Finite Population Effects for Ranking and Tournament Selection

Adam Prügel-Bennett
ECS, University of Southampton, SO17 1BJ, United Kingdom

The effect of ranking and binary tournament selection on the distribution of fitnesses and genetic correlation is calculated for a finite population. The results are different for continuous and discrete fitness functions. Results for both situations are obtained. These exact results are compared to a previously obtained approximation up to the third cumulant and shown to be in good agreement. Tournament and ranking selection are compared with Boltzmann selection for which exact finite population effects are already known.

1. Introduction

One of the important practical and intellectual challenges facing genetic algorithm researchers is to put our present heuristic understanding on a more secure theoretical basis. One approach to this has been to model the dynamics of genetic algorithms using tools from statistical mechanics [1–3]. In this formalism the state of the system is modelled by a number of macroscopic variables. The dynamics is then found by calculating how the macroscopic variables are changed by operators such as selection, mutation and crossover. The approach is approximate as the macroscopic variables are not sufficient to describe all the details of the system. However, in practice, it has been found that this approach can give a very accurate description of the dynamics. More significantly, the model can capture the most important features of a genetic algorithm, thus revealing how various operators contribute to the search. Although, our goal in modelling is to obtain an understanding, rather than to be accurate, it is, nevertheless important to understand the sources of errors in our models as this may reveal some new process at work. In this paper, we focus on the role of selection in changing the distribution in the fitness of members of the population.

Binary tournament selection and ranking selection are among the most popular selection schemes used in genetic algorithms. It is well known that binary tournament selection and ranking selection (using roulette wheel selection) give the same results on average. Their effect on an infinite population, starting from a Gaussian, was calculated by Blickle and Theile [4]. More recently an approximation valid for finite populations was given by Rogers and Prügel-Bennett [5,6]. This allowed comparison of different selection strategies, such as, steady-state, generation gap [7], CHC models [8], evolutionary strategies and stochastic universal sampling [9]. However, the approximation, although physically plausible, is not systematic. That is, there are no bounds on the errors, nor can we obtain higher order corrections.

In this paper we calculate the effect of ranking (or tournament) selection for finite populations exactly. The derivation depends on whether the fitnesses are continuous valued or discrete. We calculate the corrections in both cases and show that there is a term by term correspondence between the two cases. We briefly re-derive the approximation given in [5, 6] and extend it to include the third cumulant. For the first cumulant (*i.e.* the average fitness) the approximation agrees with the exact result. The effect of selection on the second cumulant (*i.e.* the variance) and the third cumulant starting from a Gaussian are shown in figure 1. The approximation is seen to capture the qualitative behaviour of the exact results even for very small populations. For any reasonable sized population the approximation is within a few percent of the exact result.

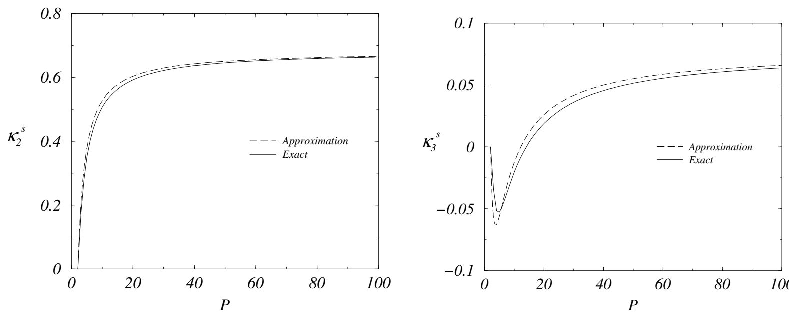

Figure 1. The exact and approximate second ($\bar{\kappa}_2^s$) and third cumulant ($\bar{\kappa}_3^2$) are shown versus P starting from a zero mean unit Gaussian.

As well as changing the distribution of fitnesses, selection correlates the population. We can compute the effect of selection on the correlation directly. This provides another perspective into the effect of selection. We conclude the paper with a brief comparison between ranking selection and Boltzmann selection for which the finite population corrections have been calculated previously [2, 3].

The paper is organized as follows. Section 2 sets up the framework for discussing the effect of selection. Section 3 presents the exact results for continuous and discrete fitness functions. The approximation scheme developed in reference [5, 6] is presented in section 4 and compared with the exact results. In section 5 we discuss the correlation produced by selection. Section 6 gives a comparison of ranking and Boltzmann selection. Conclusions are given is section 7.

2. Effects of Selection

Before going into the calculation, we must decide the form of the answer we seek. Selection depends only on the fitness of the member of the population and not on the details of the string they represent. We will concern ourselves mainly with what happens to the fitnesses of the members in the population, although in section 5 we discuss correlations. For convenience we assume that an initial population is drawn from a fitness distribution $\rho(F)$. The distribution can be described by a set of statistics such as it mean and variance. For a more accurate model we can include higher order statistics such as the third cumulant (related to the skewness) and the fourth cumulant (related to the kurtosis). Our aim is to calculate the statistical properties of the population after selection.

For an infinite population selection is deterministic and it is relatively straightforward to calculate the statistics of the distribution of fitnesses after selection. However, in a finite population, selection will be stochastic so the effect of selection is not so simple to describe. To incorporate this stochasticity into our description we consider an ensemble of populations. It simplifies the analysis to describe the populations in terms of an infinite population distribution from which a finite population is drawn. We can envision the selection process as a four stage process shown in figure 2.

These stages are

1. start from some distribution $\rho(F)$
2. draw a finite population at random from $\rho(F)$
3. a new finite population is drawn from the finite distribution
4. to complete the cycle return to a distribution $\rho^s(F)$, such that a randomly drawn sample from $\rho^s(F)$ would have the same statistics on average as the ensemble of finite populations in stage 3.

There are two sources of stochasticity in this process. The first comes in drawing a finite population from $\rho(F)$. As each sample will typically contain a different set of fitnesses we would expect different results for each sample. The second cause of stochasticity comes from selection — each time we perform selection we obtain a slightly different result. We can eliminate this second source of stochasticity by selecting an infinite population. This is equivalent to going straight from stage 2 to stage 4. If we wish to model the finite population after selection we can draw at random a finite population from $\rho^s(F)$ (that is, we can go back from

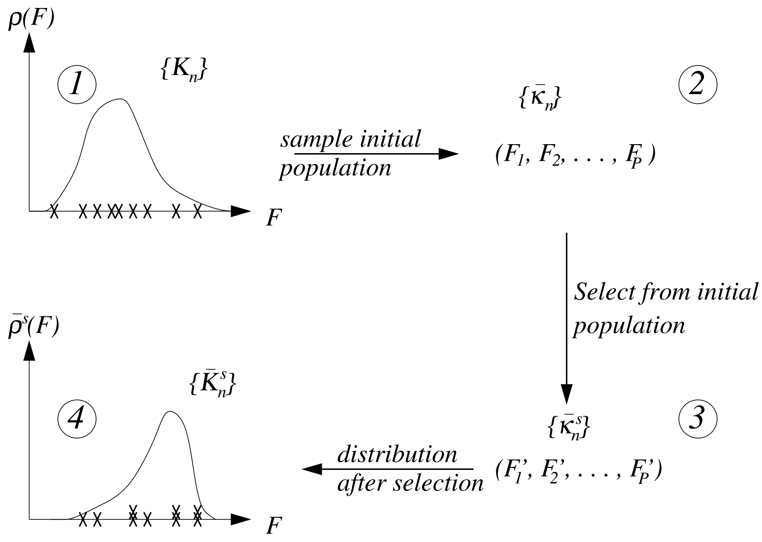

Figure 2. Schematic overview of the selection procedure. Starting from an initial distribution $\rho(F)$ we draw a finite populations at random. A new population is selected from this finite population using ranking or tournament selection. We finally go back to a distribution $\rho^s(F)$ which describes the ensemble of selected populations.

stage 4 to stage 3). We will show explicitly that we can go to stage 4 directly from stage 2 or via stage 3. For every finite population at stage 2 we have a unique distribution at stage 4, but we can have many different finite populations at stage 3. In this scheme the real source of stochasticity is in going from stage 1 to stage 2, that is in drawing a random sample from the distribution $\rho(F)$. Because of this, there will be many different distributions at stage 4 corresponding to the different finite population we draw as stage 2. If we wish to accurately model selection we must describe this ensemble of distributions. In this paper we will just consider the ensemble average of the first few cumulants of $\rho^s(F)$. A more complete analysis including the ensemble covariance has been carried out for Boltzmann selection in [3].

We denote the the cumulants of the original distribution, $\rho(F)$, by K_n and those of the distribution after selection by K_n^s. We need also to distinguish between statistics describing a finite population and those describing a distribution from which a sample is drawn. We denote the cumulants describing the finite population by κ_n. Thus

$$\kappa_1 = \frac{1}{P} \sum_{\alpha=1}^{P} F_\alpha$$

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

etc. On average a finite population drawn from a probability distribution has different cumulants than the probability distribution. The average first four cumulants for a sample of P individuals is

$$\bar{\kappa}_1(P) = K_1 \quad (1)$$

$$\bar{\kappa}_2(P) = \mathcal{P}_2 K_2 \quad (2)$$

$$\bar{\kappa}_3(P) = \mathcal{P}_3 K_3 \quad (3)$$

$$\bar{\kappa}_4(P) = \mathcal{P}_4 K_4 - \frac{6}{P}\mathcal{P}_2 K_2^2 \quad (4)$$

where $\mathcal{P}_2 = (1-1/P)$, $\mathcal{P}_3 = \mathcal{P}_2(1-2/P)$ and $\mathcal{P}_4 = \mathcal{P}_2(1-6/P+6/P^2)$. We use the bar to denote the sample average. Inverting these equations we can obtain unbiased estimates for the cumulants of a distribution from a sample. These unbiased estimates are Fisher's k-statistics [10]. (If we are to treat the fourth cumulant we must also include information on the ensemble covariance, since the average cumulants for a finite population are no longer linear in the distribution cumulants. More details can be found in [3]. In the present paper we restrict our attention to the first three cumulants.)

We will focus our investigation on ranking selection where we assign a rank of 0 to the least fit member of the population, a rank 1 to the next least fit member of the population, *etc.* The fittest member of the population will therefore be assigned a rank $P-1$. An individual α is selected with a probability

$$p_\alpha = \frac{2r_\alpha}{P(P-1)} \quad (5)$$

where r_α is the rank of member α. If some members have the same fitness we have a choice of assigning an arbitrary ranking order or else assigning them an average group rank. This choice will affect how often we select individuals with the same fitness, but will not affect the distribution of fitnesses after selection as individuals with the same fitness are identical as far as the distribution of fitness is concerned. We will consider roulette wheel selection where we draw the new population *with replacement* from the old populations. A superior selection method, leading to less fluctuations, is stochastic universal sampling [9]. This has been modelled using the approximation scheme of reference [6]. That paper also extends the analysis to the case where $p_\alpha \propto r_\alpha + c$, where c is a constant that controls the strength of selection.

Binary tournament selection can be regarded as a mechanism for achieving (roulette wheel) ranking selection. In binary tournament selection we randomly pick two different members of the population and select the fitter one. The probability of selecting a member α will be proportional to the number of individuals less fit than it, but this is just its rank r_α. Thus binary tournament selection gives raise to exactly the same probability of selection as in equation (5) for ranking selection.

3. Selection in a Finite Population

In this section we present the exact result of ranking selection on the first few cumulants. We start this section with a reminder of the infinite population result of Blickle and Theile [4] (their results are extended to include the third cumulant). We then discuss sampling corrections due to the randomness in the selection procedure. Finally we have to calculate the sampling corrections due to drawing the initial population from a distribution $\rho(F)$. The calculations differ for continuous and discrete fitness functions. We present the continuous case first and derive the exact results for the first three cumulants. For the discrete case we have calculated only the first two cumulants. Note that we are calculating the stochastic effects in the reverse order to how they occur in figure 2.

Infinite Population

In an infinite population, starting from a population with a fitness distribution $\rho(F)$, the distribution after selection would be given by

$$\rho^{\infty s}(F) = 2\rho(F)R(F) \tag{6}$$

where

$$R(F) = \int_{-\infty}^{F} \rho(F)\,\mathrm{d}F. \tag{7}$$

The cumulants after selection are given by

$$K_n^{\infty s} = \frac{\partial}{\partial\gamma^n}\log\left(\int \mathrm{e}^{\gamma F}\rho^{\infty s}(F)\,\mathrm{d}F\right)\bigg|_{\gamma=0}. \tag{8}$$

For a general distribution we would have to perform the integral numerically. If we start from a distribution where we know only the first M cumulants we need to model the distribution. One possible model is provided by the Gram-Charlier expansion

$$\rho(E) \approx \frac{\mathrm{e}^{-(E-K_1)^2/2K_2}}{\sqrt{2\pi\,K_2}}\left(1+\sum_{n=3}^{M}\frac{K_n}{K_2^{n/2}}\,h_n\left(\frac{E-K_1}{\sqrt{K_2}}\right)\right) \tag{9}$$

where $h_3(x) = (x^3 - 3x)/3!$, $h_4(x) = (x^4 - 6x^2 + 3)/4!$, *etc.* are related to the Hermite polynomials. They can be generated from the recurrence relation

$$h_n(x) = \frac{1}{n}(xh_{n-1}(x) - h_{n-2}(x))$$

with $h_0(x) = 1$ and $h_1(x) = x$. Note that for this expansion the first M cumulants are correct but the function is not guaranteed to be positive everywhere. Using a Gram-Charlier expansion with three cumulants we find

$$K_1^{\infty s} = K_1 + \sqrt{\frac{K_2}{\pi}}\left(1 - \frac{s^2}{96}\right) \tag{10}$$

$$K_2^{\infty s} = K_2\left(1 - \frac{1}{\pi} + \frac{s}{2\sqrt{\pi}} + \frac{s^2}{48\pi} - \frac{s^4}{9216\pi}\right) \tag{11}$$

$$\begin{aligned} K_3^{\infty s} = K_2^{3/2}\Bigg(&\frac{1}{2\sqrt{\pi}}\left(\frac{4}{\pi} - 1\right) + s\left(1 - \frac{3}{2\pi}\right) - \frac{s^2}{192\sqrt{\pi}}\left(\frac{12}{\pi} + 1\right) \\ &+ \frac{s^3}{64\pi} + \frac{s^4}{1536\pi^{3/2}} - \frac{s^6}{442368\pi^{3/2}}\Bigg) \end{aligned} \tag{12}$$

where $s = K_3/K_2^{3/2}$ is the skewness of the distribution. The cumulants for a Gaussian are obtained by setting $s = 0$.

One of the reasons why the statistical mechanics approach is approximate is that we do not keep sufficient statistics to describe the distribution of fitnesses exactly. Modelling the dynamics using a truncated set of cumulants is usually a good approximation provided $K_n/K_2^{n/2}$ is small for the higher cumulants. Crossover has the effect of reducing the higher cumulants, thus making the modelling process easier.

Fluctuations in Selection

We now consider the effect of selection on a finite population. The cumulants for a finite population after selection are given by

$$\kappa_1^s = \frac{1}{P}\sum_{\alpha=1}^{P} n_\alpha F_\alpha \tag{13}$$

$$\kappa_2^s = \frac{1}{P}\sum_{\alpha=1}^{P} n_\alpha F_\alpha^2 - \left(\frac{1}{P}\sum_{\alpha=1}^{P} n_\alpha F_\alpha\right)^2 \tag{14}$$

$$\kappa_3^s = \frac{1}{P}\sum_{\alpha=1}^{P} n_\alpha F_\alpha^3 - \frac{3}{P^2}\sum_{\alpha,\beta=1}^{P} n_\alpha n_\beta F_\alpha^2 F_\beta \tag{15}$$

$$+\frac{2}{P^3}\sum_{\alpha,\beta,\gamma=1}^{P} n_\alpha n_\beta n_\gamma F_\alpha F_\beta F_\gamma$$

etc., where n_α is the number of times we select member α. We assume roulette wheel sampling (or equivalently binary tournament selection) so that the probability of selecting n_1 samples of member 1, n_2 samples of member 2 *etc.* is give by the multinomial probability distribution

$$P(n_1, n_2, \ldots, n_P) = P! \prod_{\alpha=1}^{P} \frac{p_\alpha^{n_\alpha}}{n_\alpha !} \left[\sum_{\alpha=1}^{P} n_\alpha = P\right]$$

where we have used the convention

$$[predicate] = \begin{cases} 1 & \text{if } predicate \text{ is true} \\ 0 & \text{if } predicate \text{ is false} \end{cases}$$

to denote an indicator function (in this case the indicator function is a Kronecker delta). We can average over all ways of performing selection (we denote this average by $\overline{\cdots}$). For a multinomial distribution we find

$$\begin{aligned} \overline{n_\alpha} &= P p_\alpha \\ \overline{n_\alpha\, n_\beta} &= P(P-1) p_\alpha\, p_\beta + P\, p_\alpha\, [\alpha = \beta] \end{aligned}$$

etc., where $\langle \cdots \rangle$ denotes the average over all ways of performing selection. Using this we find

$$\overline{\kappa_1^s} = \sum_{\alpha=1}^{P} p_\alpha F_\alpha \tag{16}$$

$$\overline{\kappa_2^s} = \mathcal{P}_2 \left(\sum_{\alpha=1}^{P} (p_\alpha - p_\alpha^2) F_\alpha^2 - \sum_{\alpha\neq\beta} p_\alpha\, p_\beta\, F_\alpha\, F_\beta \right) \tag{17}$$

$$\begin{aligned} \overline{\kappa_3^s} = \mathcal{P}_3 \Bigg(&\sum_{\alpha=1}^{P} (p_\alpha - 3p_\alpha^2 + 2p_\alpha^2) F_\alpha^3 - 3 \sum_{\alpha\neq\beta} (p_\alpha\, p_\beta - 2p_\alpha^2\, p_\beta) F_\alpha^2\, F_\beta \\ &+ 2 \sum_{\alpha\neq\beta} p_\alpha\, p_\beta\, p_\gamma\, F_\alpha\, F_\beta F_\gamma \Bigg) \end{aligned} \tag{18}$$

where $\mathcal{P}_2$ and $\mathcal{P}_3$ are the finite population correction factors given above. From these cumulants for the finite population we can calculate k-statistics for a distribution $\rho^s(F)$. These are just $\langle K_1^s \rangle = \langle \kappa_1^s \rangle$, $\langle K_2^s \rangle = \langle \kappa_2^s \rangle / \mathcal{P}_2$ and $\langle K_3^s \rangle = \langle \kappa_3^s \rangle / \mathcal{P}_3$. We could have arrived at these results directly from equations (13)–(15). If we select an infinite population then the number of times, n_α, we select α would be $P p_\alpha$, substituting this into equations (13)–(15) we would have obtained equations (16)–(18) except without the finite population corrections $\mathcal{P}_2$, $\mathcal{P}_3$, *etc.* Referring to figure 2, this illustrates the direct way to go from stage 2 to stage 4.

So far we have calculated the result of selection starting from a particular population. We have now to average over all ways of drawing a finite population from a distribution $\rho(F)$, with cumulants K_n. This calculation depends on whether $\rho(F)$ represents a continuous or discrete distribution. We first present the case when $\rho(F)$ is continuous.

Sampling Fluctuation in a Continuous Distribution

We begin by relabelling the members of the population according to their rank. Thus the first three cumulants are given by

$$\overline{K_1^s} = \frac{1}{Z}\sum_{r=0}^{P-1} r F_r \tag{19}$$

$$\overline{K_2^s} = \sum_{r=0}^{P-1}\left(\frac{r}{Z} - \frac{r^2}{Z^2}\right) F_r^2 - \frac{1}{Z^2}\sum_{r\neq r'} r\, r'\, F_r\, F_{r'} \tag{20}$$

$$\begin{aligned}\overline{K_3^s} = \sum_{r=0}^{P-1}\left(\frac{r}{Z} - \frac{3r^2}{Z^2} + \frac{2r^3}{Z^2}\right) F_r^3 - 3\sum_{r\neq r'}\left(\frac{r\, r'}{Z^2} - \frac{2r^2 r'}{Z^3}\right) F_r^2\, F_{r'} \\ + \frac{2}{Z^3}\sum_{r\neq r'\neq r''} r\, r'\, r''\, F_r\, F_{r'}\, F_{r''}\end{aligned} \tag{21}$$

where F_r is the fitness of the rank r member of the population and $Z = P(P-1)/2$ is a normalization constant. We now average over always of drawing a finite sample from $\rho(F)$ (we denote this average by $\langle\rangle \cdots)$. The sample average for the first cumulant is therefore given by

$$\bar{K}_1^s = \langle \overline{K_1^s} \rangle = \frac{1}{Z}\sum_{r=0}^{P-1} r\,\langle F_r \rangle \tag{22}$$

where

$$\langle F_r \rangle = \int_{-\infty}^{\infty} F p_r(F)\,\mathrm{d}F. \tag{23}$$

For a finite population the fitness distribution of the r^{th} ranked individual, drawn from a population $\rho(F)$, is given by

$$p_r(F)\,\mathrm{d}F = P\binom{P-1}{r} R^r(F)\left(1 - R(F)\right)^{P-r-1}\,\mathrm{d}R(F) \tag{24}$$

where $R(F)$ is defined in equation (7), so that $\mathrm{d}R(F) = \rho(F)\,\mathrm{d}F$. The first factor of P arises because the r^{th} ranked individual may be any member of the population. The remaining members of the population consist of r with $F \leq F_r$ and $P-1-r$ with $F \geq F_r$. The number of ways of dividing the $P-1$ remainder into the two groups is given by the binomial coefficient $(P-1)!/(r!(P-1-r)!)$. Finally, $R^r(F)(1-R(F))^{P-r-1}\rho(F)\,\mathrm{d}F$ gives the probability that the r-th ranked individual has fitness F. It is simple to check that $p_r(F)$ is properly normalized when averaging over all F and over all r. In figure 3 we show $p_r(F)$ for a population of size 4, starting from a Gaussian with zero mean and unit variance.

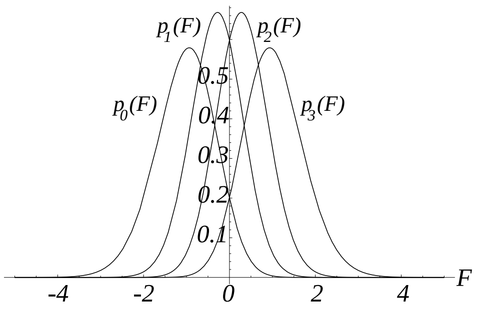

Figure 3. Probability distributions for 4 samples drawn from a Gaussian distribution with mean zero and variance 1. $p_r(F)$ shows the distribution for the r^{th} ranked sample.

From equation (22) and (23) we find

$$\bar{K}_1^s = \frac{2}{P(P-1)} \int_{-\infty}^{\infty} F \sum_{r=0}^{P-1} r p_r(F)\,\mathrm{d}F \tag{25}$$

where we have inverted the order of integration and summation. Using equation (24) we find

$$\sum_{r=0}^{P-1} r\, p_r(F) = P(P-1)\rho(F)R(F).$$

Thus equation (25) reduces to

$$\bar{K}_1^s = \int_{-\infty}^{\infty} F \rho^{\infty s}(F)\,\mathrm{d}F \tag{26}$$

where

$$\rho^{\infty s}(F) = 2\rho(F)\,R(F) \tag{27}$$

is the probability distribution after selection for an infinite population (*c.f.* equation (6)). The average effect of selection on the first cumulant (average fitness) is exactly the same for a finite population as it is for an infinite population.

The calculation for higher order cumulants follows similarly, although in this case we obtain a finite population correction. The second cumulant is given by

$$\bar{K}_2^s = \sum_{r=0}^{P-1} \left(\frac{r}{Z} - \frac{r^2}{Z^2} \right) \langle F_r^2 \rangle - \frac{1}{Z^2} \sum_{r \neq r'} r\, r' \, \langle F_r \, F_{r'} \rangle \,. \tag{28}$$

The term proportional to r is similar to what we have already calculated. The term proportional to r^2 involves the sum

$$\sum_{r=0}^{P-1} r^2 p_r(F) = \rho(F) \left(P^{\underline{3}} R^2(F) + P^{\underline{2}} R(F) \right) .$$

where we have use the notation

$$P^{\underline{n}} = P(P-1) \cdots (P-n+1) = \frac{P!}{(P-n)!}. \tag{29}$$

The first term in equations (28) gives

$$\sum_{r=0}^{P-1} \left(\frac{r}{Z} - \frac{r^2}{Z^2} \right) \langle F_r^2 \rangle = I_{2,1} - \frac{2(P-2) I_{2,2} + 2 I_{2,1}}{P(P-1)}$$

where we define the integral

$$I_{n,m} = 2 \int_{-\infty}^{\infty} F^n \, R^m(F) \, \rho(F) \, \mathrm{d}F. \tag{30}$$

Using this notation $\bar{K}_1^s = I_{1,1}$. We can write the second term in equation (28) as

$$\frac{1}{Z^2} \sum_{r \neq r'} r\, r' \, \langle F_r \, F_{r'} \rangle = \frac{2}{Z^2} \sum_{r=0}^{P-2} \sum_{s=0}^{P-2-r} r(r+s+1) \, \langle F_r \, F_{r+s+1} \rangle$$

where

$$\langle F_r \, F_{r+s+1} \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_r \, F_{r+s+1} \, p_{r,s}(F_r, F_{r+s+1}) \, \mathrm{d}F_{r+s+1} \, \mathrm{d}F_r.$$

Thus to calculate this term we need the joint probability distribution for the fitnesses of the rank r and rank $r+s+1$ member of the population. This is given by

$$p_{r,s}(F_1, F_2) = P(P-1) \, C_{r,s}(R_1, R_2) \, \rho(F_1) \, \rho(F_2) \, \left[F_1 > F_2 \right]$$

where we use the shorthand $R_n = R(F_n)$ and

$$C_{r,s}(R_1, R_2) = \frac{(P-2)! R_1^r (R_2 - R_1)^s (1-R_2)^{P-2-r-s}}{r! s! (P-2-r-s)!}.$$

Straightforward but tedious algebra gives

$$\sum_{r \neq r'} r\, r' \langle F_r\, F_{r'} \rangle = P^{\underline{4}} I_{1,1;1,1} + 2P^{\underline{3}} I_{1,1;1,0}$$

where

$$I_{i,j;k,l} = 2 \int_{-\infty}^{\infty} F_1^i\, R^j(F_1) \left(\int_{F_1}^{\infty} F_2^k\, R^l(F_2)\, \rho(F_2)\, \mathrm{d}F_2 \right) \rho(F_1)\, \mathrm{d}F_1$$

By symmetry,

$$I_{k,n;k,n} = \frac{1}{4} I_{k,n} I_{k,n}$$

so that $I_{1,1;1,1} = I_{1,1}^2/4$.

Collecting together terms we find the change in the second cumulant is give by

$$\bar{K}_2^s = I_{2,1} - \frac{P^{\underline{4}}}{(P^{\underline{2}})^2} I_{1,1}^2 - \frac{2P^{\underline{3}}}{(P^{\underline{2}})^2} \left(4I_{1,1;1,0} + I_{2,2}\right) - \frac{2I_{2,1}}{P^{\underline{2}}} \tag{31}$$

For a Gaussian we can perform the integrals exactly. Starting from a Gaussian with variance K_2, the average variance after selection is given by

$$\frac{\bar{K}_2^s}{K_2} = 1 - \frac{1}{\pi} - \left(\frac{7}{3} - \frac{5}{\pi} + \frac{6}{\sqrt{3}\pi}\right)\frac{1}{P} + \left(\frac{2}{3} - \frac{6}{\pi} + \frac{12}{\sqrt{3}\pi}\right)\frac{1}{P^2}. \tag{32}$$

The calculation of the third cumulant follows a similar pattern. The calculation is presented in the the appendix.

Sampling Fluctuation in a Discrete Distribution

The solution above is given in terms of integrals. When the fitness values are discrete these integrals should become sums, but it is not obvious how to do this. One complicating factor is that for discrete fitnesses a finite population will typically contain many individuals of the same fitness. We assume that individuals with the same fitness are assigned an arbitrary ranking. In this section we derive the change in the first two cumulants for discrete fitness functions.

We assume that the initial fitness function, $p(F)$, is given, but with $F \in \mathbb{Z}$. (To be more general we could consider a fitness function with fitnesses F_i, with $i \in \mathbb{Z}$. The generalization is straightforward, but for simplicity we restrict our attention to the case when the fitnesses take on integer values.) A finite population is now drawn from $p(F)$. We denote the number of individuals with fitness F by $n(F)$. The number of individuals with fitness less than F we denote by

$$S(F) = \sum_{F'=-\infty}^{F-1} n(F').$$

The probability of selecting an individual α is given by

$$p_\alpha = \frac{r_\alpha}{Z} = \frac{2}{P(P-1)}(S(F_\alpha) + i_\alpha)$$

where $i_\alpha \in \{0, 1, \cdots, n(F_\alpha) - 1\}$ is the ranking within the group of individuals with fitness F_α. The probability of selecting a member with fitness F is thus given by

$$\begin{aligned} p^s(F) &= \frac{2}{P(P-1)} \sum_{i=0}^{n(F)-1} (S(F) - i) \\ &= \frac{n(F)}{P(P-1)} \left(2S(F) + n(F) - 1\right). \end{aligned} \tag{33}$$

Note that we would obtain the same result if we assigned the average rank to each member of the group.

From equations (19) and (20), summing over members with the same fitness we find the first two cumulants after selection are

$$\overline{K_1^s} = \sum_{F=-\infty}^{\infty} F\, p^s(F)$$

$$\begin{aligned} \overline{K_2^s} = &\sum_{F=-\infty}^{\infty} F^2 \left(p^s(F) - p^s(F)^2\right) \\ &-2 \sum_{F=-\infty}^{\infty} \sum_{F'=F+1}^{\infty} F\, F'\, p^s(F)\, p^s(F'). \end{aligned}$$

So far we have considered a particular finite population. We now average over all ways of drawing a finite populations from a discrete distribution $p(F)$. The average fitness after selection is given by

$$\bar{K}_1^s = \sum_{F=-\infty}^{\infty} F\, \langle p^s(F) \rangle \tag{34}$$

Where $\langle p^s(F)\rangle$ is the average probability for selecting a member of fitness F. From equation (33) we find

$$\langle p^s(F) \rangle = \frac{1}{P^2} \left(2 \langle n(F)\, S(F) \rangle + \left\langle n(F)^2 \right\rangle - \langle n(F) \rangle\right).$$

But the probability of drawing $n(F)$ individuals is given by a multinomial distribution

$$P(\{n(F)\}) = P! \left(\prod_{F=-\infty}^{\infty} \frac{p(F)^{n(F)}}{n(F)!} \right) \left[\sum_F n(F) = P \right]$$

so that

$$\langle n(F) \rangle = P p(F)$$

$$\left\langle n(F)^2 \right\rangle = P^{\underline{2}} p(F)^2 + P p(F)$$

etc. To compute the average $\langle n(F)\, S(F) \rangle$ we expand $S(F)$

$$\langle n(F)\, S(F) \rangle = \sum_{F'=-\infty}^{F-1} \langle n(F)\, n(F') \rangle = P^{\underline{2}} p(F)\, P(F)$$

where

$$P(F) = \sum_{F'=-\infty}^{F-1} p(F').$$

Combining these averages together we find

$$\langle p^s(F) \rangle = p^{\infty s} = p(F) \left(2P(F) + p(F) \right).$$

Substituting this back into equation (34) we find

$$\bar{K}_1^s = \sum_{F=-\infty}^{\infty} F p(F) \left(2P(F) + p(F) \right).$$

This is the infinite population result for a discrete population. As with the continuous case there is no finite population correction in the first cumulant.

The calculation of the second cumulant follows a similar pattern to that of the first cumulant, although it involves many more terms. We find

$$\bar{K}_2^s = S_{2,1} - \frac{P^{\underline{4}}}{(P^{\underline{2}})^2} S_{1,1}^2 - \frac{2P^{\underline{3}}}{(P^{\underline{2}})^2} \left(4S_{1,1;1,0} + S_{2,2} \right) - \frac{2S_{2,1}}{P^{\underline{2}}} \tag{35}$$

where

$$S_{n,1} = \sum_{F=-\infty}^{\infty} F^n p^{\infty s}(F)$$

$$S_{n,2} = \sum_{F=-\infty}^{\infty} F^n p(F) \left(2P(F)^2 + 6p(F)P(F) + 2p(F)^2 \right)$$

$$S_{1,1;1,0} = \sum_{F=-\infty}^{\infty} F\, p^{\infty s}(F) \sum_{F'=F+1}^{\infty} F' p(F')$$

Comparing this with with equation (31) for a continuous fitness distribution we see a term by term correspondence. The terms $S_{n,1}$ and $S_{1,1;1,0}$ are exactly what we would get by replacing the sums by integrals and replacing $R(F)$ by $P(F)+p(F)/2$. The term $S_{n,2}$ is, however, not obtained by a simple discretization of $I_{n,2}$.

4. Approximation Scheme

An approximation for ranking selection has been introduced in [5, 6]. The idea behind the approximation is to model ranking selection as a two stage process.

1. Starting from $\rho(F)$ we generate the distribution

$$\rho^{\infty s}(F) = 2\rho(F) \int_{-\infty}^{F} \rho(F')\,\mathrm{d}F'$$

which would arise from an infinite population.

2. We draw a random sample of P individuals from $\rho^{\infty s}(F)$ and perform tournament or ranking selection on this finite population, but where we assign a random rank to each member of the population.

On average each member of the population will be drawn with a probability proportional to its expected rank. However, errors arise because the fluctuations due to selection in a finite population are assumed to be uncorrelated with the fitness values. The approximation appears physically plausible, but it is uncontrolled.

In contrast to the exact expressions, the approximation is relatively simple to calculate and gives simple correction factors. Furthermore it is straightforward to extend the calculation to include variations which are often used in practice. Many extensions are given in [5, 6]. Here we re-derive the approximation, extending the published results to the third cumulant. This allows us to compare the approximation to the exact results given above.

Derivation

We start from equations (13)–(15), but now we assume F_α is drawn from $\rho^{\infty s}(F)$ and n_α is independent of F_α. We can now average over all possible selections (again we denote this average by $\overline{\cdots}$). The expressions now involve terms such as $\overline{n_\alpha}$, $\overline{n_\alpha^2}$, *etc.* These can be simplified using the identities (arising from the conservation of population size)

$$\begin{aligned} \overline{n_\alpha} &= 1 \\ \overline{n_\alpha n_\beta} &= \frac{P - \overline{n_\alpha^2}}{P-1} \end{aligned} \qquad (36)$$

etc. The equations then simplify to

$$\overline{\kappa_1^s} = K_1^{\infty s} \tag{37}$$

$$\overline{\kappa_2^s} = \left(1 - \frac{\overline{n_\alpha^2}}{P}\right) K_2^{\infty s} \tag{38}$$

$$\overline{\kappa_3^s} = \left(1 - \frac{3\overline{n_\alpha^2}}{P} + \frac{2\overline{n_\alpha^3}}{P^2}\right) K_3^{\infty s} \tag{39}$$

where $K_n^{\infty s}$ is the infinite population result, equations (10)–(12). The first cumulant is identical to the infinite population result, in agreement with the exact calculations given above.

For roulette wheel selection the number of times we draw an individual α is given by the binomial probability

$$P(n_\alpha) = \binom{P}{n_\alpha} p_\alpha^{n_\alpha} (1 - p_\alpha)^{P-n_\alpha}.$$

so that

$$\overline{n_\alpha} = P p_\alpha$$

$$\overline{n_\alpha^2} = P^{\underline{2}} p_\alpha^2 + P p_\alpha$$

$$\overline{n_\alpha^3} = P^{\underline{3}} p_\alpha^3 + 3P^{\underline{2}} p_\alpha^2 + P\, p_\alpha.$$

These results are for a particular p_α corresponding to a particular rank. In addition we need to average over all assignments of a rank to an individual α. Substituting these averages into equations (37)– (39) we obtain

$$\left\langle \overline{\kappa_2^s} \right\rangle = \mathcal{P}_2 \left(1 - P \left\langle p_\alpha^2 \right\rangle\right) K_2^{\infty s}$$

$$\left\langle \overline{\kappa_3^s} \right\rangle = \mathcal{P}_3 \left(1 - 3P \left\langle p_\alpha^2 \right\rangle + 2P \left\langle p_\alpha^3 \right\rangle\right) K_3^{\infty s}$$

where

$$\langle p_\alpha^n \rangle = \frac{1}{P} \sum_{r=0}^{P-1} \left(\frac{r}{Z}\right)^n.$$

These are straightforward to compute giving

$$\bar{K}_1^s = K_1^{\infty s}$$

$$\bar{K}_2^s = \left(1 - \frac{4P-2}{3P(P-1)}\right) K_2^{\infty s}$$

$$\bar{K}_3^s = \left(1 - \frac{12P - 10}{P(P-1)}\right) K_3^{\infty s}.$$

Comparison of these results with the exact results starting from a Gaussian with mean zero and variance 1 is shown in figure 1 at the beginning of this paper. The qualitative behaviour of this approximation is the same as the exact result. For large populations the approximation over estimates the variance and the third cumulant. As we would expect the approximation becomes increasingly accurate as the population increases.

5. Correlations

In a finite population selection produces a genetic correlation due to duplication of the fitter individuals. The correlation is in addition to the natural correlation we would expect because fit strings tend to be correlated. The correlation plays an important role in determining the dynamics of genetic algorithms with recombination. The effect of combining two members of the population will depend on the correlation of the strings. Crossover will restore the variance in the fitness of the population provided the strings are not too correlated.

Modelling the dynamics of Genetic Algorithms using the genetic correlation as a macroscopic variable was first carried for Boltzmann selection in [11,12]. For tournament selection it has be calculated in [6]. We briefly re-derive this result.

The correlation between binary strings $\vec{S}^\alpha = (S_1^\alpha, S_2^\alpha, \ldots, S_L^\alpha)$ with $S_i^\alpha \in \{1, -1\}$ is defined by

$$q_{\alpha\beta} = \frac{1}{L} \sum_{i=1}^{L} S_i^\alpha S_i^\beta. \tag{40}$$

The mean correlation of the population is given by

$$q = \frac{2}{P(P-1)} \sum_{\alpha<\beta} q_{\alpha\beta}.$$

The mean correlation after selection is given by

$$\begin{aligned} q' &= \frac{1}{P(P-1)} \sum_{\alpha,\beta} q'_{\alpha\beta} - \frac{1}{P} \\ &= \frac{1}{P(P-1)} \sum_{\mu,\nu} n_\mu n_\nu q_{\mu\nu} - \frac{1}{P} \\ &= \frac{1}{P(P-1)} \left(\sum_\nu n_\nu^2 - P\right) + \frac{2}{P(P-1)} \sum_{\mu<\nu} n_\mu n_\nu q_{\mu\nu} \end{aligned}$$

where n_μ is the number of times we select member μ. Using equations (36) we can write these equations as

$$1 - q' = \sigma(1-q) - \frac{2}{P(P-1)} \sum_{\mu<\nu} n_\mu n_\nu (q_{\mu\nu} - q) \tag{41}$$

where

$$\sigma = \frac{P - \langle n^2 \rangle}{P-1}, \qquad \langle n^2 \rangle = \frac{1}{P} \sum_\mu n_\mu^2.$$

The first term in equation (41) represents the correlation due to duplication, while the second term represents the natural correlation. Note that, in the limit $P \to \infty$, the duplication factor σ goes to one so that the only correlation comes from the second term. On the other hand, if we perform neutral selection, where the number of times we select an individual does not depend on its fitness (and hence on $q_{\mu\nu}$) then the second term vanishes, so the only correlation comes from the duplication term.

The natural correlation will depend on the improvement in the average fitness. Fitter individuals tend to be correlated with each other. It will, however, depend on the particular problem being tackled. In contrast, the duplication term is problem independent and is an intrinsic property of the selection scheme. For ranking selection or tournament selection we can calculate the duplication factor exactly; it is

$$\sigma = \frac{3P^2 - 7P + 2}{3P(P-1)}.$$

Generalization of this formula for ranking selection with a parameter controlling the strength of selection is given in [6].

6. Boltzmann Selection

In Boltzmann selection the probability of selecting an individual with fitness F_α is

$$p_\alpha = \frac{\mathrm{e}^{-\beta F_\alpha}}{Z}, \qquad Z = \sum_{\alpha=1}^{P} F_\alpha$$

where β is a parameter which controls the selection strength. If $\beta = 0$ all members of the population are chosen with equal probability, while if $\beta = \infty$ only the fittest member of the population is chosen.

In an infinite population Boltzmann selection takes a Gaussian to a Gaussian of the same width, but shifted by βK_2. In contrast to ranking selection there is no loss in the variance. This is not true in a finite population. Finite population corrections are much more significant in

Boltzmann selection than in ranking selection because it is the only source of convergence.

Finite size corrections for Boltzmann selection have been calculated previously in [1–3]. The calculation can be perform exactly, although the solution is in terms of a double integral which has to be performed numerically. Figure 4 shows the change in the first two cumulants as a function of the selection strength β for $P = 2^5$, 2^{10} and 2^{20}.

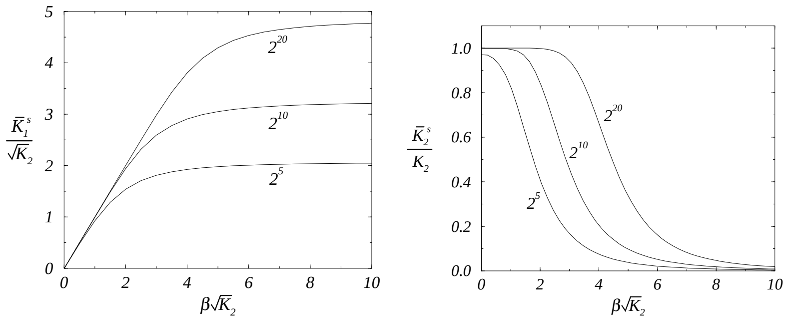

Figure 4. The curves show changes in the first and second cumulant, starting from a Gaussian, as a function of β for $P = 2^5$, 2^{10} and 2^{20}.

When $\beta\sqrt{K_2}$ is small we can find an expansion for the cumulants after selection. Two types of expansion are possible depending on how we sum up the terms. The first is an expansion in $1/P$. This has been given previously in the literature [1–3]

$$\bar{K}_n^s = \frac{\partial^n}{\partial \beta^n}\left(\sum_{m=1} \frac{K_m \beta^m}{m} - \frac{1}{2P}\exp\left(\sum_{l=2}\frac{(2^l-2)K_l\beta^l}{l!}\right)\right) \tag{42}$$

Keeping together terms of order β we obtain a second expansion

$$\bar{K}_n^s = \sum_{m=0} \frac{\bar{\kappa}_{n+m}(P)\beta^m}{m!} \tag{43}$$

where $\bar{\kappa}_n(P)$ is the expected n^{th} cumulant for a random sample of P individuals drawn from a distribution $\rho(F)$. They are just what we gave in equations (1-4) above. (Equation (43) has not appeared before in the literature.)

We can calculate the correlation due to duplications for Boltzmann selection. Unlike ranking and tournament selection this will depend on the distribution of fitnesses. The correlation term

$$\sigma = \frac{P - \langle n^2 \rangle}{P-1}$$

can be calculated in terms of multiple integrals. We can compute these integrals numerically. Figure 5 shows σ starting from a Gaussian. To leading order in $1/P$

$$\sigma = 1 - \frac{\mathrm{e}^{\kappa_2 \beta^2}}{P}.$$

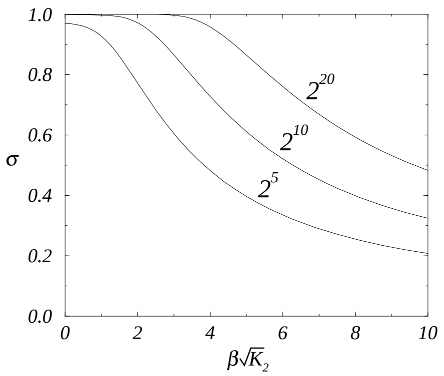

Figure 5. The duplication term σ is shown as a function of β for populations of size $P = 2^n$, with $n = 5$, 10 and 20.

To compare Boltzmann selection with ranking or tournament selection we should take the gain in the first cumulant to be the same. For a Gaussian this gain is $\sqrt{K2/\pi}$. For a reasonable sized population this gain can be achieved using Boltzmann selection with a selection strength of $\beta = 1/\sqrt{\pi K2}$. From equation (42) we see that the second cumulant is reduced by a factor of approximately $1 - (1 + 1/\pi)\exp(1/\pi)/P$. This compares with a reduction factor given by equation (32). Note that the reduction in the second cumulant is a finite size effect for Boltzmann selection, while, even in an infinite population, ranking selection reduces the variance by a factor of $1 - 1/\pi$. The correlation from the duplication term is approximately $\sigma \approx 1 - \exp(1/\pi)/P \approx 1 - 1.375/P$ for Boltzmann selection compared to $1 - 7/3P$ for ranking or tournament selection. By all these measures Boltzmann selection seems superior to ranking and tournament selection. However, Boltzmann selection causes a negative third cumulant or skewness. Ranking selection by contrast causes a small positive skewness. Negative skewnesses reduces the gain in the fitness. As the evolution proceeds negative skewness can build up reducing the effective of Boltzmann selection. Thus, to compare Boltzmann and ranking selection we must consider the full evolution.

7. Conclusions

The statistical mechanics approach to studying evolving populations has been developed over the last six years. Many of the important intuitions that have been obtained from this formalism comes through using crude approximations [1, 5, 6, 11–17]. However, the validity of these approximations rely on more careful analysis [2, 3]. This careful analysis can lead to important results, such as the exact solution of a genetic algorithm assuming linkage equilibrium [18].

This paper presents a number of new results. In particular it gives exact finite size corrections for ranking or tournament selection. Unfortunately the results are not in a particularly simple form. However, the results have been used to validate the approximation first developed in [5, 6]. The approximation is much easier to generalize and would appear to be sufficiently accurate for most modelling purposes.

We have also compared ranking selection with Boltzmann selection. A complete comparison is not possible since selection affects the higher cumulants which will influence the evolution of the lower cumulants. Thus it would be necessary to model the complete dynamics to perform a comprehensive comparison, but this would introduce a problem dependence. Nevertheless, the evidence seems to suggest that Boltzmann selection may be superior ranking and tournament selection.

Acknowledgments

I would like to thank Alex Rogers for many fruitful discussions.

Appendix

A. Exact Third Cumulant Calculation

The third cumulant is made up of three terms

$$K_3^s = \frac{\overline{\mathcal{K}_3^s}}{\mathcal{P}_3} = T_1 + T_2 + T_3$$

where

$$T_1 = \sum_{r=0}^{P-1} \left(\frac{r}{Z} - \frac{3r^2}{Z^2} + \frac{2r^3}{Z^3} \right) \overline{F_r^3}$$

$$T_2 = -3 \sum_{r \neq r'} \left(\frac{r\, r'}{Z^2} - \frac{2r^2 r'}{Z^3} \right) \overline{F_r^2\, F_{r'}}$$

$$T_3 = \frac{2}{Z^3} \sum_{r \neq r' \neq r''} r\, r'\, r'' \overline{F_r\, F_{r'}\, F_{r''}}$$

Following the same procedure as for the first two cumulants we find

$$T_1 = \frac{1}{P^3(P-1)^3}\Big(P^{\underline{3}}\left(P^2 - P - 4\right)\big((P+1)I_{3,1} - 6I_{3,2}\big)$$
$$+8\,P^{\underline{4}}\,I_{3,3}\Big)\,.$$

The second term is

$$T_2 = \frac{-3}{P^3(P-1)^3}\Big(P^{\underline{6}}I_{1,1}\,I_{2,1} - 4P^{\underline{5}}I_{1,1}\,I_{2,2}$$
$$+8P^{\underline{4}}(4I_{2,1;1,1} - 3I_{2,2;1,0})$$
$$+\,4P^{\underline{3}}(P^2 - P - 8)I_{1,1;2,0}\,+$$
$$4P^{\underline{3}}(P^2 - P - 4)I_{2,1;1,0}\Big)$$

while the third term is

$$T_3 = \frac{2}{P^3(P-1)^3}\Big(P^{\underline{6}}I_{1,1}^3 + 48P^{\underline{5}}(I_{1,1;1,0;1,1} + 2I_{1,1;1,1;1,0})$$
$$+\,144P^{\underline{4}}I_{1,1;1,0;1,0}\Big)$$

where

$$I_{i,j;k,l;m,n} = 2\int_{-\infty}^{\infty}\int_{F_1}^{\infty}\int_{F_2}^{\infty} F_1^i\,R_1^j\,F_2^k\,R_2^l\,F_3^m\,R_3^n\,\mathrm{d}R_1\,\mathrm{d}R_2\,\mathrm{d}R_3$$

and where we have used $I_{1,1;1,1;1,1} = I_{1,1}^3/6$. The third cumulant after selection, starting from a Gaussian, is plotted on the right hand graph in figure 1.

References

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

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

[3] A. Prügel-Bennett. Modelling evolving populations. *Journal of Theoretical Biology*, 185:81–95, 1997.

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

[5] A. Rogers and A. Prügel-Bennett. Genetic drift in genetic algorithm selection schemes. *IEEE Transactions on Evolutionary Computation*, 3(4):298–303, 1999. accepted for publication.

[6] A. Rogers and A. Prügel-Bennett. The dynamics of a genetic algorithm on a model hard optimization problem. *Complex Systems*, 1999. accepted for publication.

[7] K. A De Jong and J. Sarma. Generation gaps revisited. In L. D. Whitley, editor, *Foundation of Genetic Algorithms 2*, pages 19–28, San Mateo, 1993. Morgan Kaufmann.

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

[10] R. A. Fisher. Moments and product moments of sampling distributions. *Proceedings of the London Mathematical Society,* series 2, 30:199–238, 1929.

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

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

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

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

[15] 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.

[16] 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*, pages 57–68, San Francisco, 1999. Morgan Kaufmann.

[17] A. Prügel-Bennett. On the long string limit. In W. Banzhaf and C. Reeves, editors, *Foundations of Genetic Algorithms 5*, pages 45–56, San Francisco, 1999. Morgan Kaufmann.

[18] M. Rattray and J. Shapiro. Cumulant dynamics in a finite population: linkage equilibrium theory. LANL eprint archive, 1999.