

Comparing Selection Methods of Evolutionary Algorithms Using the Distribution of Fitness *

Erick Cantú-Paz

Center for Applied Scientific Computing
Lawrence Livermore National Laboratory
Livermore, CA 94551
cantupaz@llnl.gov

Abstract

Selection methods are essential components of evolutionary algorithms. To have a better chance of success, the selection method must be balanced with the operators that create new individuals, and therefore, understanding the effect of selection on the population is critical for the development of robust problem solvers. The objective of this paper is to compare selection algorithms using the cumulants of the fitness distribution of the selected individuals. The cumulants are calculated using the expected values of the order statistics of the fitness values. The calculations show important differences in the selection methods, even when they are configured to have the same selection intensity.

Keywords: linear ranking, exponential ranking, tournaments, Boltzmann selection, truncation selection, selection intensity, fitness distributions, order statistics.

1 Introduction

Evolutionary algorithms (EAs) have two basic components: selection of the fittest and randomized operators such as crossover and mutation that create new solutions from the selected ones. For EAs to succeed, it is necessary to balance the exploitation (by selection) of the information already gathered about the problem with the exploration of new solutions by the genetic operators (e.g., (Holland, 1973; Goldberg, 1998)). To achieve this balance, we must understand how selection affects the composition of the population. That is the purpose of this paper.

Selection algorithms have been studied using different methodologies. Some of the earliest attempts to quantify the effects of selection deal with the takeover time (Goldberg & Deb, 1991), which is the number of generations that it takes the selection algorithm to reproduce a single representative of the optimal solution to occupy the entire population. The takeover time is a simple and effective way to compare selection algorithms, and it has been used by Goldberg et al. (1993) to obtain a control map that relates the selection strength and the probability of crossover with the success of the algorithm. Goldberg and Deb's basic model has been extended to consider selection methods used in evolution strategies (Bäck, 1994), noisy environments (Sakamoto & Goldberg, 1997) and parallel GAs with multiple populations (Cantú-Paz, 1999).

The takeover time has also been studied with Markov chain models. For example, Chakraborty et al. (1996) studied GAs with generation gaps and obtained the probability that a specific class of individuals takes over the population at each iteration. They considered random, worst-fit,

*Submitted to IEEE Transactions on Evolutionary Computation

and exponential ranking deletion, but their framework has been used to study other replacement strategies (Smith & Vavak, 1999). Rudolph (2000) used some properties of the transition matrix to derive closed-form expressions for the takeover time of some deletion methods.

Others studies characterize the strength of selection methods by calculating the selection intensity (Bäck, 1995; Miller & Goldberg, 1995; Mühlenbein & Schlierkamp-Voosen, 1993; Thierens, 1997). The selection intensity is the increase of the mean fitness of the population after selection, normalized by the standard deviation. The selection intensity is related to the optimal mutation rate (Bäck, 1996), the population size (Mühlenbein & Schlierkamp-Voosen, 1994; Harik et al., 1999), and under some conditions the speed of convergence of GAs is inversely proportional to the intensity (Mühlenbein & Schlierkamp-Voosen, 1993). Despite of its usefulness to compare selection methods, the selection intensity provides very limited information about the fitness distribution.

More sophisticated characterizations of selection methods involve the calculation of the cumulants (or moments) of the distribution of fitness (Prügel-Bennett & Shapiro, 1994). This approach yields very detailed information of selection methods. It has been extended to study noisy fitness environments (Ratray & Shapiro, 1997) and GAs with generation gaps (Rogers & Prügel-Bennett, 1999). Most of the work in this area has dealt with Boltzmann selection, because it has certain theoretical properties that, under some assumptions, allow to obtain closed-form equations to describe the fitness distribution.

The objective of this paper is to compare several popular selection methods using the cumulants of the fitness distribution after the population undergoes selection. The emphasis is on numerical computations, not on obtaining closed-form expressions. The calculations consider finite populations of arbitrary size, and show important differences among the selection methods considered, even when they are configured to have the same selection intensity.

The observations in this paper may explain many results reported in the literature regarding the success or failure of a particular combination of a selection algorithm and variation operators. The results may also be used to choose rationally a combination of selection and variation operators with a good balance of exploitation and exploration that may lead to ultimate success.

Blickle and Thiele (1996) studied the effect of selection on the mean and the standard deviation of the selected individuals, and considered some of the selection methods studied in this paper. The present work extends Blickle and Thiele's in two main directions: it considers additional selection methods (Boltzmann, tournaments without replacement, and proportional selection), and considers the higher cumulants of the fitness distribution that give additional information about the effects of selection.

The remainder of this paper is organized as follows. The next section presents the method used to calculate the expected cumulants of the distribution of fitness. Then, section 3 examines five popular selection methods. Section 4 compares the selection methods and discusses the importance of balancing selection with the variation operators. Finally, section 5 presents a brief summary and the conclusions of this paper.

2 Describing the Fitness Distribution

The approach of this paper is to describe the distribution of fitness using its first four cumulants. The cumulants of a distribution are related to its central moments. The r -th central moment of the distribution of fitness of a population of size n is

$$\mu_r = \frac{1}{n} \sum_{i=1}^n (f_i - \bar{f})^r, \quad (1)$$

where f_i is the fitness of the i -th individual, and $\bar{f} = \frac{1}{n} \sum_{i=1}^n f_i$ is the mean population fitness. The first three cumulants are equal to the first central moments. The fourth cumulant is $\kappa_4 = \mu_4 - 3\mu_2^2$.

The first cumulant is the mean ($\kappa_1 = \mu_1 = \bar{f}$), and the second is the variance ($\kappa_2 = \mu_2 = \sigma_f^2$). The third and fourth cumulants give additional information about the shape of the distribution, and sometimes they are normalized by dividing them by $\kappa_2^{r/2}$ to obtain the skewness and kurtosis coefficients. The skewness measures the asymmetry of the distribution; it is negative if a distribution is skewed to the left, and it is positive if the distribution is skewed to the right. The kurtosis measures the ‘‘peakedness’’ of the distribution; a negative kurtosis indicates that the distribution is flatter than a normal, and a positive kurtosis indicates that the distribution is more peaked than a normal.

One possibility to calculate the expected value of the cumulants is to integrate over all possible populations after selection. This is the method used by Shapiro et al. (1994), and for some selection methods and under some assumptions, it is possible to obtain closed-form equations for the cumulants. Our approach is different. Instead of calculating the expectation over all possible outcomes of selection, we calculate the expected fitness of each individual, and then use equation 1 to obtain the cumulants.

The method of this paper is based on order statistics. The critical observation is that we may interpret the fitness values f_i as samples of random variables F_i with a common distribution, say $N(\mu, \sigma)$, but most of the calculations below apply to arbitrary distributions. We may arrange the random variables in increasing order as

$$F_{1:n} \leq F_{2:n} \leq \dots \leq F_{n:n}.$$

These are the order statistics of the F_i variables. Without loss of generality in the remainder we assume a maximization problem, and we normalize the random variables as follows:

$$Z_{i:n} = \frac{F_{i:n} - \bar{F}}{\sigma_F}. \quad (2)$$

The expected value of the i -th order statistic is defined as

$$\mu_{i:n} = \mathbb{E}(Z_{i:n}) = n \binom{n-1}{i-1} \int_{-\infty}^{\infty} z \phi(z) [\Phi(z)]^{i-1} [1 - \Phi(z)]^{n-i} dz, \quad (3)$$

where $\phi(z)$ and $\Phi(z)$ are the PDF and CDF of the distribution of fitness, respectively.

Each individual has a certain probability of survival p_i that depends on the selection method. With these probabilities we calculate the mean of the population after selection as

$$\bar{Z}_{\text{sel}} = \sum_{i=1}^n p_i \mu_{i:n}, \quad (4)$$

with $\sum_{j=1}^n p_j = 1$. The equation above also gives the selection intensity of the selection method. We can now calculate the r -th moment after selection as:

$$\mu_r^{\text{sel}} = \sum_{i=1}^n p_i (\mu_{i:n} - \bar{Z}_{\text{sel}})^r. \quad (5)$$

For some distributions, the expected values of the normalized order statistics can be found in tables (Harter, 1970), but if we make the assumption that the fitness has a normal distribution, we can use the following approximation (Harter, 1970):

$$\mu_{i:n} \approx \Phi^{-1} \left(\frac{1 - \alpha_i}{n - 2\alpha_i + 1} \right), \quad (6)$$

where $\Phi^{-1}(x)$ is the inverse of the CDF of a unit normal distribution (i.e., it returns the value of z such that $\Phi(z) = x$), and α_i is defined as

$$\alpha_i = \begin{cases} 0.315065 + 0.057974 \log n - 0.009776(\log n)^2 & \text{if } i = 1, \\ 0.327511 + 0.058212 \log n - 0.007909(\log n)^2 & \text{otherwise.} \end{cases} \quad (7)$$

Harter (1970) discourages the use of the approximation above for $n > 400$. However, the next section shows that the approximation is sufficiently accurate for our purposes for much larger populations.

3 Selection Methods

This section examines in detail several selection methods. The results presented here are numeric calculations of the cumulants, and assume that the fitness has a normal distribution with zero mean and unit standard deviation.

The results of experiments using a GA with selection only are plotted alongside the theoretical calculations. The experiments demonstrate that the approximations used to calculate the theoretical values of the moments are sufficiently accurate for our purposes. The fitness function used was a simple $l = 500$ bits onemax $F = \sum_{j=1}^l x_j$, where $x_j \in \{0, 1\}$ are the individual bits in the chromosome. The population was initialized randomly, so F has a binomial distribution with mean $\bar{F} = l/2 = 250$ and variance $\sigma^2 = l/4 = 125$. The fitness values were normalized to obtain a fitness distribution with zero mean and unit standard deviation, which approximates the unit Gaussian that the calculations assume. The experiments used populations with $n = 1000$ individuals, and were repeated 1000 times with different random seeds. The error bars represent 5% confidence intervals, and in many cases there are so small that they are not visible in the plots. Except for tournament selection, in all the experiments, the individuals were chosen using stochastic universal sampling (Baker, 1985). Tournament selection was implemented directly from its description.

3.1 Linear Ranking

In linear ranking selection, individuals are sorted according to their fitness, and the probability of being selected is linearly proportional to their rank in the population (Baker, 1985). The desired expected number of copies of the best (n^+) and worst (n^-) individuals are supplied as parameters to the algorithm.

The probability that an individual with rank i will be selected is

$$p_i = \left(n^- + \frac{n^+ - n^-}{n - 1}(i - 1) \right) / Z, \quad (8)$$

where $Z = n$ is a normalizing factor, and to ensure that that the probabilities add to 1, $n^- + n^+$ must equal 2.

Figure 1 shows the first four cumulants of the selected individuals varying the parameter n^- . The plots show that linear ranking is very weak. At its strongest setting with $n^- = 0$, the mean of the selected individuals is $\frac{1}{\sqrt{\pi}} = 0.5642$, and the variance is reduced to approximately 0.68 from its original of 1. At all possible settings of n^- , the higher cumulants remain close to their original values.

It is well known that linear ranking with $n^- = 0$ produces the exact same results as pairwise tournament selection (Goldberg & Deb, 1991; Bickel & Thiele, 1996). The calculations in section 3.5 agree with this observation.

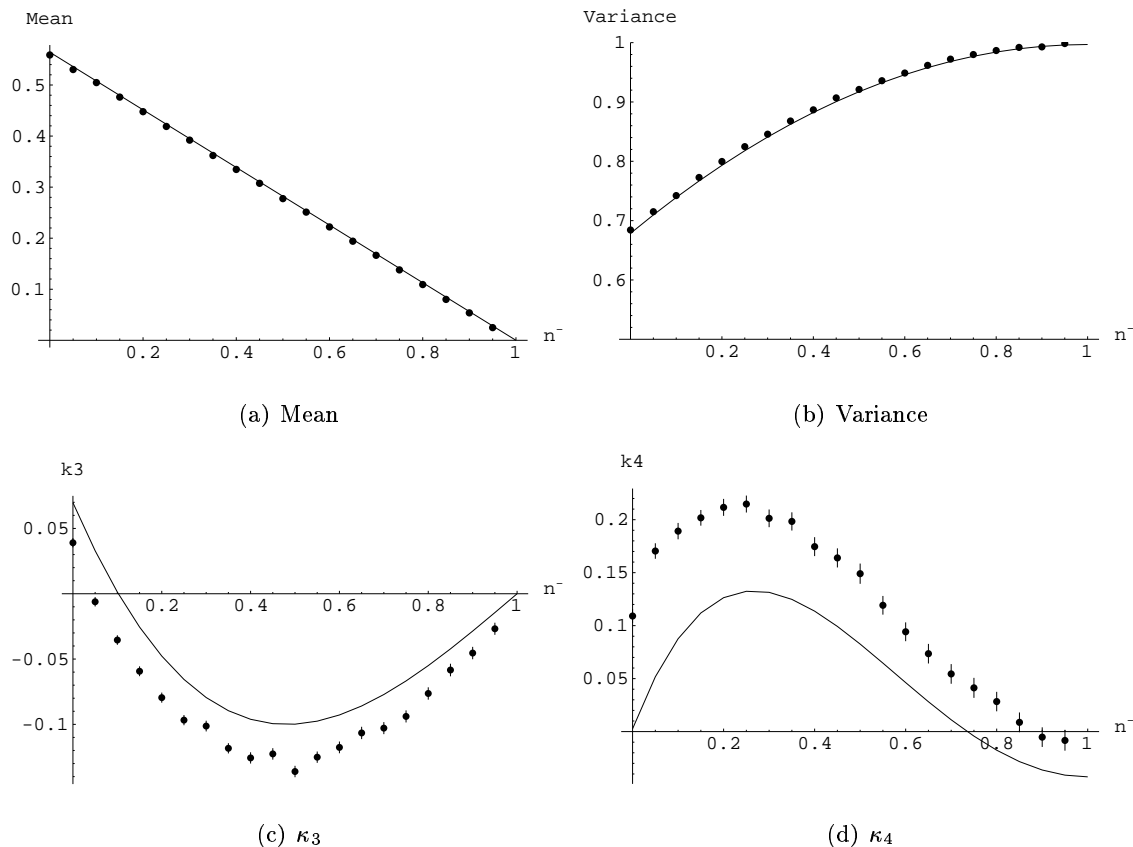


Figure 1: First four cumulants of the distribution of fitness after linear ranking selection varying n^- .

3.2 Exponential Ranking

An alternative to the weak linear ranking is to assign survival probabilities to the sorted individuals using an exponential function. In exponential ranking selection (Baker, 1985), p_i is defined as

$$p_i = \frac{c^{n-i}}{Z}, \quad (9)$$

where $c \in [0, 1]$ is a parameter that determines the behavior of the algorithm. Since $Z = \sum_{j=1}^n c^{n-j} = \frac{c-1}{c^n-1}$, p_i can be simplified to

$$p_i = \frac{c-1}{c^n-1} c^{n-i}. \quad (10)$$

Hancock (1994) observes that “for equivalent growth rates, exponential ranking ought to give a more diverse population [than linear ranking.]” because “it gives more chance to the worst individuals, at the expense of those above average”. Certainly, having a more diverse population is desirable, because the variation operators have less of a burden to produce diverse individuals. We will explore these observations in more detail in section 4.

Figure 2a shows that for most of the range of c , exponential ranking has a much higher intensity than linear ranking. It is difficult to confirm Hancock’s observation from figure 2, but it will become

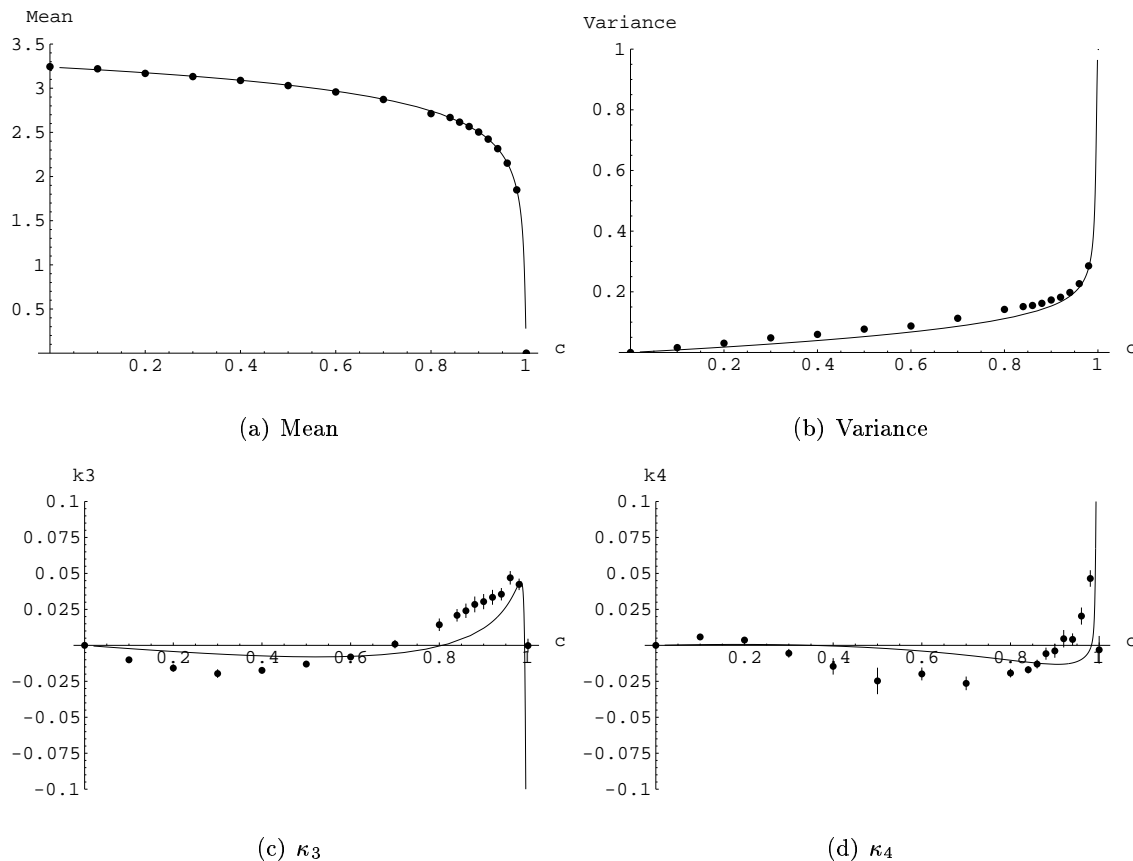


Figure 2: First four cumulants of the distribution of fitness after exponential ranking selection varying c .

clearer in figure 6. Note however, that for values of c that produce a high selection intensity, exponential ranking also produces a distribution with a low variance. This may be problematic, unless the variation operators (crossover and mutation) reintroduce diversity into the population. This may explain why in practice exponential ranking is used with values of c close to 1.

3.3 Boltzmann Selection

In this form of selection, the probability of selection is (de la Maza & Tidor, 1993):

$$p_i = \frac{\exp(\beta f_i)}{Z}, \quad (11)$$

where β controls the selection intensity, and the denominator $Z = \sum_{j=1}^n \exp(\beta f_j)$ is a normalizing factor¹.

Figure 3 has plots of the first four cumulants of the fitness distribution varying β . Rogers and Prügel-Bennett (1999) determined analytically that for weak selection, the selection intensity can be approximated by β . This can be confirmed in figure 3a, where the plot appears linear for $\beta < 2$.

¹This form of selection should not be confused with Boltzmann tournament selection (Goldberg, 1990; Mahfoud, 1993), which is a “niching” selection method intended to maintain multiple good solutions coexisting in the population.

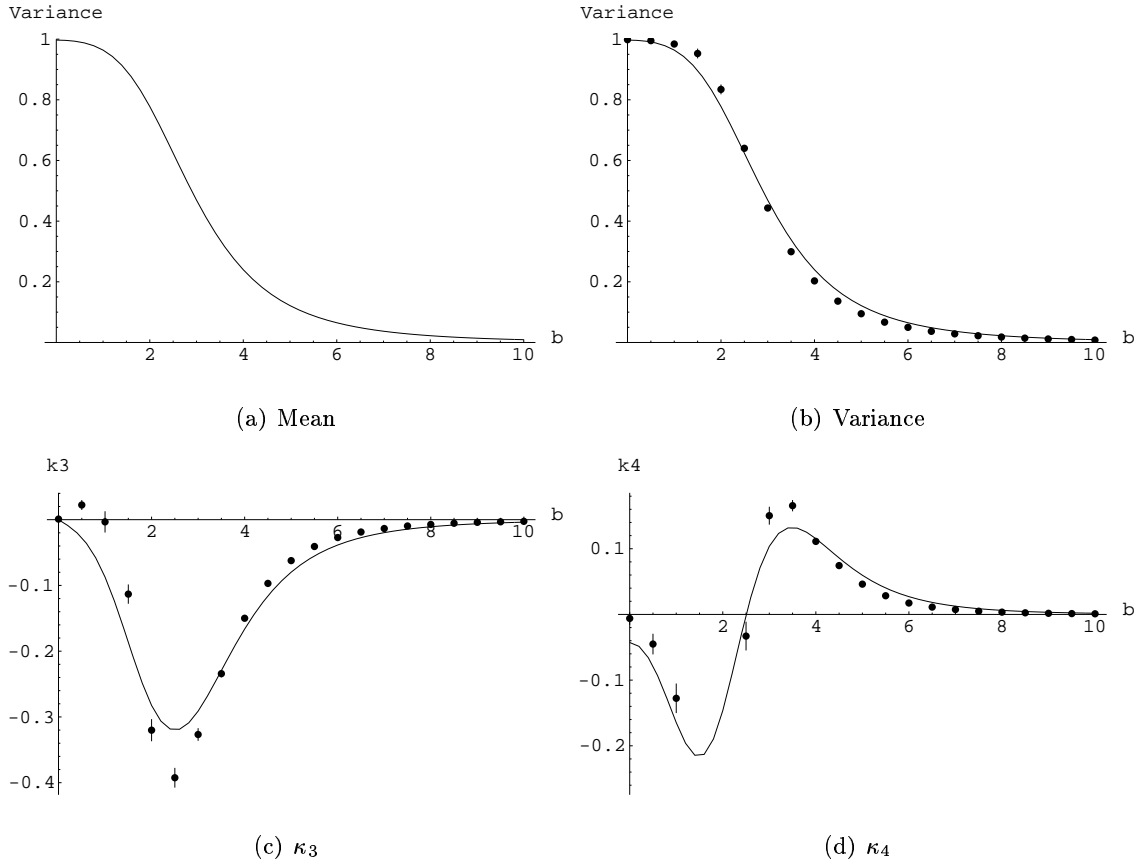


Figure 3: First four cumulants of the distribution of fitness after Boltzmann selection varying β .

The figure also shows that under strong Boltzmann selection the distribution of fitness is heavily skewed to the left and that it has a sharp peak and long tails. Notice that the mean fitness of the selected individuals does not increase much after $\beta > 3$. Thus, it may not be advisable to use Boltzmann selection with $\beta > 3$, unless it is used with mutation or recombination operators that reintroduce some diversity.

An important observation about this form of Boltzmann selection is that the probabilities of survival depend on the fitness values, and not on the rank of the individuals. A consequence of this is that the probabilities of selection are sensitive to the scale of the fitness. Thus, if the fitness values f_i are multiplied by a constant k , the p_i would change. de la Maza and Tidor (1993) described this problem, and noted that it can be compensated by using $\beta' = \beta/k$ in equation 11. A consequence of this is that if the fitness values are not normalized, we would have to use $\beta' = \beta/\sigma_F$ to obtain the response to selection described in this section. de la Maza and Tidor also noted that Boltzmann selection is translation invariant (i.e., p_i does not change if a constant is added to the fitness values), which is a desirable property. It should be noted that ranking methods and tournaments are scale and translation invariant.

3.4 Truncation Selection

Truncation selection deterministically selects the top τ individuals in the population (Mühlenbein & Schlierkamp-Voosen, 1993). It is equivalent to the (μ, λ) selection used in evolution strategies

(ES) with $\mu = \lambda/\tau$ (Schwefel, 1981). In (μ, λ) -ES a population with μ individuals is used to create λ individuals ($\lambda \geq \mu$), and the μ best offspring are selected to survive.

We can cast this deterministic selection method into our probabilistic framework by defining p_i as

$$p_i = \begin{cases} 0 & \text{if } i < n - \tau, \\ 1/\tau & \text{if } i \geq n - \tau. \end{cases} \quad (12)$$

As always, $\sum_{j=1}^n p_j = 1$. Figure 4 shows the first four cumulants of the selected individuals varying the selection threshold τ . From the figures it is evident that choosing a small value of τ increases the mean fitness of the population significantly, but it also reduces the variance greatly. This is probably not desirable, because the population would lose diversity very rapidly and it is more likely that the algorithm will converge to a suboptimal solution. Also, the shape of the distribution is most affected at low values of τ .

In evolution strategies it is common to use values of τ close to 0.15 (1/7) (Bäck, 1996; Schwefel, 1981), but it is also common to use stronger mutations than in GAs and aggressive recombination operators. We discuss the implications of this in section 4. For now, we also note that in genetic algorithms with less aggressive recombinations, weaker settings of truncation selection are used with $\tau \approx 0.5$ (Mühlenbein & Schlierkamp-Voosen, 1993).

Approximate closed-form expressions for the mean and the variance of the selected individuals have been derived for truncation selection (Burrows, 1972; 1975). Bäck (1995) introduced Burrows' approximation to evolutionary computation and confirmed its accuracy. For a unit normal distribution, the selection intensity is

$$I_\tau \approx \frac{1}{\tau} \phi(\Phi^{-1}(1 - \tau)). \quad (13)$$

Later, Blickle and Thiele (1996) re-derived Burrows' approximation for the variance using a different method. The variance of the selected individuals is:

$$\sigma_\tau^2 \approx 1 - I_\tau(I_\tau - \Phi^{-1}(1 - \tau)). \quad (14)$$

The two approximations are indistinguishable from the values presented in figure 4.

3.5 Tournament Selection

In tournament selection the best individual from a random sample of s individuals is chosen for the next generation (Brindle, 1981). The samples may be drawn with or without replacement. It is known that the selection intensity of tournament selection is $\mu_{s,s}$ (Bäck, 1995), which increases very slowly with respect to s ($\mu_{s,s} \approx \sqrt{\sqrt{2} \ln s}$).

In this form of selection, an individual will be selected only if its fitness is greater than the fitness of the other $s - 1$ individuals that participate in the tournament. If the individuals are sampled without replacement, the probability that the fitness of a randomly chosen individual is lower than the fitness of the i -th individual is just the probability that a uniform random number in $[1, n]$ is less than or equal to i : $P(f_j \leq f_i) = P(j \leq i) = \frac{i-1}{n-1}$. In a tournament with s participants, the probability that all of the opponents of the i -th individual are ranked lower than i is $P(j \leq i)^{s-1}$, so the probability that the i -th individual is selected is

$$p_i = \left(\frac{i-1}{n-1} \right)^{s-1} / Z \quad (15)$$

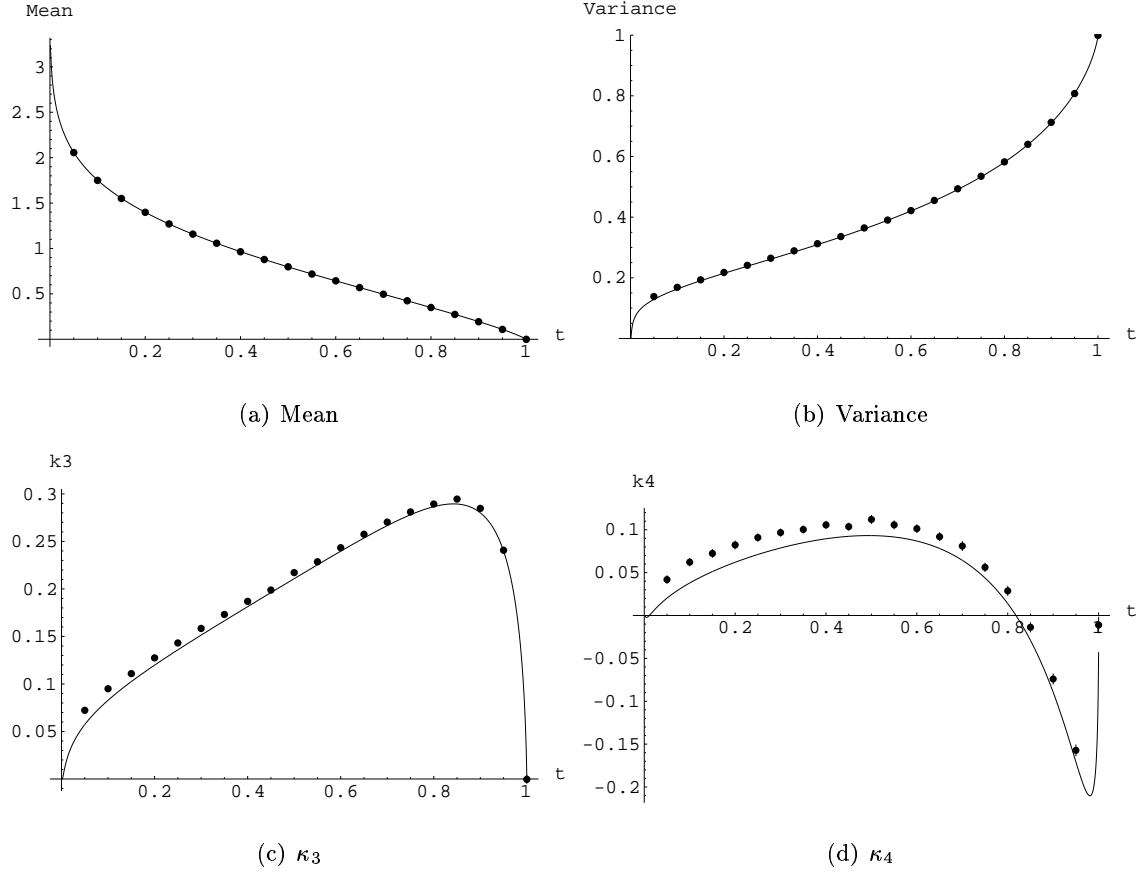


Figure 4: First four cumulants of the distribution of fitness after truncation selection varying the truncation threshold τ .

where the denominator $Z = \sum_{j=1}^n \binom{j-1}{n-1} s^{-1}$ normalizes the probabilities so they add to 1. Note that the worst individual in the population is never selected ($p_1 = 0$) and the best individual always wins in all the tournaments it participates ($p_n = 1$). However, when sampling with replacement, it is possible that the best individual will never be selected to participate in a tournament.

If selection is done without replacement, the calculations are a bit more complicated. The probability that exactly $x = s - 1$ individuals are chosen among those with lower fitness than individual i has a hypergeometric distribution:

$$P(x = s - 1) = \frac{\binom{i-1}{s-1} \binom{n-i-1}{0}}{\binom{n-1}{s-1}} = \frac{(i-1)!(n-s)!}{(n-1)!(i-s)!}$$

and normalizing gives the probability that the i -th individual is selected as

$$p_i = \frac{(i-1)!(n-s)!}{(n-1)!(i-s)!} / Z. \quad (16)$$

When $s = 2$ there is no difference between selecting with or without replacement, as should be expected because the p_i are the same. The differences become larger with bigger tournaments and smaller populations, but, in many practical situations, the differences are so small that they may be

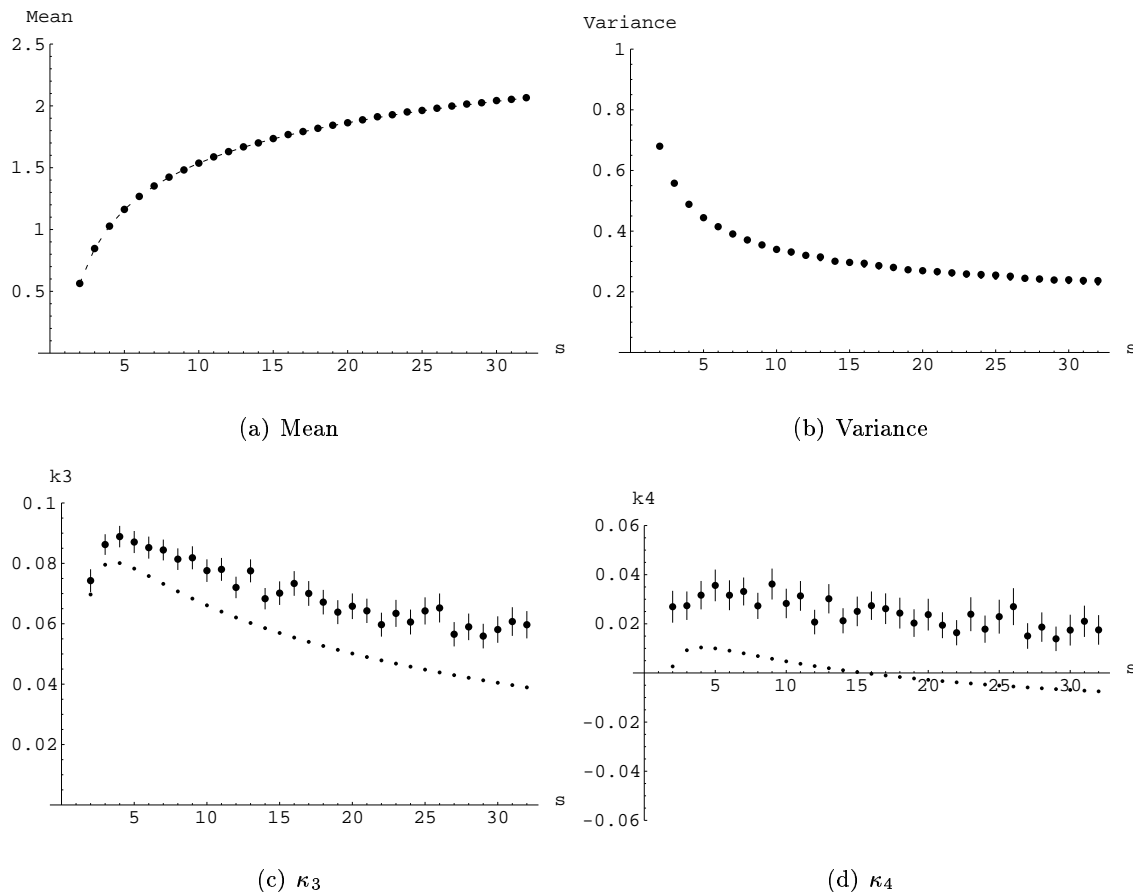


Figure 5: First four cumulants of the distribution of fitness after tournament selection without replacement varying s .

considered negligible. Blickle and Thiele (1996) considered tournament selection in their study, and verified analytically that the distribution of individuals selected by pair-wise tournaments ($s = 2$) is identical to the distribution of individuals selected by linear ranking with $n^- = 0$. Our calculations yield the same results.

The plots of the first four cumulants of the distribution of fitness varying s from 2 to 32 are displayed in figure 5. From these plots, we can observe that the skewness and kurtosis of the original distribution are not changed much by tournament selection. The variance decreases more slowly than with truncation selection, but faster than with exponential ranking or with Boltzmann weights.

3.6 Proportionate Selection

Possibly the most well-known selection method in evolutionary algorithms is fitness proportionate selection (Holland, 1975). This method assigns to each individual a probability of survival that is proportional to its fitness:

$$p_i = \frac{f_i}{\sum_{j=1}^n f_j} = \frac{f_i}{n\bar{f}}. \quad (17)$$

Note that for this to be valid, all the fitness values must be positive, and the mean must be different from zero. In practice, this is accomplished by adding a constant to the fitness, but this has the undesirable consequence of changing the probabilities of selecting the individuals (i.e., this method is not translation invariant).

Making useful numerical calculations is impossible for this method, because they would necessarily depend on the location of the fitness distribution. However, we can still use the general approach of the paper to study proportionate selection by obtaining closed-form expressions for the selection intensity.

Substituting p_i into equation 4 yields the mean of the selected individuals:

$$\begin{aligned}\bar{F}_s &= \sum_{i=1}^n p_i F_{i:n} \\ &= \frac{1}{n\bar{F}} \sum_{i=1}^n F_{i:n}^2 \\ &= \frac{1}{n\bar{F}} (n\sigma^2 + n\bar{F}) \\ &= \frac{\sigma^2}{\bar{F}} + \bar{F}.\end{aligned}$$

The third step is possible because $\sigma_F^2 = \frac{1}{n} \sum_{i=1}^n F_i^2 - (\bar{F})^2$. Subtracting the mean and dividing by the standard deviation results in the selection intensity:

$$I = \frac{\sigma}{\bar{F}}. \tag{18}$$

This result coincides with the result of Mühlenbein and Schlierkamp-Voosen (1993), and it clearly shows that the intensity depends on the location of the distribution. Also, as the run progresses and σ becomes smaller, proportionate selection becomes weaker. The traditional way to avoid this effect is to use a scaling method that assigns probabilities of selection to individuals considering some measure of the dispersion of the distribution (the standard deviation or the range).

4 Discussion

To better understand the consequences of this work, this section discusses several topics. First, we present comparisons of the selection methods that show the limitations of characterizing them only by their selection intensities. Next, we identify three sources of failure in an evolutionary algorithm which are related to selection, and comment on how this research can be applied to alleviate them. The toughest problem is to balance selection with the variation operators, so we discuss this at length.

4.1 Comparison of Selection Methods

Although it is convenient to compare selection methods by their takeover time or selection intensity, these may not be sufficiently complete representations. The calculations in section 3 reveal how selection methods modify the distribution in very different ways, even when their parameters are chosen to produce the same intensity. To facilitate comparisons figures 6, 7, and 8 plot the second, third, and fourth cumulants of the selected individuals vs. the selection intensity.

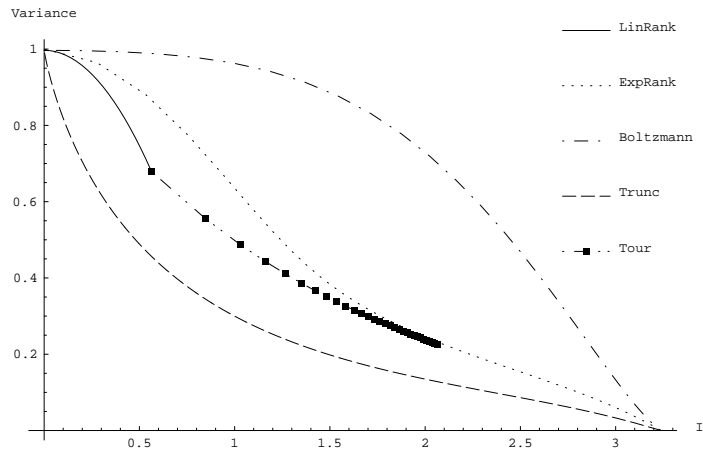


Figure 6: Comparison of the variance varying the selection intensity.

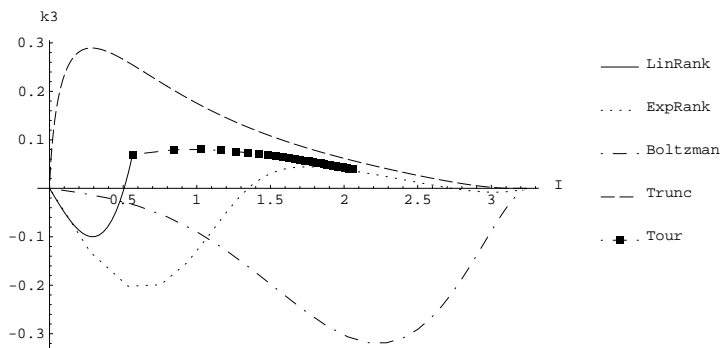


Figure 7: Comparison of the third cumulant varying the selection intensity.

From figure 6 it is evident that for equivalent selection intensities, Boltzmann selection preserves the original fitness variance of the population more than the other methods, while truncation selection results in the largest reduction of variance. Although it is a good heuristic to prefer the method that preserves the greater diversity (Bäck, 1995), we must be cautious before dismissing the other selection methods. Perhaps severely reducing diversity is useful if the variation operators produce very diverse individuals. Consider, for example, that many successful applications of evolution strategies use (μ, λ) (truncation) selection along with aggressive mutation and recombination operators.

Figure 7 shows that some methods always produce a distribution that is skewed either to the left ($\kappa_3 < 0$) or to the right ($\kappa_3 > 0$), but for the linear and exponential ranking the direction of the skewness depends on the selection intensity. The sign of the third moment is interesting, because it clearly shows a bias in the way the individuals are sampled. A method that results in a distribution with a positive skewness has fewer samples of the fitter individuals than a distribution with negative skewness. One could argue that having many highly-fit individuals is an advantage, because it is more likely that they will remain intact after mutation or recombination. However, it is not clear that this is an advantage in all problems or with arbitrary variation operators. Similar

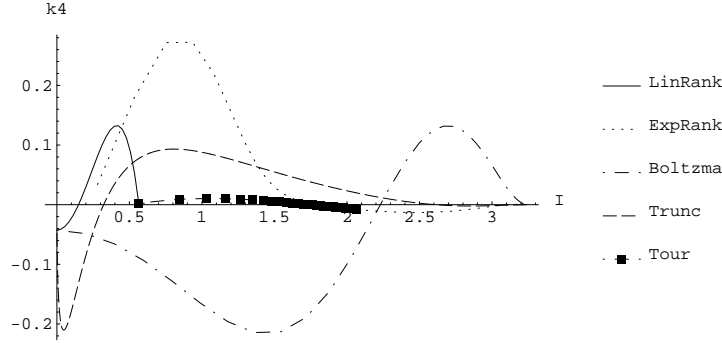


Figure 8: Comparison of the fourth cumulant varying the selection intensity.

arguments can be made for preferring flat or peaky distributions.

These differences between the algorithms can be explored further by comparing the probabilities of selection p_i . Figure 9 compares tournaments, exponential ranking, and Boltzmann selection at $I = 0.5642$, $I = 1.0$ and $I = 2.0$. Linear ranking with $n^- = 0$ is equivalent to pairwise tournaments. The figure shows that at $I = 0.5642$ and $I = 1.0$, exponential ranking and Boltzmann selection give more chance to the low-ranked individuals than tournaments at the expense of some above-average individuals. Also, tournaments give less chance to the highest-ranked individuals than exponential ranking or Boltzmann. These plots also illustrate why it is important to consider the higher cumulants. For example, the difference between tournaments and exponential ranking does not appear very large at $I = 1$, but they produce distributions skewed in opposite directions (see figure 7). At $I = 2$ exponential ranking and tournaments become more like each other, and very different from Boltzmann.

These biases should be taken into consideration when choosing or designing variation operators. For example, consider the repetitive application of selection on a population with low variance (or highly skewed). If the operators do not produce individuals that restore the diversity (or symmetry) of the distribution, after selection the population will be even less diverse (or more skewed), and it is likely that it will not explore many points. This is discussed next.

4.2 Premature Convergence and Other Problems

As was mentioned in the introduction, it has long been recognized that for EAs to succeed, there must be a balance between exploring new solutions and exploiting the information already gathered about the problem. We can identify three sources of failure in an EA related to this balance: random drift, premature convergence, and cross-competition (Goldberg, Deb, & Thierens, 1993; Goldberg, 1998). Drift is caused by an excessively weak selection method. If there is not a sufficiently strong preference for the most fit individuals, the algorithm will be extremely sensitive to sampling errors and will ultimately converge to a more or less random point that is very unlikely to be a good solution. On the other extreme, cross competition is caused by excessively strong selection. Intuitively, cross-competition occurs when an individual possesses a trait that gives it an advantage over others and the strong selection causes this individual to dominate the population very quickly without allowing time to examine other traits. Even with an intermediate selection intensity, an algorithm may converge too fast, preventing crossover or mutation to assemble or discover better solutions. This is the so-called ‘premature convergence’ problem (De Jong, 1975; Goldberg, 1998).

There are several (non-exclusive) ways to avoid or alleviate these problems. Drift or cross-

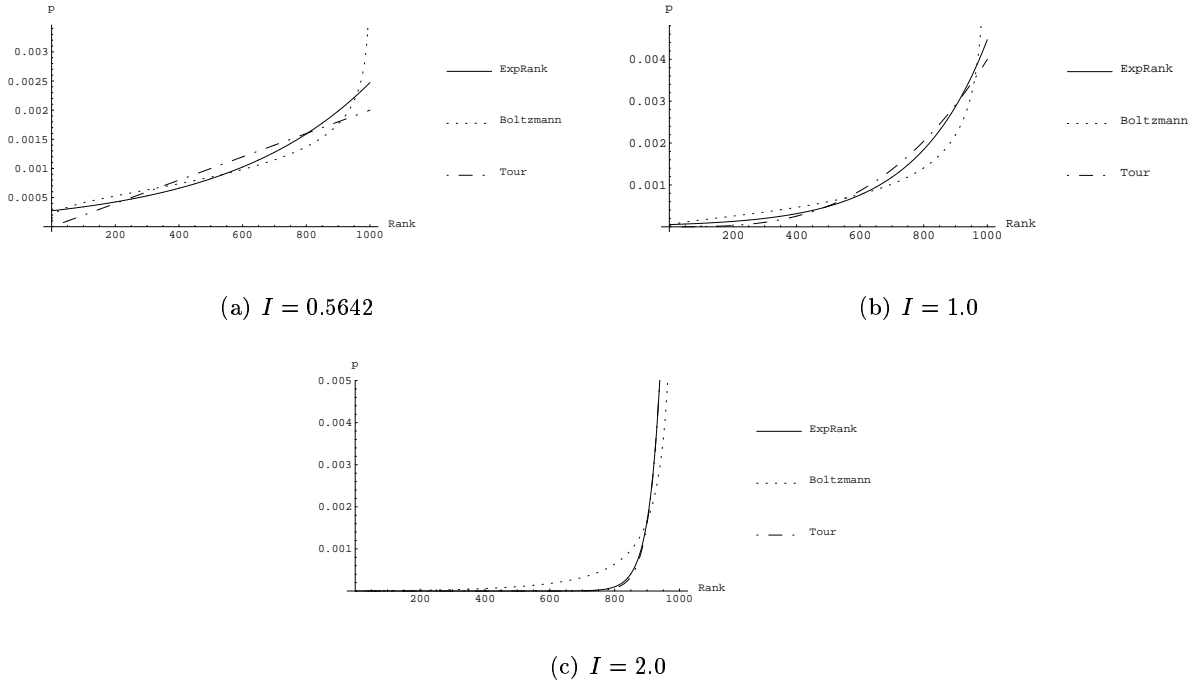


Figure 9: Probability of being selected (p_i) vs. rank for different selection methods.

competition may be addressed simply by lowering or raising the selection intensity; and to some extent, cross-competition can be alleviated by using a larger population. Another solution would be to use a selection method that preserves as much diversity as possible. This paper can help directly to implement these solutions.

A related solution is to increase diversity either by raising the mutation or crossover probabilities or by using more aggressive operators. Uniform crossover is a more aggressive form of mixing than two-points crossover, which in turn is more aggressive than one-point crossover. In addition, we may use recombination operators with more than two parents (Eiben, Raué, & Ruttkay, 1994; Eiben & Bäck, 1997; Beyer, 1995) or we may apply crossover repeatedly on the selected individuals.

There are algorithms where mixing is taken to an extreme. Algorithms such as PBIL (Baluja, 1994), the compact GA (Harik, Lobo, & Goldberg, 1998), and $(1, \lambda)$ -ES use a vector to represent the marginal distribution of alleles in the population. The new population is obtained by randomly generating individuals using the joint distribution represented by the vector. This process is analogous to crossover in a traditional algorithm, and, in a sense, the entire previous population indirectly participates in generating the new individuals. Qi and Palmieri (1993) proved that this mixing method is equivalent to using multiple two-parent recombinations on the selected individuals. These algorithms assume that the loci are independent from one another, and so they use a simple univariate probability distribution, but there are other algorithms such as the extended compact GA (Harik, 1999), the BOA (Pelikan, Goldberg, & Cantú-Paz, 1999) and the FDA (Mühlenbein & Mahnig, 1999) that use more sophisticated distributions to represent relationships between variables. In any case, these probabilistic modeling algorithms have the effect of mixing individuals more aggressively than the traditional two-parent recombinations, and are commonly used with strong selection methods that reduce the variance significantly. For example, the extended compact GA typically uses tournaments between 16 individuals, and the BOA and

FDA use truncation selection with $\tau \approx 0.5$ and $\tau \approx 0.3$, respectively.

5 Conclusions

Selection methods are essential components of evolutionary algorithms, and a deeper understanding of selection can lead to improve the robustness and efficiency of EAs. This paper presented numerical calculations that describe how different selection methods affect the fitness distributions of the population. These calculations were used to compare the various methods when they are configured to have the equivalent selection intensities.

The comparisons show that Boltzmann selection is very promising. It preserves the variance of the original population more than any other method with the same selection intensity, and it preserves the shape of the original distribution for a wide range of settings. However, it is not scale invariant, and care must be taken to choose an appropriate value of β .

There are multiple possibilities to extend this work. Possible extension include exploring selection in algorithms with generation gaps and in parallel EAs with multiple populations. Recent work on parallel GAs has established that when the individuals that migrate between populations are selected according to their fitness, the selection intensity increases (Cantú-Paz, In press), but there is no characterization of the variance or the higher cumulants.

We must emphasize that there is no single best selection method. The success of evolutionary algorithms depends on the balance between selection and the variation operators. It is quite possible that a selection method that works well in combination with certain operators on a particular problem may be a poor choice in a different setting. The calculations in this paper are a step toward making rational choices because they can suggest the use of operators that produce greater or lesser variation.

Acknowledgments

UCRL-JC-138582. This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

References

- Bäck, T. (1994). Selective pressure in evolutionary algorithms: A characterization of selection mechanisms. In *Proceedings of the First IEEE Conference on Evolutionary Computation*, Volume 1 (pp. 57–62). Piscataway, NJ: IEEE Service Center.
- Bäck, T. (1995). Generalized convergence models for tournament- and (μ, λ) -selection. In Eschelman, L. (Ed.), *Proceedings of the Sixth International Conference on Genetic Algorithms* (pp. 2–8). San Francisco, CA: Morgan Kaufmann.
- Bäck, T. (1996). *Evolutionary algorithms in theory and practice*. New York: Oxford University Press.
- Baker, J. E. (1985). Adaptive selection methods for genetic algorithms. In Grefenstette, J. J. (Ed.), *Proceedings of an International Conference on Genetic Algorithms and Their Applications* (pp. 101–111). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Baluja, S. (1994). *Population-based incremental learning: A method for integrating genetic search based function optimization and competitive learning* (Tech. Rep. No. CMU-CS-94-163). Pittsburgh, PA: Carnegie Mellon University.
- Beyer, H.-G. (1995). Toward a theory of evolution strategies: On the benefits of sex - the $(\mu/\mu,\lambda)$ theory. *Evolutionary Computation*, 3(1), 81–111.
- Blickle, T., & Thiele, L. (1996). A comparison of selection schemes used in evolutionary algorithms. *Evolutionary Computation*, 4(4), 361–394.
- Brindle, A. (1981). *Genetic algorithms for function optimization*. Unpublished doctoral dissertation, University of Alberta, Edmonton, Canada.
- Burrows, P. (1972). Expected selection differentials for directional selection. *Biometrics*, 23, 1091–1100.
- Burrows, P. (1975). Variances of selection differentials in normal samples. *Biometrics*, 31, 125–133.
- Cantú-Paz, E. (1999). Migration policies and takeover times in parallel genetic algorithms. In Banzhaf, W., Daida, J., Eiben, A. E., Garzon, M. H., Honavar, V., Jakiela, M., & Smith, R. E. (Eds.), *GECCO-99: Proceedings of the 1999 Genetic and Evolutionary Computation Conference* (pp. 775). San Francisco, CA: Morgan Kaufmann Publishers.
- Cantú-Paz, E. (In press). Migration policies, selection pressure, and parallel evolutionary algorithms. *Journal of Heuristics*.
- Chakraborty, U. K., Deb, K., & Chakraborty, M. (1996). Analysis of selection algorithms: A markov chain approach. *Evolutionary Computation*, 4(2), 133–167.
- De Jong, K. A. (1975). *An analysis of the behavior of a class of genetic adaptive systems*. Doctoral dissertation, University of Michigan, Ann Arbor. (University Microfilms No. 76-9381).
- de la Maza, M., & Tidor, B. (1993). An analysis of selection procedures with particular attention paid to proportional and boltzmann selection. In Forrest, S. (Ed.), *Proceedings of the Fifth International Conference on Genetic Algorithms* (pp. 124–131). San Mateo, CA: Morgan Kaufmann.
- Eiben, A. E., & Bäck, T. (1997). Empirical investigation of multiparent recombination operators in evolution strategies. *Evolutionary Computation*, 5(3), 347–365.
- Eiben, A. E., Raué, P.-E., & Ruttkay, Z. (1994). Genetic algorithms with multi-parent recombination. In Davidor, Y., Schwefel, H.-P., & Männer, R. (Eds.), *Parallel Problem Solving from Nature, PPSN III* (pp. 78–87). Berlin: Springer-Verlag.
- Goldberg, D. E. (1990). A note on Boltzmann tournament selection for genetic algorithms and population-oriented simulated annealing. *Complex Systems*, 4(4), 445–460. (Also TCGA Report No. 90003).
- Goldberg, D. E. (1998). *The race, the hurdle, and the sweet spot: Lessons from the genetic algorithms for the automation of design innovation and creativity* (IlliGAL Report No. 98007). Urbana, IL: University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory.
- Goldberg, D. E., & Deb, K. (1991). A comparative analysis of selection schemes used in genetic algorithms. *Foundations of Genetic Algorithms*, 1, 69–93. (Also TCGA Report 90007).
- Goldberg, D. E., Deb, K., & Thierens, D. (1993). Toward a better understanding of mixing in genetic algorithms. *Journal of the Society of Instrument and Control Engineers*, 32(1), 10–16.

- Hancock, P. J. B. (1994). An empirical comparison of selection methods in evolutionary algorithms. In Fogarty, T. C. (Ed.), *Evolutionary Computing: AISB Workshop 1994* (pp. 80–94). Springer-Verlag.
- Harik, G. (1999). *Linkage learning via probabilistic modeling in the ECGA* (IlligAL Report No. 99010). Urbana, IL: University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory.
- Harik, G., Cantú-Paz, E., Goldberg, D., & Miller, B. L. (1999). The gambler's ruin problem, genetic algorithms, and the sizing of populations. *Evolutionary Computation*, 7(3), 231–253.
- Harik, G. R., Lobo, F. G., & Goldberg, D. E. (1998). The compact genetic algorithm. In of Electrical, I., & Engineers, E. (Eds.), *Proceedings of 1998 IEEE International Conference on Evolutionary Computation* (pp. 523–528). Piscataway, NJ: IEEE Service Center.
- Harter, H. L. (1970). *Order statistics and their use in testing and estimation*. Washington, D.C.: U.S. Government Printing Office.
- Holland, J. H. (1973). Genetic algorithms and the optimal allocation of trials. *SIAM Journal on Computing*, 2(2), 88–105.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.
- Mahfoud, S. W. (1993). Finite Markov chain models of an alternative selection strategy for the genetic algorithm. *Complex Systems*, 7(2), 155–170.
- Miller, B. L., & Goldberg, D. E. (1995). Genetic algorithms, tournament selection, and the effects of noise. *Complex Systems*, 9(3), 193–212.
- Mühlenbein, H., & Mahnig, T. (1999). FDA—a scalable evolutionary algorithm for the optimization of additively decomposed functions. *Evolutionary Computation*, 7(4), 353–376.
- Mühlenbein, H., & Schlierkamp-Voosen, D. (1993). Predictive models for the breeder genetic algorithm: I. Continuous parameter optimization. *Evolutionary Computation*, 1(1), 25–49.
- Mühlenbein, H., & Schlierkamp-Voosen, D. (1994). The science of breeding and its application to the breeder genetic algorithm (BGA). *Evolutionary Computation*, 1(4), 335–360.
- Pelikan, M., Goldberg, D. E., & Cantú-Paz, E. (1999). BOA: The Bayesian optimization algorithm. In Banzhaf, W., Daida, J., Eiben, A. E., Garzon, M. H., Honavar, V., Jakiela, M., & Smith, R. E. (Eds.), *GECCO-99: Proceedings of the 1999 Genetic and Evolutionary Computation Conference* (pp. 525–532). San Francisco, CA: Morgan Kaufmann Publishers.
- Prügel-Bennett, A., & Shapiro, J. L. (1994). An analysis of a genetic algorithm using statistical mechanics. *Physical Review Letters*, 72(9), 1305–1309.
- Qi, X., & Palmieri, F. (1993). The diversification role of crossover in the genetic algorithms. In Forrest, S. (Ed.), *Proceedings of the Fifth International Conference on Genetic Algorithms* (pp. 132–137). San Mateo, CA: Morgan Kaufmann.
- Ratray, M., & Shapiro, J. (1997). Noisy fitness evaluation in genetic algorithms and the dynamics of learning. In *Foundations of Genetic Algorithms 4* (pp. 117–139). San Francisco: Morgan Kaufmann.
- Rogers, A., & Prügel-Bennett, A. (1999). Modelling the dynamics of a steady state genetic algorithm. In *Foundations of Genetic Algorithms 5* (pp. 57–68). San Francisco, CA: Morgan Kaufmann.

- Rudolph, G. (2000). Takeover times and probabilities of non-generational selection rules. In Beyer, H.-G., Cantú-Paz, E., Goldberg, D. E., Parmee, I., & Spector, L. (Eds.), *GECCO-2000: Proceedings of the 2000 Genetic and Evolutionary Computation Conference*. San Francisco, CA: Morgan Kaufmann Publishers.
- Sakamoto, Y., & Goldberg, D. E. (1997). Takeover time in a noisy environment. In Bäck, T. (Ed.), *Proceedings of the Seventh International Conference on Genetic Algorithms* (pp. 160–165). San Francisco: Morgan Kaufmann.
- Schwefel, H. (1981). *Numerical optimization of computer models*. Chichester: John Wiley and Sons.
- Shapiro, J., Prügel-Bennett, & Rattray, M. (1994). A statistical mechanical formulation of the dynamics of genetic algorithms. In Fogarty, T. C. (Ed.), *Evolutionary Computing: AISB Workshop 1994* (pp. 17–27). Springer-Verlag.
- Smith, J., & Vavak, F. (1999). Replacement strategies in steady state genetic algorithms: Static environments. In *Foundations of Genetic Algorithms 5* (pp. 219–234). San Francisco, CA: Morgan Kaufmann.
- Thierens, D. (1997). Selection schemes, elitist recombination, and selection intensity. In Bäck, T. (Ed.), *Proceedings of the Seventh International Conference on Genetic Algorithms* (pp. 152–159). San Francisco: Morgan Kaufmann.