A Probabilistic Model of the Multiple-Choice Genetic Algorithm

James C. Bean
Department of Industrial & Operations Engineering
University of Michigan
Ann Arbor, MI 48109

A Probabilistic Model of the
Multiple-Choice Genetic Algorithm

James C. Bean

University of Michigan

Ann Arbor, MI 48109-2117

September 9, 1996

ABSTRACT

We derive a probabilistic model of the multiple-choice genetic algorithm. Specifically, we give the distribution of the number of correct alleles in the best element of the population, and the expected number of correct alleles over the entire population following each generation. We then compare the theoretical performance against empirical runs. The results challenge the traditional explanation for efficacy of genetic algorithms.

A great deal of research over the past twenty years has shown the efficacy of genetic algorithms in studying or solving complex problems. Mathematical programming constitutes one important subclass of the use of genetic algorithms (see Goldberg [1989] or Michalewicz [1994]). There has been less work explaining how and why the approach works. Papers along these lines included Markov chain models such as Goldberg and Secrest [1987], Horn [1993] and Suzuki [1993]. Probabilistic models include Ankenbrandt [1991] and Rabinovich and Widgerson [1991]. Much of this literature concentrates on characteristics of the population rather than of the individual chromosome. A classic example is the Schema Lemma (Holland [1975]).

The objective of this paper is to present a probabilistic, transient, chromosome-level analysis of specific genetic algorithm implementations to gain a better understanding of their mechanisms.

Consider the class of discrete optimization problems that can be encoded for a genetic algorithm such that each gene may take any of a finite set of alleles. An example is the multiple-choice integer program (Hadj-Alouane and Bean [1992]). Problems of this

class can be solved by a particular genetic algorithm that we define as the multiple-choice genetic algorithm (MCGA). In this paper we will predict theoretically the behavior of this algorithm by deriving the distribution of the best element of the population and the population fitness average following each generation. We make a number of assumptions that allow this analysis. Empirical tests will then evaluate the reasonability of these assumptions and value of the understanding drawn from the model.

MCGAs have been used successfully in Hadj-Alouane and Bean [1992], Bean and Hadj-Alouane [1992] and Hadj-Alouane, Bean and Murty [1993]. They begin with the concepts of general genetic algorithms as described in Holland [1975] and Goldberg [1989]. The MCGAs use specific operators selected to facilitate the probabilistic analysis that follows.

Reproduction is not carried out by the traditional roulette wheel. The roulette wheel introduces complex dependencies between the probability of being selected as a parent and the fitness values of all members of the population. Rather, we choose parents randomly over all elements of the prior generation. The computational versions of the algorithm in previous papers employ an elitist strategy. This operator will not be included in this paper.

Traditional one-point crossover introduces dependencies among genes within a chromosome based on index. This dependence is often arbitrary and complicates analysis. We employ a *uniform crossover* (Syswerda [1989]). After two parents are chosen, a coin is flipped for each gene. Heads selects the allele from the first parent and tails selects from the second. Analysis of the mechanics of this operator are core to this paper. We present a new model of the operator relative to existing work such as Syswerda [1989], and Spears and De Jong [1991a, 1991b].

Uniform crossover of two parents can create two offspring, the second being constructed from the alleles not selected above. In *post-tournament selection* (PTS), we construct and evaluate both offspring and retain only the superior offspring in the next generation. Results show that this approach strongly promotes a "survival-of-the-fittest" upward pressure.

Rather than the traditional mutation operation, we use immigration. In each gen-

eration an entirely new member of the population is generated. During subsequent generations, its genetic material is spread throughout the population. Hence, immigration prevents premature convergence and loss of genetic material as traditional mutation does.

Immigration also leads to a trivial proof of convergence of the algorithm to an optimal solution based on random sampling of a finite space. However, empirical tests have shown that the algorithm converges much faster than would be predicted by sampling. Crossover is making a major contribution.

1. Problem Definition

We have a population of m chromosomes. Each chromosome has n genes. In generation k, the allele of chromosome i, gene j, is a random variable $x_{ij}^k \in X_j$, where X_j is the finite set of alleles that gene j may take. Assume that the cardinality of X_j is a, independent of j. If cardinalities differ, let $a = \max_j |X_j|$ and pad other genes with dummy alleles.

Assumption 1: For the purpose of the following analysis, we assume a *unique optimal solution* so that each gene has a unique correct allele.

Assumption 2: In order to develop results with some generality, we assume a *surrogate fitness measure* that counts the number of correct alleles in a chromosome.

These assumptions are not reasonable in most cases. Most problems have multiple optima. The surrogate fitness measure ignores nonlinear effects. However, the assumptions allow an analysis of the algorithm that hopefully will lead to insights that generalize beyond the assumptions.

By Assumption 1, there is a single correct allele for each gene. Let b_{ij}^k be the indicator random variable of the event that x_{ij}^k has the correct allele. Let $S_i^k = \sum_{j=1}^n b_{ij}^k$ be the number of correct alleles in chromosome i of generation k. By Assumption 2, S_i^k is the fitness function we will analyze. The population average in generation k is defined as $G^k = \frac{1}{m} \sum_{i=1}^m S_i^k$. The best element of the population is $A^k = \max_i S_i^k$. Given a generation, $\{S_i^{k-1}\}_{i=1}^m$, we construct the next generation by $m - 1$ independent applications of the uniform operator with PTS. The last position is filled with a randomly generated

3

immigrant.

The multiple-choice genetic algorithm to be studied can be formally stated:

MCGA

(Initialize) Generate x_{ij}^0 uniformly from $\{1, 2, \ldots, a\}$ for $i = 1, \ldots, m$; $j = 1, \ldots, n$. Calculate all S_i^0 and $A^0 = \max_i S_i^0$. $k \leftarrow 1$.

(Iterate) While $A^{k-1} < n$ do

      (Crossover) Generate $m - 1$ offspring via the uniform operator with PTS.

      (Immigration) Randomly generate one chromosome as in the initialization step.

      (Evaluate) Evaluate each S_i^k. Let $A^k = \max_i S_i^k$.

      (Increment) $k \leftarrow k + 1$.

(end iteration)

The goal of this paper is to derive the distribution of A^k and the population average, G^k.

2. Derivation of the Distribution

In Section 2.1 we will analyze the initialization step of the algorithm. Section 2.2 models the inheritance of correct alleles from parents to child. Independence of offspring is discussed in Section 2.3 leading to a distribution of the best element of a generation given the distribution of the previous generation.

2.1 Initialization

For the initial generation, x_{ij}^0 is taken uniformly from the set of potential alleles, represented by $\{1, 2, \ldots, a\}$. Recalling that exactly one of these alleles is correct, the probability of randomly choosing the correct allele is $p = 1/a$. Hence, b_{ij}^0 is a Bernoulli random variable with parameter p. By construction, all x_{ij}^0 and b_{ij}^0 are independent random variables.

Then S_i^0 has a binomial distribution with parameters (n, p). Further, all S_i^0 are independent from each other, though clearly not independent of the corresponding x_{ij}^0 and

4

b_{ij}^0. The initial population average is $E[S_i^0] = np = n/a$.

With m independent members of the population, we have

$$P(A^0 \leq J) = P(\max_i S_i^0 \leq J) = P(S_i^0 \leq J, \forall i) = [F(J)]^m \equiv F^m(J),$$

where $F(J)$ is the cumulative distribution function of a binomial (n, p) random variable. The expectation of A^0 can be found easily as $E[A^0] = \sum_{J=0}^{n-1}[1 - F^m(J)]$.

2.2 Inheritance

Given two parents selected at random, indexed p_1, p_2, with fitness $S_{p_1}^{k-1} = J_1, S_{p_2}^{k-1} = J_2$, we will derive the distribution of the the fitness of their offspring, $S_i^k, i < m$ ($i = m$ is the immigrant). We begin without PTS and add it below. To facilitate this analysis we make one substantive assumption.

Assumption 3: Given that the parents have J_1 and J_2 correct alleles, respectively, the location of the correct alleles are distributed uniformly within the parent's chromosomes and are independent of each other.

Assumption 3 will be true in early generations, but becomes less accurate as the population converges and building blocks appear. The impact of this assumption will be judged empirically.

gene	1	2	3	4	5	
	0	0	1	0	1	parent 1, $J_1 = 2$
	1	0	0	1	1	parent 2, $J_2 = 3$

Figure 1: Five gene example of correct allele indicators

From Assumption 3, the b_{ij}^k for the two parents can be viewed as zeros and ones randomly placed in a $2 \times n$ grid. See Figure 1 for an example with chromosome length five. For each gene there are three possible cases:

1. Both parents have the correct allele (as in Figure 1, gene 5)

2. Exactly one parent has the correct allele (as in Figure 1, gene 1)

5

3. Neither parent has a correct allele (as in Figure 1, gene 2)

Define η_1, η_2 and η_3 as the number of genes falling into each of these three cases. Then $n = \eta_1 + \eta_2 + \eta_3$.

For any gene in case 1, the offspring will certainly have the correct allele. For a gene in case 2, there is a 1/2 chance of having the correct allele, since we assume an unbiased coin is tossed during uniform crossover. For a gene in case 3, there is no chance of generating a correct allele. Then, for the offspring of these parents,

$$S_i^k = \eta_1 + \sum_{j=1}^{\eta_2} w_j,$$

where the w_j are $i.i.d.$ Bernoulli (1/2) random variables.

Lemma 1: Under Assumption 3, η_1 has a hypergeometric distribution with parameters n, J_1, J_2, that is, for $\max[0, J_1 + J_2 - n] \leq y \leq \min[J_1, J_2]$,

$$P(\eta_1 = y | S_{p_1}^{k-1} = J_1, S_{p_2}^{k-1} = J_2) = \frac{\binom{J_1}{y}\binom{n-J_1}{J_2-y}}{\binom{n}{J_2}}.$$

Proof: Without loss of generality, fix the correct allele indicators for the second parent in the grid. Then we can begin crossover by randomly placing the indicators of parent 1 so that there are exactly J_1 ones. This is probabilistically equivalent to random selection without replacement. Consider an urn with n balls of which J_2 are red. Select J_1 balls from the urn. The number of red you select is η_1, the number of genes in case 1. Hence η_1 has a hypergeometric distribution with parameters n, J_1, J_2. The mathematical statement follows from the definition of hypergeometric. The bounds on η_1 come from the fact that if $J_1 + J_2 > n$, then there must be some genes in case 1 by the pigeon hole principle, and the fact the number of correct alleles in each parent is an upper bound on η_1. ∎

We can now condition on η_1 to find the distribution of S_i^k given its parentage.

Lemma 2: Without PTS, for $\max[0, J_1 + J_2 - n] \leq J_0 \leq \min[n, J_1 + J_2]$,

$$P(S_i^k = J_0 | S_{p_1}^{k-1} = J_1, S_{p_2}^{k-1} = J_2)$$

$$= \sum_{y=\max[0, J_1+J_2-n]}^{\min[J_1, J_2]} \binom{J_1 + J_2 - 2y}{J_0 - y} \left(\frac{1}{2}\right)^{J_1+J_2-2y} \left[\frac{\binom{J_1}{y}\binom{n-J_1}{J_2-y}}{\binom{n}{J_2}}\right], \quad (1)$$

where S_i^k is the offspring of $S_{p_1}^{k-1}$ and $S_{p_2}^{k-1}$.

Proof: We know that there are a total of $J_1 + J_2$ correct alleles between the two parents. If η_1 genes have two correct alleles, there are $J_1 + J_2 - 2\eta_1$ remaining. These must all be at genes in case 2, one per gene. Hence,

$$\eta_2 = J_1 + J_2 - 2\eta_1, \text{ and } S_i^k = \eta_1 + \sum_{j=1}^{J_1+J_2-2\eta_1} w_j.$$

Given parentage with J_1 and J_2 correct alleles, the offspring will have at least $\max[0, J_1 + J_2 - n]$ correct alleles. If $J_1 + J_2 > n$, there must be some genes in case 1. At most, the number of correct alleles is $\min[n, J_1 + J_2]$. In the derivation below, assume J_0 is in this range. From Lemma 1, $\max[0, J_1 + J_2 - n] \leq y \leq \min[J_1, J_2]$. Conditioning on η_1,

$$P(S_i^k = J_0 | S_{p_1}^{k-1} = J_1, S_{p_2}^{k-1} = J_2)$$

$$= \sum_y P(S_i^k = J_0 | \eta_1 = y, S_{p_1}^{k-1} = J_1, S_{p_2}^{k-1} = J_2) P(\eta_1 = y | S_{p_1}^{k-1} = J_1, S_{p_2}^{k-1} = J_2)$$

$$= \sum_y P(y + \sum_{j=1}^{J_1+J_2-2y} w_j = J_0) \left[\frac{\binom{J_1}{y}\binom{n-J_1}{J_2-y}}{\binom{n}{J_2}} \right]$$

$$= \sum_y \binom{J_1 + J_2 - 2y}{J_0 - y} \left(\frac{1}{2}\right)^{J_1+J_2-2y} \left[\frac{\binom{J_1}{y}\binom{n-J_1}{J_2-y}}{\binom{n}{J_2}} \right],$$

where $\binom{a}{b} = 0$ if $b < 0$ or $b > a$. ∎

Remark: From this it is simple to derive that $E[S_i^k | S_{p_1}^{k-1} = J_1, S_{p_2}^{k-1} = J_2] = \frac{J_1+J_2}{2}$.

We build upon this result to add PTS. If the parents have a total of $J_1 + J_2$ correct alleles and the offspring being analyzed has J_0, then the other offspring has $J_1 + J_2 - J_0$. PTS selects the larger of these values to add to the next generation. That is, with PTS

$$S_i^k = \eta_1 + \max\{ \sum_{j=1}^{J_1+J_2-2\eta_1} w_j, \sum_{j=1}^{J_1+J_2-2\eta_1} (1 - w_j)\}.$$

The maximum value J_0 can take given that the parents have J_1 and J_2 correct alleles remains unchanged when PTS is added. It is based on structure independent of crossover.

7

The smallest value changes upward. Without PTS, it is possible that we get unlucky such that most correct alleles are not selected by the Bernoulli coin-flips. With PTS, these lost correct alleles are simply moved to the alternative offspring and selected during the tournament.

Lemma 3: Given parents with J_1 and J_2 correct alleles and a uniform crossover with PTS, the minimum number of correct alleles in any offspring is $\lceil \frac{J_1+J_2}{2} \rceil$.

Proof: $\eta_2 = J_1 + J_2 - 2\eta_1$. At least half of these alleles are in the better of the two offspring. Hence, the number of correct alleles in the better offspring is at least $\eta_1 + \lceil \frac{J_1+J_2-2\eta_1}{2} \rceil = \lceil \frac{J_1+J_2}{2} \rceil$. ∎

We are now ready to state the offspring distribution with PTS as a Corollary to Lemma 2. The only subtlety is that if $J_1 + J_2$ is even, there is only one outcome with $J_0 = \frac{J_1+J_2}{2}$. In all other instances, either offspring may total J_0 correct alleles.

Corollary 4: With PTS, for $\lceil \frac{J_1+J_2}{2} \rceil \leq J_0 \leq \min[n, J_1 + J_2]$,

Case 1: If $J_0 = \frac{J_1+J_2}{2}$

$$
P(S_i^k = J_0 | S_{p_1}^{k-1} = J_1, S_{p_2}^{k-1} = J_2)
$$
$$
= \sum_{y=\max[0, J_1+J_2-n]}^{\min[J_1, J_2]} \binom{J_1+J_2-2y}{J_0-y} \left(\frac{1}{2}\right)^{J_1+J_2-2y} \left[\frac{\binom{J_1}{y}\binom{n-J_1}{J_2-y}}{\binom{n}{J_2}} \right].
$$

Case 2: If $J_0 > \frac{J_1+J_2}{2}$

$$
P(S_i^k = J_0 | S_{p_1}^{k-1} = J_1, S_{p_2}^{k-1} = J_2) =
$$
$$
\sum_{y=\max[0, J_1+J_2-n]}^{\min[J_1, J_2]} \left[\binom{J_1+J_2-2y}{J_0-y} + \binom{J_1+J_2-2y}{J_1+J_2-J_0-y} \right] \tag{2}
$$
$$
\times \left(\frac{1}{2}\right)^{J_1+J_2-2y} \left[\frac{\binom{J_1}{y}\binom{n-J_1}{J_2-y}}{\binom{n}{J_2}} \right],
$$

where S_i^k is the offspring of $S_{p_1}^{k-1}$ and $S_{p_2}^{k-1}$ and $\binom{a}{b} = 0$ if $b < 0$ or $b > a$.

Proof: Equation (2) follows from adding the combinatorial coefficient in (1) corresponding to the appropriate number of correct alleles being assigned to one offspring, and the
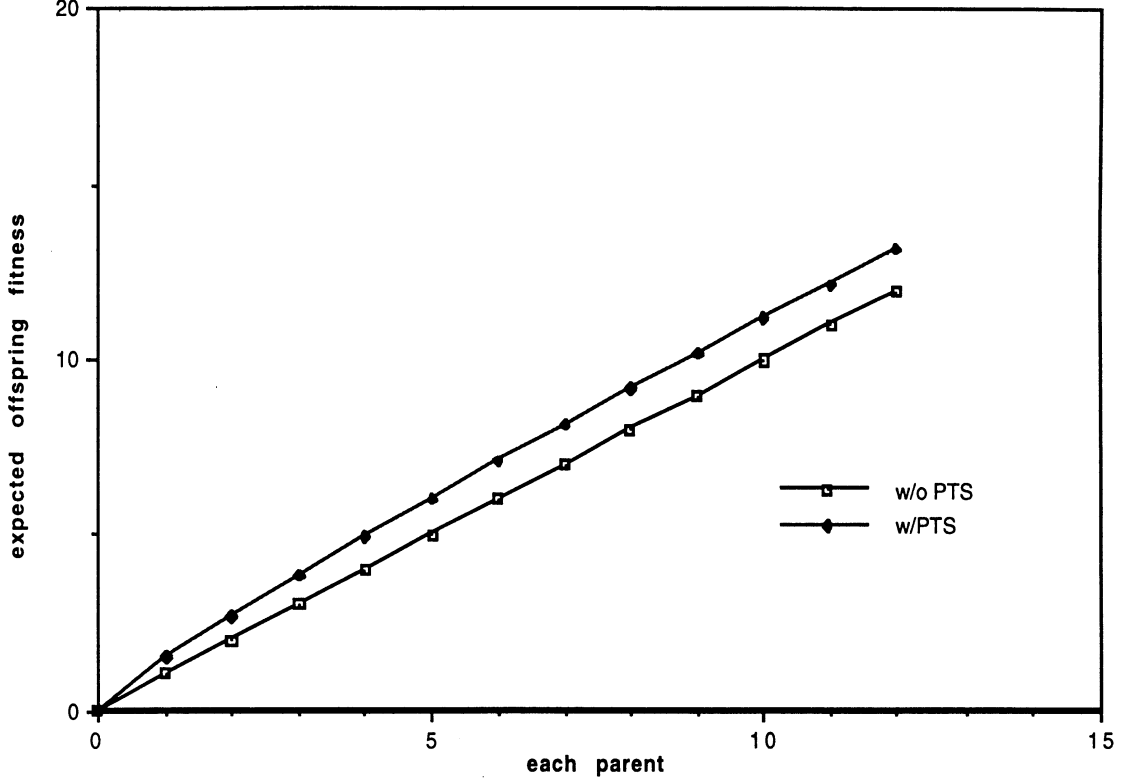
8

Figure 2: $E[S_i^k|S_{p_1}^{k-1} = J_1, S_{p_2}^{k-1} = J_2]$ with and without PTS

coefficient corresponding to the appropriate number of correct alleles being assigned to the other offspring. ∎

The difference between $E[S_i^k|S_{p_1}^{k-1} = J_1, S_{p_2}^{k-1} = J_2]$ without PTS in (1), and with PTS in (2) is crucial to understanding the MCGA. In (1) the offspring generated has the same expected fitness as the average of the parents. With PTS, the offspring is superior to the parents. Figure 2 shows a plot of this factor. For an example with $n = 20$, the x-axis shows $J_1 = J_2$. The lower plot shows $E[S_i^k|S_{p_1}^{k-1} = J_1, S_{p_2}^{k-1} = J_2] = J_1 = J_2$ for the case without PTS. The upper plot shows $E[S_i^k|S_{p_1}^{k-1} = J_1, S_{p_2}^{k-1} = J_2]$ with PTS. It is, approximately, one correct allele higher than without PTS.

The analysis above holds for $S_i^k, i < m$. The distribution of S_m^k, the immigrant, is given by $F(\cdot)$.

2.3 Independence

To move from understanding a single offspring to understanding the next generation
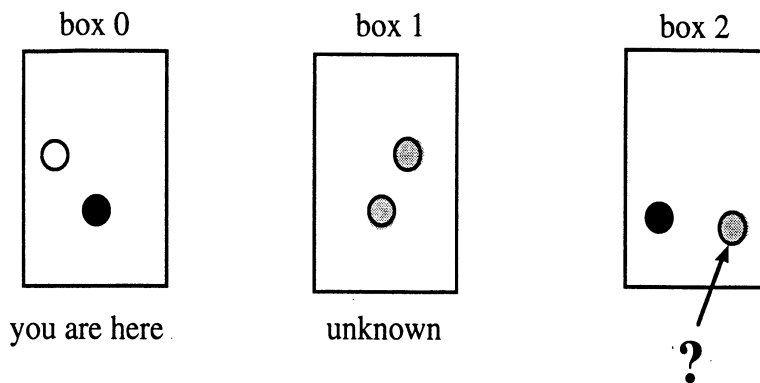
Figure 3: Independence of individuals

we must understand the stochastic interdependencies between members of a population. In generation 0, members of the population are independent by construction making such analysis simple (see Section 2.1). Viewed from time 0, members of generation k are dependent for $k \geq 2$. However, their correlation is greatly reduced given the distribution of the fitness values in population $k - 1$ and sufficiently large m. We describe a simplified experiment to communicate these issues.

The following simple experiment has the same dependence structure as the MCGA and helps us understand the intricate web of interdependence caused by this GA. Consider an infinite sequence of boxes. In box 0 place one white ball and one black ball. Randomly select a ball from box 0 and place a ball of the same color in box 1. Repeat with replacement. Now box 1 has two balls. Randomly select a ball from box 1 and place a ball of the same color in box 2, repeat with replacement. Repeat indefinitely. The result is an infinite sequence of boxes, each containing two balls. Since the experiment is perfectly symmetric between white and black balls, if we randomly take a ball from box k, $P(black) = 1/2$.

By independence of the trials used to determine the balls in box 1, they are independent. Knowing that the first ball selected is black has no impact on the likelihood of the other being black. However, viewed from the beginning of the experiment, the balls in box 2 are dependent. Note that from this perspective we do not know the contents of box 1. Hence, if the first ball selected for box 2 is known to be black, it implies that box 1 is not all white (without information, this event had a 1/4 probability). Now the likelihood of the second ball in box 2 being black increases from 1/2 to 3/4. See Figure 3 for a graphical description.

10

Suppose that you are given that box 1 is half black and half white. Now the information that the first ball in box 2 is black has no impact on the other ball. The realization of box 1 makes the two balls in box 2 conditionally independent. That is, independence depends not on knowing the distribution of possible items in box 1, but the realization of items in box 1.

The dependence of colors of balls in box 2 is extreme in this example due to the small sample sizes. Consider the same experiment with 20 balls per box. Box 0 is known to have 10 white and 10 black balls. Box 1 is filled by 20 independent replications as above, and so on. Now assume a ball has been observed in box 2 to be black. How does that change our distribution of white versus black for the next trial? Box 1 has 20 balls. One is known now to be black, and the others are equally likely to be black or white. Hence, the next ball chosen has probability $1/2 + 0.5/20$ of being black. If the experiment has m balls per box, this probability is $1/2 + 1/(2m)$. As m increases, the amount of information gleaned from one replication of the selection fades. The colors of balls in box two become increasingly uncorrelated.

From this motivation, we make the following approximation.

Approximation 1: For sufficiently large m, approximate

$$P(S_i^k \leq J_i, i = 1, 2, \dots m) \approx \prod_{i=1}^{m} P(S_i^k \leq J_i).$$

We can now approximate the important values for each generation.

There is a subtlety here since random reproduction has a $1/m$ chance of selecting the immigrant for a parent. Let $S_{p_1}^{k-1}$ and $S_{p_2}^{k-1}$ be the randomly selected parents. Then $P(S_{p_1}^{k-1} \leq J) = \left(\frac{m-1}{m}\right) P(S_i^{k-1} \leq J) + \left(\frac{1}{m}\right) F(J)$, where $P(S_i^{k-1} \leq J)$ is given in (1) or (2). The other parent has the same distribution.

Conditioning on parentage, for $i < m$,

$$P(S_i^{k'} \leq J_0) \approx$$
$$\sum_{J_1=0}^{n} \sum_{J_2=0}^{n} P(S_i^k \leq J_0 | S_{p_1}^{k-1} = J_1, S_{p_2}^{k-1} = J_2) P(S_{p_1}^{k-1} = J_1) P(S_{p_2}^{k-1} = J_2). \tag{3}$$

The first factor is known from (1) or (2). The last two factors are known from an inductive hypothesis. Then $E[S_i^k] \approx \sum_{J_0=0}^{n}[1 - P(S_i^k \leq J_0)]$.

By Approximation 1, $P(A^k \leq J_0) \approx [P(S_i^k \leq J_0)]^{m-1}P(S_i^0 \leq J_0)$, the last term coming from the immigrant. Finally, $E[A^k] \approx \sum_{J_0=0}^{n-1}[1 - P(A^k \leq J_0)]$.

To compute $E[A^k]$, we compute the anchor distribution of S_i^0 as in Section 2.1, then inductively compute the distributions of S_i^k via (3).
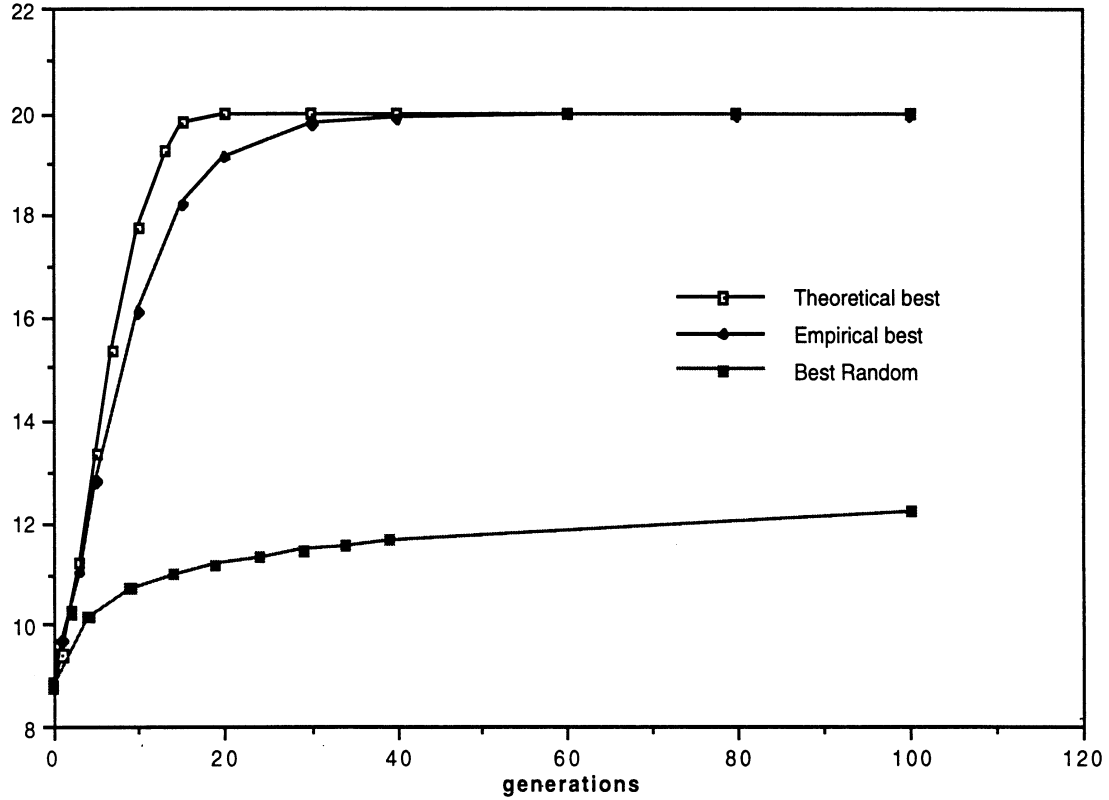

Figure 4: Best solution found by random sampling, theoretical GA and empirical GA (average of 30 runs), $n = 20, m = 20$

3. Empirical Tests

To evaluate the theory presented in Section 2 we conducted empirical tests of appropriate MCGAs to compare theory with practice. The underlying problem fits Assumptions 1 and 2. In the first test, we assume $n = 20, m = 20, a = 4$. We ran 30 random number seeds for 100 generations each. Figure 4 presents the average of the best fitness found in generation k over the 30 runs. It is compared with $E[A^k]$ as derived in Section 2 (with

12

PTS). Recall that 20 correct alleles is perfect for $n = 20$.

As a benchmark, Figure 4 also plots the theoretical progression of an algorithm which generates 20 immigrants each generation. That is, it represents pure random sampling of the space. The substantial improvement of the empirical and theoretical MCGA curves over the random sampling curve demonstrates the contribution made by crossover beyond pure mutation.
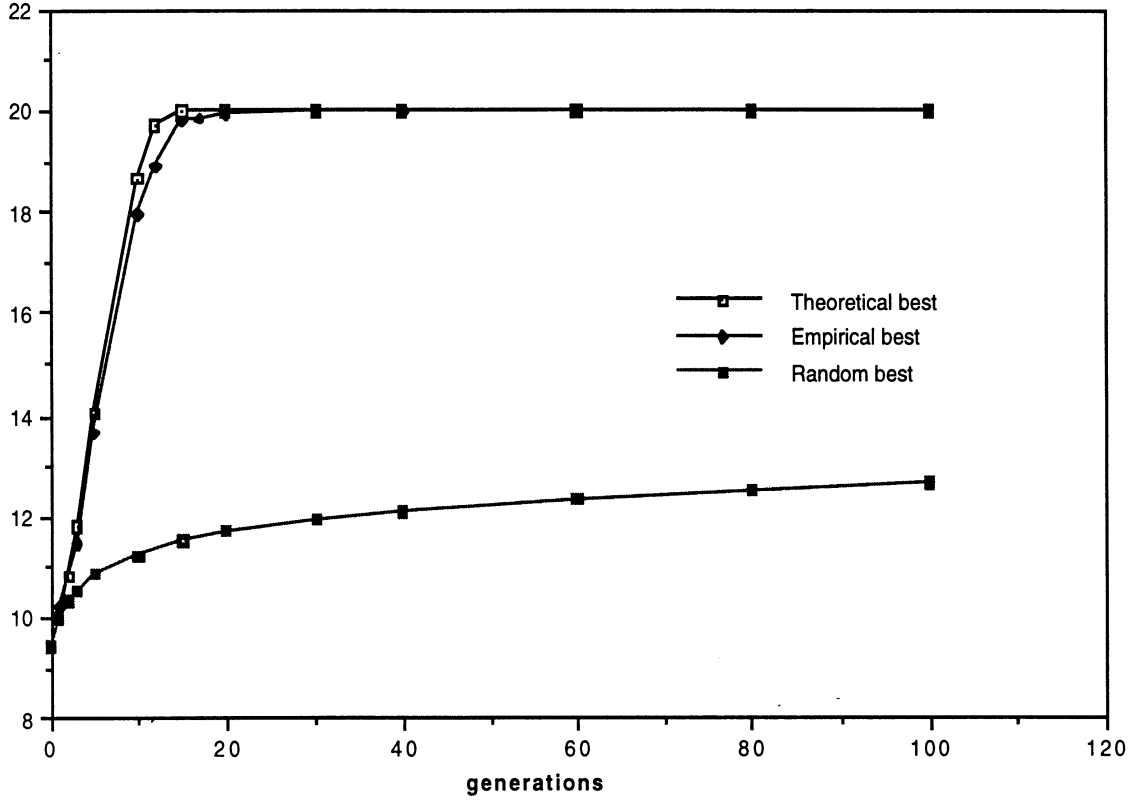

Figure 5: Best solution found by random sampling, theoretical GA
and empirical GA (average of 30 runs), $n = 20, m = 40$

The theoretical and empirical curves track well. The only deviation comes at the shoulder as the curves near perfection. We believe that this error is due to Assumption 3 and Approximation 1. The theory assumes uniform distribution of correct alleles. Near convergence of the algorithm, the empirical runs do not evidence this behavior in a problem with this small a population size. Further, Approximation 1 is less accurate with small m.

To further evaluate the theory, Assumption 3 and Approximation 1, the experiment

was rerun with $m = 40$, that is, the same chromosome length, but double the population. Figure 5 gives the random sampling benchmark, empirical average of 30 runs and theoretical $E[A^k]$. In this experiment the theory tracks very closely with the empirical results. Assumption 3 and Approximation 1 are more valid in the empirical problem with a larger population since it retains more entropy throughout the runs.
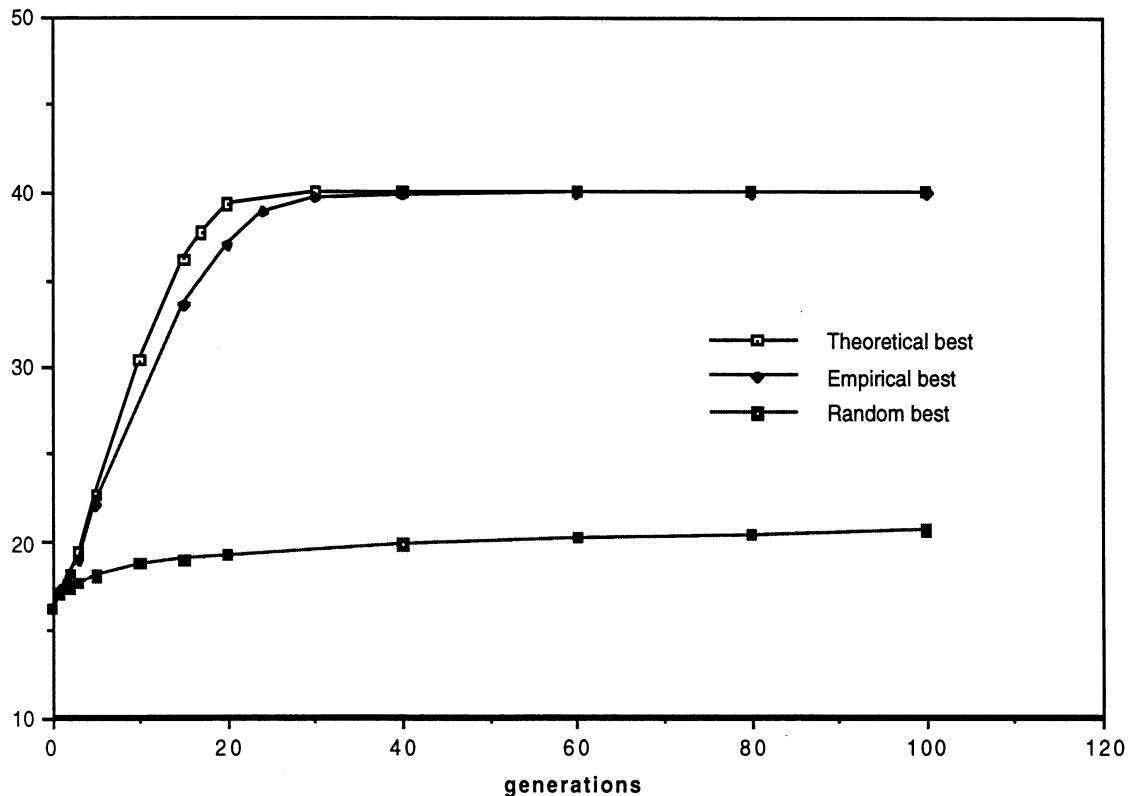

Figure 6: Best solution found by random sampling, theoretical GA
and empirical GA (average of 30 runs), $n = 40, m = 40$

To test scalability, a third experiment was run with $n = 40, m = 40$. Now the chromosomes are doubled in length as well. Figure 6 gives the results which qualitatively replicate Figure 4.

Figures 7, 8 and 9 show the theoretical and empirical population averages, $E[S_i^k]$ vs. the empirical average, for the same experiments. These results show the impact of the assumptions. In each graph, the theoretical population average converges to a higher level than the empirical. Assumption 3 and Approximation 1 allow the theory to reach perfection

14

in the chromosomes other than the immigrant. The immigrant always has expectation np. The empirical results show the effects of premature convergence in some runs that is not possible in the theory due to Assumption 3. In the best of the empirical runs, the GA graphs very close to the theoretical prediction. Again, the test with larger population (Figure 8) performed much better due to the closer adherence to the Assumptions.
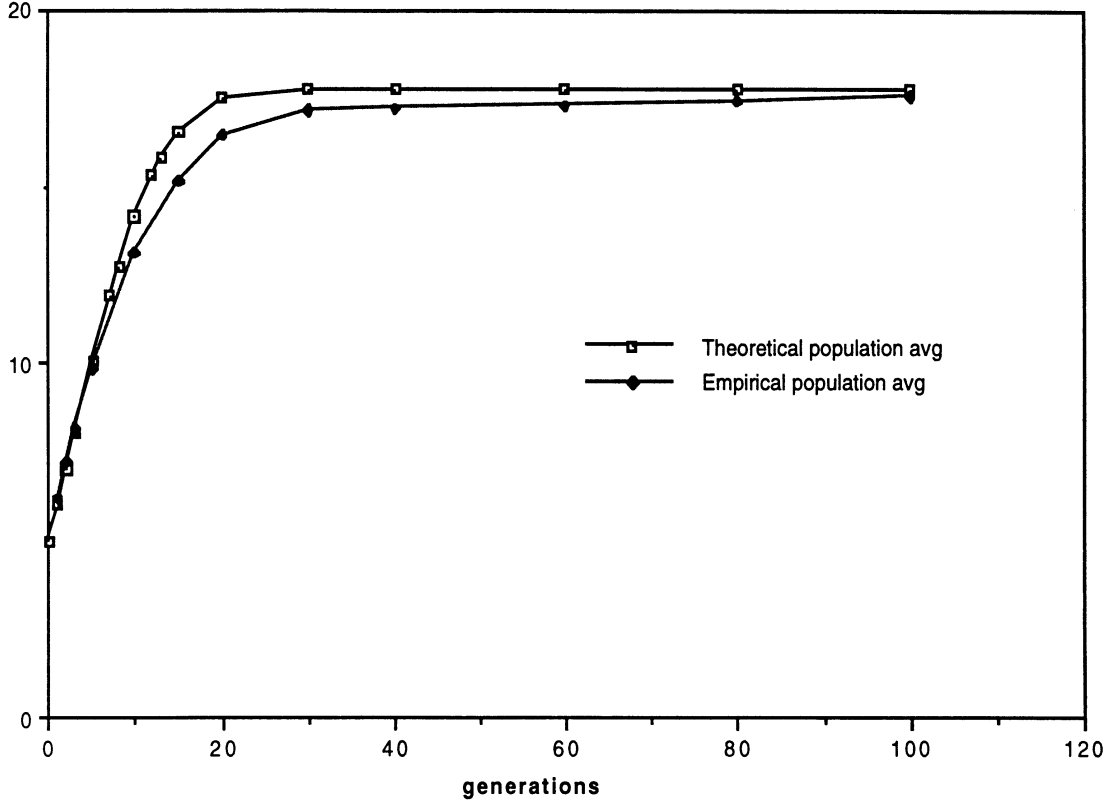

Figure 7: Theoretical and empirical population averages, $n = 20, m = 20$

4. Conclusions and Extensions

We have presented a probabilistic model for the multiple-choice genetic algorithm. The model allows recursive forward computation of the expected fitness of the best solution found and expected population average fitness for each generation. Several simplifying assumptions are made to allow construction of this model. Empirical runs are made to evaluate the predictions of the model. In all runs, the model predicts the performance of the MCGA relatively well. What error exists appears to derive from the assumption that correct alleles are uniformly distributed across a chromosome and correlation approxima-
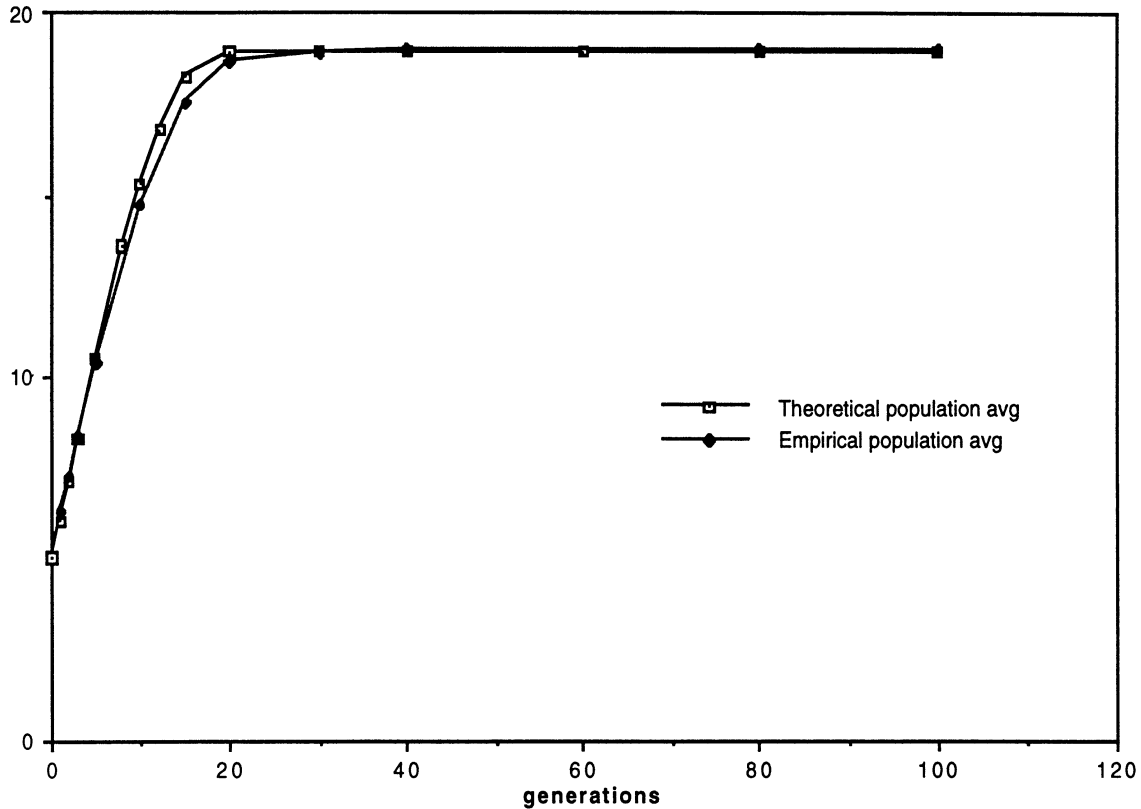
Figure 8: Theoretical and empirical population averages, $n = 20, m = 40$

tion.

The MCGA does not use traditional roulette wheel reproduction, one-point crossover, or mutation. It uses random reproduction, uniform crossover with post-tournament selection and immigration. Each of these operators is more easily modeled than the traditional operators, but still results in a computationally effective GA. The results in Figures 4, 5 and 6 show that random search is not the driving force behind this algorithm. The algorithm performs much better than random search would predict. Results shown in Figure 2 indicate that the post-tournament selection is the driving force behind the algorithm. Uniform crossover constructs two offspring. They are evaluated and only the better entered into the new population. The result is an expected offspring fitness that gains approximately one correct allele per generation.

Perhaps the most interesting observation to be made from this study is the role of entropy vs. building blocks and population convergence. In all empirical runs, the MCGA falls short of the performance predicted by the theory. We conjecture that this is primarily
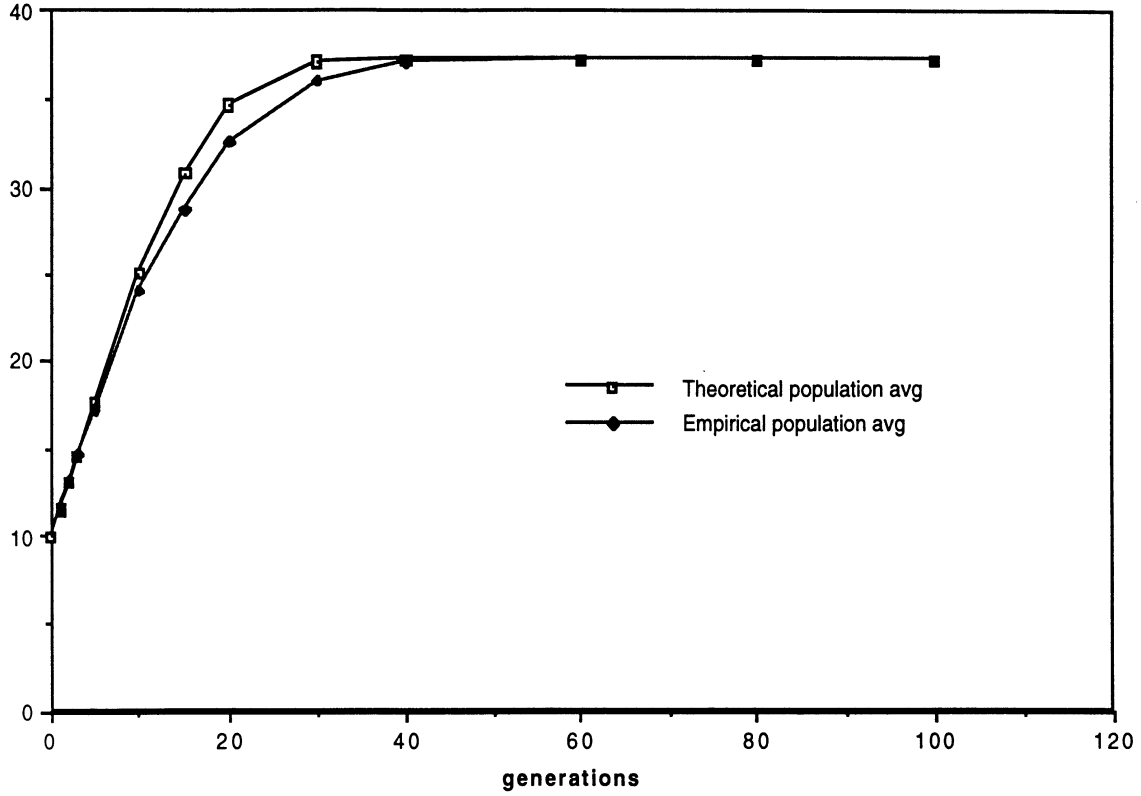
Figure 9: Theoretical and empirical population averages, $n = 40, m = 40$

due to Assumption 3 and Approximation 1. This suggests that the empirical GAs do not perform as well as the theory because of the formation of building blocks and population convergence. The theoretical MCGA uses the randomness of Assumption 3 to plug holes in the tableau of correct alleles. Further discussion and study will be necessary to validate this conjecture and to see if it extends to more common GAs.

The modeling presented above does not tell us why genetic algorithms work in general. For a specific GA, it gives substantial insight for the inner workings of the GA dynamic. It also gives an approach for analyzing genetic algorithms. In future research we will relax the uniform correct allele assumption, consider more traditional operators and consider more realistic fitness functions.

REFERENCES

Ankenbrandt, C. [1991], "An Extension to the Theory of Convergence and a Proof of the Time Complexity of Genetic Algorithms," **Foundations of Genetic Algorithms,** pp. 53-68.

Bean, J. and A. Hadj-Alouane [1992], "A Dual Genetic Algorithm for Bounded Integer Programs," To appear in **R.A.I.R.O.-R.O.**

Goldberg, D. [1989], **Genetic Algorithms in Search Optimization and Machine Learning,** Addison Wesley.

Goldberg, D. and P. Segrest [1987], "Finite Markov Chain Analysis of Genetic Algorithms," **Proceedings of the Second International Conference on Genetic Algorithms,** pp. 1-8.

Hadj-Alouane, A. and J. Bean [1992], "A Genetic Algorithm for the Multiple-Choice Integer Program," To appear in **Operations Research.**

Hadj-Alouane, A., J. Bean and K. Murty [1993], "A Hybrid Genetic/Optimization Algorithm for a Task Allocation Problem," Technical Report 93-30, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, 48109.

Holland, J. [1975], **Adaptation in Natural and Artificial Systems,** The University of Michigan Press.

Horn, J. [1993], "Finite Markov Chain Analysis of Genetic Algorithms with Niching," **Proceedings of the Fifth International Conference on Genetic Algorithms,** pp. 110-117.

Michalewicz, Z. [1994], **Genetic Algorithms + Data Structure = Evolution Programs,** Springer-Verlag.

Rabinovich, Y. and A. Wigderson [1991], "An Analysis of a Simple Genetic Algorithm," **Proceedings of the Fourth International Conference on Genetic Algorithms,** pp. 215-221.

Spears, W. and K. De Jong [1991a], "On the Virtues of Parameterized Uniform Crossover," **Proc. of the Fourth International Conference on Genetic Algorithms,** pp.

230-236.

Spears, W. and K. De Jong [1991b], "An Analysis of Multi-point Crossover," **Foundations of Genetic Algorithms**, pp. 301-315.

Suzuki, J. [1993], "A Markov Chain Analysis on a Genetic Algorithm," **Proceedings of the Fifth International Conference on Genetic Algorithms**, pp. 146-154.

Syswerda, G. [1989], "Uniform Crossover in Genetic Algorithms," **Proceedings of the Third International Conference on Genetic Algorithms**, pp. 2-9.