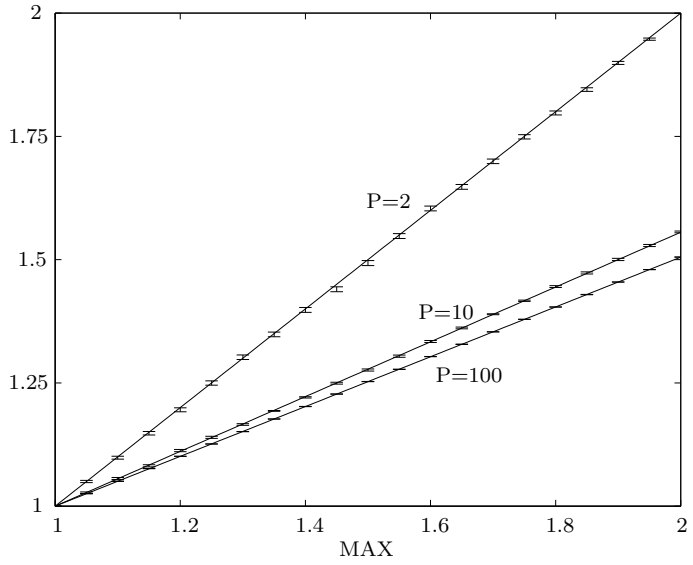


Figure 4.2: Comparison of simulation and theory calculation of $\langle n^2 \rangle$ for stochastic universal sampling. Three population sizes are shown and simulation results are averaged over 10 000 runs.

Assuming independence between individuals $\langle n^2 \rangle$ is found by calculating the distribution averaged over each ranking

$$\langle n^2 \rangle = \frac{1}{P} \sum_{i=1}^{P} \lceil n_i \rceil^2 \left(n_i - \lfloor n_i \rfloor \right) + \lfloor n_i \rfloor^2 \left(\lceil n_i \rceil - n_i \right). \tag{4.10}$$

Calculating this for selection schemes such as Boltzmann or proportional selection is far from trivial as the exact population fitness distribution is required. However for ranking selection n_i and thus $\lfloor n_i \rfloor$ and $\lceil n_i \rceil$ are known independently of the population structure. For $i \leq P/2$, $\lfloor n_i \rfloor = 1$ and $\lceil n_i \rceil = 2$ whilst when $i > P/2$, $\lfloor n_i \rfloor = 0$ and $\lceil n_i \rceil = 1$. Applying these ranges gives

$$\langle n^2 \rangle = \frac{1}{P} \sum_{i=1}^{P/2} 4 \left(n_i - 1 \right) + \left(2 - n_i \right) + \frac{1}{P} \sum_{i=\frac{P}{2}+1}^{P} n_i. \tag{4.11}$$

Again using the expression for n_i given in equation (4.1) and performing the summation gives

$$\langle n^2 \rangle = \mathrm{MAX} - \frac{(\mathrm{MAX} - 1)}{2} \frac{P - 2}{P - 1}. \tag{4.12}$$

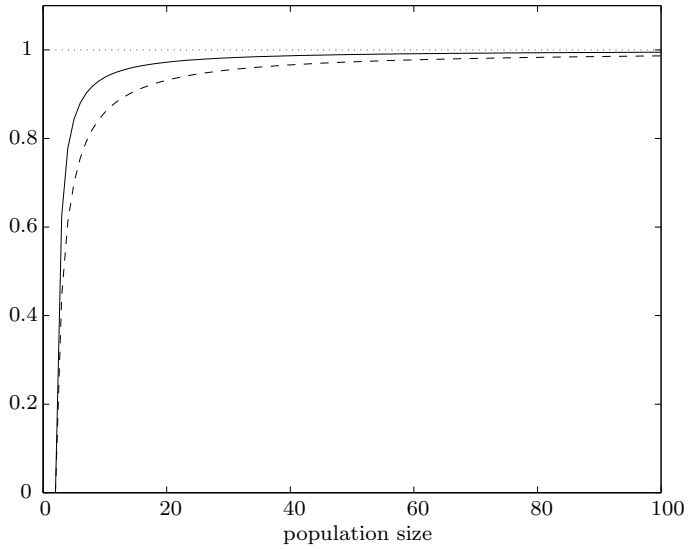


Figure 4.3: Comparison of $(P - \langle n^2 \rangle)/(P - 1)$ for roulette wheel selection (*dashed line*) and stochastic universal sampling (*solid line*).

Fig 4.2 shows a comparison of simulation results and this theory prediction.

4.3 Results

For any particular value of MAX, $\langle n^2 \rangle$ for stochastic universal sampling is less than that of roulette wheel selection, showing that the variance in the number of times any population member is selected is less. This result is expected intuitively.

The smaller value of $\langle n^2 \rangle$ results in less loss of variance at each selection scheme. In chapters two and three this has been shown to have a direct effect on the evolution of the genetic algorithm. This analysis confirms the generally held beliefs about stochastic universal sampling and roulette wheel selection.

Figure 4.3 shows the theoretical calculations of r for both roulette wheel selection and stochastic universal sampling. Both tend to unity reasonably quickly with increasing population size. For reasonable size populations, a finite population behaves both qualitatively and quantitatively similarly to an infinite population.

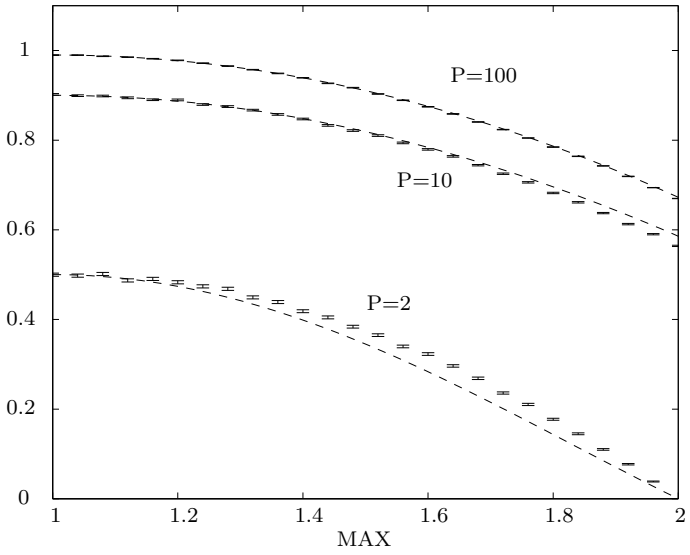


Figure 4.4: Comparison of theory (*dashed line*) and simulation for ensemble fitness variance after selection using roulette wheel selection. Three population sizes are shown drawn from a unit Gaussian and simulation results are averaged over 10 000 runs.

Under the approximation that the final result is assumed to be multiplicative combination of both factors , the final results is given by

$$\begin{aligned}\langle K_1\rangle_s &= K_1 + (\mathrm{MAX}-1)\sqrt{\frac{K_2}{\pi}} \\ \langle K_2\rangle_s &= \frac{P-\langle n^2\rangle}{P-1}\left[1-\frac{(\mathrm{MAX}-1)^2}{\pi}\right]K_2.\end{aligned} \tag{4.13}$$

Whilst the calculation of finite population effects is exact under neutral sampling, the assumption that the two factors can be applied multiplicatively has not been justified and the approximation is uncontrolled. It must be shown *a posteriori* to be a good approximation. Figures 4.4 and 4.5 show the total reduction in ensemble fitness variance after selection. The initial population is drawn from a unit Gaussian and selection applied once.

4.4 Discussion

The theoretical results are in good agreement with the simulation results even at extremes of small population sizes where finite population effects are dominant. There is clearly however some systematic error in the approximation. In comparison

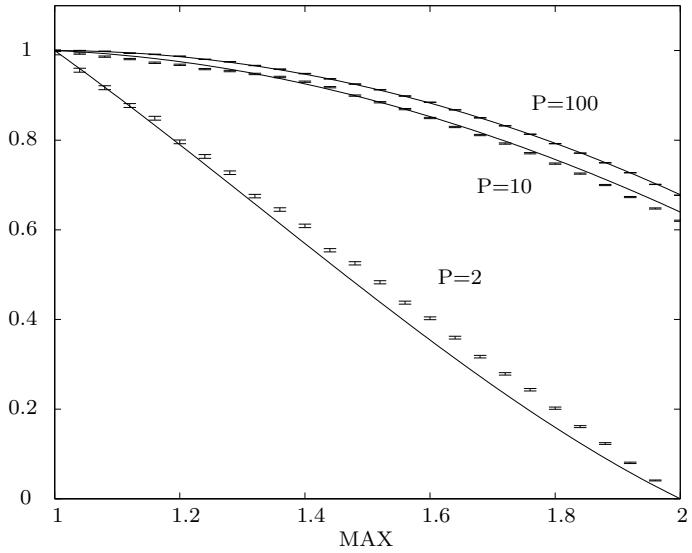


Figure 4.5: Comparison of theory (*solid line*) and simulation for ensemble fitness variance after selection using stochastic universal sampling. Three population sizes are shown drawn from a unit Gaussian and simulation results are averaged over 10 000 runs.

with an exact approach, the ease with which the results may be derived and used in future analysis justifies the small quantitative error.

The behaviour of the genetic algorithm under strong ranking or tournament selection is qualitatively different from the weak Boltzmann selection case. Finite population effects are less significant except when populations are small. However in these cases, they may be closely approximated in an algebraically simple expressions which allows both roulette wheel selection and stochastic universal sampling to be modelled.

Chapter 5

Crossover and the Onemax Problem

5.1 Introduction

The analysis of the genetic algorithm in the preceding chapters has focused on the selection operator. Whilst the results of this analysis has been independent of the actual problem space, the analysis can go no further until the other genetic operators — mutation and crossover — are included.

In this chapter the analysis is extended to a simple model of the full genetic algorithm including all three genetic operators.

5.2 Onemax

To model more accurately a genetic algorithm, the problem space must also be modelled and the interaction of the genetic operators – mutation and crossover — included. The simplest problem commonly used in the genetic algorithm literature is the onemax or ones-counting problem. Here the fitness of the individual is simply the number of ones in the binary string.

For a number of reasons there are objections to this simple model. The interactions of the bits have no spatial bearing and there are no interactions between them — no epistasis. Thus the problem has no local minima and is trivial to solve for any algorithm. Indeed, it is often said that studying the onemax problem tells us very little about the action of real genetic algorithms on real problem spaces.

These objections are valid but do no negate the value in starting any analysis here. A theory which can not predict the behaviour of the genetic algorithm in this simple case will probably be of little use elsewhere. Onemax and simple extensions of it, also show very complex and surprising behaviour which challenges intuitions about the dynamics of the genetic algorithm.

5.2.1 The Model Genetic Algorithm

The model genetic algorithm consist of a population of P individuals. Borrowing the terminology of solid state physics, each individual consists of a string of L spins whose value may be 1 or -1. The magnetization of individuals and in the case of onemax, the fitness, is given by the sum of its spins

$$M = \sum_{i=1}^{L} S_i \quad \text{where} \quad S_i = \{-1, 1\}. \tag{5.1}$$

The population is initialized with random spins and generational ranking selection used to select the new generation from the initial population. The mutation operator is then applied whereby each spin has a small probability of mutation

$$S_i \rightarrow -S_i \quad \text{with probability } \gamma. \tag{5.2}$$

Crossover is then applied to the population. Population members are randomly paired and uniform crossover [45] applied whereby spins are randomly drawn from each parent α and β

$$S_i = \chi_i S_i^{\alpha} + (1 - \chi_i)\, S_i^{\beta} \tag{5.3}$$

where

$$\chi_i = \begin{cases} 1 & \text{with probability } 1/2 \\ 0 & \text{with probability } 1/2. \end{cases} \tag{5.4}$$

The complementary offspring is also created by using $\bar{\chi}_i$ and thus the states of all the spins in the population are conserved. That completes one generation and the process is repeated.

5.2.2 Selection

The effect of selection on the first two cumulants has been calculated for a finite population in the last chapter. The analysis is used here directly. If the initial

population is created with randomly assigned spins, the initial ensemble distribution has mean of zero and variance of L.

5.2.3 Mutation

The analysis of the effect of mutation on the cumulants of the ensemble fitness distribution has previously been performed by Prügel-Bennett and Shapiro [29, 30] and the derivation is included in appendix A. The first and second cumulants after mutation are

$$\begin{aligned} \langle K_1 \rangle_m &= \Gamma K_1 \\ \langle K_2 \rangle_m &= \Gamma^2 K_2 + L\left(1 - \Gamma^2\right) \quad \text{where} \quad \Gamma = 1 - 2\gamma \end{aligned} \tag{5.5}$$

where $\langle \ldots \rangle_m$ represents the average over all ways of performing mutation.

The effect of mutation can be clearly seen in these expressions. It acts to push the ensemble distribution back to the maximum entropy state, decreasing the mean and increasing the variance. In this way it acts against selection which is reducing the variance and increasing the mean fitness.

5.2.4 Crossover

Like mutation, the analysis of uniform crossover has been performed previously by Prügel-Bennett and Shapiro [29, 30]. Details of the derivation are included in appendix B. The effects of crossover on the first two cumulants are

$$\begin{aligned} \langle K_1 \rangle_x &= K_1 \\ \langle K_2 \rangle_x &= \frac{K_2}{2} + \frac{L}{2}\left(1 - q\right) \end{aligned} \tag{5.6}$$

where q is defined as

$$q = \frac{1}{P\left(P-1\right)} \sum_{\alpha \neq \beta} \frac{1}{L} \sum_{i=1}^{L} S_i^\alpha S_i^\beta \tag{5.7}$$

and is called the correlation of the population. It describes the similarity of strings in the population. In the initial random population, the correlation is zero. In a population consisting a P identical strings, correlation is equal to one. Due to the conservation of spins under crossover, the correlation is not changed by crossover.

The first cumulant does not change under crossover. This is expected as crossover conserves the states of all the spins. Crossover does change the second cumulant however and increases the variance towards the natural variance of the population which is defined by the value of q.

The analysis of higher cumulants show that they are greatly reduced by uniform crossover. This ensures that the ensemble fitness distribution stays close to a Gaussian. In this analysis, higher order cumulants are assumed to be small and the dynamics of the genetic algorithm are predicted using just the first and second cumulants.

To fully understand the effect of crossover, how the correlation evolves under selection and mutation must also be modelled. As S_i^α and S_i^β are not independent, taking the average of q directly is not trivial. If the spins are distributed within the population with maximum entropy, the correlation will be related to the first cumulant by

$$q = \frac{K_1^2}{L^2} \tag{5.8}$$

simply by considering all spins to be independent and finding $\langle S_i \rangle^2$.

This natural correlation under estimates the correlation in small populations; selection acts by duplicating individuals and these duplicates significantly affect the correlation of the population. This effect is most significant when the mutation rate is small.

In order to model this deviation away from the maximum entropy case, a similar approach is taken to that developed by Rattray [31]. A measure of the deviation, $C_{\alpha\beta}$, away from the natural correlation of the population , $\tilde{q}$, is defined

$$q_{\alpha\beta} = C_{\alpha\beta} + (1 - C_{\alpha\beta})\,\tilde{q} \tag{5.9}$$

and averaged over the population

$$q = C + (1 - C)\,\tilde{q}. \tag{5.10}$$

As such, C represents a linear deviation away from the natural correlation of the population when the spins are arranged with maximum entropy, towards a completely correlated population. Figure 5.1 shows this schematically. Previously C was interpreted as the founder effect; a measure of the degree of duplication within

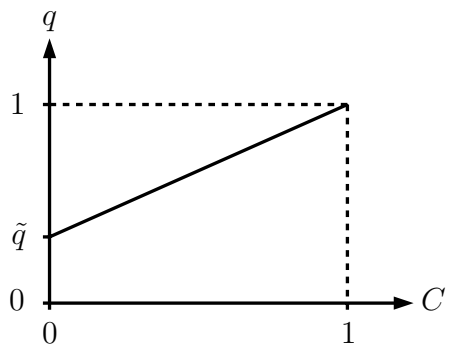


Figure 5.1: Diagram of deviation from natural correlation of a population under changes in C.

the population. This interpretation has however been shown to be incorrect and is discussed in section 5.4.

The deviation from natural correlation is used as another macroscopic variable and as such the effect of selection and mutation on C must be calculated.

Selection

The effect of selection on C may be calculated by summing over the new population where there are now n_α copies of individual α. After some algebra detailed in appendix C, the result is

$$1 - \langle q \rangle_s = \frac{P - \langle n^2 \rangle}{P - 1} (1 - \tilde{q}) (1 - C) - \frac{1 - \tilde{q}}{P(P-1)} \sum_{\alpha \neq \beta} (C_{\alpha\beta} - C) n_\alpha n_\beta. \qquad (5.11)$$

This expression has two components. The first expresses the change in correlation due to the stochastic nature of the selection scheme in a finite population and the creation of duplicates within the population. The second term determines the change in population correlation through the dependency of correlation and fitness. Initially, the first terms is of interest and the second is assumed to average to zero. This is found to be true over the expected range of population sizes and in cases where mutation is sufficient to stop the population correlating. However, in large populations with little mutation, the second term dominates and gives rise to an interesting phase transition discussed in chapter nine.

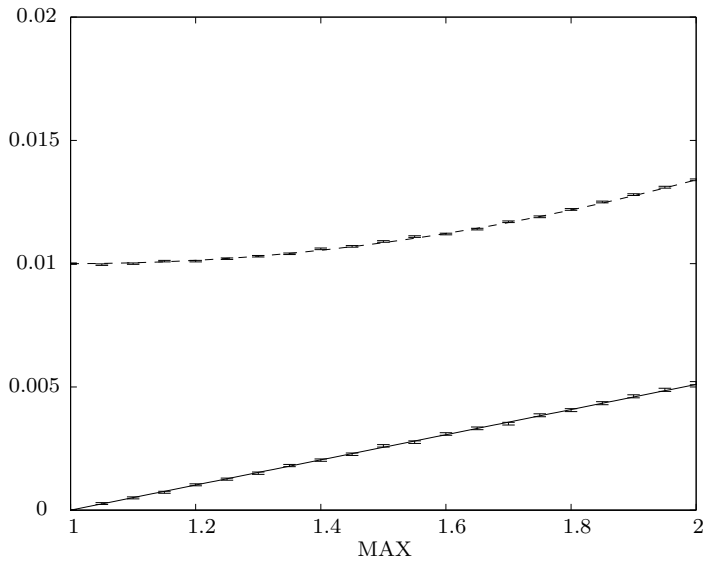


Figure 5.2: Comparison of theoretical and simulation results for C when roulette wheel selection (*dashed line*) and stochastic universal sampling (*solid line*) are used with ranking selection. Population size is 100.

Thus, ignoring the right hand expression in equation (5.11) and using the definition of C given in (5.10) gives the final result for C after selection

$$1 - \langle C \rangle_s = \frac{P - \langle n^2 \rangle}{P - 1} (1 - C) \tag{5.12}$$

where $\langle n^2 \rangle$ is the variance of the selection scheme as calculated in chapter four. Indeed the factor here is exactly the same as the factor describing the loss in population variance after selection due to finite population effects.

Figure 5.2 shows simulation and theoretical results for C after one selection step. An initial population of 100 individuals were created with a random strings. Selection was applied using both stochastic universal sampling and roulette wheel selection and C calculated. Results were averaged over 1 000 runs. The difference between the selection schemes is clear. Roulette wheel selection shows a large increase in correlation even at very low selection strengths due to its propensity to produce more duplicates than stochastic universal sampling.

Over time this correlation builds up within the population, decreasing the natural variance and thus reducing the effectiveness of crossover in restoring the population variance lost through selection. The genetic algorithm using roulette wheel selection exhibits a smaller final population variance than one using stochastic universal

sampling and thus will evolve to a lower final mean fitness.

Mutation

Mutation will affect the correlation of the population by introducing variation. The expected value of any spin after mutation is simply

$$\langle S_i \rangle_m = \Gamma S_i. \tag{5.13}$$

The correlation after selection is thus simply

$$\langle q \rangle_m = \Gamma^2 q. \tag{5.14}$$

Using the definition of q in equation (5.10) and the change in the first cumulant due to mutation, gives C after mutation as

$$\langle C \rangle_m = \frac{\Gamma^2 \left(K_1^2 - L^2\right)}{\Gamma^2 K_1^2 - L^2} C. \tag{5.15}$$

5.3 Results

Using just the first two ensemble cumulants and C, the full dynamics of a genetic algorithm using ranking selection with either roulette wheel selection or stochastic universal sampling can be modelled. Figure 5.3 shows the theory predictions compared to simulation results for the first two cumulants and the correlation. The simulation data is averaged over 1 000 runs and uses the parameters $L = 96$, $\gamma = 1/96$, $P = 100$ and MAX = 1.1. The figures show very good agreement between theory and simulations. Stochastic universal sampling shows less correlation and less loss in population variance and thus evolve to a higher final mean fitness.

The use of just two cumulants makes a Gaussian approximation to the fitness distribution. This approximation has been shown to give good quantitative results and captures the full dynamics of the genetic algorithm. Whilst mutation and selection increase the higher order cumulants, these are suppressed by crossover thus improving the accuracy of the results.

5.4 The Original Interpretation of C

The interpretation of C as the deviation of the correlation from its natural value was not the original one. The original interpretation was that C represents the founder

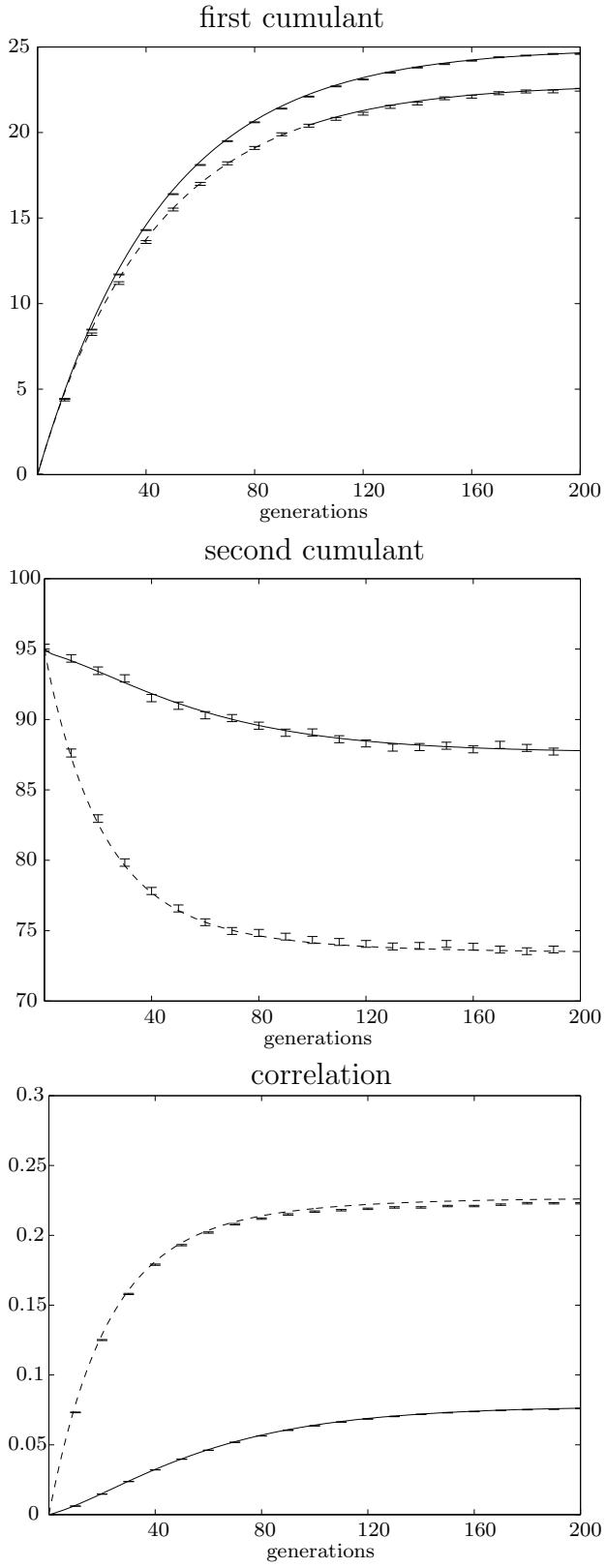


Figure 5.3: Comparison of theoretical and simulation results for the first two cumulants and correlation for a genetic algorithm with ranking selection, mutation and crossover. Roulette wheel selection (*dashed line*) and stochastic universal sampling (*solid line*) are shown. Parameters used were $L = 96$, $\gamma = 1/96$, $P = 100$ and MAX $= 1.1$.

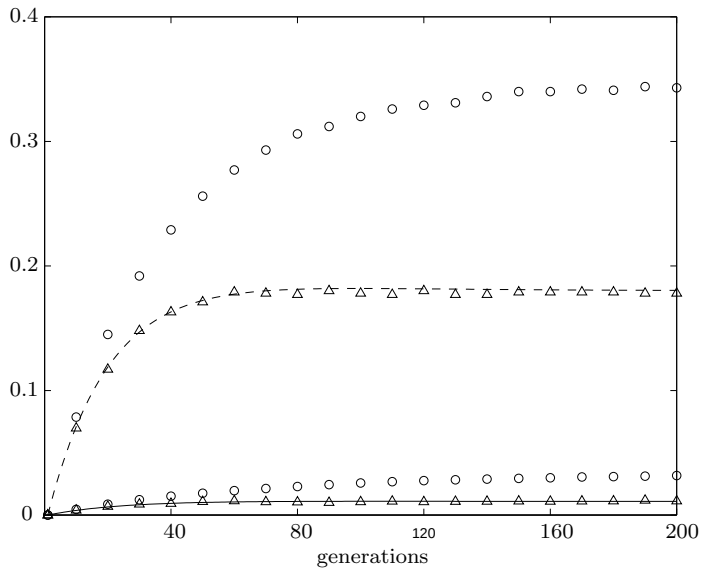


Figure 5.4: Comparison of theoretical and simulation results for the C within the population. Simulation measurements of the founder effect (*circles*) and deviation from natural correlation (*triangles*) are shown against theoretical results for roulette wheel selection (*dashed line*) and stochastic universal sampling (*solid line*). Parameters are as figure 5.3.

effect; that is, C indicates whether spins at the same site in two randomly drawn population members, originate from the same ancestor in the initial population

$$C = \frac{1}{P\,(P-1)} \sum_{\alpha} \sum_{\alpha \neq \beta} \frac{1}{L} \sum_{i} \left[S_i^{\alpha} \sim S_i^{\beta} \right] \tag{5.16}$$

where

$$\left[S_i^{\alpha} \sim S_i^{\beta} \right] = \begin{cases} 1 & \text{if both spins come from the same ancestor} \\ 0 & \text{otherwise.} \end{cases} \tag{5.17}$$

If the spins originate from the same ancestor, they will be identical and contribute $+1$ to the correlation. If they originate from different ancestors, their contribution to the correlation can be calculated from the natural correlation of the population

$$q = C + (1 - C)\ \tilde{q}. \tag{5.18}$$

The value of C after selection will be related to the number of duplicates introduced by selection and the probability that they are drawn from the population together.

If n_α is the number of copies of population member α after selection, C' is given by

$$C' = \sum_{\alpha=1}^{P} \left[\frac{n_\alpha}{P} \frac{n_\alpha - 1}{P-1} + C \frac{n_\alpha}{P} \left(1 - \frac{n_\alpha - 1}{P-1} \right) \right] \tag{5.19}$$

where the assumption has been explicitly made that the correlations are independent of the individuals fitness. Averaging over all n_α and using the fact that population size is constant and thus $\langle n \rangle = 1$, gives

$$1 - \langle C \rangle_s = \frac{P - \langle n^2 \rangle}{P-1} (1 - C) \tag{5.20}$$

where $\langle n^2 \rangle$ is the variance of the selection scheme as calculated in chapter four.

This factor is identical to that obtained previously. The calculation of the effect of mutation differs between the two interpretations and thus whilst the selection term is the same, when the two are measured in simulation runs the difference is clear. Figure 5.4 shows theoretical results of C, against simulation results of the founder effect and the deviation from natural correlation.

5.5 Linkage Equilibrium and a Closed Form Approximation

As the dynamics of the genetic algorithm can be described by a small set of coupled equations, they may be solved as a set of simultaneous equations to find the final equilibrium values of the cumulants. This was not possible for previous models using Boltzmann selection as the higher cumulants were significant in the final equilibrium point.

Solving the equations numerically by iteration is trivial under the assumption that after selection, mutation and crossover the cumulants are unchanged. Figure 5.5 shows the final equilibrium values of K_1 and K_2, denoted as K_1^* and K_2^*, for a genetic algorithm using roulette wheel selection and stochastic universal sampling. Theoretical predictions are plotted as a continuous line and simulation results at discrete values of selection strength. The parameters used are as before.

A simplified analytical expression may be derived by making use of an observation which applies when averaging over the ensemble in the onemax problem

$$K_2 \approx L\,(1 - q)\,. \tag{5.21}$$

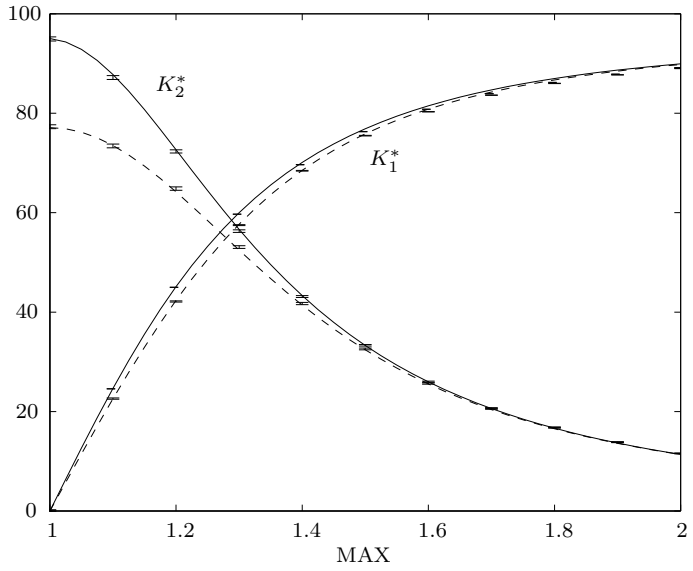


Figure 5.5: Comparison of numerical theoretical and simulation results for the end point distribution of the genetic algorithm. Roulette wheel selection (*dashed line*) and stochastic universal sampling (*solid line*) are shown. Parameters used were $L = 96$, $\gamma = 1/96$ and $P = 100$.

This identity goes by the name of linkage equilibrium in population genetics and derives from the assumption that

$$\langle S_i^\alpha S_j^\alpha \rangle = \langle S_i^\alpha S_j^\beta \rangle \tag{5.22}$$

where α and β are two independent population members. The proof of this is shown in appendix D.

Intuitively this is seen as the end point of repeated applications of crossover. Indeed, inspection of equation (5.6) describing the effect of uniform crossover on the second cumulant confirms this.

Generally the population is not in linkage equilibrium as selection acts to disrupt it. However uniform crossover acts strongly to restore the population to nearly this equilibrium state. It is quite common in biological models to suppose that crossover or recombination acts faster than selection and thus impose linkage equilibrium. This has also been done by Shapiro and Rattray [33] in a model of a population evolving under Boltzmann selection. In this case the coupled cumulant expansions become linear and are thus much easier to solve.

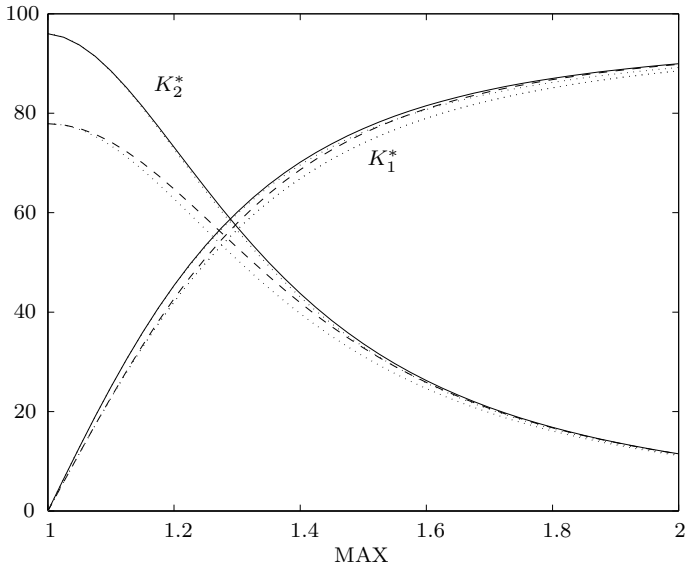


Figure 5.6: Comparison of numerical result for roulette wheel selection (*dashed line*) and stochastic universal sampling (*solid line*) and simplified analytical (*dotted lines*) results for the end point equilibrium. Parameters used were $L = 96$, $\gamma = 1/96$ and $P = 100$.

Other forms of crossover have been suggested which keep the population in linkage equilibrium. One is gene pool recombination where spins are drawn randomly from all of the population members. Here the population goes directly to linkage equilibrium after crossover

$$\langle K_2 \rangle_x = L\,(1-q)\,. \tag{5.23}$$

Another strategy is bit-simulated-crossover proposed by Syswerda [47]. Rather than actually maintain a population of binary strings, a single vector of bit probabilities (or allele frequencies) is used to generate population members. This probability vector is then updated by averaging over the population weighted by fitness. This approach has also been used by Mühlenbein in his factorisation distribution algorithm (FDA) [25].

Using the expression in equation (5.23) the analytical expressions for the equilibrium distribution are significantly simplified without a great deal of loss in accuracy

$$\begin{aligned} K_1^* &= \frac{\Gamma\,(\mathrm{MAX}-1)}{1-\Gamma}\sqrt{\frac{K_2^*}{\pi}} \\ K_2^* &= \frac{L\,(1-\Gamma^2)}{1-\Gamma^2\left(\frac{P-\langle n^2\rangle}{P-1}\right)\left(1-\frac{(\mathrm{MAX}-1)^2}{\pi}\right)}. \end{aligned} \tag{5.24}$$

Figure 5.6 shows a comparison of the previously presented numerical results and the simplified analytical results. The final equilibrium state of the genetic algorithm can thus be accurately predicted in terms of its initial parameters - population size, mutation rate, selection strength and string length.

5.6 Discussion

By considering ranking selection, the full dynamics of a genetic algorithm with mutation and crossover has been modelled using just three macroscopic variables. Indeed if linkage equilibrium is assumed, this may be done with just two variables.

Calculating more cumulants would improve the accuracy of the genetic algorithm without crossover. However when crossover is included, some information is lost in the macroscopic description and the correlation of the population must be reconstructed somehow. In this chapter it has been calculated at a deviation from the natural correlation which arises when spins are arranged in the population with maximum entropy. The change in correlation due to selection has been shown to be dependent on two terms; one describing the change due to finite populations and another describing the dependence of fitness and correlation. In a population with some mutation, the population is sufficiently mixed that the second term can be ignored.

By reducing the number of macroscopic variables to three, some approximations have been introduced. However the gain is significant as a qualitative understanding of the evolution of the population can be gained. The small number of equations of motion also means that closed form solutions can be derived for the end point equilibrium.

Chapter 6

Stabilising Selection

6.1 Introduction

The onemax problem space studied previously is very popular in theoretical studies of genetic algorithms. Whilst it is trivial to solve and unrepresentative of most real problems, it is a first step and gives some insights into the dynamics of the genetic algorithm.

In order to gain some insight into the performance of genetic algorithms on real world optimisation problems, it is necessary to consider harder problem spaces. Characterizing the hardness of a problem has been an active area of research in the genetic algorithm literature. Measures of problem difficulty such as fitness distance correlation [12] and epistasis variance [4] have been suggested but there remains much debate as to the interpretation of these measures and their applicability to real world fitness landscapes.

Despite these arguments, there is some consensus as to what features make a problem space hard to search. There may be many local minima. These local minima may be separated by a potential barrier from better solutions, resulting in the need for non-local search steps. If these local minima occupy the majority of the search space it may take a long time to generate the moves necessary to fall into the basin of attraction of the global minimum.

In this chapter, a simple model which addresses the last two requirements of a hard optimisation problem is considered. The problem is known as the 'basin with a barrier' fitness landscape [42]. It has some of the features of a hard optimisation problem but is still amenable to analysis.

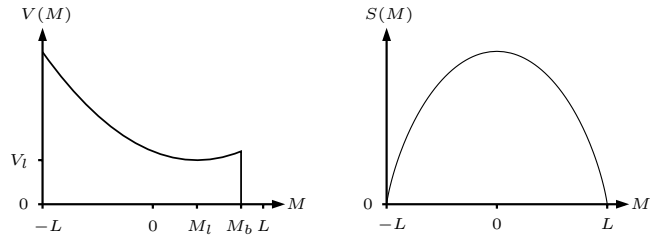


Figure 6.1: Diagram of potential and entropy for the basin with a barrier problem.

The landscape consists of a large local minimum separated from the global minimum by a potential barrier. To continue the analogy with solid state physics, a series of L spins whose value may be 1 or -1 is considered and the total magnetization, M, of the string given by

$$M = \sum_{i=1}^{L} S_i. \tag{6.1}$$

The potential or fitness, which is being minimized, is a function of this magnetization and is given generically as

$$V(M) = \begin{cases} (M - M_l)^2 + V_l & \text{if} \quad M \leq M_b \\ 0 & \text{if} \quad M > M_b. \end{cases} \tag{6.2}$$

The choice of a quadratic function is arbitrary as ranking selection effectively makes a piece-wise linear approximation of the function and all symmetric concave functions are identical. Evolution on a problem space such as this is known as stabilising selection in evolutionary genetics whilst the case of onemax is analogous to directional selection.

The entropy of the system, S, is such that the number of states in the global minimum is much smaller than that in the local minimum. The maximum entropy state is some distance from both the local and global minima. Figure 6.1 shows the landscape schematically.

Whilst being a toy problem this model holds some of the features seen in a real optimisation problem. Random search is expected to produce poor solutions most of the time and these poor solutions are expected to occupy the majority of the problem space. Good solutions are expected to be near one another in problem space but they may be separated by non-local moves.

Prügel-Bennett and Shapiro [42] analysed this landscape for a genetic algorithm

under Boltzmann selection and uniform crossover. They showed that in the infinite population limit there is a phase in which the population will move from the local minimum to the global minimum from any initial configuration. The analysis suggests that this can occur orders of magnitude faster than a stochastic hill climber can find the global minimum. In order to obtain qualitative results, much of the complexities of modeling the dynamics of the genetic algorithm were omitted and the fit between theoretical and simulation results was poor.

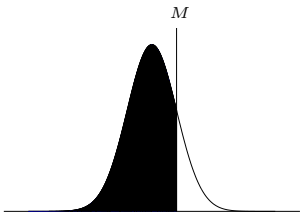
In this chapter, the analysis of ranking selection is extended to the basin with a barrier problem and the dynamics of the population is considered. In chapter seven, the time required for a genetic algorithm to find one solution in the global minimum of the basin with a barrier problem — the first passage time — is calculated and the influence of the various parameters discussed.

6.2 Stabilising Selection

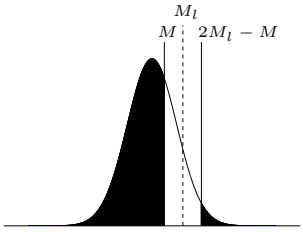
Stabilising selection was first considered under this formalism by Rattray [31]. The dynamics of a genetic algorithm using Boltzmann selection were solved on a simple model of the subset sum or knapsack problem.

The knapsack problem in weakly NP hard. It is analogous to stabilising selection if L packages of random size are considered and they must be optimally packed. A cost function, normally quadratic, is constructed to account for the excess or empty space.

In a similar way to onemax was calculated, the infinite population case is considered first. In onemax the weighting of any population member was dependent on its position within the population and was thus proportional to the area shading below.



Under stabilising selection, the weighting of a population member whose magnetization is M is not only related to its position within the population magnetization distribution but also the position of this minimum, M_l.



Algebraically it is

$$n_M = \begin{cases} \text{MAX} - 2\,(\text{MAX} - 1) \displaystyle\int_M^{2M_l - M} \rho\,(M')\,\mathrm{d}M' & \text{when } M \leq M_l \\ \\ \text{MAX} - 2\,(\text{MAX} - 1) \displaystyle\int_{2M_l - M}^{M} \rho\,(M')\,\mathrm{d}M' & \text{when } M \geq M_l. \end{cases} \tag{6.3}$$

Calculating the cumulants after selection is performed as before. The weighting is integrated over the distribution and the first two moments calculated. The cumulants are then directly calculated from the moments. The full calculation is included in appendix E and the resulting expressions are

$$\langle K_1 \rangle_s = K_1 + (\text{MAX} - 1)\sqrt{\frac{K_2}{\pi}}\text{erf}\left(\frac{M_l - K_1}{\sqrt{K_2}}\right)$$

$$\langle K_2 \rangle_s = \left[1 - \frac{2\,(\text{MAX} - 1)}{\pi}\exp\left(-\frac{(M_l - K_1)^2}{K_2}\right) - \frac{(\text{MAX} - 1)^2}{\pi}\text{erf}^2\left(\frac{M_l - K_1}{\sqrt{K_2}}\right)\right] K_2 \tag{6.4}$$

where $\text{erf}\,(x)$ represents the standard error function.

Clearly when M_l is sufficiently large that little of the distribution falls over the minimum, the expressions above simplify to those presented earlier in chapter four for onemax.

Finite population effects are again considered as an additional multiplicative function and are calculated as before.

6.3 Results

The resulting expressions can be used iteratively to solve the complete dynamics of the genetic algorithm. Figure 6.2 show the theory predictions compared to simulation results from repeated runs of a genetic algorithm using roulette wheel selection and stochastic universal sampling. The simulation data is averaged over 1 000 runs and uses the parameters $L = 48$, $\gamma = 1/48$, $P = 100$, MAX $= 1.4$ and $M_l = L/2$. The figures show very good agreement between theory and simulations.

The population evolves until the increase in mean magnetization due to selection is balanced by the decrease due to mutation. The loss in variance due to selection is balanced against the increase due to mutation and crossover.

Selection by stochastic universal sampling leads to a final distribution with less correlation and higher population variance than roulette wheel. The mean of both distributions is the same and is close to the minimum at M_l.

6.4 Equilibrium Distribution

As with the onemax problem, the equilibrium point may be solved numerically as a set of simultaneous equations. Figure 6.3 shows the simulation results and theoretical predictions for the first two cumulants of the equilibrium distribution against changing selection pressure. Simulation results are averaged over 1 000 runs.

The case of just mutation can be considered by omitting those macroscopic variables and expressions detailing crossover. Here the model fit is slightly poorer, as without crossover to suppress the higher cumulants, the distribution becomes skewed from a Gaussian. However when the distribution significantly overlaps the minimum, a Gaussian shape is restored and reasonable agreement is obtained without having to calculate higher order cumulant terms. Figure 6.4 shows the results for the same genetic algorithm without the crossover operator. The same behavior in the mean is seen whilst the variance of the distribution is greatly reduced. Without crossover acting to restore the variance to its natural value, the population evolves very rapidly to a highly correlated distribution and samples a very much smaller area of the problem space.

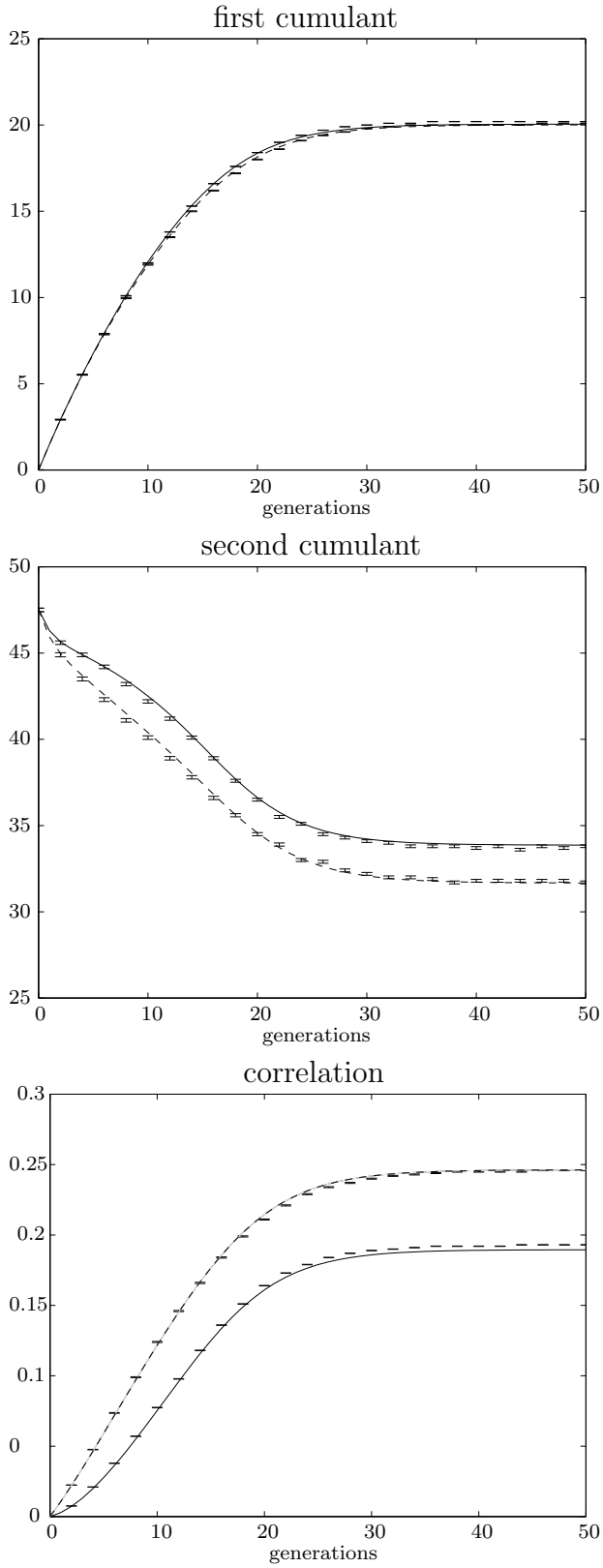


Figure 6.2: Comparison of theoretical and simulation results for the first two cumulants and correlation for a genetic algorithm with ranking selection, mutation and crossover. Roulette wheel selection (*dashed line*) and stochastic universal sampling (*solid line*) are shown. Parameters used were $L = 48$, $\gamma = 1/48$, $P = 100$, MAX $= 1.4$ and $M_l = L/2$.

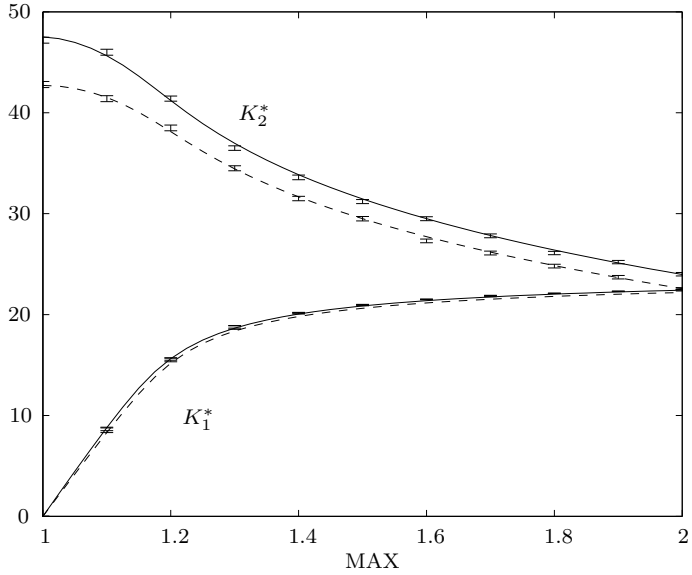


Figure 6.3: Comparison of theoretical numerical solutions and simulation results for the end point equilibrium distribution with crossover. Roulette wheel selection (*dashed line*) and stochastic universal sampling (*solid line*) are shown. Parameters used were $L = 48$, $\gamma = 1/48$, $P = 100$ and $M_l = L/2$.

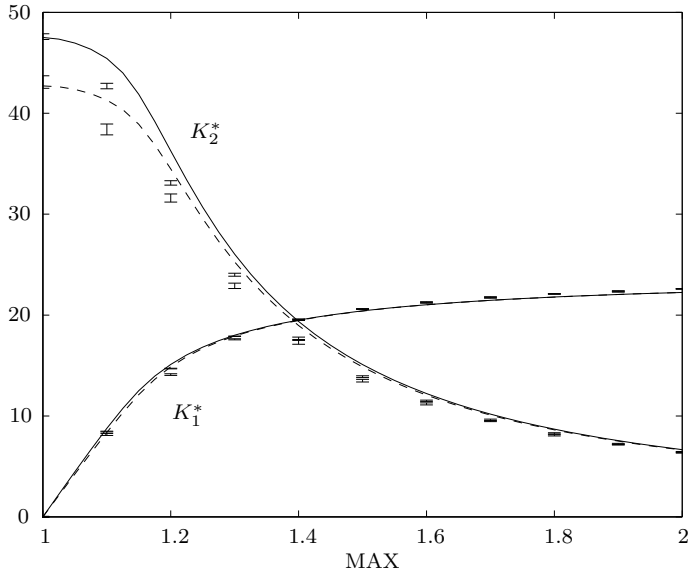


Figure 6.4: Comparison of theoretical numerical solutions and simulation results for the end point equilibrium distribution without crossover. Roulette wheel selection (*dashed line*) and stochastic universal sampling (*solid line*) are shown. Parameters used were $L = 48$, $\gamma = 1/48$, $P = 100$ and $M_l = L/2$.

6.5 A Closed Form Solution

An analytical solution for the end point equilibrium may be made by assuming that selection is strong enough for the population to approach the minimum. In this case, an approximation to the error term in the selection expressions may be made

$$\operatorname{erf}(x) \approx \frac{2x}{\sqrt{\pi}} \qquad \text{where} \quad x = \frac{M_l - K_1}{\sqrt{K_2}}. \tag{6.5}$$

The response of the first cumulant to selection simplifies significantly in this case and is no longer dependent on the variance of the distribution. The equilibrium mean K_1^* can easily be solved to give

$$K_1^* \approx \frac{2\,(\mathrm{MAX}-1)\,\Gamma M_l}{\pi\,(1-\Gamma) + 2\,(\mathrm{MAX}-1)\,\Gamma}. \tag{6.6}$$

The lack of dependence on the equilibrium variance seen in figures 6.3 and 6.4 is clearly evident in this expression. The equilibrium mean will be close to M_l. For small mutation rates, the distance from the minimum is independent of string length L and is of the same order as the number of spins per string which are expected to mutate at each generation

$$M_l - K_1^* \approx \frac{\gamma\pi}{(\mathrm{MAX}-1)} M_l. \tag{6.7}$$

The equilibrium correlation is determined by the balance of mutation and selection and the equilibrium mean. Solving for the equilibrium value of C

$$C^* = \frac{x\,(1-r)}{1-xr} \tag{6.8}$$

where

$$x = \frac{\Gamma^2\,(K_1^{*2} - L^2)}{\Gamma^2 K_1^{*2} - L^2} \quad \text{and} \quad r = \frac{P - \langle n^2 \rangle}{P-1}. \tag{6.9}$$

The equilibrium correlation is thus given by

$$q^* \approx C^* + (1-C^*)\,\frac{K_1^{*2}}{L^2}. \tag{6.10}$$

To find the equilibrium variance, some accuracy is sacrificed in order to derive a simple result. It is again assumed that the end point population distribution

significantly overlaps the local minimum. In this case the reduction in variance due to selection approximates to

$$\langle K_2 \rangle_s \approx r \left[1 - \frac{2\,(\mathrm{MAX} - 1)}{\pi} \right] K_2. \tag{6.11}$$

Thus the equilibrium variance is given by

$$K_2^* \approx \frac{L\,(1 - \Gamma^2) + L\,(1 - q^*)}{2 - \Gamma^2 r \left[1 - \frac{2(\mathrm{MAX}-1)}{\pi} \right]}. \tag{6.12}$$

In the case when no crossover is applied, the result can be derived directly from the approximation in equation (6.11) and is given by

$$K_2^* \approx \frac{L\,(1 - \Gamma^2)}{1 - \Gamma^2 r \left[1 - \frac{2(\mathrm{MAX}-1)}{\pi} \right]}. \tag{6.13}$$

For all but the largest mutation rates, $(1 - \Gamma^2)$ is small compared to the correlation term $(1 - q^*)$. Thus the correlation is significant in producing the larger equilibrium variance of the genetic algorithm with crossover. The deviation from the natural correlation at equilibrium is a function of population size and mutation rate and for mutation rates greater than $1/P$ is small. The correlation is thus defined by the equilibrium mean and thus the position of the minimum, M_l.

Figures 6.5 and 6.6 show the closed form results above plotted against the previously shown numerical solutions for the genetic algorithm with and without crossover.

As expected, the agreement is reasonable when selection is strong and the distribution is near the minimum. The accuracy is better for the case with crossover as the larger distribution variance gives the distribution a significant overlap over the minimum which improves the accuracy of the approximations made in equations (6.5) and (6.11). The closed form results provide a clear insight into the factors determining the final shape of the distribution.

6.6 Discussion

Again, the use of two cumulants provides a good agreement with theoretical results when the dynamics of the genetic algorithm are calculated. Stabilising selection is advantageous as the population is kept near to the high entropy areas and thus higher order cumulants are suppressed. However, the population is clearly not in linkage equilibrium and the correlation must be calculated to model the action of crossover.

Calculating the end point equilibrium in closed form proves to be easy, with the result for the mean becoming increasing accurate with decreasing mutation rate. Interestingly at equilibrium, the variance does not effect the mean.

In the case of no crossover, the skewness of the distribution again increases through selection, limiting the accuracy when calculating the dynamics. Calculating a third cumulant would improve the prediction of the dynamics but increase the complexity of deriving closed form expressions. When close to the optimum however, this skewness is selected against and reasonable estimates for the end point equilibrium can be derived in closed form with just two cumulants.

Of most significance is the difference between the variance of the population with and without crossover. When no crossover is applied, the variance at equilibrium is solely a result of mutation. When the mutation rate is small as is typical, this results in a very little variance at equilibrium.

When crossover is applied, the equilibrium variance is dependent on the correlation of the population. This in turn depends on the deviation from the natural correlation, C. When the mutation rate is small, mutation does not overcome the increase in duplication due to selection and the correlation of the population increase. If the mutation rate is greater than typically $1/P$, mutation prevents the correlation of the population and crossover is able to generate a large equilibrium variance.

In chapter seven this will be shown to be of significance in how the genetic algorithm is searching the problem space and how it may escape a large local minimum.

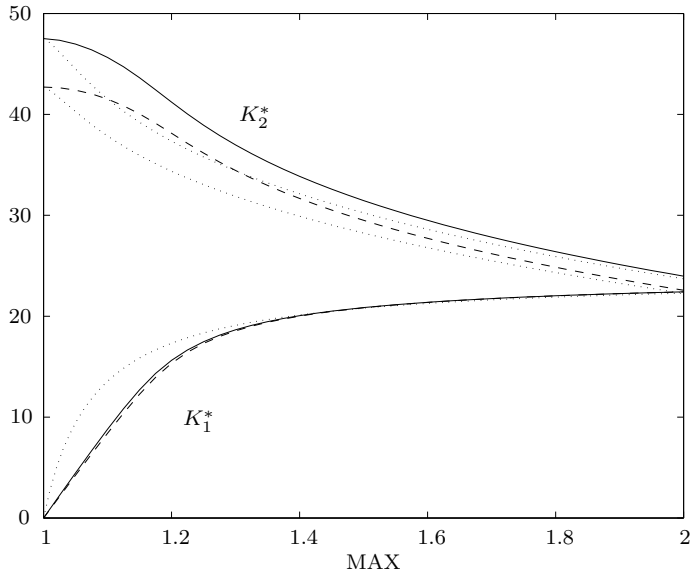


Figure 6.5: Comparison of theoretical numerical solutions for roulette wheel selection (*dashed line*) and stochastic universal sampling (*solid line*) and closed from expressions (*dotted line*) for the end point equilibrium distribution with crossover. Parameters were as before.

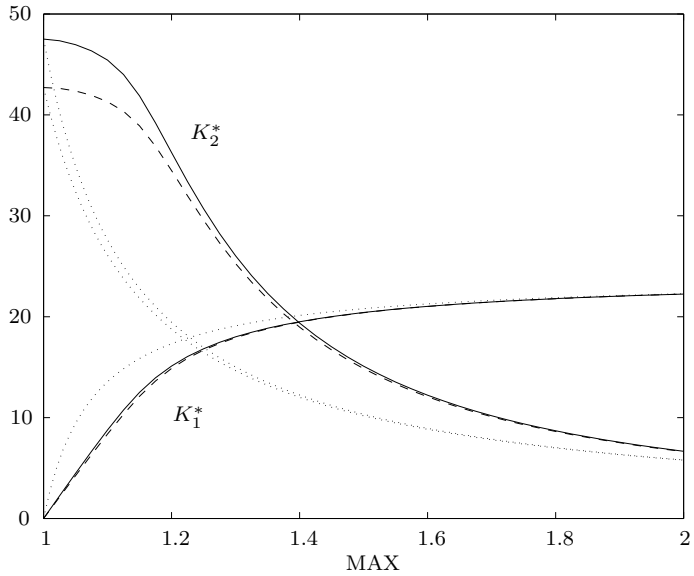


Figure 6.6: Comparison of theoretical numerical solutions for roulette wheel selection (*dashed line*) and stochastic universal sampling (*solid line*) and closed from expressions (*dotted line*) for the end point equilibrium distribution without crossover. Parameters were as before.

Chapter 7

Solving the Basin with a Barrier

7.1 Introduction

In chapter six the dynamics of a genetic algorithm under stabilising selection were solved. Besides giving an quantitatively accurate description of the dynamics, the closed form expressions provide an intuitive insight as to how the genetic algorithm is searching the fitness landscape.

The analysis shows the population evolving to a point where the forces of selection, mutation and crossover are in equilibrium. The equilibrium point is influenced by the three genetic operators.

- Selection acts to focus the population onto areas of improved fitness and thus increases the mean magnetization. In doing so however it reduces the variance of the population leading to a smaller area of the landscape being sampled and increases the correlation of the population.
- Mutation will generate new population members around the selected area but will act to push the population back to the maximum entropy state thereby increasing the variance and decreasing the correlation and mean.
- Crossover does not effect the mean but forces the variance towards a *natural* value defined by the correlation of the population, q. This acts to restore variance to the population lost through selection.

Whilst the population is in equilibrium, it is not static. New population members are continually being generated by mutation and crossover and thus searching the area of the problem space defined by the ensemble cumulants. The greater the

variance of the population, the greater area of the problem space which is being searched.

The final equilibrium variance depends on the contribution of all the various genetic algorithm parameters. Understanding how these parameters affect the time required to solve the basin with a barrier — to create one population member in the global minimum — is the aim of this chapter.

7.2 First Passage Time

The first passage time is defined as the time required for one population member to reach the global minimum. Clearly this will be related to the population size and mean and variance of the population magnetization distribution at its equilibrium point.

As the population magnetization distribution is being described by as a continuous Gaussian, the probability of finding any one population member with a magnetization less than the barrier, M_b, is simply given by

$$p = \Phi(x) \qquad \text{where} \qquad x = \frac{M_b - K_1^*}{\sqrt{K_2^*}} \tag{7.1}$$

and $\Phi(x)$ represents integration of a unit Gaussian from $-\infty$ to x. When x is large, this expression approximates to

$$p \approx 1 - \frac{\mathrm{e}^{-x^2/2}}{x\sqrt{2\pi}}. \tag{7.2}$$

The probability of finding one member above the barrier and thus in the global minimum in any generation is $(1-p)^P$ and since p is small this can be approximated to $1 - Pp$. If the probability of an event in one time step is ϵ, we expect to wait $1/\epsilon$ time steps for the event to occur. See appendix F for a derivation of this result. Thus the expected time in terms of function evaluations, n, is given by $P/(1 - Pp)$. Using the above result gives

$$n \approx x\sqrt{2\pi}\mathrm{e}^{x^2/2}. \tag{7.3}$$

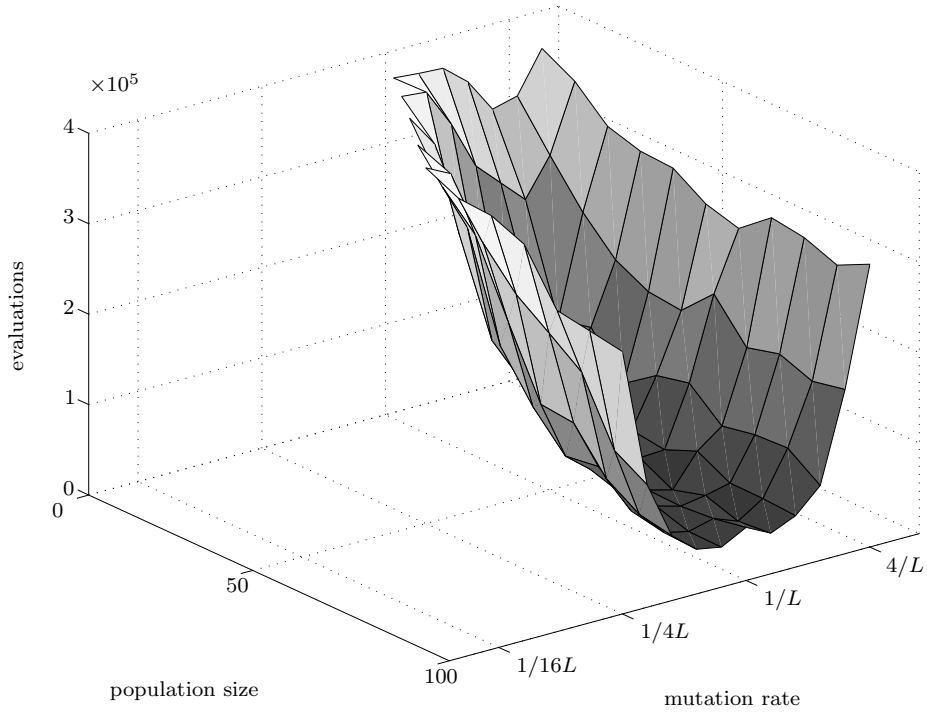


Figure 7.1: Simulation results to solve basin with a barrier using a genetic algorithm with stochastic universal sampling, mutation and uniform crossover. Parameters used were MAX = 2, $L = 48$, $M_l = L/2$ and $M_b = 7L/8$.

The most significant factor here is the exponential dependence on the equilibrium mean and variance

$$n \propto \mathrm{e}^{\frac{(M_l - K_1)^2}{2K_2}}. \tag{7.4}$$

Small changes in the mean and variance of the magnetization distribution will lead to significant changes in the first passage time.

Also interesting is the lack of any population size term in this expression. Once the population has reached the equilibrium point, the number of evaluations required to find a solution is independent of the population size.

7.3 Simulation Results

The time to solve a typical size basin with a barrier problem was found for a range of population sizes and mutation rates by simulation. The results are shown in figure 7.1 and are the results of averaging over 100 runs.

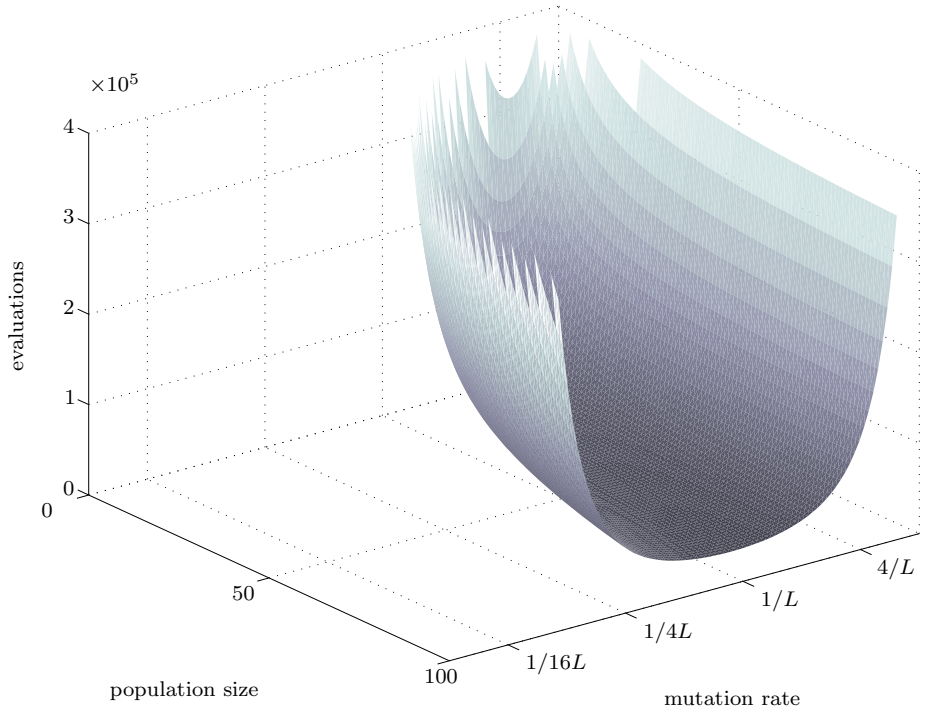


Figure 7.2: Theoretical results to solve basin with a barrier using a genetic algorithm with stochastic universal sampling, mutation and uniform crossover. Parameters used were MAX = 2, $L = 48$, $M_l = L/2$ and $M_b = 7L/8$.

The results show a very clear optimal mutation rate with some population size dependence particularly when the population is small.

7.4 Theoretical Analysis

Having derived expressions for the time to solve the problem and understanding how each of the genetic operators influence the evolution of the population, the effect that each genetic algorithm parameters such as mutation rate and population size have on search can be found.

The numerical solutions to the equilibrium dynamics solved in chapter six were used along with the analytical results derived earlier in this chapter to calculate the first passage time. Figure 7.2 shows the results over the same range of parameters simulated in figure 7.1.

The theory results agree well with the simulation results. The agreement is not expected to be particularly good as the extremes of the population are critical to this calculation and these are poorly defined in a two cumulant expansion. However

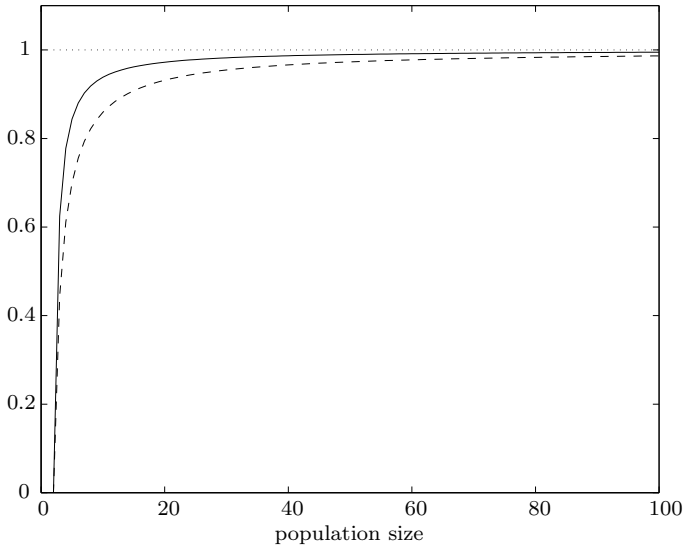


Figure 7.3: Comparison of $(P - \langle n^2 \rangle)/(P - 1)$ for roulette wheel selection (*dashed line*) and stochastic universal sampling (*solid line*).

the theory results clearly show the same influence of parameter settings and agrees well in quantitative values at the optimum mutation rate.

7.4.1 Population Size

As seen in the analysis of the first passage time, the size of the population does not enter the expressions in equation (7.3) for the time required for the population in equilibrium to produce a solution in the global minimum. The population size does however affect this time indirectly by changing the final equilibrium distribution of the population. With small populations the stochastic nature of the selection operator becomes significant and must be accounted for to accurately describe the dynamics and equilibrium distribution.

Finite population effects have been modelled by the factor r which occurs in the selection terms. In chapter five it was shown to rapidly tend to unity as P increased. Figure 7.3 shows this factor plotted against population size.

The finite population causes an extra decrease in population variance through selection and an increase in the population correlation which limits the effectiveness of crossover. A small population thus results in a smaller equilibrium variance. Again as seen in chapter six, the equilibrium mean is unaffected.

Clearly the curve approaches unity — the infinite population limit — very quickly.

For small population sizes however, the deviation from unity is significant. This feature explains the relative lack of dependence of population size on the time to solve the problem when the population is large and the very rapid decline in performance as the population size decreases.

In this analysis, it has been assumed that the initial dynamics are comparatively short in comparison to the time spent at equilibrium waiting for the mutation and crossover operators to generate a solution in the global minimum. If we consider a more realistic problem consisting of a cascade of barriers, it is clear that this initial dynamic phase favors a smaller population as it requires less function evaluations to move the population to its new equilibrium point.

This suggests an optimum population size which is a balance between the need to maintain the speed which the population can move around the landscape whilst not be so small that the area of the landscape being searched is significantly reduced by finite population effects.

7.4.2 Selection Scheme

Comparison of the finite population effects for both stochastic universal sampling and roulette wheel selection expressed in terms of the factor r, shows that there is a significant difference between the two. In chapter six this was shown to result in a significant difference in the final equilibrium variance of the population. The mean however is unchanged.

In the simulation and theory results plotted in this chapter, stochastic universal sampling has been used. When using roulette wheel selection, the small difference in equilibrium variance leads to a doubling in the time required to solve the problem.

7.4.3 Mutation Rate

Perhaps the most significant feature of the surfaces plotted is the strong dependence on mutation rate, with extremely poor performance outside the optimal range. Understanding this feature involves the interplay of all the effects previously discussed.

Mutation has been shown to increase the variance of the final population equilibrium distribution but also move the mean of the distribution away from the global minimum back towards the maximum entropy state. The second effect is most significant here and mutation has a detrimental effect on performance. Increasing

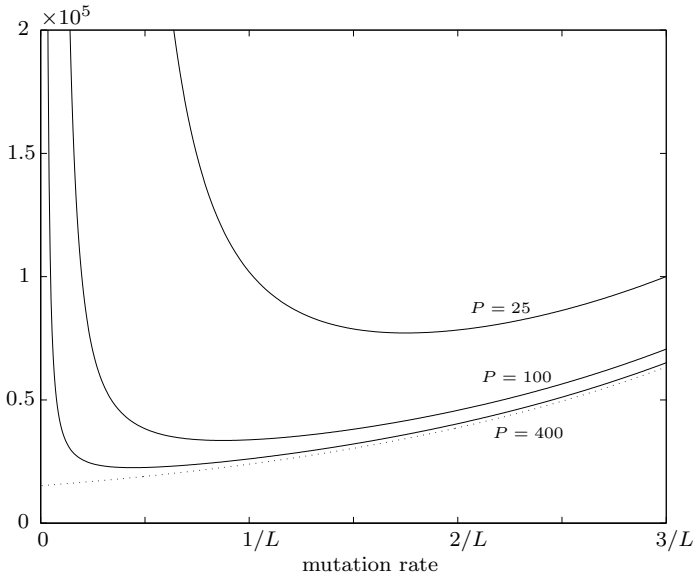


Figure 7.4: Theoretical results to solve basin with a barrier with varying mutation rate at three different population sizes — $P = 25, 100$ and 400. The dotted line is the infinite population response. Parameters used were MAX $= 2$, $L = 48$, $M_l = L/2$ and $M_b = 7L/8$.

mutation rate increases the number of function evaluations required to solve the problem.

In an infinite population, the optimum mutation rate would thus be zero. However, a finite population is being considered and the correlation of the population caused by selection can not be ignored. With no mutation, the correlation of the population will increase very rapidly, limiting crossovers ability to restore variance to the population. This will result in a very small equilibrium variance which searches a very small area of the problem space and thus takes a very long time to reach the global minimum.

A balance is achieved when mutation is large enough to prevent the correlation of the population but not so large as to disrupt the search.

Figure 7.4 shows theoretical results for the time to solve the basin with a barrier for varying mutations rates. Three different population sizes are shown along with the infinite population case as a dotted line. Clearly as P increases, the finite population effects decrease and the finite populations approach the infinite case. However as the mutation rate becomes small, the correlation of the population through finite population effects becomes significant.

The optimum mutation rate is seen to be dependent on the population size. A large

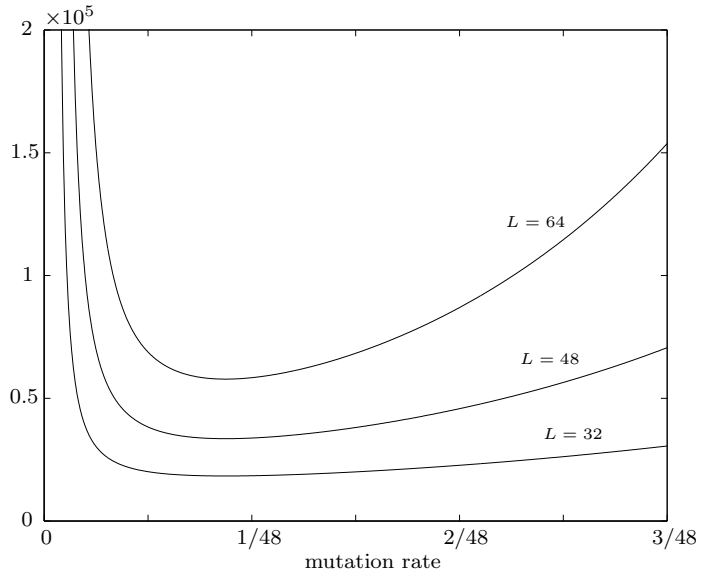


Figure 7.5: Theoretical results to solve basin with a barrier with varying mutation rate at three string lengths — $L = 32, 48$ and 64. Vertical scales for $L = 32$ and $L = 64$ are multiplied and divided by ten respectively to enable easy comparison. Parameters used were MAX $= 2$, $P = 100$, $M_l = L/2$ and $M_b = 7L/8$.

population suffers less correlation due to finite population effects and thus will not suffer the same correlation until the mutation rate is very small.

Using the closed form expressions derived in chapter six for the behaviour of the population close to the equilibrium, the optimum mutation rate can by shown to be approximately proportional to $1/\sqrt{P}$. This can clearly be seen in in figure 7.4, where slices are made through the theoretically calculated surface shown in figure 7.2.

7.4.4 String Length

As the effects discussed here are independent of the string length, and the optimum mutation rate is thus also independent of the string length. Figure 7.5 shows the results of varying the length of the string for various mutation rates. As expected, after performing this analysis, there is no dependence on the optimum mutation rate with string length but the time to solve the problem increases as the global minimum occupies an exponentially decreasing section of the problem space.

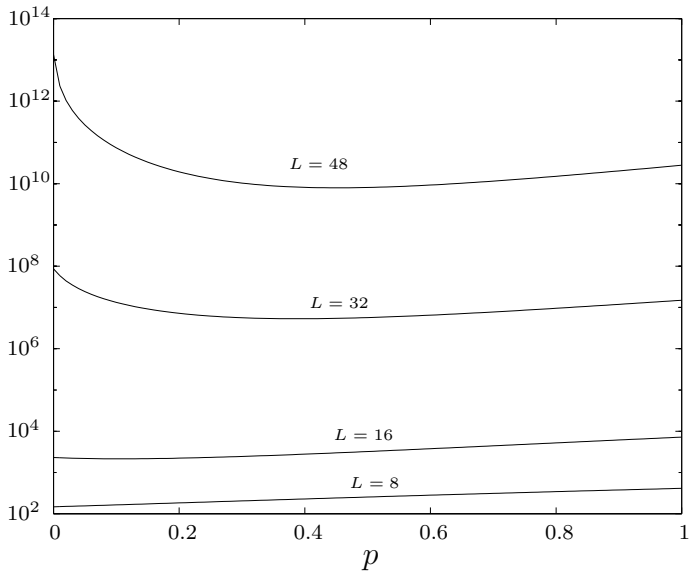


Figure 7.6: Theoretical first passage times for a single stochastic walker to solve the basin with a barrier. Problem sizes are $L = 8, 16, 32$ and 48.

7.5 Stochastic Walker

It is interesting to compare the results of the genetic algorithm to that of a single stochastic walker. A simple stochastic search algorithm is considered whereby at each time step a new move is generated by allowing each spin to mutate with probability $1/L$ — equivalent to the genetic algorithm mutation rate. Strictly speaking this allows global and not just local moves. However as L increases, the probability of this becomes small. A simple generic Metropolis algorithm is used, whereby steps which increase fitness are always accepted and steps which will decrease fitness are accepted with some probability, p. As there is only one barrier in this problem, there is no need to anneal this probability as is done in simulated annealing.

The transition times to reach the global minima are calculated directly from the transition matrix describing the probability of changing from one state to another. A randomly assigned starting string is assumed. Figure 7.6 shows the results for each problem size on a logarithmic scale. When $p = 0$ the hill climber performs simple steepest descent and sits at the local minima waiting for the correct mutation to jump straight into the global minimum. When $p = 1$ the hill climber performs a random walk and most time is spent in the maximum entropy area away from both the local and global minima. Between the two an optimum is reached.

As the problem size increases, the time to reach the global minimum increases rapidly. For problem sizes greater than L=16, the single walker is orders of magnitude slower at finding a solution in the global minimum than the time predicted and observed for the genetic algorithm.

7.6 Conclusions

The basin with a barrier problem is a caricature of a real world optimisation problem. Unlike simpler models such as onemax, it has local minima and thus shows some of the features of a hard optimisation problem. For this problem, the effects of all the genetic algorithm parameters have been modelled and the model shown to give good quantitative results and allow qualitative insights. In this case, mutation acts as a disruptive force. Unlike crossover it has no knowledge of which parts of the strings are shared by many population members and thus disrupts parts of the string which are beneficial to fitness.

Crossover has been shown to be the dominant search operator on this landscape. Indeed without crossover, the extremely small equilibrium variance results in the genetic algorithm taking many orders of magnitude longer to solve the problem. By mixing those parts of the strings which are not identical, it is able to produce new population members without disrupting what has already been gained. However in the absence of mutation, selection very rapidly produces a highly correlated population which prevents crossover from operating. Thus a minimum level of mutation is required to overcome this correlation without disrupting the search.

At larger population sizes, the increase in correlation of the population is slower and thus the optimum mutation rate is lower. This optimum is seen to be independent of the length of the string.

Relating this work to real problems however is still some way off. It is unlikely that the relationships between the optimum mutation rate with population size and string length will hold on more general problems. However the techniques developed here have enabled a model problem to be analysed and definite statements made about the influence of parameters. In this way it represents a first step towards understanding the influences these parameters may have on real world problems.

Chapter 8

Biological Models

8.1 Introduction

The model genetic algorithms in this thesis are very similar to models of evolving populations developed in the field of population genetics. Despite the similarity, there is little crossover of ideas between the two fields. Beside the issue of terminology, this is probably because the two fields are interested in different aspects.

In quantitative genetics, the allele frequency are of interest as this is what can be measured in a real population. The models tend thus to be of few loci. Linkage equilibrium is often assumed as it renders the allele frequencies independent and the equilibrium solution may be solved.

There are however two areas where the models developed in this thesis are of direct relevance to research in quantitative genetics. These are the cases of the comparison between overlapping and non-overlapping populations considered in chapter two and the comparison of sexual and asexual population in stabilising selection.

8.2 Overlapping and Non-Overlapping Generations

A standard model of an evolving population used in population genetics is to consider a population of P haploid individuals consisting of a single genetic string of L loci. Each locus contributes multiplicatively to the fitness of the individual with a factor $1 + s$. This is commonly known as evolution on a multiplicative fitness landscape.

It is easy to recast the model presented in chapter two into this form. The efficacy of an individual is given by the sum of the alleles

$$E = \sum_{i=1}^{L} S_i \qquad \text{where} \qquad S_i = \{-1, 1\}. \tag{8.1}$$

This was previously referred to as the fitness but as this term has a specific meaning in quantitative genetics, efficacy is used in its place. Under Boltzmann selection, the weighting of any individual is given by

$$w_\alpha = \frac{\mathrm{e}^{\beta E_\alpha}}{Z} \qquad \text{where} \qquad Z = \sum_{\alpha=1}^{P} \mathrm{e}^{\beta E_\alpha} \tag{8.2}$$

and β is the selection pressure. The summation of the alleles in the exponential makes their contributions multiplicative. In terms of the more commonly used measure of the selective advantage of a favourable allele, s, the selection strength is given by

$$\beta = \ln(1 + s) \tag{8.3}$$

and thus for small β

$$s \approx \beta. \tag{8.4}$$

Moran [21] considered two single-loci models of this type with overlapping and non-overlapping generations — equivalent to generational Boltzmann selection and steady state Boltzmann selection with random deletion. Under neutral selection he showed that the rate of genetic drift in the population with over-lapping generations is twice that of the population with non-overlapping generations. This result was shown in the current analysis in chapters two and three.

In a more detailed comparison, Moran [20, 19] went on to compare populations subject to selection and mutation. He used a diffusion theory result to approximate the distribution when the population is in equilibrium. The expressions for overlapping and non-overlapping generations were shown to be approximately equal if the mutation rate and selection strength of the overlapping populations were doubled. No explanation was given as to the reason for this.

This observation is clearly understood from the rescaling observed in chapter two. The overlapping population exhibits twice the rate of genetic drift as the non-overlapping case and thus doubling the selection strength and mutation rate offsets

this increase and the population evolves to the same end point equilibrium but at twice the rate.

The rescaling of the mutation rate is easily understood under the formalism developed here as the dynamics of the evolving population are solved and not just the final equilibrium point.

8.3 Stabilising Selection-Mutation Balance

The selection in the above example is directional. Of more interest in biology is stabilising selection where the trait being modeled has some optimum value. The model of stabilising selection developed in the study of the basin with a barrier problem is directly applicable to this case and is used in a comparison of sexual and asexual population i.e. with and without crossover.

The model considers a population of P haploid individuals whose fitness is determined by one quantitative trait. This trait is affected by L loci, each with two alleles which are denoted by A_i and a_i. All loci are assumed to have equal contributions and an indicator variable l_i is used such that $l_i = 1$ if the gamete contains A_i and $l_i = -1$ if the gamete contains a_i at the ith position. The phenotype value of the character is computed additively as

$$x = \sum_{i=1}^{L} l_i. \tag{8.5}$$

Thus x varies between $-L$ and L in value. Stabilising selection is consider to occur with the fitness of an individual being some function of its distance from an optimal phenotype αL, where $-1 < \alpha < 1$. In conventional models of stabilising selection, this function is often taken to be either quadratic or Gaussian. Tournament selection is considered so the exact fitness function is not significant and all functions which are symmetrical about the optimum, αL, are equivalent.

The population evolves subject to mutation and populations with and without recombination are considered.

Population size in biological models tend to be large and thus an infinite population model may be considered. This significantly reduces the complexity of the model

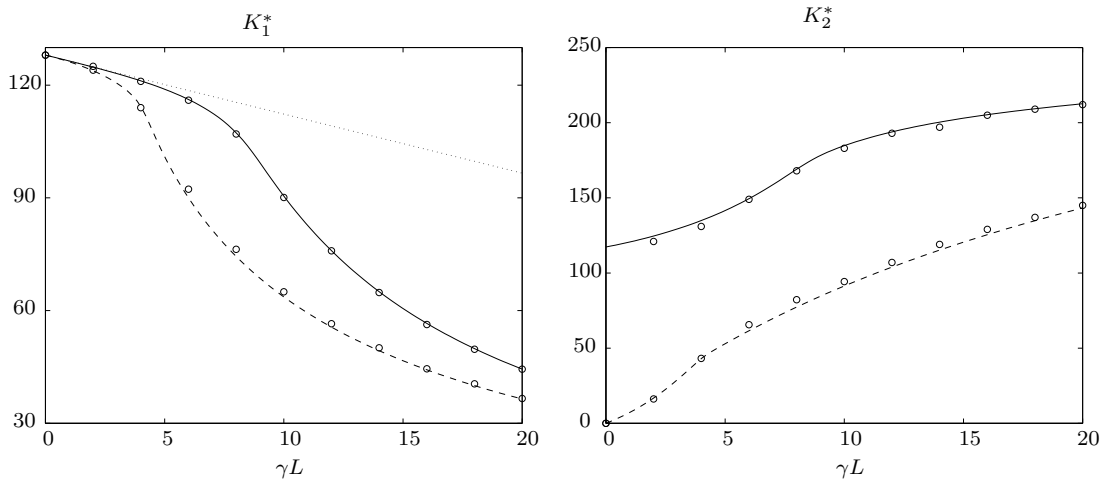


Figure 8.1: Comparison of simulation and theory results for equilibrium mean, K_1^*, and variance, K_2^*, for sexual (*solid line*) and asexual (*dashed line*) populations with 256 loci and α equal to 1/2. Simulation results are for a population size of 1 000 individuals and are averaged over 100 runs. Error bars are the size of the symbols.

as the correlation may be assumed to be the natural correlation of the population

$$q = \frac{K_1^2}{L^2}. \tag{8.6}$$

The model may thus be described by just two macroscopic variables, K_1 and K_2. Agreement will be better for the sexual population as the higher order cumulants are suppressed but good quantitative results are obtained. Figure 8.1 shows the final equilibrium mean and variance for a sexual and asexual population with 256 loci and α equal to 1/2. The simulation results are for a population of 1 000 individuals.

When the mutation rate is small, both sexual and asexual populations evolve to an equilibrium with phenotype mean very close to the optimum at αL and increasing linearly with mutation rate. In chapter six, this distance from the optimum was approximated in closed form

$$\alpha L - K_1^* \approx \gamma \pi \alpha L. \tag{8.7}$$

This analytical result is shown in figure 8.1 as a dotted line. As the mutation rate increases, the mean decreases almost linearly until a threshold is reached when the fitness decreases rapidly. This threshold occurs later in the sexual population. In the regime beyond the threshold, the population is effectively subject to directional selection as no part of the population reaches the optimum phenotype.

The variance of the sexual population is very much greater than that of the asexual population due to the effect of recombination to restore variance lost through selection.

The fit of theoretical results to simulation results is generally very good. The main departure between simulation and theory occurs at very small mutation rates in the sexual population. With very little mutation, the correlation deviates from the natural correlation. This factor has been neglected and thus the theory predicts some variance at zero mutation whilst in reality, the variance is zero.

8.3.1 Mutation Rate Threshold

The mutation rate threshold observed is a function of the changing variance of the population with mutation rate. Whilst the variance is great enough that the distribution overlaps the optimum, the approximation made to the error function in chapter six is valid. As the variance decreases and the mean moves further from the optimum, the approximation underlining the derivation (that $\alpha L - K_1^*/\sqrt{K_2^*}$ is small) no longer holds. At this point the population is effectively subject to directional selection and the mean decreases rapidly.

Under the large population limit, the equilibrium variance may be approximated as

$$\begin{aligned} K_2^* &\approx 2\pi\gamma L && \text{asexual} \\ K_2^* &\approx \frac{1-\alpha^2\left(1-2\pi\gamma\right)+4\gamma}{1+\frac{2}{\pi}}L && \text{sexual.} \end{aligned} \tag{8.8}$$

The threshold mutation rate at which the population fitness rapidly declines is found by considering when the assumption that $\left(\alpha L - K_1^*\right)/\sqrt{K_2^*}$ is small, is no longer valid. This will certainly be true when $\left(\alpha L - K_1^*\right)/\sqrt{K_2^*}$ is greater than one. Using this limit and the value $\alpha = 1/2$ gives the result

$$\begin{aligned} \gamma &\approx \frac{2}{L} && \text{asexual} \\ \gamma &\approx 0.4\sqrt{\frac{1}{L}} && \text{sexual.} \end{aligned} \tag{8.9}$$

Figure 8.2 shows the equilibrium mean and variance for sexual and asexual populations for five different numbers of loci. The mutation rates are scaled as $\sqrt{L}$ for sexual populations and L for asexual populations to show the scaling of the threshold with string length.

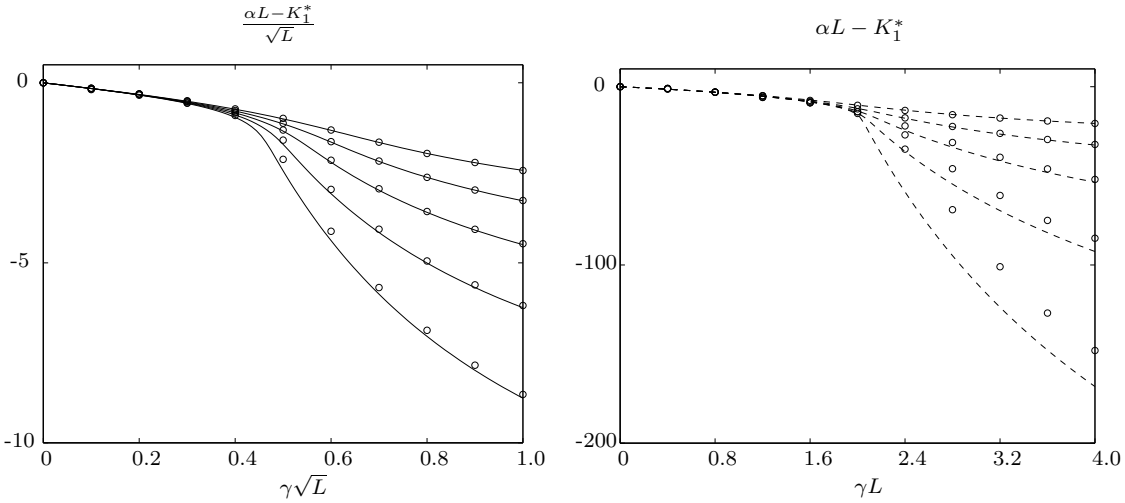


Figure 8.2: Comparison of sexual (*solid line*) and asexual (*dashed line*) populations with 64, 128, 256, 512 and 1024 loci. Horizontal axis are scaled as $\sqrt{L}$ for sexual populations and L for asexual populations to show the threshold scaling.

The simulation and theory results show very good agreement. The scaling calculated in the analysis is clearly present in the simulation results and the calculated position of the threshold agrees very accurately with that found from simulation. The results of the sexual population analysis are quantitatively more accurate due to the effect of recombination to reduce the higher order cumulants.

The mutation rate threshold shows that the mutation rate which a sexual population can withstand is $\sqrt{L}$ times greater than the equivalent asexual population. In quantitative genetics, the distance from the optimum phenotype due to mutation is known as mutation load and thus the sexual population exhibits lower mutation load than the equivalent sexual population.

This fact appears to have first been noted by Kondrashov [15] and was recently calculated independently by Mackay [16] in an analysis of a single generation of truncation selection.

8.4 Discussion

The comparison of sexual and asexual populations is a common theme in theoretical studies. There is much debate as to the value of sex. Besides the increase in complexity which sex involves, the fitness of each individual is effectively half that of an individual which can reproduce asexually. Despite this cost, all organisms of sufficient complexity reproduce sexually.

This analysis shows that under stabilising selection with a low mutation rate, the most significant difference between sexual and asexual populations is the phenotype variance at equilibrium. Recombination results in a significantly greater variance in the sexual population. As the rate at which the population can move under directional selection is proportional to the width of the distribution, this increase in variance allows the sexual population to follow a changing environment.

In the high mutation rate regime, the sexual population can withstand a mutation rate $\sqrt{L}$ times greater than the equivalent asexual population.

Whether either of these mechanisms is an explanation of the efficacy of sexual populations is still an open question in biology. However it is interesting that differences between asexual and sexual populations can be found even in simple models of evolving systems.

Chapter 9

Conclusions and Future Directions

9.1 Introduction

The formalism presented in this thesis describes the behaviour of a simple model of the genetic algorithm very well. In the course of developing a macroscopic model however, some information about the population is lost.

In the genetic algorithm without crossover, the population can be better represented by using more cumulants to describe the fitness population. However in the case of the genetic algorithm with crossover, information about the correlation of the population is lost and must be recovered in some way. In chapter five, the correlation was calculated as a deviation away from the natural correlation of the population which occurs when the spins are distributed with maximum entropy.

The effect of selection on crossover was shown to be reduced to two terms; one depending on the stochastic nature of the selection scheme on a finite population and another related to the dependence of fitness with correlation. In a well mixed population subject to mutation it was shown that the second term can be ignored. However in the low mutation regime, the second term becomes significant and gives rise to a phase transition.

9.2 Low Mutation Phase Transition

In a large population, the correlation due to the stochastic nature of the selection scheme is very small and thus the correlation is assumed to be close to the the natural correlation of the population. For a population subject to stabilising selection,

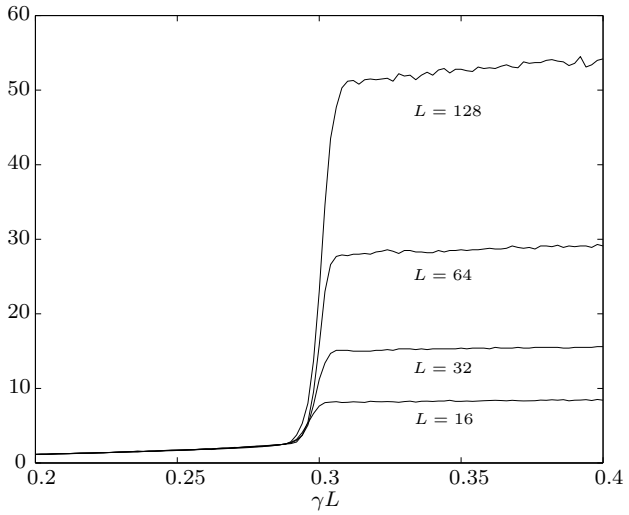


Figure 9.1: Simulation results of equilibrium variance, K_2^*, at the low mutation threshold for a sexual population under stabilising selection. Four different string lengths are shown. Results are for a population size of 2 000 and are averaged over 100 runs.

the variance at equilibrium was shown in chapter eight to be

$$K_2^* \approx \frac{1 - \alpha^2 (1 - 2\pi\gamma) + 4\gamma}{1 + \frac{2}{\pi}} L \tag{9.1}$$

where $M_l = \alpha L$. However, when simulations are performed at mutation rates below $1/L$ a remarkable change in variance is seen to occur at a point which scales with L. Figure 9.1 shows the equilibrium variance of a population evolving under stabilising tournament selection with a population size of 2 000 and four different lengths of string.

Below the threshold the variance is very similar to that of the population without crossover. However when mutation exceeds a threshold value, crossover is able to produce the large equilibrium variance predicted in equation (9.1).

In the previous analysis of the basin with a barrier problem, this will result in very high first passage times at small mutation rates regardless of population size. In chapter seven, the mutation rates are higher than this threshold and the effect was not observed.

The change is a phase transition which can be seen in its purest form by considering stabilising selection with the optimum at zero. Figure 9.2 shows the equilibrium correlation of the population under changing mutation rates for three different string lengths.

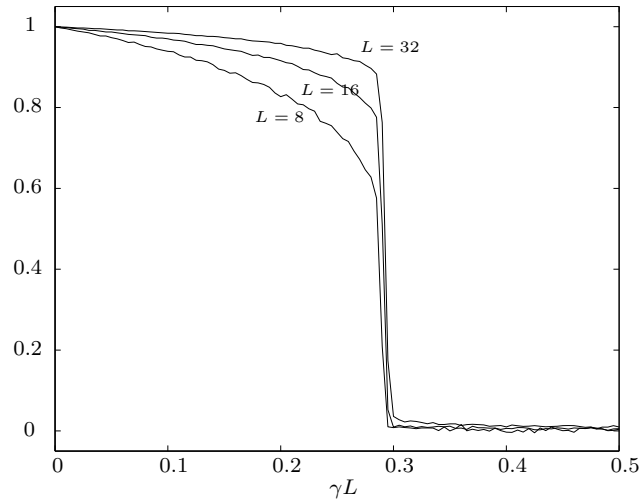


Figure 9.2: Simulation results showing phase transition in equilibrium correlation, q^*, under stabilising selection. Three different string lengths are shown. Results are for a population size of 2 000 and are averaged over 100 runs.

In this case, the natural correlation of the population, $\tilde{q}$, is zero and thus

$$q = C. \tag{9.2}$$

The change of correlation was shown in chapter five to be

$$1 - \langle C \rangle_s = \frac{P - \langle n^2 \rangle}{P - 1} (1 - C) - \frac{1}{P(P-1)} \sum_{\alpha \neq \beta} (C_{\alpha\beta} - C)\, n_\alpha n_\beta. \tag{9.3}$$

Assuming the population is large, the change is dominated by the second term

$$\langle q \rangle_s = q + \frac{1}{P(P-1)} \sum_{\alpha \neq \beta} (q_{\alpha\beta} - q)\, n_\alpha n_\beta. \tag{9.4}$$

Here the change in correlation is directly related to the dependence of correlation and fitness. The highly correlated population consists of near identical individuals. Mutations are rare and when they do occur, they result in an individual of less than optimal fitness. These individuals are selected against and selection will increase the correlation of the population.

The optimum fitness however may be represented by many states, only one of which is actually represented within the population. When the mutation rate increases, there is an increased probability that another form of the optimum fitness is generated. These individuals are not selected against and thus the correlation of the

population decreases towards the natural correlation. In the uncorrelated state, each of the forms of the optimum is equally likely and thus the composition of the population is very different.

This change in composition of the population occurs very rapidly and appears to be a first order phase transition. Although the correlated state is energetically preferable, and has a higher mean fitness due to the smaller variance, the entropy of the mixed state is such that beyond a threshold, the population moves between the two.

It would appear that this phase transition has not been previously noted and would be of interest in biology as it marks a mutation rate at which the sexual population is distinguished from the asexual population.

9.3 Conclusions

Although the model genetic algorithm presented in this thesis is really a caricature of the real world case, it displays complex and often surprising behaviour. It appears that the calculation of the phase transition observed at low mutation rates represents a challenge to the formalism as presented. The macroscopic variables do not contain enough information about the distribution of spins within the population to enable the dynamics to be calculated at this point.

Whilst it is likely that the phase transition may be calculated by a different approach, extending the formalism to cover this case would clearly be of great interest. Such an extension would hopefully contribute towards attempts to describe the dynamics of the genetic algorithm on more complex problem spaces.

Appendix A

Mutation

When the mutation operator is applied, each spin of the population has a small probability of mutation

$$S_i \to -S_i \quad \text{with probability } \gamma.$$

The effect of mutation on the ensemble fitness distribution in the onemax problem is calculated by first considering the effect on any individual spin. The expected value of any spin after mutation is easily shown to be

$$\langle S_i \rangle_m = \Gamma S_i \quad \text{where} \quad \Gamma = (1 - 2\gamma) \tag{A.1}$$

and $\langle \ldots \rangle_m$ represents the average over all mutations. Applying this to the expected fitness of any individual after mutation gives

$$\begin{aligned} \langle F \rangle_m &= \sum_{i=1}^{L} \langle S_i \rangle_m \\ &= \Gamma F. \end{aligned} \tag{A.2}$$

For second order terms the expression is more complicated as dependent terms must be collected together

$$\begin{aligned}\langle F^2\rangle_m &= \sum_{i\neq j}\langle S_i\rangle_m\langle S_j\rangle_m + \sum_{i=1}^{L}\langle S_i^2\rangle_m \\ &= \Gamma^2\left(F^2 - L\right) + L \\ &= \Gamma^2 F^2 + L\left(1-\Gamma^2\right). \end{aligned} \tag{A.3}$$

Applying these expressions to the definitions of the cumulants and averaging over the ensemble gives

$$\begin{aligned}\langle K_1\rangle_m &= \Gamma K_1 \\ \langle K_2\rangle_m &= \Gamma^2 K_2 + L\left(1-\Gamma^2\right). \end{aligned} \tag{A.4}$$

Appendix B

Crossover

When uniform crossover [45] is applied, the spins of any offspring are drawn from each parent, α and β, at random

$$S_i = \chi_i S_i^{\alpha} + (1 - \chi_i)\, S_i^{\beta}. \tag{B.1}$$

where

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

The effect of crossover is calculated by considering each spin. The expected value of any spin, averaged over all ways of drawing bits from each parent is simply

$$\langle S_i \rangle_x = \frac{S_i^{\alpha}}{2} + \frac{S_i^{\beta}}{2}. \tag{B.3}$$

Clearly then the expected fitness of any offspring produced through crossover is

$$\langle F \rangle_x = \frac{F_{\alpha}}{2} + \frac{F_{\beta}}{2}. \tag{B.4}$$

As α and β are drawn independently from the population, the average over the ensemble can be taken to give the mean ensemble fitness after crossover

$$\langle F \rangle_x = \langle F \rangle. \tag{B.5}$$

For second order terms the analysis is slightly more complicated. For any offspring

$$\begin{aligned}\langle F^2\rangle_x &= \sum_{i\neq j}\langle S_i\rangle_x\langle S_j\rangle_x + \sum_{i=1}^{L}\langle S_i^2\rangle_x \\ &= \sum_{i\neq j}\langle S_i\rangle_x\langle S_j\rangle_x + L \\ &= \left(\sum_{i=1}^{L}\langle S_i\rangle_x\right)^2 - \sum_{i=1}^{L}\langle S_i\rangle_x^2 + L. \end{aligned} \tag{B.6}$$

Now considering all ways of drawing the parents and the preceding results gives

$$\begin{aligned}\langle F^2\rangle_x &= \left(\frac{F_\alpha}{2}+\frac{F_\beta}{2}\right)^2 - \sum_{i=1}^{L}\left(\frac{S_i^\alpha}{2}+\frac{S_i^\beta}{2}\right)^2 + L \\ &= \left(\frac{F_\alpha^2}{4}+\frac{F_\beta^2}{4}+\frac{F_\alpha F_\beta}{2}\right) - \sum_{i=1}^{L}\left(\frac{1}{2}+\frac{S_i^\alpha S_i^\beta}{2}\right) + L. \end{aligned} \tag{B.7}$$

Averaging over all ways of drawing α and β independently from the population gives the second moment of the ensemble fitness

$$\langle F^2\rangle_x = \frac{\langle F^2\rangle}{2} + \frac{\langle F\rangle^2}{2} + \frac{L}{2}(1-q) \tag{B.8}$$

where q is defined as

$$q = \frac{1}{P(P-1)}\sum_{\alpha\neq\beta}\frac{1}{L}\sum_{i=1}^{L} S_i^\alpha S_i^\beta. \tag{B.9}$$

The cumulant terms are simply found from the ensemble moments derived above

$$\begin{aligned}\langle K_1\rangle_x &= K_1 \\ \langle K_2\rangle_x &= \frac{K_2}{2} + \frac{L}{2}(1-q). \end{aligned} \tag{B.10}$$

Appendix C

Correlation under Selection

A measure of the deviation, $C_{\alpha\beta}$, away from the natural correlation, $\tilde{q}$, is calculated

$$q_{\alpha\beta} = C_{\alpha\beta} + (1 - C_{\alpha\beta})\,\tilde{q}_{\alpha\beta} \tag{C.1}$$

and averaged over the population

$$q = C + (1 - C)\,\tilde{q}. \tag{C.2}$$

After selection, there will be n_α copies of individual α in the new population. Thus

$$\sum_{\mu,\nu} q_{\mu\nu} = \sum_{\alpha,\beta} n_\alpha n_\beta q_{\alpha\beta}. \tag{C.3}$$

Separating independent terms gives

$$\begin{aligned}\sum_{\mu\neq\nu} q_{\mu\nu} + P &= \sum_{\alpha=1}^{P} n_\alpha^2 + \sum_{\alpha\neq\beta} n_\alpha n_\beta q_{\alpha\beta} \\ &= \sum_{\alpha=1}^{P} n_\alpha^2 + \sum_{\alpha\neq\beta} n_\alpha n_\beta \left(C_{\alpha\beta} + (1 - C_{\alpha\beta})\,\tilde{q}\right) \\ &= \sum_{\alpha=1}^{P} n_\alpha^2 + \tilde{q} \sum_{\alpha\neq\beta} n_\alpha n_\beta + (1-\tilde{q}) \sum_{\alpha\neq\beta} C_{\alpha\beta} n_\alpha n_\beta \\ &= \sum_{\alpha=1}^{P} n_\alpha^2 + \tilde{q} \sum_{\alpha\neq\beta} n_\alpha n_\beta + (1-\tilde{q})\, C \sum_{\alpha\neq\beta} n_\alpha n_\beta \\ &\qquad + (1-\tilde{q}) \sum_{\alpha\neq\beta} (C_{\alpha\beta} - C)\, n_\alpha n_\beta. \end{aligned} \tag{C.4}$$

As the selection scheme maintains a constant population size

$$\sum_{\alpha\neq\beta} n_\alpha n_\beta = P^2 - \sum_{\alpha=1}^{P} n_\alpha^2. \tag{C.5}$$

Subtracting both sides from P^2 gives

$$\begin{aligned} P\,(P-1) - \sum_{\mu\neq\mu} q_{\mu\nu} &= \left(P^2 - \sum_{\alpha} n_\alpha^2\right)(1-C)\,(1-\tilde{q}) \\ &\qquad - (1-\tilde{q}) \sum_{\alpha\neq\beta} (C_{\alpha\beta} - C)\, n_\alpha n_\beta. \end{aligned} \tag{C.6}$$

The correlation after selection is defined as

$$\langle q \rangle_s = \frac{1}{P(P-1)} \sum_{\mu\neq\nu} q_{\mu\nu} \tag{C.7}$$

and thus the effect of selection on correlation is given by

$$\begin{aligned} 1 - \langle q \rangle_s &= \frac{P - \langle n^2 \rangle}{P-1}\,(1-\tilde{q})\,(1-C) \\ &\qquad - \frac{(1-\tilde{q})}{P(P-1)} \sum_{\alpha\neq\beta} (C_{\alpha\beta} - C)\, n_\alpha n_\beta. \end{aligned} \tag{C.8}$$

Assuming that the second term is negligible give the result for C after selection as

$$1 - \langle C \rangle_s = \frac{P - \langle n^2 \rangle}{P - 1} (1 - C) \tag{C.9}$$

where $\langle n^2 \rangle$ is the variance of the selection scheme as calculated in chapter four.

Appendix D

Linkage Equilibrium

The definition of correlation is

$$q = \frac{1}{P\,(P-1)} \sum_{\alpha \neq \beta} \frac{1}{L} \sum_{i=1}^{L} S_i^{\alpha} S_i^{\beta}. \tag{D.1}$$

The population variance can be expanded in terms of the bit sums as

$$\begin{aligned}
\kappa_2 &= \frac{1}{P} \sum_{\alpha=1}^{P} F_{\alpha}^2 - \left(\sum_{\alpha=1}^{P} F_{\alpha} \right)^2 \\
&= \left(\frac{1}{P} - \frac{1}{P^2} \right) \sum_{\alpha=1}^{P} F_{\alpha}^2 - \frac{1}{P^2} \sum_{\alpha \neq \beta} F_{\alpha} F_{\beta} \\
&= \left(\frac{1}{P} - \frac{1}{P^2} \right) \sum_{\alpha=1}^{P} \sum_{i=1}^{L} S_i^{\alpha} \sum_{j=1}^{L} S_j^{\alpha} - \frac{1}{P^2} \sum_{\alpha \neq \beta} \sum_{i=1}^{L} S_i^{\alpha} \sum_{j=1}^{L} S_j^{\beta} \\
&= \left(\frac{1}{P} - \frac{1}{P^2} \right) \sum_{\alpha=1}^{P} \left[\sum_{i=1}^{L} S_i^{\alpha} S_i^{\alpha} + \sum_{i \neq j} S_i^{\alpha} S_j^{\alpha} \right] . \\
&\qquad\qquad - \frac{1}{P^2} \sum_{\alpha \neq \beta} \left[\sum_{i=1}^{L} S_i^{\alpha} S_i^{\beta} + \sum_{i \neq j} S_i^{\alpha} S_j^{\beta} \right] .
\end{aligned} \tag{D.2}$$

Under the assumption of linkage equilibrium, $\langle S_i^{\alpha} S_j^{\alpha} \rangle = \langle S_i^{\alpha} S_j^{\beta} \rangle$. Thus two terms in the expression cancel and after averaging over the binary string gives

$$\kappa_2 = \left(\frac{1}{P} - \frac{1}{P^2} \right) \sum_{\alpha=1}^{P} L \langle S^2 \rangle - \frac{1}{P^2} \sum_{\alpha \neq \beta} \sum_{i=1}^{L} S_i^{\alpha} S_i^{\beta}. \tag{D.3}$$

Substituting in from equation (D.1) gives

$$\kappa_2 = \left(1 - \frac{1}{P}\right) L\,(1 - q)\,. \tag{D.4}$$

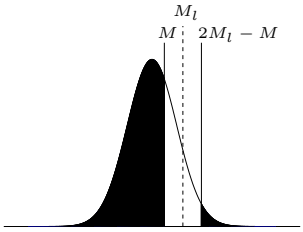
Using the finite population effect, the ensemble variance is thus

$$K_2 = L\,(1 - q)\,. \tag{D.5}$$

Appendix E

Stabilising Selection

When the optimum is introduced to stabilising selection, the ranking and thus the expected number of times an individual is selected, depends on its position within the population and distance from the optimum.



Algebraically it is

$$n_M = \begin{cases} 2 - 2\left[\Phi\left(2M_l - M\right) - \Phi\left(M\right)\right] & \text{when } M \leq M_l \\ 2 - 2\left[\Phi\left(M\right) - \Phi\left(2M_l - M\right)\right] & \text{when } M \geq M_l \end{cases}$$

where $\Phi\left(x\right)$ represents the integral of a unit Gaussian from negative infinity to x. The first and second moments are calculated by integrating the weighting over the population distribution. A unit Gaussian of zero mean is assumed initially. The

results is

$$\begin{aligned}\langle M\rangle &= 2\int_{-\infty}^{\infty} MDM - 2\int_{-\infty}^{M_l} M\left[\Phi\left(2M_l - M\right) - \Phi\left(M\right)\right]DM \\ &\qquad - 2\int_{M_l}^{\infty} M\left[\Phi\left(M\right) - \Phi\left(2M_l - M\right)\right]DM \\ &= \frac{\mathrm{erf}\left(M_l\right)}{\sqrt{\pi}}\end{aligned}$$

$$\begin{aligned}\langle M^2\rangle &= 2\int_{-\infty}^{\infty} M^2DM - 2\int_{-\infty}^{M_l} M^2\left[\Phi\left(2M_l - M\right) - \Phi\left(M\right)\right]DM \\ &\qquad - 2\int_{M_l}^{\infty} M^2\left[\Phi\left(M\right) - \Phi\left(2M_l - M\right)\right]DM \\ &= 1 - \frac{2\mathrm{e}^{-M_l^2}}{\pi}\end{aligned} \tag{E.1}$$

where the term $\mathrm{erf}\left(M_l\right)$ refers to the standard error function which can be translated into this notation as

$$\mathrm{erf}\left(M_l\right) = 2\Phi\left(M_l\sqrt{2}\right) - 1. \tag{E.2}$$

The first and second cumulant after selection are thus

$$\begin{aligned}K_1 &= \langle M\rangle = \frac{\mathrm{erf}\left(M_l\right)}{\sqrt{\pi}} \\ K_2 &= \langle M^2\rangle - \langle M\rangle^2 = 1 - \frac{\mathrm{erf}^2\left(M_l\right)}{\pi} - \frac{2\mathrm{e}^{-M_l^2}}{\pi}.\end{aligned} \tag{E.3}$$

Extending this to a Gaussian of mean K_1 and variance K_2 is straightforward

$$K_1 = K_1 + \sqrt{\frac{K_2}{\pi}}\mathrm{erf}\left(\frac{M_l - K_1}{\sqrt{K_2}}\right)$$

$$\begin{aligned}K_2 = \Bigg[1 - \frac{2}{\pi}\exp\left(-\frac{\left(M_l - K_1\right)^2}{K_2}\right) \\ -\frac{1}{\pi}\mathrm{erf}^2\left(\frac{M_l - K_1}{\sqrt{K_2}}\right)\Bigg] K_2.\end{aligned} \tag{E.4}$$

Appendix F

First Passage Time

If the probability of some event in a single time step is ϵ, the expected number of time steps to observe that event is given by

$$\langle t \rangle = \sum_{n=1}^{\infty} n\epsilon \left(1-\epsilon\right)^{n-1}. \tag{F.1}$$

Rearranging the expression in the summation gives

$$\langle t \rangle = -\epsilon \frac{\delta}{\delta \epsilon} \sum_{n=1}^{\infty} \left(1-\epsilon\right)^{n}. \tag{F.2}$$

The summation is a standard one if epsilon is small and gives the result

$$\langle t \rangle = -\epsilon \frac{\delta}{\delta \epsilon} \left(\frac{1}{\epsilon} - 1 \right). \tag{F.3}$$

Performing the differentiation gives the final result

$$\langle t \rangle = 1/\epsilon. \tag{F.4}$$

Bibliography

[1] J. Baker. Adaptive selection methods for genetic algorithms. In J.J. Grefenstette, editor, *Proceedings of the First International Conference on Genetic Algorithms*, pages 101–111. Erlbaum, 1985.

[2] J. Baker. Reducing bias and inefficiency in the selection algorithm. In J.J. Grefenstette, editor, *Proceedings of the Second International Conference on Genetic Algorithms*. Erlbaum, 1987.

[3] T. Blickle and L. Theile. A mathematical analysis of tournament selection. In L. Eshelman, editor, *Proceedings of the Sixth International Conference on Genetic Algorithms*, pages 9–16. Erlbaum, 1995.

[4] Y. Davidor. Epistasis variance: A viewpoint on ga-hardness. In G.J.E. Rawlins, editor, *Foundations of Genetic Algorithms*, pages 23–35. Morgan Kaufmann, 1991.

[5] K. A. De Jong. *An Analysis of the Behaviour of a Class of Genetic Adaptive Systems*. PhD thesis, University of Michigan, 1975.

[6] K. DeJong and J. Sarma. Generation Gaps Revisited. In L. Darrel Whitley, editor, *Foundations of Genetic Algorithms 2*, pages 19–28, San Mateo, 1993. Morgan Kaufmann.

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

[8] L. Eshelman. The CHC Adaptive Search Algorithm: How to Have Safe Search When Engaging in Nontraditional Genetic Recombination. In G. Rawlins, editor, *Foundations of Genetic Algorithms 1*, pages 265–283, San Mateo, 1991. Morgan Kaufmann.

[9] W. J. Ewens. *Mathematical Population Genetics*. Springer-Verlag, 1979.

[10] R. A. Fisher. *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford, 1930.

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

[12] T. Jones and S. Forrest. Fitness distance correlation as a measure of problem difficulty in genetic algorithms. In L. Eshelman, editor, *Proceedings of the Sixth International Conference on Genetic Algorithms*, pages 184–192. Morgan Kaufmann, 1995.

[13] M. Kimura. *Diffusion Models in Population Genetics*. Applied Probability. Methuen's Review Series in Applied Probability, 1964.

[14] S. Kirkpatrick, C.D. Gelatt, and M.D. Vecchi. Optimisation by simulated annealing. *Science*, 220(4598):671–680, 1983.

[15] A. S. Kondrashov. Deleterious mutations and the evolution of sexual reproduction. *Nature*, 336(6198):435–440, 1988.

[16] D. Mackay. Rate of information acquisition by a species subjected to natural selection. Available on the world wide web at `http://wol.ra.phy.cam.ac.uk/mackay/abstracts/gene.html`, 1999.

[17] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[18] M. Mitchell, J. Holland, and S. Forrest. When Will a Genetic Algorithm Outperform Hill Climbing? In J. Cowan, G. Tesauro, and J. Alspector, editors, *Advances in Neural Information Processing Systems*, pages 51–58, San Francisco, CA., 1994. Morgan Kauffman.

[19] P. A. P. Moran. A general theory of the distribution of gene frequencies I. Non-overlapping generations. *Proceedings of Royal Society B*, 149:113–116, 1958.

[20] P. A. P. Moran. A general theory of the distribution of gene frequencies I. Overlapping generations. *Proceedings of Royal Society B*, 149:102–112, 1958.

[21] P. A. P. Moran. Random Processes in Genetics. *Proceedings of the Cambridge Philosophical Society*, 54:60–71, 1958.

[22] P.A.P. Moran. *An Introduction to Probability Thoery*. Clarendon Press, Oxford, 1968.

[23] H. Mühlenbein. Parallel genetic algorithms, population genetics and combinatorial optimization. In *Proceedings of the Third International Conference on Genetic Algorithms*, pages 416–421. Morgan Kaufmann (San Mateo), 1989.

[24] H. Mühlenbein. Genetic Algorithms. In E. Aarts and J. K. Lenstra, editors, *Local Search in Combinatorial Optimization*, pages 137–171, New York, 1997. Jon Wiley and Sons.

[25] H. Mühlenbein and Th. Mahnig. FDA - A Scalable Evolutionary Algorithm for the Optimization of Additively Decomposed Functions. *Evolutionary Computation*, 7(4), 1999.

[26] 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.

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

[28] A. Prügel-Bennett. Modelling Evolving Populations. *J. Theor. Biol.*, 185:81–95, 1997.

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

[30] A. Prügel-Bennett and J. L. Shapiro. The Dynamics of a Genetic Algorithm for Simple Random Ising Systems. *Physica D*, 104:75–114, 1997.

[31] M. Rattray. The Dynamics of a Genetic Algorithm under Stabilizing Selection. *Complex Systems*, 9(3):213–234, 1995.

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

[33] M. Rattray and J. L. Shapiro. Cumulant Dynamics in a Finite Population: Linkage Equilibrium Theory. In preparation, 2000.

[34] A. Rogers and A. Prügel-Bennett. Genetic Drift in Genetic Algorithm Selection Schemes. *IEEE Transactions on Evolutionary Computation*, 3(4):298–303, 1999.

[35] 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, 1999. Morgan Kaufmann.

[36] A. Rogers and A. Prügel-Bennett. A Solvable Model of a Hard Optimisation Problem. In L. Kallel, B. Naudts, and A. Rogers, editors, *The Second Evonet Summer School on Theoretical Aspects of Evolutionary Computing*, Heidelberg, 2000. Springer-Verlag.

[37] A. Rogers and A. Prügel-Bennett. Evolving Populations with Overlapping Generations. *Theoretical Population Biology*, 57(2):121–129, 2000.

[38] A. Rogers and A. Prügel-Bennett. Sexual and Asexual Populations in Stabilising Selection-Mutation Balance. *Genetics*, 2000. Submitted.

[39] A. Rogers and A. Prügel-Bennett. The Dynamics of a Genetic Algorithm on a Model Hard Optimization Problem. *Complex Systems*, 11(6):437–464, 2000.

[40] J. Schaffer, M. Mani, L. Eshelman, and K. Mathias. The Effect of Incest Prevention on Genetic Drift. In W. Banzhaf and C. Reeves, editors, *Foundations of Genetic Algorithms 5*, San Francisco, 1999. Morgan Kaufmann.

[41] J. L. Shapiro and A. Prügel-Bennett. Maximum Entropy Analysis of Genetic Algorithm Operators. In T. C. Fogarty, editor, *Lecture Notes in Computer Science* **993**, pages 14–24, Berlin, 1995. Springer-Verlag.

[42] 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.

[43] J. L. Shapiro, 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.

[44] C. Stephens, H Walbroeck, and R. Aguirre. Schemata as Building Blocks: Does Size Matter? In W. Banzhaf and C. Reeves, editors, *Foundations of Genetic Algorithms 5*, San Francisco, 1999. Morgan Kaufmann.

[45] G. Syswerda. Uniform crossover in genetic algorithms. In *Proceedings of the Third International Conference on Genetic Algorithms*. Morgan Kaufmann (San Mateo), 1989.

[46] G. Syswerda. A Study of Reproduction in Generational and Steady-State Genetic Algorithms. In L. Darrel Whitley, editor, *Foundations of Genetic Algorithms*, pages 19–28, San Mateo, 1991. Morgan Kaufmann.

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

[48] E. Van Nimvegen and J. Crutchfield. Optimizing Epochal Evolutionary Search: Population-Size Dependent Theory. Technical report, Santa Fe Institute, 1998.

[49] E. Van Nimvegen and J. Crutchfield. Optimizing Epochal Evolutionary Search: Population-Size Independent Theory. Technical report, Santa Fe Institute, 1998.

[50] M. D. Vose. Generalizing the notion of schema in genetic algorithms. *Artificial Intelligence*, 1991.

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

[52] M. D. Vose. *The Simple Genetic Algorithm: Foundations and Theory*. MIT Press, 1999.

[53] M. D. Vose and G. E. Liepins. Punctuated equilibria in genetic search. *Complex Systems*, 5:31–44, 1991.

[54] S. Wright. *Evolution and the Genetics of Populations*. Chicago University Press, 1968.