

Infanticide Considered Harmful?

The Effect of Discarding Offspring on Schema Frequency Variance

Theodore C. Belding
Center for the Study of Complex Systems
University of Michigan
Ann Arbor, MI 48109-1120, USA
Ted.Belding@umich.edu
<http://www-personal.umich.edu/~streak/>

July 20, 2004

Abstract

Most implementations of genetic algorithms (GAs) keep both of the offspring that result from crossover, while many theoretical results (such as the schema theorem) assume that one child is discarded from each mating pair, to simplify the mathematics. Since the expected frequency of a given schema is the same in both cases, this has no effect on the derivation of the schema theorem. However, Booker has given empirical evidence that discarding a child can increase allele loss. Holland has suggested that this would increase the variance in schema frequency. This paper investigates whether discarding a child does in fact increase allele loss and variance. Theoretical and experimental results are presented showing that under fitness-proportionate selection, the number of parents and the number of offspring kept have no effect on the mean or variance of a schema's frequency. The presence or absence of crossover also has no effect on the mean or variance of the frequency of individual alleles. However, discarding a child does increase allele loss and schema variance under remainder-stochastic selection, replicating Booker's results.

1 Introduction

Genetic algorithms (GAs) are computational search methods loosely based on biological evolution [6]. Most current GA implementations keep both of the offspring that result from crossover, while many theoretical results (such as the schema theorem [5, 4]) assume that one child is discarded, in order to simplify the mathematics. Since the expected frequency of a given schema is the same in both cases, this has no effect on the derivation of the schema theorem. However, Booker [1] has given empirical evidence that discarding a child can increase allele loss. Holland (J. H. Holland, personal communication) has suggested that this would increase the variance in schema frequency.

Booker [1] argued that discarding one child from every mating pair would increase allele loss (i.e., the number of loci fixed to all 1s or all 0s):

Discarding a string might be an unwarranted source of allele loss when the string is the last instance of some first order schema. It could also be a source of sampling error when the discarded string is a sample point for under-represented regions of the search space. This would lead, of course, to even more allele loss. [1, pp. 301–302]

To test this, he ran two variants of the GA, using remainder-stochastic selection [2], on De Jong’s [3] test function $F1$ and compared the allele loss in each. In one variant, one child was discarded from each pair, and in the other, both children were kept. He found that the variant in which both children were kept had sharply reduced allele loss (an average of about 13 lost alleles after 120 generations, compared with about 20 lost alleles when one child was thrown away). He also compared online and offline performance on each of the functions in De Jong’s test suite, $F1$ – $F5$, and showed that both online and offline performance were improved when both children were kept, except that offline performance was the same on $F1$ and worse on $F5$. In each case, he calculated the statistics over 10 runs.

This paper investigates under what circumstances discarding a child increases allele loss and variance. Theoretical and experimental results are presented showing that under fitness-proportionate selection the number of parents and the number of offspring kept have no effect on the mean or variance of a schema’s frequency. The presence or absence of crossover also has no effect on the mean or variance of the frequency of individual alleles. However, discarding a child does increase allele loss and schema variance under remainder-stochastic selection [2], confirming Booker’s [1] results.

2 Fitness-Proportionate Selection

2.1 Theoretical Argument

2.1.1 Case 1: One Parent, with Cloning

Consider an individual x in a GA population. Assuming fitness-proportionate selection on a flat fitness landscape (or random selection) and no crossover or mutation, the process of selecting the next generation’s population is a sequence of n Bernoulli trials, each having a probability p of selecting x , where n is the population size and $p = P(x, t)$ is the proportion of x in the current population. Hence, the expected frequency m of x in the next generation is simply

$$\begin{aligned} E(m(x, t + 1)) &= np \\ &= nP(x, t), \end{aligned} \tag{1}$$

and the variance of the frequency is

$$\begin{aligned} V(m(x, t + 1)) &= np(1 - p) \\ &= nP(x, t)(1 - P(x, t)). \end{aligned} \tag{2}$$

The same analysis holds for the frequency of a schema H , and for fitness-proportionate selection on other fitness landscapes (substituting the probability of selecting H for p).

Table 1: Variance in genotype frequency for the case of one parent and no crossover or mutation under fitness proportionate selection on a flat fitness landscape (Case 1). x is the genotype, $E(V)$ is the expected value (mean) of the variance of x 's frequency, \bar{v} is the sample average of the variance, s_v is the sample standard deviation of the variance, and Pr is the two-sided normal p -value that the difference between $E(V)$ and \bar{v} is at least as great as observed, assuming H_0 is true. Statistics are calculated over 100 sets of runs. Population size n was 100.

x	$E(V)$	\bar{v}	s_v	Pr
00	18.75	18.96	2.52	0.40
01	18.75	18.45	2.80	0.28
10	18.75	18.79	2.66	0.88
11	18.75	18.38	2.89	0.20

To verify these results, a simple simulation was performed. A population of 100 individuals was created, with 25 instances of each of 4 genotypes (00, 01, 10, and 11). A new population of 100 individuals was created by drawing from the initial population 100 times, with replacement, and the frequency of each genotype was recorded. This was performed 100 times, and the variance of each genotype was calculated over these 100 runs. This process was in turn repeated for a total of 100 sets of runs, and the average and standard deviation of the variance was calculated, for each genotype, over these 100 sets. The results are shown in Table 1. As can be seen, the experimental results agree very well with the expected value of the variance from the theory. A two-sided z -test shows that the differences are consistent with random variation (the significance level α was taken to be 0.01 in all statistical tests in this paper).

2.1.2 Case 2: Two Parents

Now let us examine the case where each pair of offspring has two parents but there is no crossover. (In other words, the canonical simple GA with crossover and mutation turned off.) Then,

$$\begin{aligned}
 E(m(x, t + 1)) &= E(M_1) + E(M_2) \\
 &= \frac{n}{2}p + \frac{n}{2}p \\
 &= np \\
 &= nP(x, t),
 \end{aligned} \tag{3}$$

and

$$\begin{aligned}
 V(m(x, t + 1)) &= V(M_1) + V(M_2) \\
 &= \frac{n}{2}p(1 - p) + \frac{n}{2}p(1 - p) \\
 &= np(1 - p) \\
 &= nP(x, t)(1 - P(x, t)).
 \end{aligned} \tag{4}$$

Here, M_1 and M_2 are the number of instances of x originating in the first and second parent, respectively. Since the parents are drawn with replacement from the population, these are independent random variables, and the variance of the sum is simply the sum of the variances. The

Table 2: Variance in genotype frequency for the case of two parents, two children, and no crossover or mutation, using fitness proportionate selection on a flat fitness landscape (Case 2). x is the genotype, $E(V)$ is the expected value (mean) of the variance of x 's frequency, \bar{v} is the sample average of the variance, s_v is the sample standard deviation of the variance, and Pr is the two-sided normal p -value that the difference between $E(V)$ and \bar{v} is at least as great as observed, assuming H_0 is true. Statistics are calculated over 100 sets of runs. Population size n was 100.

x	$E(V)$	\bar{v}	s_v	Pr
00	18.75	18.76	3.06	0.97
01	18.75	18.56	2.62	0.47
10	18.75	19.01	2.37	0.27
11	18.75	18.74	2.63	0.97

results are the same as in Equations 1 and 2. Hence, adding a parent without crossover has no effect on the mean or variance, which is intuitive: Simply grouping the draws of parents into pairs should have no effect on the outcome.

Another set of experiments was performed to confirm this result, with the same setup as in Section 2.1.1, except that pairs of individuals were drawn 50 times (each individual was still drawn with replacement). The results are shown in Table 2. The results here, too, are in close agreement with the theory. A two-sided z -test shows that the differences are consistent with random variation.

2.1.3 Case 3: Two Parents, One Child

What happens if we change the scenario from Section 2.1.2 so that one child is discarded at random from each pair, and the number of pairs of parents selected is doubled? Then,

$$\begin{aligned}
 p &= \frac{1}{2}p_1 + \frac{1}{2}p_2 \\
 &= \frac{1}{2}P(x, t) + \frac{1}{2}P(x, t) \\
 &= P(x, t),
 \end{aligned} \tag{5}$$

where p_1 and p_2 are the (independent) success probabilities for the first and second parent, respectively. The remaining analysis remains the same as in Section 2.1.1, with the same results:

$$\begin{aligned}
 E(m(x, t + 1)) &= np \\
 &= nP(x, t),
 \end{aligned} \tag{6}$$

and

$$\begin{aligned}
 V(m(x, t + 1)) &= np(1 - p) \\
 &= nP(x, t)(1 - P(x, t)).
 \end{aligned} \tag{7}$$

Intuitively, throwing out the result of every other Bernoulli trial should not have any effect on the results, as long as the number of trials is doubled, so that the results of n trials are kept.

Table 3: Variance in genotype frequency for the case of two parents, one child, and no crossover or mutation, using fitness-proportionate selection on a flat fitness landscape (Case 3). x is the genotype, $E(V)$ is the expected value (mean) of the variance of x 's frequency, \bar{v} is the sample average of the variance, and s_v is the sample standard deviation of the variance, and Pr is the two-sided normal p -value that the difference between $E(V)$ and \bar{v} is at least as great as observed, assuming H_0 is true. Statistics are calculated over 100 sets of runs. Population size n was 100.

x	$E(V)$	\bar{v}	s_v	Pr
00	18.75	18.85	2.74	0.72
01	18.75	18.75	2.56	1.00
10	18.75	19.06	2.82	0.27
11	18.75	19.03	2.73	0.31

Once again, a set of experiments was performed. In this case, two individuals were drawn with replacement 100 times, with one individual from each pair discarded at random. The results are shown in Table 3, and are again in close agreement with the predicted values. A two-sided z -test shows that the differences are consistent with random variation.

2.1.4 Case 4: Two Parents, with Crossover

What happens if crossover is added to the scenario from Section 2.1.2? Then,

$$\begin{aligned}
 E(m(H, t + 1)) &= E(N_1 + N_2) \\
 &= E([(1 - p_c(H))M_1 + p_c(H)M_2] + [(1 - p_c(H))M_2 + p_c(H)M_1]) \\
 &= E(M_1 + M_2) \\
 &= E(M_1) + E(M_2) \\
 &= \frac{n}{2}p + \frac{n}{2}p \\
 &= np \\
 &= nP(H, t),
 \end{aligned} \tag{8}$$

and

$$\begin{aligned}
 V(m(H, t + 1)) &= V(N_1 + N_2) \\
 &= V([(1 - p_c(H))M_1 + p_c(H)M_2] + [(1 - p_c(H))M_2 + p_c(H)M_1]) \\
 &= V(M_1 + M_2) \\
 &= V(M_1) + V(M_2) \\
 &= \frac{n}{2}p(1 - p) + \frac{n}{2}p(1 - p) \\
 &= np(1 - p) \\
 &= nP(H, t)(1 - P(H, t)).
 \end{aligned} \tag{9}$$

Here, M_1 and M_2 are the number of instances of schema H originating in the first and second parent, respectively, and N_1 and N_2 are the number of instances in the first and second offspring,

Table 4: Variance in allele frequency for the case of two parents, two children, with crossover but without mutation, using fitness-proportionate selection on a flat fitness landscape (Case 4). H is the allele, $E(V)$ is the expected value (mean) of the variance of H 's frequency, \bar{v} is the sample average of the variance, s_v is the sample standard deviation of the variance, and Pr is the two-sided normal p -value that the difference between $E(V)$ and \bar{v} is at least as great as observed, assuming H_0 is true. Statistics are calculated over 100 sets of runs. The crossover probability p_c was 0.5 per pair, and the crossover point was always between the two loci. The population size n was 100.

H	$E(V)$	\bar{v}	s_v	Pr
*0	25.0	24.46	3.08	0.08
*1	25.0	24.46	3.08	0.08
0*	25.0	25.64	3.48	0.07
1*	25.0	25.64	3.48	0.07

after crossover. $p_c(H)$ is the probability that crossover exchanges the schema's defining loci. It is assumed that crossover does not disrupt or create schemata; this holds if the schemata in question are single alleles.

The results again remain the same (substituting H for x). Intuitively, since there is no chance of H being disrupted by mutation or crossover, if a parent contains H , one of its offspring will also, and this occurs with a probability of p , with n parents being chosen in total. In other words, the number of copies of an allele in a pair of individuals remains the same after crossover.

A set of experiments was performed to confirm these results. The setup was the same as in Section 2.1.2, except that crossover was performed with a probability $p_c = 0.5$ per pair (the crossover point was always between the two loci), and allele (schema) frequencies were counted instead of genotype frequencies. The results are shown in Table 4. The results are once again in close agreement with the predicted values. A two-sided z -test shows that the differences are consistent with random variation.

2.1.5 Case 5: Two Parents, One Child, with Crossover

Finally, what happens if we change the scenario from Section 2.1.4 so that one child is discarded at random from each pair, and the number of draws from the population is doubled? Then,

$$\begin{aligned}
 p &= \frac{1}{2}q_1 + \frac{1}{2}q_2 \\
 &= \frac{1}{2}[(1 - p_c(H))p_1 + p_c(H)p_2] + \frac{1}{2}[(1 - p_c(H))p_2 + p_c(H)p_1] \\
 &= \frac{1}{2}p_1 + \frac{1}{2}p_2 \\
 &= \frac{1}{2}P(H, t) + \frac{1}{2}P(H, t) \\
 &= P(H, t),
 \end{aligned} \tag{10}$$

where p_1 and p_2 are the success probabilities for the first and second parent, respectively, and q_1 and q_2 are those for the first and second offspring. $p_c(H)$ is the probability that crossover exchanges

Table 5: Variance in allele frequency for the case of two parents, one child, with crossover but without mutation, using fitness-proportionate selection on a flat fitness landscape (Case 6). H is the allele, $E(V)$ is the expected value (mean) of the variance of H 's frequency, \bar{v} is the sample average of the variance, s_v is the sample standard deviation of the variance, and Pr is the two-sided normal p -value that the difference between $E(V)$ and \bar{v} is at least as great as observed, assuming H_0 is true. Statistics are calculated over 100 sets of runs. The crossover probability p_c was 0.5 per pair, and the crossover point was always between the two loci. The population size n was 100.

H	$E(V)$	\bar{v}	s_v	Pr
*0	25.0	24.88	3.04	0.69
*1	25.0	24.88	3.04	0.69
0*	25.0	25.20	3.85	0.60
1*	25.0	25.20	3.85	0.60

the schema's defining loci. As in the previous case, it is assumed that crossover does not disrupt or create schemata. The remaining analysis remains the same as in Section 2.1.1, with the same results (substituting H for x):

$$\begin{aligned}
 E(m(H, t + 1)) &= np \\
 &= nP(H, t),
 \end{aligned}
 \tag{11}$$

and

$$\begin{aligned}
 V(m(H, t + 1)) &= np(1 - p) \\
 &= nP(H, t)(1 - P(H, t)).
 \end{aligned}
 \tag{12}$$

Once again, the results remain the same.

As before, a set of experiments was performed to confirm these results. The setup was the same as in Section 2.1.3, except that crossover was performed with a probability $p_c = 0.5$ per pair (the crossover point was always between the two loci), and allele (schema) frequencies were counted instead of genotype frequencies. The results are shown in Table 5. The results are once again in close agreement with the predicted values. A two-sided z -test shows that the differences are consistent with random variation.

2.1.6 Summary of Theoretical Results for Fitness-Proportionate Selection

Thus, under fitness-proportionate selection, the variance in each case remains $np(1 - p)$, regardless of the number of parents, the number of offspring kept, or the presence or absence of crossover (under the assumption that crossover does not disrupt or create schemata, which holds when the schemata in question are alleles).

In general, note that for any GA variant where the process of generating the population for the next generation can be modeled as a sequence of n Bernoulli trials, with some constant probability p of including an instance of schema H in the $t + 1$ th population, the exact schema theorem has the form

$$E(m(H, t + 1)) = np(H, t),
 \tag{13}$$

and the schema variance can be calculated by the equation

$$V(m(H, t + 1)) = np(H, t)(1 - p(H, t)). \quad (14)$$

Furthermore, for *any* schema theorem variant of the form

$$E(m(H, t + 1)) \geq np(H, t), \quad (15)$$

the corresponding equation for schema variance is

$$V(m(H, t + 1)) \geq np(H, t)(1 - p(H, t)), \quad (16)$$

by the same reasoning.

This observation is implicit in the analysis of Poli et al. [7], but it is worthwhile to emphasize it explicitly. However, this result does not hold for cases where p is a function of variables other than H and t .

2.2 Fitness-Proportionate Selection: Empirical Results

The results of Section 2.1 are intended to model the behavior of a GA under fitness-proportionate selection. However, since the object of study is the GA, experiments should be performed to see whether the theoretical conclusions match the GA’s actual behavior.

I performed two similar experiments measuring allele loss. The first one was conducted on a flat, or neutral, fitness landscape without mutation, to examine the amount of fixation under drift alone for each GA variant. The second experiment was performed on $F1$, to more closely match Booker’s [1] original experiment.

2.2.1 Flat Landscape under Fitness-Proportionate Selection

In the first experiment, the GA was run on a fitness function that simply returned a value of 1 for every individual. This allows us to examine the GA’s behavior in the absence of selective pressure, where drift (sampling error) alone is active. A simple GA (SGA) was run 100 times for 120 generations each, with a population size of 50 and a chromosome length of 30. The crossover rate p_c was 0.6. These are the parameter settings used by Booker [1]. However, mutation was turned off, since mutation acts to generate new alleles, or restore alleles that have been lost in the population. Also, in contrast to Booker’s implementation, no scaling or sampling correction was performed, and the GA was non-elitist. The number of loci that were fixed to either all 0s or all 1s in the population was recorded for each generation and averaged over the 100 runs. In the first set of runs, both children resulting from a pair of parents were kept. In the second set of runs, one child of each pair was discarded at random. The results are plotted in Figure 1, sampled every 10 generations to make the plot readable. The length of the error bars in each direction is the standard deviation. As can be seen, there is almost no variation between the two sets of runs. The two-sided normal p -value that the difference between the two means at generation 120 is at least as great as observed, assuming H_0 is true, is 0.35. Hence, the difference is consistent with random variation.

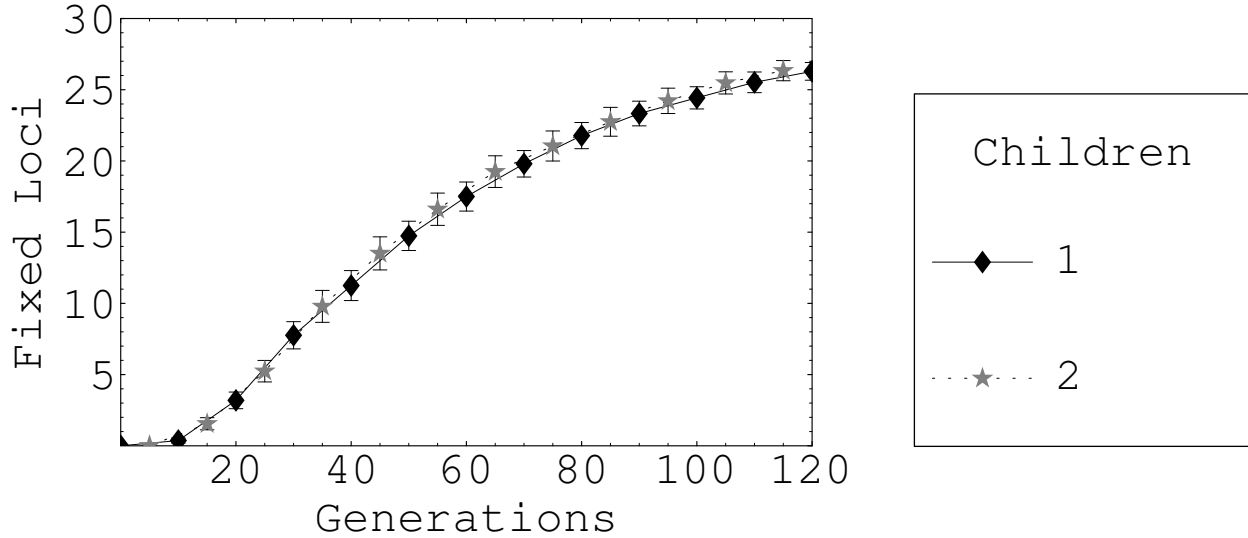


Figure 1: Number of fixed loci (lost alleles) for two variants of the simple GA on a flat fitness landscape, averaged over 100 runs. In the first variant, both children from each pair of parents are kept; in the second, one child is discarded at random. There is no mutation in either variant. Error bars extend three standard errors in each direction. Data points are sampled every 10 generations for readability.

2.2.2 $F1$ under Fitness-Proportionate Selection

Another experiment was performed, this time running an SGA on the fitness function $F1$, which Booker [1] used in his experiment. (The function was actually inverted by calculating the value $f(x) = 80 - F1(x)$, thereby converting it into a maximization problem. Booker performed a similar transformation.) The parameters remained the same as in Section 2.2.1, except that the mutation rate p_m was set to 0.001 per locus, as in Booker’s experiments. (Booker used remainder-stochastic selection [2] instead of fitness-proportionate selection. His implementation was also elitist and used a form of scaling; none of this was done in these experiments.) In one set of runs, both offspring of each pair were kept; in the other, one child was discarded at random. Once again, the number of fixed loci was recorded for each generation, and the results are plotted in Figure 2. Once again, there is very little difference between the two variants. The two-sided normal p -value that that the difference between the two means at generation 120 is at least as great as observed, assuming H_0 is true, is 0.078. Hence, the difference is consistent with random variation.

3 Remainder-Stochastic Selection

I have shown that discarding one child at random does not change the variance in a schema’s frequency under fitness-proportionate selection. However, Booker [1] used remainder-stochastic selection [2] in his experiments, to reduce the difference between expected and actual sampling rates. In that selection method, each individual x is first allocated a number of offspring equal to $\lfloor np(x, t) \rfloor$. The remaining $n - \sum_x \lfloor np(x, t) \rfloor$ offspring are allocated by roulette wheel selection, where each individual’s slice of the roulette wheel is $r(x) = np(x, t) - \lfloor np(x, t) \rfloor$, or the fractional

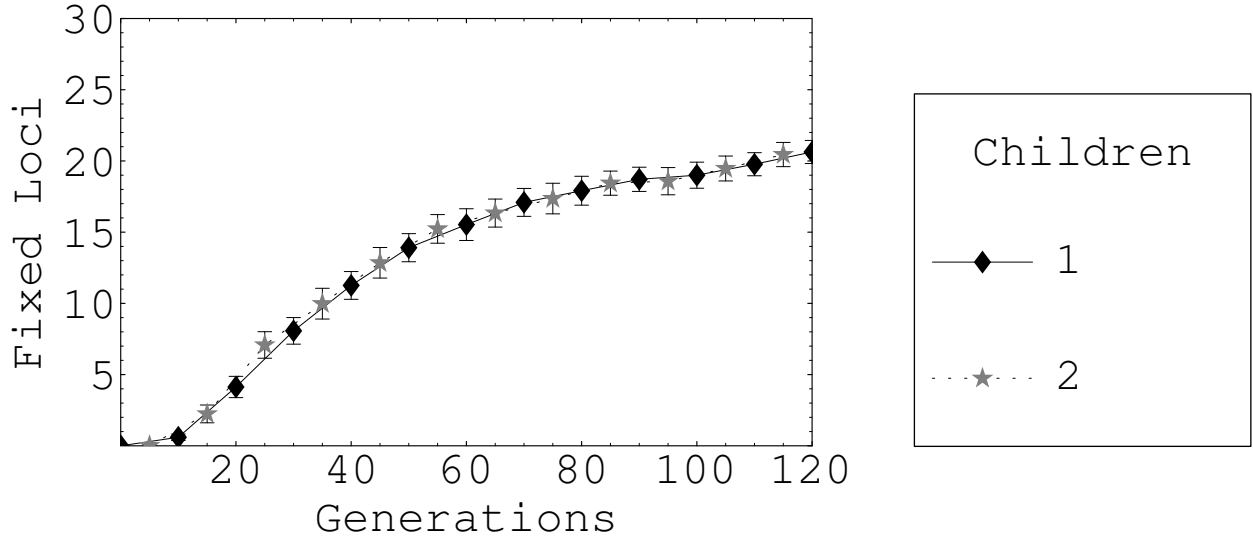


Figure 2: Number of fixed loci (lost alleles) for two variants of the simple GA on $F1$, averaged over 100 runs. In the first variant, both children from each pair of parents are kept; in the second, one child is discarded at random. The mutation rate p_m is 0.001 per locus. Error bars extend three standard errors in each direction. Data points are sampled every 10 generations for readability.

portion of $np(x, t)$. This section presents theoretical and empirical results for remainder-stochastic selection.

3.1 Remainder-Stochastic Selection: Theoretical Results

3.1.1 Case 6: Remainder-Stochastic Selection, Two Parents, Two Children

Brindle and Sampson [2] showed that the expected number of times a parent x was chosen by remainder-stochastic selection was

$$E(M(x, t + 1)) = np(x, t), \quad (17)$$

and the variance is

$$V(M(x, t + 1)) = r(x, t) \left(1 - \frac{r(x, t)}{\sum_y r(y, t)} \right). \quad (18)$$

This is equal to the variance in the number of times x is chosen in the stochastic portion of the selection algorithm, since the deterministic portion does not contribute any variance.

3.1.2 Case 7: Remainder-Stochastic Selection, Two Parents, One Child

What happens if one child is discarded under remainder-stochastic selection? There are two plausible ways of implementing this. One is to construct the deterministic table and roulette wheel as if the population size were $2n$. The other is to set up the algorithm for a population size of n , but then run the deterministic and stochastic portions of the algorithm twice each. The expected

number of offspring is the same in both cases, but the variance is easier to calculate for the second implementation, so this is what will be derived here.

$$\begin{aligned}
E(m(x, t + 1)) &= E(w(x, t + 1)) + E(g(x, t + 1)) \\
&= \frac{2\lfloor np(x, t) \rfloor}{2} + \frac{2r(x, t) \sum_y r(y, t)}{2 \sum_y r(y, t)} \\
&= \lfloor np(x, t) \rfloor + r(x, t) \\
&= np(x, t)
\end{aligned} \tag{19}$$

and

$$\begin{aligned}
V(m(x, t + 1)) &= V(w(x, t + 1)) + V(g(x, t + 1)) \\
&= \frac{2\lfloor np(x, t) \rfloor}{2} \left(1 - \frac{1}{2}\right) + \frac{2r(x, t) \sum_y r(y, t)}{2 \sum_y r(y, t)} \left(1 - \frac{r(x, t)}{2 \sum_y r(y, t)}\right) \\
&= \frac{\lfloor np(x, t) \rfloor}{2} + r(x, t) \left(1 - \frac{r(x, t)}{2 \sum_y r(y, t)}\right).
\end{aligned} \tag{20}$$

Here $w(x, t + 1)$ is the number of times a parent is chosen by the deterministic portion of the algorithm and $g(x, t + 1)$ is the number of times it is chosen in the stochastic portion. Note that the result from Equation 20 is greater than or equal to that from Equation 18, since

$$r(x, t) \left(1 - \frac{r(x, t)}{2 \sum_y r(y, t)}\right) \geq r(x, t) \left(1 - \frac{r(x, t)}{\sum_y r(y, t)}\right). \tag{21}$$

Hence, discarding one child under remainder-stochastic selection increases the variation in individual frequencies, and thus schema frequency variance. To confirm this, a set of experiments were run to measure allele loss under remainder-stochastic selection.

3.2 Remainder-Stochastic Selection: Empirical Results on $F1$

In order to replicate Booker's [1] original results, another experiment on $F1$ was performed, just as in Section 2.2.2 but using remainder-stochastic selection [2] instead of fitness-proportionate selection. As before, elitism and scaling were not used, in contrast to Booker's experiment. Three variants were compared. In the first two variants, one child is discarded at random. In the first, the remainder-stochastic selection algorithm is set up for selecting $2n$ parents; in the second, it is set up as if n parents were selected and then used two times. In the third variant, both children from each pair of parents are kept, and the selection algorithm is set up for a population size of n . Once again, the number of fixed loci was recorded for each generation, and the results are plotted in Figure 3. In contrast to the results for fitness-proportionate selection, discarding one child has a large effect on the rate of allele loss. The two-sided normal p -value that that the difference between either of the one-child variants and the two-child variant at generation 120 is at least as great as observed, assuming H_0 is true, is 0 (to 16 decimal places). Hence, the difference is highly statistically significant. This qualitatively replicates Booker's result. The actual numbers of alleles lost for each set of runs do not match his results; this is presumably due to the scaling and elitism that he used. The two-sided normal p -value that that the difference between the one-child variants at generation 120 is at least as great as observed, assuming H_0 is true, is 0.0239296. Under the

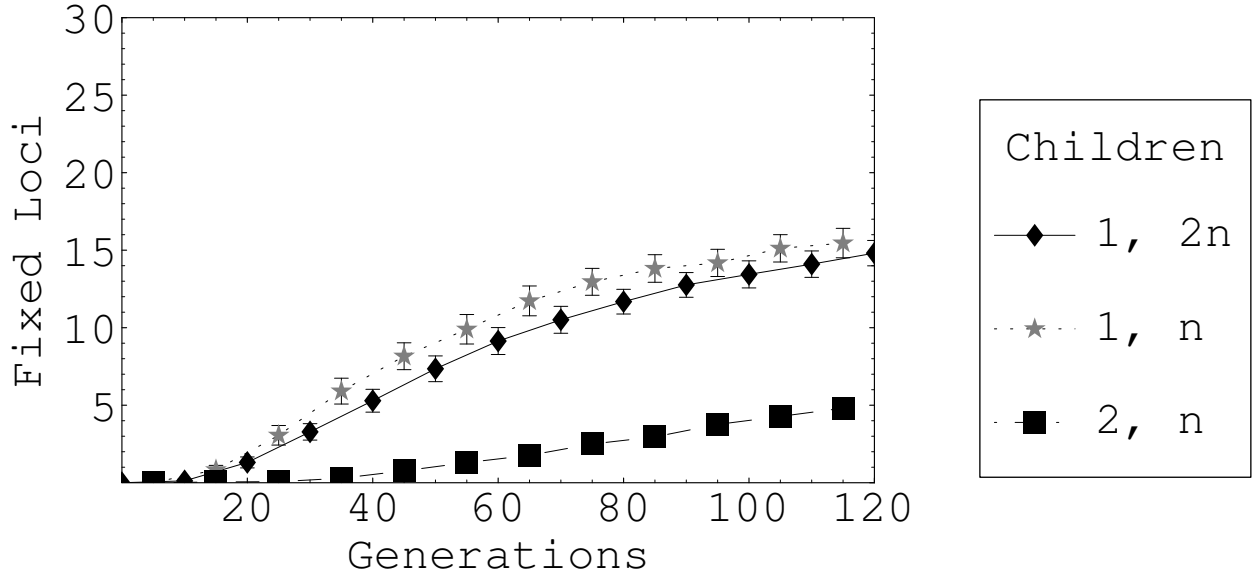


Figure 3: Number of fixed loci (lost alleles) for three variants of the simple GA on $F1$, using remainder-stochastic selection, averaged over 100 runs. In the first two variants, one child is discarded at random. In the first, the remainder-stochastic selection algorithm is set up for selecting $2n$ parents; in the second, it is set up as if n parents were selected and then used two times. In the third variant, both children from each pair of parents are kept, and the selection algorithm is set up for a population size of n . The mutation rate p_m is 0.001 per locus. Error bars extend three standard errors in each direction. Data points are sampled every 10 generations for readability.

criteria used in this paper, this difference is not significantly significant. but it seems likely that there is in fact a small difference between those two variants. Thus, these experimental results confirm the theoretical results from Section 3.1.2: Discarding offspring increases allele loss and schema variance if remainder-stochastic selection is used.

4 Conclusion

The theoretical results in Section 2.1 show that under fitness-proportionate selection, the number of parents and the number of offspring kept have no effect on the mean or variance of a schema's frequency. Even though a schema is only half as likely to survive if one child is discarded, it has twice as many chances to be sampled, which cancels out this apparent disadvantage. The presence or absence of crossover also has no effect on the mean or variance of the frequency of individual alleles. (Crossover does have an effect on the mean frequency of higher-order schemata [5, 8]. Presumably it has an effect on the variance as well, though this has not been calculated in this paper.) Section 2.2 confirmed empirically that discarding one child does not increase allele loss under fitness-proportionate selection.

However, Section 3.1 demonstrated that discarding one child increases the variance of a schema's frequency, and Section 3.2 confirmed empirically that this increases allele loss, replicating Booker's [1] results. This is because doing so makes the deterministic portion of the selection algorithm stochas-

tic, thus increasing the variance.

Future extensions include extending the theoretical results from alleles to arbitrary schemata, as well as calculating the variance when a child is discarded under the first variant of the remainder-stochastic algorithm.

Acknowledgments

I am grateful to Lashon Booker for his generous and patient answers to my questions. His suggestion that remainder-stochastic selection might produce different variance and allele loss rates from fitness-proportionate selection was particularly helpful. I would also like to thank the members of the University of Michigan Royal Road Group for their comments and support.

References

- [1] Lashon B. Booker. *Intelligent Behavior as an Adaptation to the Task Environment*. PhD thesis, University of Michigan, Ann Arbor, 1982. Appendix A.
- [2] Anne Brindle and Jeffrey Sampson. Analysis of frequency error in three sampling algorithms. Unpublished paper, Department of Computing Science, University of Alberta, Edmonton, 1979.
- [3] Kenneth A. De Jong. *An Analysis of the Behavior of a Class of Genetic Adaptive Systems*. PhD thesis, University of Michigan, Ann Arbor, 1975.
- [4] David E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA, USA, 1989.
- [5] John H. Holland. *Adaptation in Natural and Artificial Systems*. MIT Press, Cambridge, MA, USA, second edition, 1992. First edition: University of Michigan Press, Ann Arbor, 1975.
- [6] Melanie Mitchell. *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, MA, USA, 1996.
- [7] Riccardo Poli, W. B. Langdon, and Una-May O'Reilly. Analysis of schema variance and short term extinction likelihoods. In John R. Koza, Wolfgang Banzhaf, Kumar Chellapilla, Kalyanmoy Deb, Marco Dorigo, David B. Fogel, Max H. Garzon, David E. Goldberg, Hitoshi Iba, and Rick Riolo, editors, *Genetic Programming 1998: Proceedings of the Third Annual Conference*, pages 284–292. Morgan Kaufmann, 1998. University of Wisconsin, Madison, 22–25 July 1998.
- [8] Chris Stephens and Henri Waelbroeck. Schemata evolution and building blocks. *Evolutionary Computation*, 7(2):109–124, 1999.