
Analysis of Selection Algorithms: A Markov Chain Approach^{1,2}

Uday Kumar Chakraborty

Reader
Department of Computer Science
and Engineering
Jadavpur University
Calcutta 700032, India
uday@jadav.ernet.in

Kalyanmoy Deb

Assistant Professor
Department of Mechanical Engineering
Indian Institute of Technology
Kanpur, UP 208016, India
deb@iitk.ernet.in

Mandira Chakraborty

Lecturer
Department of Computer Science
and Engineering
Jadavpur University
Calcutta 700032, India
mc@jadav.ernet.in

Abstract

A Markov chain framework is developed for analyzing a wide variety of selection techniques used in genetic algorithms (GAs) and evolution strategies (ESs). Specifically, we consider linear ranking selection, probabilistic binary tournament selection, deterministic s -ary ($s = 3, 4, \dots$) tournament selection, fitness-proportionate selection, selection in Whitley's GENITOR, selection in (μ, λ) -ES, selection in $(\mu + \lambda)$ -ES, (μ, λ) -linear ranking selection in GAs, $(\mu + \lambda)$ -linear ranking selection in GAs, and selection in Eshelman's CHC algorithm. The analysis enables us to compare and contrast the various selection algorithms with respect to several performance measures based on the probability of takeover. Our analysis is exact—we do not make any assumptions or approximations. Finite population sizes are considered. Our approach is perfectly general, and following the methods of this paper, it is possible to analyze any selection strategy in evolutionary algorithms.

Keywords

Genetic algorithms, evolution strategies, selection, takeover, Markov chain.

1. Introduction

Selection is a central component of the genetic algorithm (GA) (Holland, 1975; Goldberg 1989a). In all variants of the GA, some form of the selection operator must be present. Selection pressure directly controls the exploitation factor in the “exploitation-versus-exploration” trade-off that is believed to be critically important in the working of the GA.

¹ A part of this work appeared in Chakraborty (1995c).

² Correspondence to: U. K. Chakraborty, 19 Kalibari Road, Santoshpur, Calcutta 700075, India.

A wide variety of selection algorithms have been proposed in the GA literature. In addition, selection methods that were traditionally used in evolution strategies (ESs) (Schwefel, 1981) have now been extended (Bäck & Hoffmeister, 1991) to GAs. Despite the existence of an impressive body of empirical studies establishing the superiority of one selection algorithm over another, there is a dearth of exact probabilistic analysis of the behavior of different selection algorithms. Goldberg and Segrest (1987) have analyzed fitness-proportionate selection in a finite-population GA using Markov chain theory under the assumption that the GA manipulates single-locus individuals (1 and 0) only. Baker (1989) provides extensive empirical comparisons of GAs that use a broad range of selection schemes, including fitness-proportionate, linear ranking, and hybrid strategies. The interaction between selection pressure and implicit parallelism has been examined by Grefenstette and Baker (1989). Whitley (1989) empirically demonstrates the effectiveness of linear ranking coupled with one-at-a-time reproduction in the context of his GENITOR system. Syswerda (1991) has studied generational and steady-state reproduction mechanisms experimentally. Eshelman (1991) has experimentally demonstrated the usefulness of a population-elitist selection strategy in his CHC algorithm. Deterministic analyses of the expected behavior of proportionate, linear ranking, tournament, and GENITOR selections have been presented by Goldberg and Deb (1991). In that study the above-mentioned four selection algorithms have been compared on the basis of the following metrics: (1) expected growth ratio, (2) takeover time, and (3) time complexity of the selection algorithms. Building on earlier work (De Jong, 1975; Holland, 1975; Goldberg & Deb, 1991), De Jong and Sarma (1993) have analyzed the effect of generation gap on selection using deterministic, expected-value equations. In a study by de la Maza and Tidor (1993), proportional and Boltzmann selections have been discussed with reference to the translation- and scale-invariance properties.

In this paper we present stochastic analyses of a wide spectrum of selection algorithms used in GAs and ESs. Specifically, we examine the following selection schemes:

- linear ranking selection,
- binary tournament selection (probabilistic),
- s -ary tournament selection (deterministic), $s = 3, 4, \dots$,
- fitness-proportionate selection,
- selection in GENITOR (or “steady-state” GA),
- (μ, λ) -selection in ESs,
- $(\mu + \lambda)$ -selection in ESs,
- (μ, λ) -linear ranking selection in GAs,
- $(\mu + \lambda)$ -linear ranking selection in GAs,
- selection in CHC.

The usefulness of the Markov chain approach in analyzing GA behavior has long been appreciated by researchers (De Jong, 1975; Goldberg & Segrest, 1987). Finite Markov chains have been used to study specific features, such as niching in a single-locus GA (Horn, 1993) and Boltzmann tournament selection (Mahfoud, 1993). Vose (1993), and Nix and Vose (1992) have investigated convergence properties of the simple GA by analyzing the asymptotic behavior of finite-population GAs with reference to the fixed points of an infinite-population

model. The Vose model incorporates all three operators—proportionate selection, single-point crossover, and bitwise mutation—in a single transition matrix. A Markov model of the simple three-operator GA has also been developed by Davis and Principe (1993), who consider the asymptotics of steady-state distributions as the mutation rate decreases. Suzuki (1993) investigates GA convergence by developing a Markov model for a GA with a “modified” elitist strategy. Asoh and Mühlenbein (1994) have developed a number of empirical formulas for the mean time to absorption in genetic drift (no selection pressure and uniform crossover) for the multiple-loci case.

The focus of the present paper is on selection alone (no crossover or mutation). We develop a uniform Markov chain framework and introduce a number of measures of selection pressure for analyzing and comparing the selection schemes. This analysis covers almost all types of selection algorithms used in GAs and ESs, ranging from generational to steady-state, deterministic to probabilistic, elitist to nonelitist, dynamic to static, and preservative to extinctive. The approach is perfectly general, and following the methods of this paper, it should be possible to analyze any selection strategy in evolutionary algorithms. The analysis is exact—we do not make any assumptions or approximations. Finite population sizes are considered.

The remainder of this paper is organized as follows. In Section 2, a framework for analysis is developed by defining the takeover probability and the states of the Markov model and by describing the experimental setup for simulations. Sections 3 through 8 analyze various selection algorithms. In Section 9, we compare and contrast different selection algorithms with reference to a number of (probabilistic) performance measures. Finally, in Section 10 we present the summary and conclusions.

2. Framework for Analysis

In this section we define the probability of takeover, set up our Markov chain model, and describe the experimental procedure for carrying out the simulations.

2.1 The Markov Chain Model and the Probability of Takeover

Takeover (Goldberg & Deb, 1991) is the situation when the entire population is filled with members of a particular class, at least one member of the class being present in the initial population. In Chakraborty and Dastidar (1993) and Chakraborty (1995a,b), the term *takeover* has been used in a similar sense. The term *convergence* has been used in a related context (e.g., Goldberg & Segrest, 1987; Goldberg, 1989b; Mühlenbein & Schlierkamp-Voosen, 1993). Convergence thus used means approaching a homogeneous population, and should not be confused with “tending to the global optimum” (e.g., Eiben, Aarts, & Van Hee, 1991).

Let N denote the population size. We consider those individuals in the population whose fitness values are equal to or better than a predetermined value. To investigate the effect of various selection algorithms, we will study the growth/decay of individuals of this designated class in the population.

Let $X^{(t)}$ be the random variable representing the number of members of the designated class present in the population at time (generation) t . The stochastic process $\{X^{(t)}\}$ is a discrete-parameter, discrete-state process with the state space $\{0, 1, 2, \dots, N\}$ and the index set $\{0, 1, 2, 3, \dots\}$. For the GA, the successive random variables $X^{(0)}, X^{(1)}, X^{(2)}, \dots$, form a Markov chain.

Under selection alone (no crossover or mutation), for all $k \geq 0$,

$$\text{Prob}(X^{(n+k)} = 0 | X^{(n)} = 0) = 1, \quad \text{Prob}(X^{(n+k)} = N | X^{(n)} = N) = 1.$$

In other words, the states 0 and N are absorbing, and this is true for all selection algorithms.

The behavior of any selection algorithm can be completely known if we have (1) the transition probability matrix $\mathcal{P} = [p_{ij}]$, and (2) the initial distribution, $\text{Prob}(X^{(0)} = i)$ for $i = 0, 1, \dots, N$. The t -step transition probabilities $p_{ij}^{(t)}$ are obtained as the elements of \mathcal{P}^t .

Note that the “designated class” does not necessarily refer to the best fitness in the search space. Given (1) any arbitrary fitness function, and (2) any arbitrary method (e.g., random) of creating the initial population, one can generate the initial population (generation 0) of individuals. Now, given an arbitrary cutoff value of the fitness, one can count how many individuals in the initial population possess a fitness better than or equal to this cutoff value. Let this number be i . The designated class now consists of these i individuals, and the takeover probabilities will be obtained from

$$p_{\text{takeover}}(t) \equiv p_{iN}^{(t)}.$$

However, for selection in (μ, λ) -ES and $(\mu + \lambda)$ -ES (Section 7.1), we take

$$p_{\text{takeover}}(t) \equiv p_{i\mu}^{(t)}.$$

When one considers the growth of the best individual in the initial population (e.g., Bäck, 1994), takeover is equivalent to convergence to a homogeneous population (i.e., a population with N copies of the best individual in the initial generation). The present paper uses a more general notion (the designated class consists of individuals of better than a given fitness), and thus at takeover the population composition need not necessarily be homogeneous.

Note that the present definition of “state” is totally different from that used in the Markov models of Nix and Vose (1992), Davis and Principe (1993), and De Jong, Spears, and Gordon (1995). The fundamental difference between those approaches and ours is that our model considers selection alone, and without crossover or mutation, no new points can be generated.

DEFINITION 1: Let $p_{\text{takeover}}(t)$ stand for the probability that takeover occurs at or before generation t ; $p_{\text{takeover}}(0) = 0$.

DEFINITION 2: Let $f_{\text{takeover}}(t)$ denote the probability that the first occurrence of takeover is at generation t ; $f_{\text{takeover}}(0) = 0$.

Because the state N is absorbing, we have

$$p_{\text{takeover}}(t) = \sum_{i=0}^t f_{\text{takeover}}(i)$$

and

$$f_{\text{takeover}}(t) = p_{\text{takeover}}(t) - p_{\text{takeover}}(t-1)$$

for $t = 1, 2, \dots$.

Later in this paper we will see that two situations are possible:

$$1. \quad \sum_{t=0}^{\infty} f_{\text{takeover}}(t) = 1$$

In this case, the $\{f_{\text{takeover}}(t)\}$ is a proper probability distribution, and we refer to it as the *takeover time distribution*. The *mean takeover time* is then given by $\sum_{t=0}^{\infty} t f_{\text{takeover}}(t)$.

$$2. \quad \sum_{t=0}^{\infty} f_{\text{takeover}}(t) < 1$$

In this case, the $\{f_{\text{takeover}}(t)\}$ is not a proper probability distribution, and therefore the mean takeover time is not meaningful.

Most previous studies on takeover have used the *takeover time* as the metric for comparison. A comment on why we do not use it here is in order. Most of the existing analyses make an (implicit or explicit) assumption that the best individual never vanishes from the population. This is true only for the elitist strategies [e.g., GENITOR, $(\mu + \lambda)$ -selection] (or in infinite-population models). For the nonelitist schemes (e.g., linear ranking, tournament), there is a nonzero probability that the best individual will disappear from the (finite) population, and as discussed above, it does not make sense to discuss the mean (expected) takeover time for these algorithms.

The nature of the change of $p_{\text{takeover}}(t)$ with t tells us a great deal about the selection “push” of any given selection algorithm. The different selection procedures are characterized by their respective p_{takeover} versus t behaviors. In the sections that follow, we derive the transition probability expressions for each of the selection algorithms under consideration and compute the p_{takeover} values. Later in this paper we introduce several measures (based on p_{takeover}) of selection pressure, and using these measures, we compare and contrast the selection techniques.

In this paper the Markov chain calculations and the simulations were performed with a single member of the designated class in the initial population (this is the case considered in most previous studies), but the present analysis is completely general and holds for any number ($\leq N$) of initial members.

2.2 The Experiment

To obtain empirical support for our model, experiments were conducted by running the GA on a simulation problem where the fitness of each binary-coded individual was its decoded decimal value plus 1. This problem is similar to those used by Goldberg and Deb (1991) and De Jong, Spears, and Gordon (1995). No crossover or mutation was applied. The problem was constructed as a maximization function. The simulation was started with a single member of the best individual in the initial generation. Each GA run was continued until the best individual either filled the entire population or vanished from the population. One thousand independent runs—with as many different initial seeds for the pseudorandom number generator—were taken. All 1000 runs were started with the same initial population. For each run, the generation number at which the best individual took over or vanished was noted. The takeover probability was obtained from the relation

$$p_{\text{takeover}}(t) = \frac{n(t)}{1000}$$

where $n(t)$ = total number of runs (of 1000) in which takeover occurred at or before generation t .

Note that the theory presented in this paper is independent of the experimental setup. The same results would be obtained if, for example, a nonlinear fitness function were used or the initial population were randomly generated. The actual fitness values do not affect the takeover probabilities (except in the case of proportional selection). We chose this particular problem for the simulations because it is one of the simplest possible fitness functions in the literature.

3. Linear Ranking Selection

In ranking selection (Baker, 1985; Grefenstette & Baker, 1989; Goldberg & Deb, 1991) at each generation, the individuals in the population are sorted according to their fitness, and each individual is assigned a rank in the sorted population. The worst individual gets the rank 1, whereas the best receives the rank N . The selection probabilities of the individuals x_k ($k = 1, \dots, N$) are given by some function (most commonly linear) of their rank.

Let $\{x_1^{(t)}, x_2^{(t)}, \dots, x_N^{(t)}\}$ denote the population at generation t . Then, in linear ranking selection, the probability of selecting individual x_k ($k = 1, 2, \dots, N$) is given by

$$p_{\text{sel}}(x_k^{(t)}) = \frac{1}{N} \left(\min + \frac{(\max - \min)(\text{rank}(x_k^{(t)}) - 1)}{N - 1} \right)$$

where $\max + \min = 2$, and $1 \leq \max \leq 2$. The $\{p_{\text{sel}}(x_k^{(t)})\}$ is a proper probability distribution ($\sum_{k=1}^N p_{\text{sel}}(x_k^{(t)}) = 1$ for each t), and sampling N individuals according to this probability distribution yields the next generation.

Let i denote the total number of individuals in the designated class (Section 2) at any generation. Then, the ranks of these individuals (at that generation) are N (the best), $N - 1$ (the second best), $N - 2$ (the third best), \dots , $N - i + 1$ (the i th best). Now the probability that, in a single trial, a member of this class would be selected is given by

$$p_{\text{select}}(i) = \sum_{j=N-i+1}^N \frac{1}{N} \left(\min + \frac{(\max - \min)(j - 1)}{N - 1} \right)$$

On substitution of $2 - \max$ for \min and simplification, we get

$$p_{\text{select}}(i) = \frac{i}{N} \left(\max - \frac{(\max - 1)(i - 1)}{N - 1} \right) \quad (1)$$

The reproduction process for the entire population at each generation can be thought of as a sequence of N independent Bernoulli trials (with replacement) where, in each trial, the “probability of success” is $p_{\text{select}}(i)$. It follows that the transition probabilities p_{ij} ($i, j = 0, 1, \dots, N$) are given by

$$p_{ij} = \begin{cases} 1 & \text{for } p_{\text{select}}(i) = 0; j = 0 \\ 0 & \text{for } p_{\text{select}}(i) = 0; j = 1, 2, \dots, N \\ 1 & \text{for } p_{\text{select}}(i) = 1; j = N \\ 0 & \text{for } p_{\text{select}}(i) = 1; j = 0, 1, \dots, N - 1 \\ \binom{N}{j} (p_{\text{select}}(i))^j (1 - p_{\text{select}}(i))^{N-j} & \text{for } 0 < p_{\text{select}}(i) < 1; j = 0, 1, \dots, N \end{cases} \quad (2)$$

where $p_{\text{select}}(i)$ is given by equation 1.

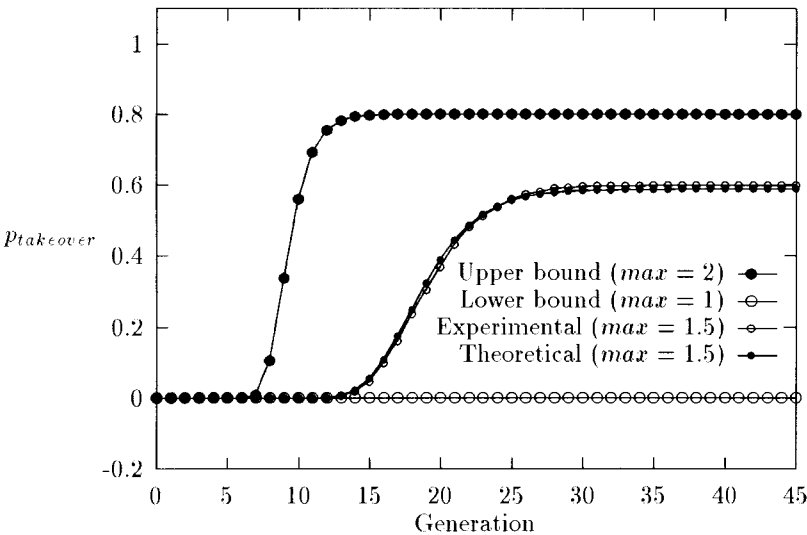


Figure 1. Growth of $p_{takeover}$ with generation in linear ranking selection ($N = 128$).

Noting that $p_{select}(i)$ is independent of time, we see from Equations 1 and 2 that the Markov chain for generation-based linear ranking selection is finite, time homogeneous, aperiodic, and absorbing with two absorbing states and $N - 1$ transient states.

It is evident from Equation 2 that \max ($1 \leq \max \leq 2$) and N are the two parameters affecting the performance of linear ranking selection. For a fixed population size, \max is the only parameter that controls the selection pressure. For $\max = 1$, there is no selection pressure in favor of the individuals of the designated class, and linear ranking selection then reduces to uniform random selection. For $\max = 2$, on the other hand, one attains the highest possible pressure. In Figure 1 we show the variation of $p_{takeover}$ values corresponding to $\max = 1$ and $\max = 2$. This pair of plots defines the upper and lower bounds of the selection pressure that is ever possible to achieve with linear ranking (at population size = 128).

In Figure 1 we also show a representative plot (corresponding to $\max = 1.5$ and population size = 128) of the theoretical $p_{takeover}$ values, along with a plot of the experimentally obtained values. Because of finite sampling, the two plots, although very similar, are not exactly identical. The theoretical and empirical $f_{takeover}$ values for this pair are shown in Table 1, where the standard deviations and the 95% confidence intervals are also presented (number of samples = 1000; sample mean = relative frequency). The empirical probabilities are not statistically significantly different from the corresponding true values at the 5% level of significance (i.e., with 95% confidence). (This is true for all the selection algorithms discussed in this paper.) Most figures in this paper have been rounded to three decimal places. Confidence intervals indicated by “—” in Table 1 (and also in Tables 2–7) correspond to values of $f_{takeover}$ close to zero for which the product of the number of samples and the probability is less than 5. In these cases, the normal distribution approximation (to the binomial distribution) underlying the confidence interval calculation process is not applicable (Bhattacharyya & Johnson, 1977; Marascuilo & McSweeney, 1977).

Table 1. Theoretical and experimental f_{takeover} values in linear ranking selection ($N = 128$, $\max = 1.5$).

Gen.	Theoretical	Sample Mean	Sample Std. Dev.	95% Conf. Intvl.
1	0.000	0		
⋮	⋮	⋮		
11	0.000	0		
12	0.001	0.001	0.032	—
13	0.006	0.006	0.077	(0.001, 0.011)
14	0.016	0.013	0.113	(0.006, 0.020)
15	0.033	0.026	0.159	(0.016, 0.036)
16	0.052	0.053	0.224	(0.039, 0.067)
17	0.067	0.062	0.241	(0.047, 0.077)
18	0.075	0.077	0.267	(0.060, 0.094)
19	0.074	0.068	0.252	(0.052, 0.084)
20	0.065	0.063	0.243	(0.048, 0.078)
21	0.054	0.063	0.243	(0.048, 0.078)
22	0.042	0.048	0.214	(0.035, 0.061)
23	0.032	0.032	0.176	(0.021, 0.043)
24	0.022	0.025	0.156	(0.015, 0.035)
25	0.016	0.022	0.147	(0.013, 0.031)
26	0.011	0.013	0.113	(0.006, 0.020)
27	0.007	0.007	0.083	(0.002, 0.012)
28	0.005	0.011	0.104	(0.005, 0.017)
29	0.003	0.002	0.045	—
30	0.002	0.003	0.055	—
31	0.001	0.001	0.032	—
32	0.001	0		—
33	0.001	0.001	0.032	—
34	0.000	0		
35	0.000	0.001	0.032	—
36	0.000	0		
⋮	⋮	⋮		

4. Tournament Selection

Although there are a number of variants, the basic mechanism of tournament selection (Brindle, 1981; Goldberg, Korb, & Deb, 1989) consists of

- Randomly choosing, with or without replacement, a predetermined number of individuals from the population and picking, probabilistically or deterministically, the best of these individuals.

- Repeating the above step N times (N = population size) to fill the next generation.

In Sections 4.1 and 4.2, binary and larger tournaments are considered.

4.1 Binary Tournament

Let the population at any generation contain i members of the designated class. Then, under binary tournament with replacement, in each tournament any one of the following four mutually exclusive and exhaustive events will occur:

1. Both the individuals participating in the tournament belong to the designated class.
2. Neither of the two individuals belongs to the designated class.
3. The first individual is from the designated class, but the second is not.
4. The second individual is from the designated class, but the first is not.

It is easy to see that

$$\begin{aligned} \text{Prob(event 1)} &= \left(\frac{i}{N}\right) \cdot \left(\frac{i}{N}\right) \\ \text{Prob(event 2)} &= \left(1 - \frac{i}{N}\right) \cdot \left(1 - \frac{i}{N}\right) \\ \text{Prob(event 3)} &= \left(\frac{i}{N}\right) \cdot \left(1 - \frac{i}{N}\right) \\ \text{Prob(event 4)} &= \left(1 - \frac{i}{N}\right) \cdot \left(\frac{i}{N}\right) \end{aligned}$$

If p denotes the probability with which the better of the two individuals is chosen in any single tournament, then the probability that in any single tournament a member of the designated class is picked is given by

$$p_{\text{select}}(i) = \left(\frac{i}{N}\right)^2 + 2\frac{i}{N}\left(1 - \frac{i}{N}\right)p \quad (3)$$

where $0.5 < p \leq 1$. Noting that the reproduction process for the entire population is complete after N such tournaments have been held, we have the following transition probabilities p_{ij} ($i, j = 0, 1, \dots, N$):

$$p_{ij} = \begin{cases} 1 & \text{for } i = 0; \quad j = 0 \\ 0 & \text{for } i = 0; \quad j = 1, 2, \dots, N \\ 1 & \text{for } i = N; \quad j = N \\ 0 & \text{for } i = N; \quad j = 0, 1, \dots, N - 1 \\ \binom{N}{j} (p_{\text{select}}(i))^j (1 - p_{\text{select}}(i))^{N-j} & \text{otherwise} \end{cases} \quad (4)$$

where $p_{\text{select}}(i)$ is given by Equation 3. As in the case of linear ranking, the Markov chain for binary tournament is finite, aperiodic, and absorbing with two absorbing states (0 and N) and $N - 1$ transient states. Equations 3 and 4 show that the p_{ij} 's are time invariant. Thus, the chain is time homogeneous.

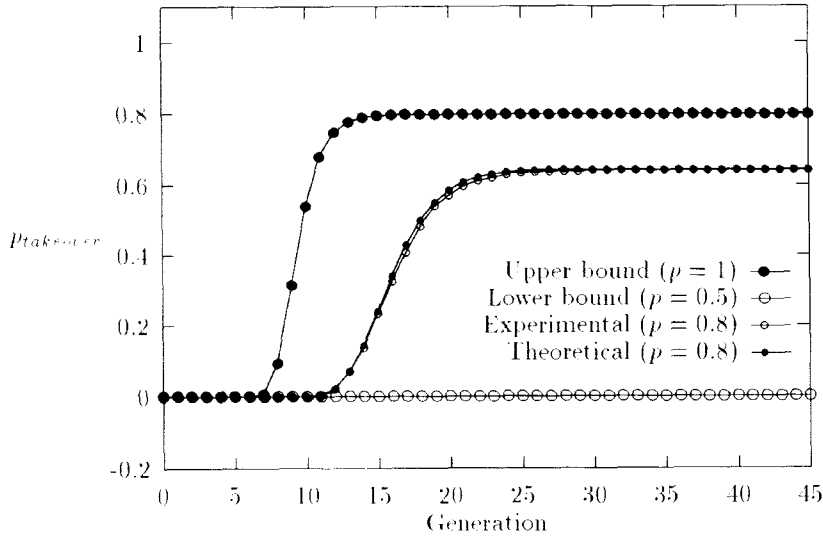


Figure 2. Growth of p_{takeover} with generation in binary tournament selection ($N = 128$).

From Equations 3 and 4 we see that the behavior of binary tournament selection is controlled by two external parameters: p ($0.5 < p \leq 1$) and N . In Figure 2 we show (for $N = 128$) the lower and upper bounds of p_{takeover} values corresponding, respectively, to $p = 0.5$ and $p = 1$. Also shown in Figure 2 is one theoretical–experimental pair for a representative case ($p = 0.8$ and $N = 128$). Table 2 shows the finite sampling effects for this case.

4.2 Deterministic s -ary Tournaments ($s > 2$)

We consider deterministic tournaments (Goldberg & Deb, 1991) of an arbitrary size s ($s > 2$). If at any generation the designated class contains i members, then the probability that in any single tournament (with replacement) none of the s participating individuals is from this class is

$$\left(\frac{N-i}{N}\right)^s$$

and thus,

$$p_{\text{select}}(i) = 1 - \left(\frac{N-i}{N}\right)^s \quad (5)$$

With $p_{\text{select}}(i)$ given by the above equation, the transition probabilities p_{ij} ($i, j = 0, 1, \dots, N$) are given by Equation 6:

$$p_{ij} = \begin{cases} 1 & \text{for } i = 0; \quad j = 0 \\ 0 & \text{for } i = 0; \quad j = 1, 2, \dots, N \\ 1 & \text{for } i = N; \quad j = N \\ 0 & \text{for } i = N; \quad j = 0, 1, \dots, N-1 \\ \binom{N}{j} (p_{\text{select}}(i))^j (1 - p_{\text{select}}(i))^{N-j} & \text{otherwise} \end{cases} \quad (6)$$

Table 2. Theoretical and experimental f_{takeover} values in binary tournament selection ($N = 128, p = 0.8$).

Gen.	Theoretical	Sample Mean	Sample Std. Dev.	95% Conf. Intvl.
1	0.000	0		
⋮	⋮	⋮		
9	0.000	0		
10	0.001	0		—
11	0.004	0.002	0.045	—
12	0.018	0.020	0.140	(0.011, 0.029)
13	0.045	0.048	0.214	(0.035, 0.061)
14	0.076	0.067	0.250	(0.052, 0.082)
15	0.096	0.097	0.296	(0.079, 0.115)
16	0.101	0.092	0.289	(0.074, 0.110)
17	0.088	0.081	0.273	(0.064, 0.098)
18	0.070	0.076	0.265	(0.060, 0.092)
19	0.051	0.057	0.232	(0.043, 0.071)
20	0.035	0.029	0.168	(0.019, 0.039)
21	0.023	0.027	0.162	(0.017, 0.037)
22	0.014	0.015	0.122	(0.007, 0.023)
23	0.009	0.009	0.094	(0.003, 0.015)
24	0.005	0.009	0.094	(0.003, 0.015)
25	0.003	0.005	0.071	(0.001, 0.009)
26	0.002	0.002	0.045	—
27	0.001	0.001	0.032	—
28	0.001	0.001	0.032	—
29	0.000	0		
30	0.000	0.002	0.045	—
31	0.000	0		
⋮	⋮	⋮		

From Equations 5 and 6 it is clear that the parameter affecting the performance of s -ary deterministic tournament selection is N (population size). As a cross-check, one can easily verify that for deterministic binary tournaments ($p = 1, s = 2$), Equations 3 and 5 become identical. Figure 3 (and Table 3) and Figure 4 show several representative cases.

5. GENITOR

In GENITOR (Whitley & Kauth, 1988; Whitley, 1989), an individual is chosen for offspring generation according to linear ranking, and the offspring replaces the current worst individual in the population.

As before, let there be i individuals in the designated class. When an individual is chosen from the population following the linear ranking scheme, the probability that it would be

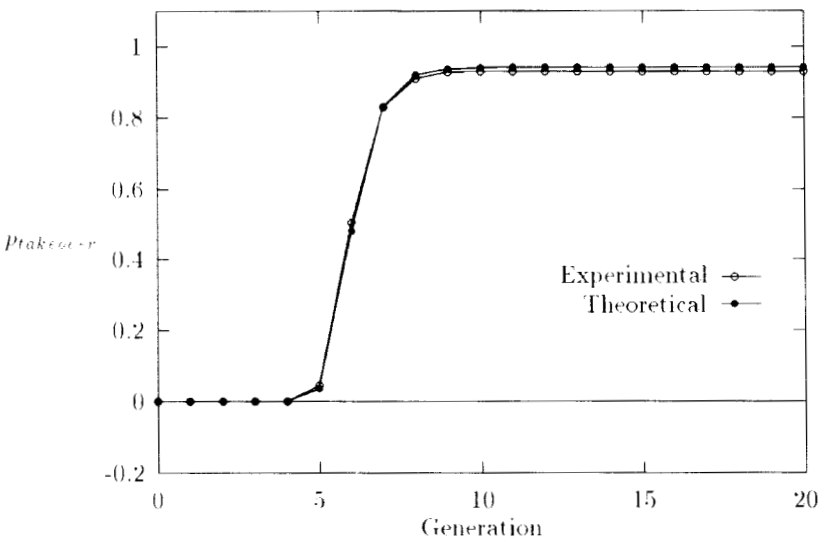


Figure 3. Comparison of theoretical and experimental $p_{takeover}$ for ternary tournament selection ($N = 128$).

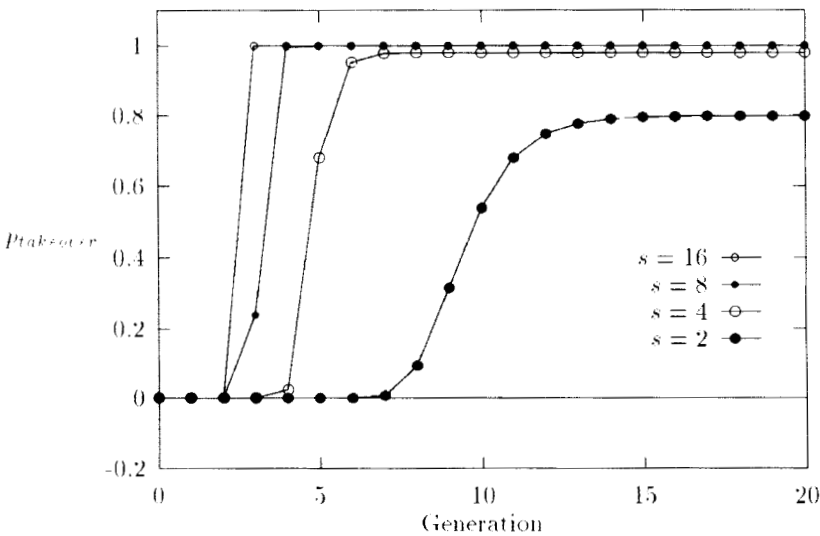


Figure 4. Growth of $p_{takeover}$ for tournaments of various sizes ($N = 128$).

Table 3. Theoretical and experimental f_{takeover} values in deterministic ternary tournament ($N = 128$).

Gen.	Theoretical	Sample Mean	Sample Std. Dev.	95% Conf. Intvl.
1	0.000	0		
\vdots	\vdots	\vdots		
4	0.000	0		
5	0.036	0.045	0.207	(0.032, 0.058)
6	0.444	0.459	0.498	(0.428, 0.490)
8	0.089	0.080	0.271	(0.063, 0.097)
9	0.017	0.019	0.137	(0.011, 0.027)
10	0.003	0.002	0.045	—
11	0.001	0		—
12	0.000	0		
\vdots	\vdots	\vdots		

from this class is given by Equation 1 (see Section 3). We can think of the reproduction process of GENITOR as a sequence of choose-and-insert iterations, where each choose-and-insert iteration consists of a probabilistic choice operation followed by a deterministic replacement operation. We recognize that corresponding to each choose-and-insert iteration, any one of the following three exhaustive and mutually exclusive events may occur (we use the phrase “class member” to denote any individual belonging to the designated class):

1. A class member is chosen, and a nonclass member is removed.
2. A nonclass member is chosen, and a nonclass member is removed.
3. A class member is chosen, and a class member is removed (this can happen only when the entire population has been filled with members of the designated class).

Note that a fourth possibility—a nonclass member is picked, and a class member is removed—can never materialize because removal is deterministic and of the current worst. It follows that the transition probabilities for transition to the next iteration (not to be confused with the transition to the next generation; for GENITOR, one generation = N iterations) are given by

$$p_{ij, \text{iteration}} = \begin{cases} 1 & \text{for } i = 0; \quad j = 0 \\ 0 & \text{for } i = 0; \quad j = 1, 2, \dots, N \\ 1 & \text{for } i = N; \quad j = N \\ 0 & \text{for } i = N; \quad j = 0, 1, \dots, N - 1 \\ 1 - p_{\text{select}}(i) & \text{for } i \neq 0, N; \quad j = i \\ p_{\text{select}}(i) & \text{for } i \neq 0, N; \quad j = i + 1 \\ 0 & \text{for } i \neq 0, N; \quad j \neq i, i + 1 \end{cases} \quad (7)$$

where $p_{\text{select}}(i)$ is given by Equation 1.

LEMMA 1: $\mathcal{P}_{\text{iteration}} = [p_{ij, \text{iteration}}]$ is a band matrix.

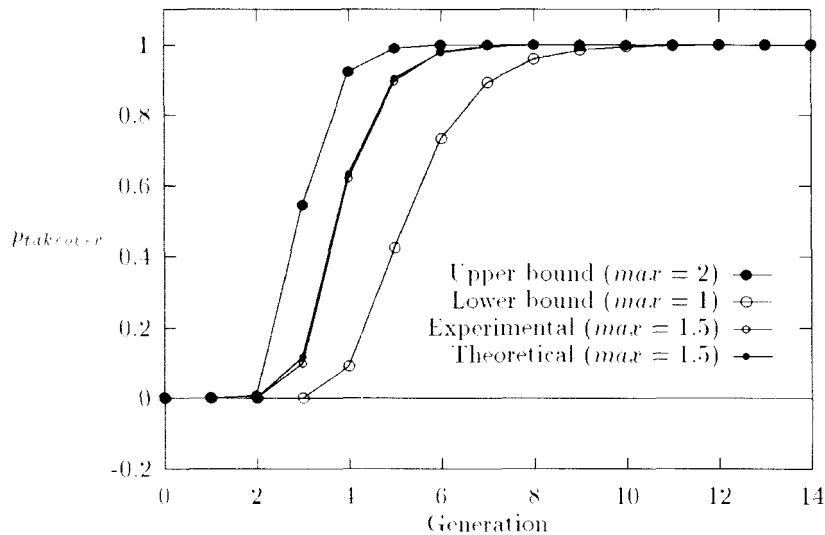


Figure 5. Growth of p_{takeover} with generation in selection in GENITOR ($N = 128$).

PROOF: Follows from Equation 7.

The generational transition probability matrix of GENITOR is given by

$$\mathcal{P}_{\text{generation}} = \mathcal{P}_{\text{iteration}}^N \quad (8)$$

We now have the following lemma:

LEMMA 2: *The (one-step) transition probability matrix $\mathcal{P}_{\text{generation}}$ is upper triangular.*

PROOF: Follows from Equation 8 and Lemma 1.

GENITOR has a finite, time-homogeneous, aperiodic, absorbing Markov chain with two absorbing states and $N - 1$ transient states. Equations 1 and 7 show that \max ($1 \leq \max \leq 2$) and N are the two parameters controlling GENITOR's performance. In Figure 5 we show the plots corresponding to $\max = 2$ and $\max = 1$. These two plots determine the upper and lower bounds of selection pressure that are possible to obtain with GENITOR (at population size = 128). Figure 5 also shows a representative pair of theoretical and empirical p_{takeover} plots (the corresponding f_{takeover} values are presented in Table 4).

Following the methods of this section it is easy to analyze the steady-state genetic algorithm (SSGA) of Syswerda (1989, 1991), which is very similar to GENITOR, except that (1) instead of ranking, the SSGA uses the fitness-proportionate scheme; and (2) in addition to the principle of deleting the worst individual, SSGA variants may use a number of alternative deletion methods [e.g., random (Chakraborty, 1995c), exponential ranking (Chakraborty, 1995c), reverse fitness].

6. Proportional Selection

Proportional selection (also called “proportionate” or “fitness-proportionate” selection) (Goldberg, 1989a) is perhaps the most popular selection algorithm in the literature. In this scheme, the selection probabilities of the individuals are given by their relative fitness. If

Table 4. Theoretical and experimental f_{takeover} values in GENITOR ($N = 128$, $\max = 1.5$).

Gen.	Theoretical	Sample Mean	Sample Std. Dev.	95% Conf. Intvl.
1	0.000	0		
2	0.000	0		
3	0.118	0.099	0.299	(0.080, 0.118)
4	0.514	0.521	0.500	(0.490, 0.552)
5	0.272	0.276	0.447	(0.248, 0.304)
6	0.074	0.083	0.276	(0.066, 0.100)
7	0.017	0.018	0.133	(0.010, 0.026)
8	0.004	0.003	0.055	—
9	0.001	0		—
10	0.000	0		
\vdots	\vdots	\vdots		

$\{x_1^{(t)}, x_2^{(t)}, \dots, x_N^{(t)}\}$ denotes the population at generation t , then the probability that a particular individual, say x_k , is selected in a single trial is given by

$$p_{\text{sel}}(x_k^{(t)}) = \frac{f_k^{(t)}}{\sum_{k=1}^N f_k^{(t)}}$$

where $f_k^{(t)}$ stands for the fitness of individual x_k at the t th generation.

If there are i individuals in the designated class at generation t , then the probability that in a single trial a member of this class would be picked is given by

$$p_{\text{select}}(i) = \frac{\sum_{k \in \text{class}} f_k^{(t)}}{\sum_{k=1}^N f_k^{(t)}} \quad (9)$$

where the sum in the numerator of the right-hand side is over the i individuals of the designated class. The transition probabilities p_{ij} ($i, j = 0, 1, \dots, N$) are then given by

$$p_{ij} = \begin{cases} 1 & \text{for } i = 0; \quad j = 0 \\ 0 & \text{for } i = 0; \quad j = 1, 2, \dots, N \\ 1 & \text{for } i = N; \quad j = N \\ 0 & \text{for } i = N; \quad j = 0, 1, \dots, N - 1 \\ \binom{N}{j} (p_{\text{select}}(i))^j (1 - p_{\text{select}}(i))^{N-j} & \text{otherwise} \end{cases} \quad (10)$$

with $p_{\text{select}}(i)$ given by Equation 9. Equation 9 shows that $p_{\text{select}}(i)$ is dependent on time, and consequently, the Markov chain defined by Equation 10 is time inhomogeneous (nonstationary).

To obtain a time-homogeneous chain, we introduce a restriction: We confine our attention to a GA having only two types of individuals³ with a fitness ratio

$$\gamma = f_1/f_0 \quad (11)$$

where f_1 and f_0 are the fitnesses of the two individuals. Assuming the designated class to be the class of 1's, Equation 9 becomes

$$\begin{aligned} p_{\text{select}}(i) &= \frac{if_1}{if_1 + (N-i)f_0} \\ &= \frac{i\gamma}{i\gamma + N-i} \end{aligned} \quad (12)$$

by Equation 11.

Equations 10 and 12 show that for a fixed population size, the fitness ratio γ is the only parameter controlling proportional selection. The case $\gamma = 1$ corresponds to no selection pressure in favor of the designated class. Selection pressure can be increased indefinitely by making γ arbitrarily large. As $\gamma \rightarrow \infty$, $p_{\text{select}}(i) \rightarrow 1$ for all $i = 1, \dots, N$, and the upper bound of selection pressure, irrespective of population size, is given by

$$p_{\text{takeover}} = \begin{cases} 0 & \text{for } t = 0 \\ 1 & \text{otherwise} \end{cases} \quad (13)$$

The fitness ratio γ that affects the behavior of proportional selection is not a parameter of the selection algorithm per se; rather, it is a parameter characterizing the problem itself. In other words, the parameter controlling proportional selection is problem dependent, whereas the parameters affecting ranking, tournament, and GENITOR are problem independent.

Figure 6 and Table 5 show some representative cases.

7. (μ, λ) - and $(\mu + \lambda)$ -Selection

In this section we analyze (1) the selection method used in ESs, and (2) selection methods in GAs that have been inspired by selection in ESs.

7.1 Selection in (μ, λ) -ES and $(\mu + \lambda)$ -ES

In both (μ, λ) -selection and $(\mu + \lambda)$ -selection in ESs (Bäck & Schwefel, 1993; T. Bäck, personal communication, March 1994), λ offspring are created from μ parents ($1 \leq \mu < \lambda$) by uniformly randomly picking individuals from the parent pool. In (μ, λ) -ES, the best μ individuals from the offspring pool (of size λ) are used as the parents of the next generation. Figure 7 shows the working scheme of (μ, λ) -selection in ESs. In $(\mu + \lambda)$ -ES, on the other hand, the best μ individuals from the combined pool of μ parents and λ offspring are taken as the parents of the next generation. The working principle of $(\mu + \lambda)$ -selection in ESs is shown in Figure 8.

Traditionally, in ESs, the term “population” is used to refer to the set of μ parents. We therefore compute the $(\mu + 1) \times (\mu + 1)$ transition probability matrix for transition from parents at generation t to parents at generation $t + 1$ for both (μ, λ) - and $(\mu + \lambda)$ -selection in ESs. Note, however, that following the present approach, the $(\lambda + 1) \times (\lambda + 1)$ matrix for offspring-offspring transition can also be determined.

³ Such a single-locus GA has been considered in Goldberg and Segrest (1987), Goldberg (1989b), Goldberg and Deb (1991), and Horn (1993).

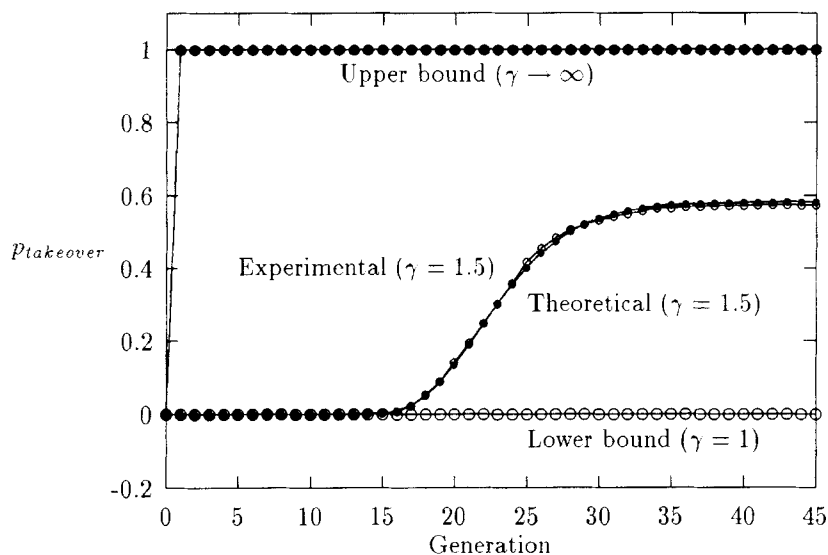


Figure 6. Growth of p_{takeover} with generation in fitness-proportionate selection ($N = 128$).

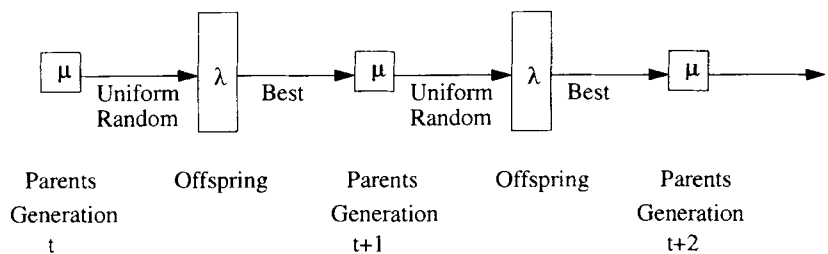


Figure 7. Schematic diagram of (μ, λ) -selection in ESs.

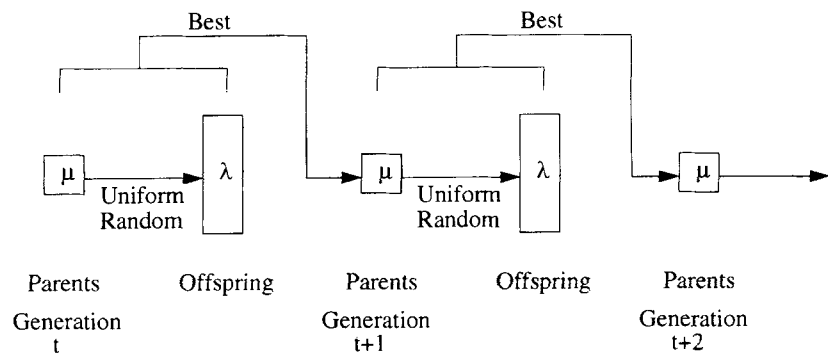


Figure 8. Schematic diagram of $(\mu + \lambda)$ -selection in ESs.

Table 5. Theoretical and experimental f_{takeover} values in proportional selection ($N = 128, \gamma = 1.5$).

Gen.	Theoretical	Sample Mean	Sample Std. Dev.	95% Conf. Intvl.
1	0.000	0		
\vdots	\vdots	\vdots		
13	0.000	0		
14	0.001	0.002	0.045	—
15	0.003	0.002	0.045	—
16	0.007	0.006	0.077	(0.001, 0.011)
17	0.016	0.015	0.122	(0.007, 0.023)
18	0.025	0.031	0.173	(0.020, 0.042)
19	0.037	0.035	0.184	(0.024, 0.046)
20	0.047	0.051	0.220	(0.037, 0.065)
21	0.054	0.052	0.222	(0.038, 0.066)
22	0.057	0.056	0.230	(0.042, 0.070)
23	0.057	0.054	0.226	(0.040, 0.068)
24	0.053	0.055	0.228	(0.041, 0.069)
25	0.046	0.060	0.237	(0.045, 0.075)
26	0.040	0.037	0.189	(0.025, 0.049)
27	0.033	0.030	0.171	(0.019, 0.041)
28	0.026	0.022	0.147	(0.013, 0.031)
29	0.021	0.014	0.117	(0.007, 0.021)
30	0.015	0.011	0.104	(0.005, 0.017)
31	0.012	0.011	0.104	(0.005, 0.017)
32	0.009	0.007	0.083	(0.002, 0.012)
33	0.007	0.008	0.089	(0.002, 0.014)
34	0.004	0.006	0.077	(0.001, 0.011)
35	0.004	0.001	0.032	—
36	0.002	0.003	0.055	—
37	0.002	0.002	0.045	—
38	0.001	0.001	0.032	—
39	0.001	0		—
40	0.001	0.001	0.032	—
41	0.000	0.001	0.032	—
42	0.000	0		
\vdots	\vdots	\vdots		

First, we analyze selection in (μ, λ) -ESs. Let i be the number of members of the designated class in the parent population at generation t . When individuals are chosen uniformly randomly (and with replacement) from the parent population, the probability that

in a single trial a member of the designated class would be picked is given by

$$p_{\text{select}}(i) = i/\mu, \quad 0 \leq i \leq \mu \quad (14)$$

After λ such (independent) trials have been completed, the offspring pool (of size λ) would be filled. Let the offspring pool thus created contain k members of the designated class. When $0 \leq k < \mu$, the parent pool of generation $t + 1$ will have k class members and $\mu - k$ nonclass members (see Figure 7). If, on the other hand, $\mu \leq k \leq \lambda$, all the μ individuals of the parent pool of generation $t + 1$ would be class members. The transition probabilities p_{ij} ($i, j = 0, 1, \dots, \mu$) are then given by

$$p_{ij} = \begin{cases} 1 & \text{for } i = 0; j = 0 \\ 0 & \text{for } i = 0; j = 1, 2, \dots, \mu \\ 1 & \text{for } i = \mu; j = \mu \\ 0 & \text{for } i = \mu; j = 0, 1, \dots, \mu - 1 \\ \binom{\lambda}{j} (p_{\text{select}}(i))^j (1 - p_{\text{select}}(i))^{\lambda-j} & \text{for } i = 1, 2, \dots, \mu - 1; j < \mu \\ \sum_{k=\mu}^{\lambda} \binom{\lambda}{k} (p_{\text{select}}(i))^k (1 - p_{\text{select}}(i))^{\lambda-k} & \text{for } i = 1, 2, \dots, \mu - 1; j = \mu \end{cases} \quad (15)$$

where $p_{\text{select}}(i)$ is given by Equation 14.

Next we consider selection in $(\mu + \lambda)$ -ES. In this scheme, if there are i members of the designated class in the parent population at generation t , then the parent population at generation $t + 1$ is guaranteed to contain at least i members of the class (see Figure 8). Here the probability laws governing the creation of λ offspring from μ parents are the same as in (μ, λ) -selection, and Equation 14 holds for $(\mu + \lambda)$ -selection, too. Note that the parent pool at generation $t + 1$ can contain μ members of the designated class in one of the following two situations:

- The parent pool at generation t contains μ class members.
- The parent pool at generation t contains i ($0 < i < \mu$) class members and produces *at least* $\mu - i$ class members as its offspring.

We now have the following transition probabilities p_{ij} ($i, j = 0, 1, \dots, \mu$) for selection in $(\mu + \lambda)$ -ES:

$$p_{ij} = \begin{cases} 1 & \text{for } i = 0; j = 0 \\ 0 & \text{for } i = 0; j = 1, 2, \dots, \mu \\ 1 & \text{for } i = \mu; j = \mu \\ 0 & \text{for } i = \mu; j = 0, 1, \dots, \mu - 1 \\ 0 & \text{for } i = 1, 2, \dots, \mu - 1; j < i \\ \binom{\lambda}{j-i} (p_{\text{select}}(i))^{j-i} (1 - p_{\text{select}}(i))^{\lambda-j+i} & \text{for } i = 1, 2, \dots, \mu - 1; \\ & j = i, i + 1, \dots, \mu - 1 \\ \sum_{k=\mu-i}^{\lambda} \binom{\lambda}{k} (p_{\text{select}}(i))^k (1 - p_{\text{select}}(i))^{\lambda-k} & \text{for } i = 1, 2, \dots, \mu - 1; j = \mu \end{cases} \quad (16)$$

with $p_{\text{select}}(i)$ given by Equation 14.

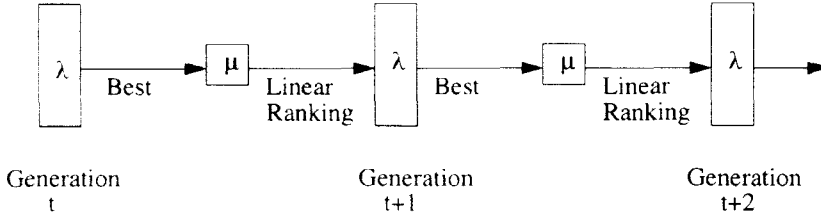


Figure 9. Schematic diagram of (μ, λ) -linear ranking selection in GAs.

LEMMA 3: *The $(\mu + 1) \times (\mu + 1)$ transition probability matrix for selection in $(\mu + \lambda)$ -ES is upper triangular.*

PROOF: Follows from Equation 16.

7.2 (μ, λ) -Linear Ranking Selection in GAs

In (μ, λ) -linear ranking selection in GAs (Bäck & Hoffmeister, 1991; Hoffmeister & Bäck, 1992), the new population (of generation $t + 1$) is formed by running linear ranking on the best μ individuals of the old population (of generation t). Here, as is customary in GAs, by population we mean the (offspring) population of size λ ($= N$). The mechanism is shown pictorially in Figure 9.

Given i class members in the population at any generation, the probability that, in the linear ranking scheme (Fig. 9), a class member would be picked in a single trial is (see Section 3)

$$p_{\text{select}}(i) = \begin{cases} \frac{i}{\mu} \left(\max - \frac{(\max - 1)(i - 1)}{\mu - 1} \right) & \text{for } 0 \leq i \leq \mu \\ 1 & \text{for } \mu < i \leq \lambda \end{cases} \quad (17)$$

where rank μ corresponds to the best individual in a population of size λ and rank 1 to the μ th best. Equation 17 holds for $\mu > 1$. For $\mu = 1$, we have

$$p_{\text{select}}(i) = \begin{cases} 0 & \text{for } i = 0 \\ 1 & \text{for } i \geq 1 \end{cases} \quad (18)$$

A special case in Equation 17 is worth noting: For all μ ($1 \leq \mu < \lambda$), $p_{\text{select}}(\mu - 1) < 1$ for $\max < 2$, but $p_{\text{select}}(\mu - 1) = 1$ for $\max = 2$. The transition probabilities p_{ij} ($i, j = 0, 1, \dots, \lambda$) are given by

$$p_{ij} = \begin{cases} 1 & \text{for } p_{\text{select}}(i) = 0; j = 0 \\ 0 & \text{for } p_{\text{select}}(i) = 0; j = 1, 2, \dots, \lambda \\ 1 & \text{for } p_{\text{select}}(i) = 1; j = \lambda \\ 0 & \text{for } p_{\text{select}}(i) = 1; j = 0, 1, \dots, \lambda - 1 \\ \binom{\lambda}{j} (p_{\text{select}}(i))^j (1 - p_{\text{select}}(i))^{\lambda - j} & \text{for } 0 < p_{\text{select}}(i) < 1; j = 0, 1, \dots, \lambda \end{cases} \quad (19)$$

where $p_{\text{select}}(i)$ is given by Equations 17 and 18.

Equations 17–19 show that the Markov chain for (μ, λ) -linear ranking selection in GAs is finite, aperiodic, and time homogeneous, with two absorbing states and $\lambda - 1$ transient states. For a fixed population size λ , μ and \max are the two parameters affecting the performance

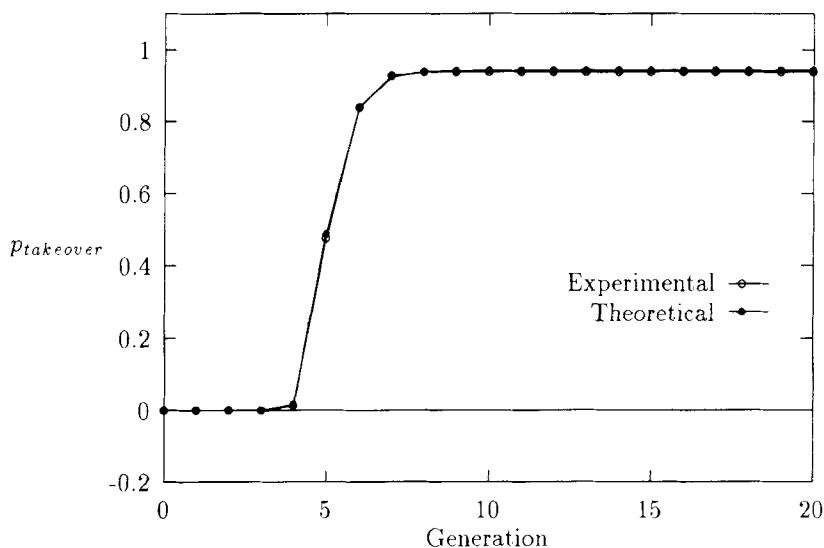


Figure 10. Theoretical and experimental $p_{takeover}$ for (μ, λ) -linear ranking selection in GAs ($\lambda = 128, \mu = 64, \text{max} = 1.5$).

Table 6. Theoretical and experimental $f_{takeover}$ values for (μ, λ) -linear ranking selection in GAs ($\lambda = 128, \mu = 64, \text{max} = 1.5$).

Gen.	Theoretical	Sample Mean	Sample Std. Dev.	95% Conf. Intvl.
1	0.000	0		
2	0.000	0		
3	0.000	0		
4	0.017	0.015	0.122	(0.007, 0.023)
5	0.471	0.460	0.498	(0.429, 0.491)
6	0.354	0.364	0.481	(0.334, 0.394)
7	0.082	0.089	0.285	(0.071, 0.107)
8	0.016	0.010	0.099	(0.004, 0.016)
9	0.002	0.001	0.032	—
10	0.001	0		—
11	0.000	0		
⋮	⋮	⋮		

of this selection scheme. Figure 10 shows the growth of $p_{takeover}$ for a typical case (see also Table 6). Figure 11 shows the upper bound ($\text{max} = 2$) and the lower bound ($\text{max} = 1$) of $p_{takeover}$ for a particular value of μ . In Figure 12 we show two other bounds—corresponding to the minimum and maximum values of μ —when max is kept fixed at a typical value.

By inserting $\text{max} = 1$ in Equation 17, one obtains the transition probabilities for (μ, λ) -uniform ranking selection in GAs (Bäck & Hoffmeister, 1991) from Equations 17 and 19. Again,

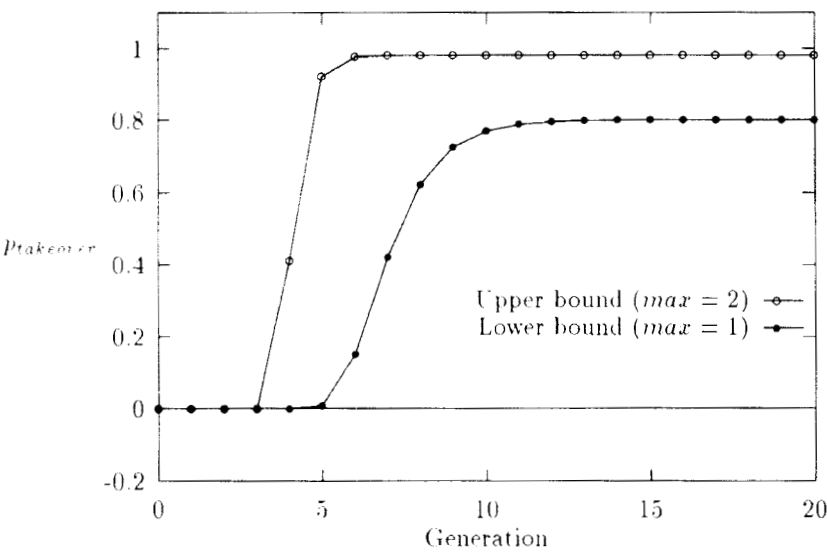


Figure 11. Upper and lower bounds in (μ, λ) -linear ranking selection in GAs ($\lambda = 128, \mu = 64$).

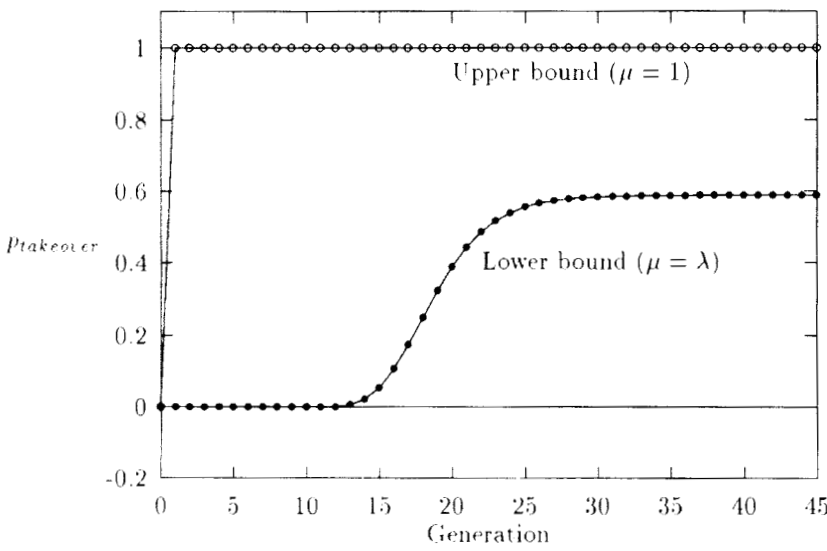


Figure 12. Upper and lower bounds for a fixed $max = 1.5$, in (μ, λ) -linear ranking selection in GAs ($\lambda = 128$).

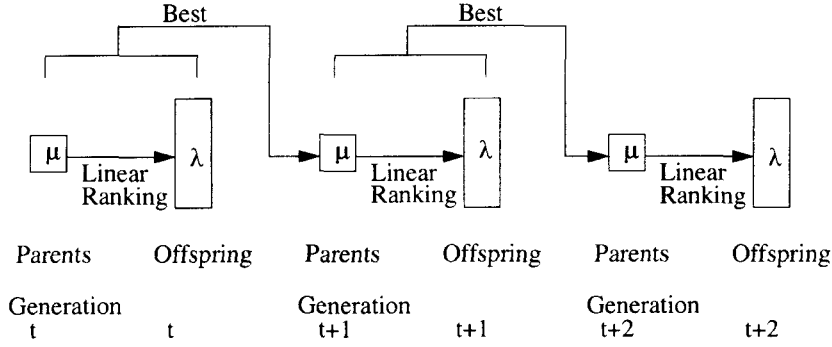


Figure 13. Schematic diagram of $(\mu + \lambda)$ -linear ranking selection in GAs.

following the present approach, it is not difficult to analyze (μ, λ) -proportional selection in GAs (Bäck & Hoffmeister, 1991).

7.3 $(\mu + \lambda)$ -Linear Ranking Selection in GAs

In this selection method, the offspring population of generation $t + 1$ is created by running linear ranking on the best μ individuals from the combined pool of μ ($1 \leq \mu < \lambda$) parents of generation t and λ ($= N$) offspring of generation t (see Fig. 13).

Let $\mathcal{P}_{\text{parent-offspring}} = [p_{ij}]_{(\mu+1) \times (\lambda+1)}$ be the transition probability matrix for transition from parents at generation t to offspring at generation t . It is easy to see that the entries of this matrix are given by

$$p_{ij} = \begin{cases} 1 & \text{for } p_{\text{select}}(i) = 0; j = 0 \\ 0 & \text{for } p_{\text{select}}(i) = 0; j = 1, 2, \dots, \lambda \\ 1 & \text{for } p_{\text{select}}(i) = 1; j = \lambda \\ 0 & \text{for } p_{\text{select}}(i) = 1; j = 0, 1, \dots, \lambda - 1 \\ \binom{\lambda}{j} (p_{\text{select}}(i))^j (1 - p_{\text{select}}(i))^{\lambda-j} & \text{for } 0 < p_{\text{select}}(i) < 1; j = 0, 1, \dots, \lambda \end{cases} \quad (20)$$

where $p_{\text{select}}(i)$, $i = 0, 1, \dots, \mu$, is given by

$$p_{\text{select}}(i) = \begin{cases} \frac{i}{\mu} \left(\max - \frac{(\max - 1)(i - 1)}{\mu - 1} \right) & \text{for } \mu > 1 \\ i & \text{for } \mu = 1 \end{cases} \quad (21)$$

Let $\mathcal{P}_{\text{parent-parent}} = [p_{ij}]_{(\mu+1) \times (\mu+1)}$ be the transition probability matrix for transition from parents at generation t to parents at generation $t + 1$. Then the elements of $\mathcal{P}_{\text{parent-parent}}$

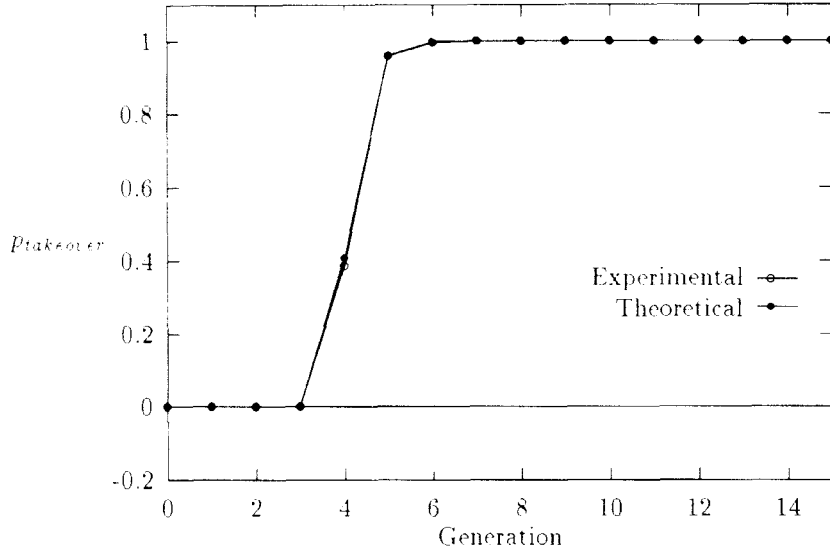


Figure 14. Theoretical and empirical p_{takeover} for $(\mu + \lambda)$ -linear ranking selection in GAs ($\lambda = 128$, $\mu = 64$, $\text{max} = 1.5$).

are given by

$$p_{ij} = \begin{cases} 1 & \text{for } i = 0; j = 0 \\ 0 & \text{for } i = 0; j = 1, 2, \dots, \mu \\ 1 & \text{for } i = \mu; j = \mu \\ 0 & \text{for } i = \mu; j = 0, 1, \dots, \mu - 1 \\ 0 & \text{for } i = 1, 2, \dots, \mu - 1; j < i \\ \binom{\lambda}{j-i} (p_{\text{select}}(i))^{j-i} (1 - p_{\text{select}}(i))^{\lambda-j+1} & \text{for } i = 1, 2, \dots, \mu - 1; \\ & j = i, i + 1, \dots, \mu - 1 \\ \sum_{k=\mu-i}^{\lambda} \binom{\lambda}{k} (p_{\text{select}}(i))^k (1 - p_{\text{select}}(i))^{\lambda-k} & \text{for } i = 1, 2, \dots, \mu - 1; j = \mu \end{cases} \quad (22)$$

where $p_{\text{select}}(i)$ is given by Equation 21.

LEMMA 4: $\mathcal{P}_{\text{parent} \rightarrow \text{parent}}$ is upper triangular.

PROOF: Follows from Equation 22.

The transition probability matrix for transition from parents at generation t to offspring at generation $t + k$ (for any t), $k \geq 1$, is obtained from

$$\mathcal{P}_{\text{parent} \rightarrow \text{parent}}^k \mathcal{P}_{\text{parent} \rightarrow \text{offspring}}$$

From Equations 20–22 it is clear that for a fixed λ , this selection algorithm is controlled by two parameters: μ and max . Figure 14 shows the growth of p_{takeover} with time for a typical case (see also Table 7). Figures 15 and 16 show the p_{takeover} bounds.

Table 7. Theoretical and experimental $f_{takeover}$ values for $(\mu + \lambda)$ -linear ranking selection in GAs ($\lambda = 128, \mu = 64, \max = 1.5$).

Gen.	Theoretical	Sample Mean	Sample Std. Dev.	95% Conf. Intvl.
1	0.000	0		
2	0.000	0		
3	0.000	0		
4	0.409	0.387	0.487	(0.357, 0.417)
5	0.552	0.572	0.495	(0.541, 0.603)
6	0.037	0.038	0.191	(0.026, 0.050)
7	0.002	0.002	0.045	—
8	0.000	0.001	0.032	—
9	0.000	0		
\vdots	\vdots	\vdots		

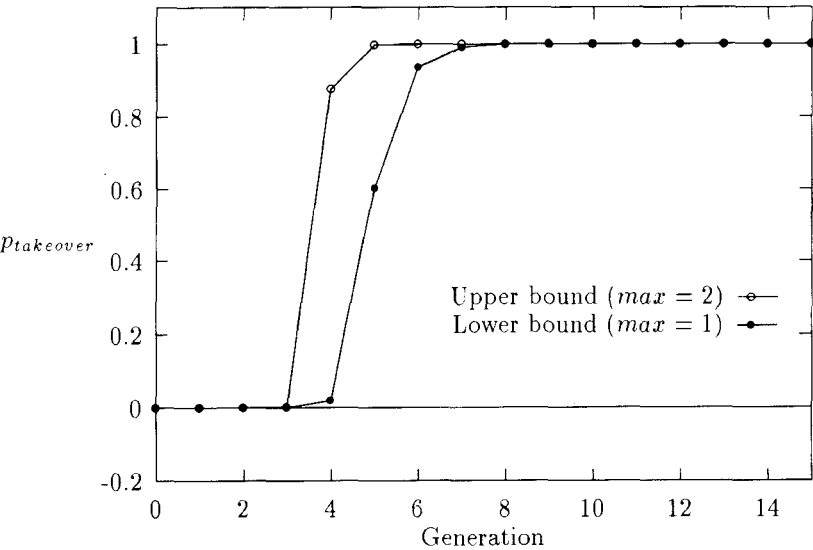


Figure 15. $(\mu + \lambda)$ -linear ranking selection in GAs ($\lambda = 128, \mu = 64$).

8. Selection in CHC

Selection in the CHC algorithm (Eshelman, 1991) is “population elitist” in the sense that in this selection strategy, the best N ($N =$ population size) individuals obtained so far are guaranteed to survive.

If P_t and C_t represent, respectively, the parent and the child populations at generation t (the sizes of the two populations are identical and equal to N), then in CHC, P_{t+1} is produced by deterministically picking the best N individuals from the joint pool of P_t and C_{t+1} (Eshelman, 1991). Under selection alone, C_{t+1} is identical to P_t , and we then have the

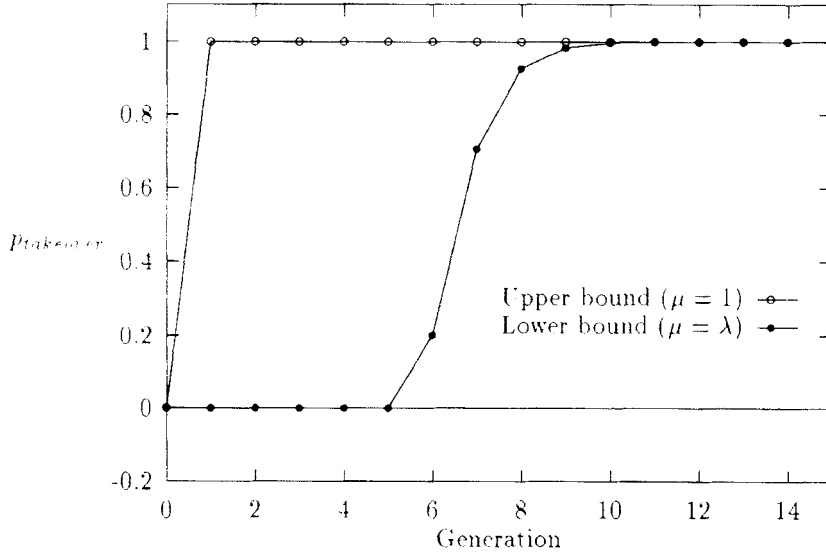


Figure 16. $(\mu + \lambda)$ -linear ranking selection in GAs ($\lambda = 128$, $\max = 1.5$).

following transition probabilities for transition from P_t to P_{t+1} :

$$p_{ij} = \begin{cases} 1 & \text{for } i = 0; & j = 0 \\ 0 & \text{for } i = 0; & j = 1, 2, \dots, N \\ 1 & \text{for } i = N; & j = N \\ 0 & \text{for } i = N; & j = 0, 1, \dots, N-1 \\ 1 & \text{for } i = 1, 2, \dots, N_1; & j = 2i \\ 1 & \text{for } i > N_1; & j = N \\ 0 & \text{otherwise} \end{cases} \quad (23)$$

where

$$N_1 = \begin{cases} \frac{N}{2} & \text{if } N \text{ is even} \\ \frac{N-1}{2} & \text{if } N \text{ is odd} \end{cases}$$

9. Comparison

In this section we present a comparative analysis of the selection algorithms using the following:

- a maximum–minimum range (upper and lower bounds) of possible p_{takeover} values;
- the f_{takeover} versus t behavior;
- the nature of change of ultimate p_{takeover} values with the parameter(s) of selection algorithms;
- the highest ultimate p_{takeover} values.

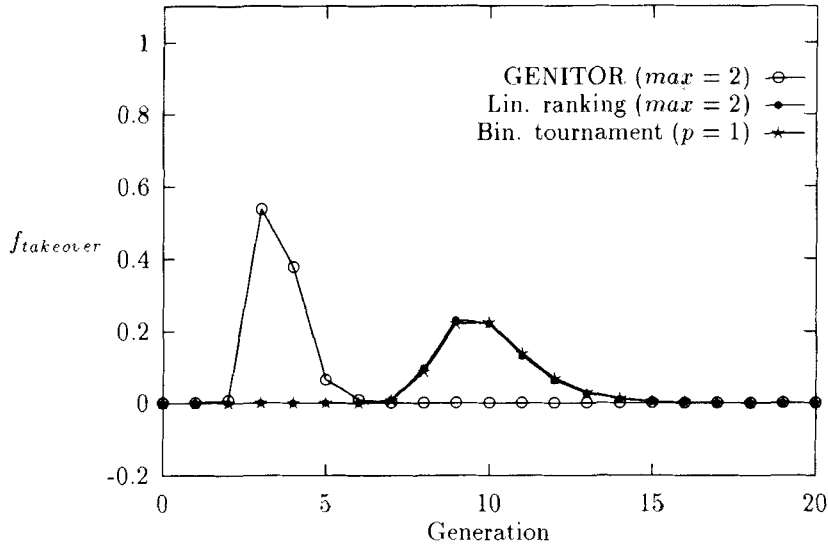


Figure 17. Comparison of f_{takeover} versus t for three selection algorithms ($N = 128$). Lin. = linear. Bin. = binary.

The lower and upper bounds of $p_{\text{takeover}}(t)$ versus t plots were shown in the previous sections. The lower bounds in linear ranking, binary tournament, and proportionate⁴ selection correspond to genetic drift. Drift is also represented by the cases $\mu = \lambda$ in (μ, λ) -selection in ESs and $\mu = \lambda, \max = 1$ in (μ, λ) -linear ranking selection in GAs. The figures in Sections 3–6 show that the range is much narrower for GENITOR than for proportionate, ranking, or tournament selection. Again, the lower bound of GENITOR is remarkably higher than that of these three schemes.

The $f_{\text{takeover}}(t)$ versus t graphs (corresponding to $N = 128$) for the upper bound in GENITOR, binary tournament, and linear ranking are shown in Figure 17. Figure 17 brings out two important facts:

- GENITOR starts pushing much earlier than linear ranking or binary tournament;
- the hardest push of GENITOR is harder than that of the other two schemes.

This corroborates the finding of Goldberg and Deb (1991) that GENITOR selection is much harder than ranking.

DEFINITION 3: The ultimate p_{takeover} is defined as the p_{takeover} value as the number of generations tends to infinity.

THEOREM 1: Regardless of the selection algorithm's parameter settings, for GENITOR, CHC, $(\mu + \lambda)$ -selection in ESs, and $(\mu + \lambda)$ -linear ranking selection in GAs, the ultimate p_{takeover} is always unity, whereas for binary tournament selection and linear ranking selection, this probability is less than 1.

4 Only the two-alternative case of proportionate selection has been considered in this paper.

PROOF: Using (1) Lemma 2, (2) Lemma 3, and (3) Equations 2, 4, 6, and 23, it is easy to see from Theorem A.2 (Appendix) that the $(\text{pop} - 1) \times 2$ matrix $\mathcal{B} = \mathcal{M} \cdot \mathcal{R}$ [where $\text{pop} = N$ for GENITOR, ranking, tournament, and CHC and μ for $(\mu + \lambda)$ -ES] is of the form

$$\begin{pmatrix} 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}$$

for GENITOR, CHC, and $(\mu + \lambda)$ -selection in ESs and that this matrix is

$$\begin{pmatrix} \alpha_{\text{pop}-1} & 1 - \alpha_{\text{pop}-1} \\ \alpha_{\text{pop}-2} & 1 - \alpha_{\text{pop}-2} \\ \vdots & \vdots \\ \alpha_1 & 1 - \alpha_1 \end{pmatrix}$$

(where $0 < \alpha_i < 1, i = 1, 2, \dots, \text{pop} - 1$) for tournament and linear ranking selection algorithms. Again, for $(\mu + \lambda)$ -linear ranking in GAs, we see that (by Lemma 4) the $\mathcal{B} = \mathcal{M} \cdot \mathcal{R}$ matrix for $\mathcal{P}_{\text{parent}-\text{parent}}$ is of the form

$$\begin{pmatrix} 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}$$

and hence $\mathcal{P}_{\text{parent}-\text{parent}}^k \cdot \mathcal{P}_{\text{parent}-\text{offspring}}$ as $k \rightarrow \infty$ is the $(\mu + 1) \times (\lambda + 1)$ matrix

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ & & \cdots & \\ \vdots & \vdots & & \vdots \\ & & \cdots & \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Q.E.D. □

THEOREM 2: *For (1) (μ, λ) -selection in ESs with $\mu = 1$, and (2) (μ, λ) -linear ranking selection in GAs with $\mu = 1$, the ultimate p_{takeover} is unity. This probability is < 1 for any other parameter values in these algorithms.*

PROOF: Noting from Equations 15 and 19 that the transition probability matrix in both the cases is upper triangular, one can complete the proof by proceeding as in the proof of Theorem 1. Q.E.D. □

From Theorems 1 and 2 we see that $\{f_{\text{takeover}}\}$ is the takeover time distribution in the case of GENITOR, (μ, λ) -selection with $\mu = 1$ and $(\mu + \lambda)$ -selection, and consequently the mean (expected) takeover time can be computed for these algorithms. In the case of tournament and ranking, on the other hand, the mean takeover time does not make sense

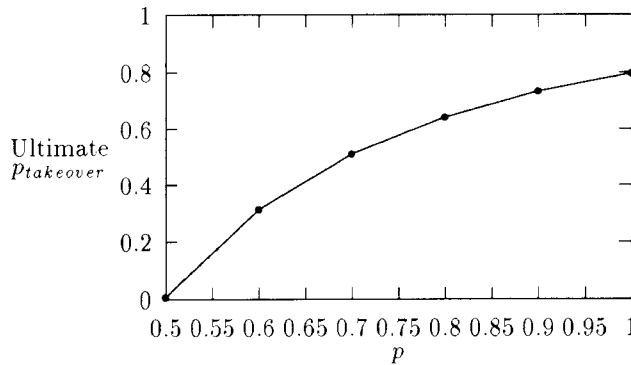


Figure 18. Binary tournament selection.

(see Section 2.1). In the former case, for i ($1 \leq i \leq N - 1$) initial members, the expected takeover time is given by the row-sum corresponding to state i in the \mathcal{M} matrix. In the latter, one can compute the *conditional* mean takeover time; that is, the mean takeover time relative to the hypothesis that the process ends up in the absorbing state N . To obtain this conditional mean, we have to consider a modified Markov chain with $N - 1$ transient states ($1, 2, \dots, N - 1$) and a single absorbing state (N) and recalculate all conditional transition probabilities. Using absorbing Markov chain theory (Kemeny & Snell, 1960), it can be shown that the conditional mean takeover time corresponding to i initial members is given by the row-sum of row i of the matrix

$$\mathcal{D}^{-1} \mathcal{M} \mathcal{D}$$

where \mathcal{D} is an $(N - 1) \times (N - 1)$ diagonal matrix with diagonal entries b_{jN} , for non-absorbing state j (recall that b 's are elements of the \mathcal{B} matrix).

In Figures 18–21 we have shown how the ultimate p_{takeover} value changes with changes in the value of the selection algorithm's parameter (in all these figures, population size = 128). A comparison of the plots for linear ranking and binary tournament shows that when $\max = 2p$, these two algorithms behave similarly, with the ultimate p_{takeover} values being nearly identical for $\max = 2p$. This observation, coupled with the fact that the upper bound plots of p_{takeover} for linear ranking and binary tournament (Figs. 1 and 2) are almost identical, corroborates a similar observation in Goldberg and Deb (1991) about these two strategies.⁵ Again, as expected from Theorem 1, for GENITOR and $(\mu + \lambda)$ -selection, the ultimate p_{takeover} remains 1.0 throughout, independent of the parameter settings.

⁵ One of the referees pointed out a conflict with the paper by Goldberg and Deb (1991). Goldberg and Deb (1991, page 80) state that the two strategies “are identical” when $\max = 2p$. Our analysis (Equations 1–4) shows that the two algorithms, although very similar, are *not* exactly identical for $\max = 2p$. The reason for this contradiction is that, in the analysis of linear ranking selection, the Goldberg–Deb paper uses a continuous approximation to the discrete distribution, and Equation 18 on page 77 of that paper is an approximation that improves with increasing population size. In contrast, the present analysis is exact. The difference between the two approaches can easily be seen by taking the following example. If, at any generation, the population contains $N - 1$ class members, then linear ranking selection with $\max = 2$ dictates that the number of class members in the next generation must always be N (because with $\max = 2$, the worst individual gets no copies). Now setting $i = N - 1$, $\max = 2$ on the right-hand side of Equation 1 of the present paper yields $p_{\text{select}}(i) = 1$. However, setting $P_{i,t} = (N - 1)/N$, $c_0 = 2$ on the right side of Equation 18 of the Goldberg–Deb paper does *not* give $P_{i,t+1} = 1$; instead, $P_{i,t+1}$ becomes $(N^2 - 1)/N^2$, which of course approaches 1 as $N \rightarrow \infty$. In view of the fact that the Goldberg–Deb paper uses $(N - 1)/N$ (and not 1) as the final proportion at takeover, this is quite expected.

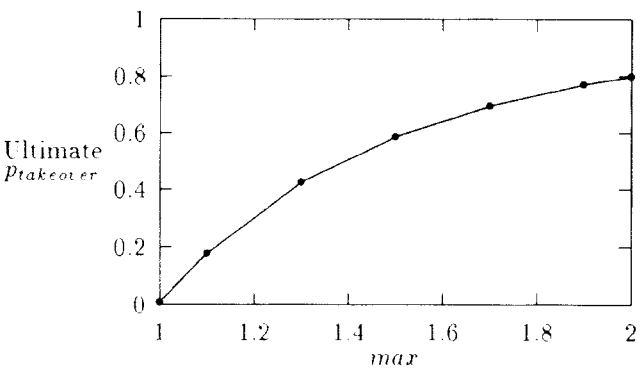


Figure 19. Linear ranking selection.

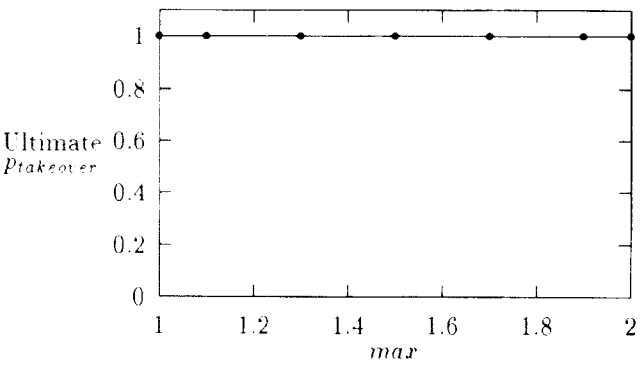


Figure 20. Selection in GENITOR.

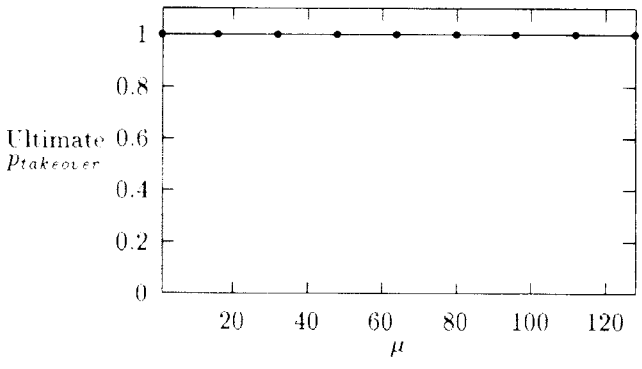


Figure 21. $(\mu + \lambda)$ -selection.

Table 8. Highest ultimate p_{takeover} at different population sizes.

Population Size	20	30	50	100	128	250	500	1000
Linear ranking	0.8203	0.8125	0.8062	0.8015	0.8005	0.7987	0.7977	0.7973
Binary tournament	0.8066	0.8033	0.8007	0.7988	0.7983	0.7976	0.7972	0.7970
Other algorithms ^a	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

^aSelection in GENITOR, selection in $(\mu + \lambda)$ -ES, $(\mu + \lambda)$ -linear ranking selection in GAs, selection in CHC, and proportional selection.

DEFINITION 4: *The highest ultimate p_{takeover} (for a particular population size) of any selection algorithm is the ultimate p_{takeover} value (at that population size) corresponding to the maximum selection pressure.*

Table 8 shows the highest ultimate p_{takeover} values, at various population sizes, for the different selection strategies. It is interesting to note from these data that the highest ultimate p_{takeover} is not very sensitive to population size changes. We capture the main results of Table 8 in the following fact:

FACT 1: The highest ultimate p_{takeover} value, regardless of the population size, is about 0.8 for linear ranking and binary tournament selection. This value is 1 for GENITOR, proportional, CHC, and $(\mu + \lambda)$ -selection.

10. Summary and Conclusions

Genetic algorithms are a nonlinear, complex, probabilistic, adaptive system, and knowledge of the selection mechanism is of fundamental importance in understanding this system. The results of this paper sharpen our insight into the working of selection. A Markov chain framework has been used to analyze the major types of selection algorithms.

This paper has introduced the following probabilistic measures of selection pressure: (1) upper and lower bounds of p_{takeover} values that can ever be achieved with a given selection algorithm; (2) f_{takeover} versus t behavior; (3) upper bound of the ultimate probability of takeover; and (4) variation of ultimate takeover probability with changes in the value of the selection algorithm’s parameter. We have analyzed and compared the selection schemes with reference to these measures. This analysis explains aspects of the selection mechanism that were hitherto observed empirically or explained by (approximate) deterministic models. The present approach is simple yet powerful and permits generalization to cover any selection scheme in evolutionary algorithms.

The limitation of the present approach (in fact, this is a limitation of any analysis that uses the concept of takeover) is that it tells us nothing about the creation of new points in the search space. The takeover analysis must necessarily ignore genetic operators like crossover and mutation because the primary aim is to investigate selection pressure. What is now needed is an approach that can analyze selection pressure in the presence of crossover. New measures of convergence are to be developed. The *response to selection* approach (Mühlenbein & Schlierkamp-Voosen, 1993, 1994; Mühlenbein, 1996) is a step in this direction. Further work is needed to extend the present analysis into these areas.

Acknowledgments

We are grateful to the anonymous referees for their extensive comments and suggestions. The first author wishes to thank Dr. Heinz Mühlenbein for many valuable discussions and for pointing out an embarrassing mistake (due to floating-point errors) in Table 1 in an earlier version of this paper, and Ashutosh Malaviya and Christine Crisan for their help with GNUPLOT and FrameMaker. Thanks to Dr. D. Ghosh Dastidar for access to the IMPACT Lab where the work was done during October–December 1993. The first author is supported by an AICTE Career Award. The first and the third authors are partially supported by grants from the Jadavpur University Research Fund. The final version of this paper was prepared at GMD, Sankt Augustin, Germany, while the first author was working there as a guest researcher.

Appendix: Some Results from Markov Chain Theory

DEFINITION A.1: *The Markov chain is time homogeneous (stationary) if the state transition matrix is independent of the time index. Otherwise, it is time inhomogeneous (nonstationary).*

DEFINITION A.2: *An ergodic set of states is a set in which every state can be reached from every other state and which, once entered, cannot be left.*

DEFINITION A.3: *An ergodic state is an element of an ergodic set.*

DEFINITION A.4: *A transient set of states is a set in which every state can be reached from every other state and which can be left.*

DEFINITION A.5: *A transient state is an element of a transient set.*

DEFINITION A.6: *An absorbing state is a state which once entered is never left.*

DEFINITION A.7: *The period of any state i is defined as the greatest common divisor of all integers k (≥ 1) for which $p_{ii}^{(k)} > 0$. When this greatest common divisor is 1, the state i is aperiodic.*

Let the Markov chain have s transient states and $r - s$ ergodic states. Then, the canonical form of the transition probability matrix becomes

$$P = \begin{pmatrix} S & \mathcal{O} \\ \mathcal{R} & Q \end{pmatrix}$$

The region \mathcal{O} consists of zeros. The $(r - s) \times (r - s)$ matrix S concerns the process after it has reached an ergodic set. The $s \times s$ submatrix Q corresponds to the transition probabilities only among the transient states. \mathcal{R} is an $s \times (r - s)$ matrix whose elements are the probabilities of transition from the transient states to the ergodic states. It can be shown that as $k \rightarrow \infty$, Q^k tends to \mathcal{O} (zero matrix). For an absorbing chain, S is an $(r - s) \times (r - s)$ identity matrix, $\mathcal{I}_{(r-s) \times (r-s)}$.

DEFINITION A.8: *For an absorbing Markov chain, the fundamental matrix is defined as $\mathcal{M} \equiv (\mathcal{I} - Q)^{-1}$.*

The fundamental matrix plays a useful role in absorbing Markov chain theory. The existence of the inverse of the matrix $(\mathcal{I} - Q)$ is established by Theorem A.1, which we state without proof.

THEOREM A.1: For an absorbing Markov chain partitioned as shown above, the inverse $(\mathcal{I} - \mathcal{Q})^{-1}$ exists, and

$$(\mathcal{I} - \mathcal{Q})^{-1} = \mathcal{I} + \mathcal{Q} + \mathcal{Q}^2 + \cdots = \sum_{k=0}^{\infty} \mathcal{Q}^k$$

THEOREM A.2: If b_{ij} denotes the probability that the process starting in transient state i ends up in absorbing state j , then

$$[b_{ij}] = \mathcal{B} = \mathcal{M} \cdot \mathcal{R}, \quad i \in T, \quad j \in \bar{T}$$

where T and \bar{T} represent, respectively, the transient and absorbing sets.

PROOF: See Kemeny and Snell (1960).

References

- Asoh, H., & Mühlenbein, H. (1994). On the mean convergence time of evolutionary algorithms without selection and mutation. In Y. Davidor, H.-P. Schwefel, & R. Männer (Eds.), *Parallel problem solving from nature 3* (pp. 88–97). Berlin: Springer-Verlag.
- Bäck, T. (1994). Selective pressure in evolutionary algorithms: a characterization of selection mechanisms. In *Proceedings of the First IEEE Conference on Evolutionary Computation*, Vol. 1 (pp. 57–62). Piscataway, NJ: IEEE Press.
- Bäck, T., & Hoffmeister, F. (1991). Extended selection mechanisms in genetic algorithms. In R. K. Belew & L. B. Booker (Eds.), *Proceedings of the Fourth International Conference on Genetic Algorithms* (pp. 92–99). San Mateo, CA: Morgan Kaufmann.
- Bäck, T., & Schwefel, H.-P. (1993). An overview of evolutionary algorithms for parameter optimization. *Evolutionary Computation* 1, 1–23.
- Baker, J. E. (1985). Adaptive selection methods for genetic algorithms. In J. J. Grefenstette (Ed.), *Proceedings of the First International Conference on Genetic Algorithms* (pp. 101–111). Hillsdale, NJ: Lawrence Erlbaum.
- Baker, J. E. (1989). *Analysis of the effects of selection in genetic algorithms*. Unpublished doctoral dissertation. Vanderbilt University, Nashville, Tennessee.
- Bhattacharyya, G. K., & Johnson, R. A. (1977). *Statistical concepts and methods*. New York: Wiley.
- Brindle, A. (1981). *Genetic algorithms for function optimization*. Unpublished doctoral dissertation. University of Alberta, Edmonton, Canada.
- Chakraborty, U. K. (1995a). A simpler derivation of schema hazard in genetic algorithms. *Information Processing Letters* 56, 77–78.
- Chakraborty, U. K. (1995b). A branching process model for genetic algorithms. *Information Processing Letters* 56, 281–292.
- Chakraborty, U. K. (1995c). An analysis of selection in generational and steady-state genetic algorithms: A Markov chain approach. In *Proceedings of the National Conference on Molecular Electronics, NERIST (A.P.), India*.
- Chakraborty, U. K., & Dastidar, D. G. (1993). Using reliability analysis to estimate the number of generations to convergence in genetic algorithms. *Information Processing Letters* 46, 199–209.
- Davis, T. E., & Principe, J. C. (1993). A Markov chain framework for the simple genetic algorithm. *Evolutionary Computation* 1, 269–288.
- De Jong, K. A. (1975). An analysis of the behavior of a class of genetic adaptive systems. *Dissertation Abstracts International* 36 (10), 5140B. (University Microfilms No. 76-9381.)

- De Jong, K. A., & Sarma, J. (1993). Generation gaps revisited. In L. D. Whitley (Ed.), *Foundations of genetic algorithms 2* (pp. 19–28). San Mateo, CA: Morgan Kaufmann.
- De Jong, K. A., Spears, W. M., & Gordon, D. F. (1995). Using Markov chains to analyze GAFOs. In L. D. Whitley & M. D. Vose (Eds.), *Foundations of genetic algorithms 3* (pp. 115–137). San Mateo, CA: Morgan Kaufmann.
- Eiben, A. E., Aarts, E. H. L., & Van Hee, K. M. (1991). Global convergence of genetic algorithms: A Markov chain analysis. In H.-P. Schwefel & R. Männer (Eds.), *Parallel problem solving from nature 1* (pp. 4–12). Berlin: Springer-Verlag.
- Eshelman, L. J. (1991). The CHC adaptive search algorithm: how to have safe search when engaging in nontraditional genetic recombination. In G. J. E. Rawlins (Ed.), *Foundations of genetic algorithms 1* (pp. 265–283). San Mateo, CA: Morgan Kaufmann.
- Goldberg, D. E. (1989a). *Genetic algorithms in search, optimization, and machine learning*. Reading, MA: Addison-Wesley.
- Goldberg, D. E. (1989b). Sizing populations for serial and parallel genetic algorithms. In J. D. Schaffer (Ed.), *Proceedings of the Third International Conference on Genetic Algorithms* (pp. 70–79). San Mateo, CA: Morgan Kaufmann.
- Goldberg, D. E., & Deb, K. (1991). A comparative analysis of selection schemes in genetic algorithms. In G. J. E. Rawlins (Ed.), *Foundations of genetic algorithms 1* (pp. 69–93). San Mateo, CA: Morgan Kaufmann.
- Goldberg, D. E., & Segrest, P. (1987). Finite Markov chain analysis of genetic algorithms. In J. J. Grefenstette (Ed.), *Proceedings of the Second International Conference on Genetic Algorithms* (pp. 1–8). Hillsdale, NJ: Lawrence Erlbaum.
- Goldberg, D. E., Korb, B., & Deb, K. (1989). Messy genetic algorithms: motivation, analysis, and first results. *Complex Systems* 3, 493–530.
- Grefenstette, J. J., & Baker, J. E. (1989). How genetic algorithms work: A critical look at implicit parallelism. In J. D. Schaffer (Ed.), *Proceedings of the Third International Conference on Genetic Algorithms* (pp. 20–27). San Mateo, CA: Morgan Kaufmann.
- Hoffmeister, F., & Bäck, T. (1992). *Genetic algorithms and evolution strategies: Similarities and differences* (Report of the Systems Analysis Research Group SYS-1/92). Dortmund, Germany: University of Dortmund, Department of Computer Science.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.
- Horn, J. (1993). *Finite Markov chain analysis of genetic algorithms with niching* (IlligAL Report No. 93002). Urbana, IL: University of Illinois at Urbana-Champaign, Department of General Engineering.
- Kemeny, J. G., & Snell, J. L. (1960). *Finite Markov chains*. Princeton, NJ: Van Nostrand.
- Mahfoud, S. M. (1993). Finite Markov chain models of an alternative selection strategy for the genetic algorithm. *Complex Systems* 7, 155–170.
- Marascuilo, L. A., & McSweeney, M. (1977). *Nonparametric and distribution-free methods for the social sciences*. Monterey, CA: Brooks/Cole.
- de la Maza, M., & Tidor, B. (1993). An analysis of selection procedures with particular attention paid to proportional and Boltzmann selection. In S. Forrest (Ed.), *Proceedings of the Fifth International Conference on Genetic Algorithms* (pp. 124–131). San Mateo, CA: Morgan Kaufmann.
- Mühlenbein, H. (1996). *The breeder genetic algorithm: From recombination of genes to the estimation of fitness distributions*. Manuscript submitted for publication.
- Mühlenbein, H., & Schlierkamp-Voosen, D. (1993). Predictive models for the breeder genetic algorithm I. Continuous parameter optimization. *Evolutionary Computation* 1, 25–49.

- Mühlenbein, H., & Schlierkamp-Voosen, D. (1994). The science of breeding and its application to the breeder genetic algorithm. *Evolutionary Computation* 1, 335–360.
- Nix, A. E., & Vose, M. D. (1992). Modeling genetic algorithms with Markov chains. *Annals of Mathematics and Artificial Intelligence* 5, 79–88.
- Schwefel, H.-P. (1981). *Numerical optimization of computer models*. Chichester: Wiley.
- Suzuki, J. (1993). A Markov chain analysis on a genetic algorithm. In S. Forrest (Ed.), *Proceedings of the Fifth International Conference on Genetic Algorithms* (pp. 146–153). San Mateo, CA: Morgan Kaufmann.
- Syswerda, G. (1989). Uniform crossover in genetic algorithms. In J. D. Schaffer (Ed.), *Proceedings of the Third International Conference on Genetic Algorithms* (pp. 2–9). San Mateo, CA: Morgan Kaufmann.
- Syswerda, G. (1991). A study of reproduction in generational and steady state genetic algorithms. In G. J. E. Rawlins (Ed.), *Foundations of genetic algorithms 1* (pp. 94–101). San Mateo, CA: Morgan Kaufmann.
- Vose, M. D. (1993). Modeling simple genetic algorithms. In L. D. Whitley (Ed.), *Foundations of genetic algorithms 2* (pp. 63–74). San Mateo, CA: Morgan Kaufmann.
- Whitley, L. D. (1989). The GENITOR algorithm and selection pressure: Why rank-based allocation of reproductive trials is best. In J. D. Schaffer (Ed.), *Proceedings of the Third International Conference on Genetic Algorithms* (pp. 116–121). San Mateo, CA: Morgan Kaufmann.
- Whitley, L. D., & Kauth, J. (1988). GENITOR: A different genetic algorithm. In *Proceedings of the Rocky Mountain Conference on Artificial Intelligence* (pp. 118–130).