

Genetic Drift in Genetic Algorithm Selection Schemes

Alex Rogers and Adam Prügel-Bennett

Abstract—A method for calculating genetic drift in terms of changing population fitness variance is presented. The method allows for an easy comparison of different selection schemes and exact analytical results are derived for traditional generational selection, steady-state selection with varying generation gap, a simple model of Eshelman's CHC algorithm, and $(\mu + \lambda)$ evolution strategies. The effects of changing genetic drift on the convergence of a GA are demonstrated empirically.

Index Terms—Evolution strategy, genetic algorithm, genetic drift, selection operator.

I. INTRODUCTION

GENETIC drift is a term borrowed from population genetics where it is used to explain changes in gene frequency through random sampling of the population. It is a phenomenon observed in genetic algorithms (GA) due to the stochastic nature of the selection operator, and is one of the mechanisms by which the population converges to a single member. Analysis of genetic drift is often performed by calculating the Markov chain transition matrices and hence finding the time for the system to reach an absorption state where all population members are identical. Comparisons in the genetic algorithm literature are often performed numerically in this fashion [1], [2]. In population genetics some work has been done to solve this analytically [3]–[5]; however, the results are approximations and are difficult to generalize to other cases.

Analysis of selection schemes such as those by Prügel-Bennett and Shapiro [6]–[8], Rattray [9], and Mühlenbein [10] show that the change in mean fitness at each generation is a function of the population fitness variance. At each generation this variance is reduced due to two factors. One factor is selection pressure producing multiple copies of fitter population members while the other factor is independent of population member fitness and is due to the stochastic nature of the selection operator—genetic drift. The loss in population fitness variance due to genetic drift thus has a direct effect on the performance of the genetic algorithm. By considering neutral selection we decouple the effect of selection pressure and can see the effect of genetic drift directly.

This paper presents a method of calculating the rate of genetic drift in terms of this change in population fitness variance. Unlike calculations in terms of convergence time, this approach lends itself to an exact analytical solution. We are able to derive a general expression for the change in

population fitness variance due to genetic drift and apply it to the range of selection schemes used in evolutionary algorithms. We first consider the generational GA and then compare it to steady-state selection where one member is drawn from the population, replicated, and replaces another population member chosen at random.

To generalize between these two extremes, De Jong [1], [11] introduced the term *generation gap* G which describes the percentage of the population selected from the initial population at each time step. For generational selection $G = 1$ and for steady-state selection $G = 1/P$. We follow this generalization and calculate the change in variance for any value of generation gap.

The formalism can also be extended to other nontraditional selection schemes such as that used in Eshelman's CHC algorithm [12]. Here we confirm analytically an observation made by Schaffer *et al.* [2] that shows using a numerical Markov chain analysis that a simple model of CHC style selection exhibits half the rate of genetic drift of the traditional genetic algorithm. The simple model of the CHC algorithm is equivalent to selection schemes in evolution strategies and we can generalize the approach for these selection schemes.

In Section II we derive the result which enables us to calculate the rate of genetic drift. In Section III we present the analytical results for different selection schemes and compare them to simulation results. Section IV contains details of the calculations and in Section V we discuss the results and their implications on the performance of genetic algorithms.

II. POPULATION FITNESS VARIANCE

If we consider an initial population of P discrete members each with fitness F_α , the variance (κ_2) of the population fitness distribution is simply given by

$$\begin{aligned}\kappa_2 &= E[F^2] - E[F]^2 \\ &= \frac{1}{P} \sum_{\alpha=1}^P F_\alpha^2 - \left(\frac{1}{P} \sum_{\alpha=1}^P F_\alpha \right)^2.\end{aligned}\quad (1)$$

We can separate out terms that are not independent to give

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

We now apply some selection scheme to this population and draw from it a new population of P individuals. In this new population there are now n_α copies of population member F_α and the variance of the new population fitness distribution is

Manuscript received November 30, 1998; revised March 24, 1999. The work of A. Rogers was supported by an award from the EPSRC—Engineering and Physical Science Research Council.

The authors are with the ISIS Research Group, Department of Electronics and Computer Science, University of Southampton, Highfield, Southampton, SO17 1BJ, U.K. (e-mail: A.Rogers@ecs.soton.ac.uk).

Publisher Item Identifier S 1089-778X(99)08072-8.

given by

$$\kappa'_2 = \frac{1}{P} \sum_{\alpha=1}^P n_{\alpha} F_{\alpha}^2 - \left(\frac{1}{P} \sum_{\alpha=1}^P n_{\alpha} F_{\alpha} \right)^2. \quad (3)$$

Again we can separate out terms that are not independent

$$\kappa'_2 = \sum_{\alpha=1}^P \left(\frac{n_{\alpha}}{P} - \frac{n_{\alpha}^2}{P^2} \right) F_{\alpha}^2 - \sum_{\alpha \neq \beta} \frac{n_{\alpha} n_{\beta}}{P^2} F_{\alpha} F_{\beta}. \quad (4)$$

To consider the average case, we average over all ways of performing selection. In the case of neutral selection, n_{α} is independent of F_{α} and these terms may be taken outside the summation and the expected population fitness variance considered

$$E[\kappa'_2] = \left(\frac{E[n]}{P} - \frac{E[n^2]}{P^2} \right) \sum_{\alpha=1}^P F_{\alpha}^2 - \frac{E[n_{\alpha} n_{\beta}]}{P^2} \sum_{\alpha \neq \beta} F_{\alpha} F_{\beta}. \quad (5)$$

To simplify this result further, we use the fact that population size is kept constant and thus $E[n] = 1$. We can use this to derive the identity

$$\left(\sum_{\alpha=1}^P n_{\alpha} \right)^2 = P^2 = \sum_{\alpha=1}^P n_{\alpha}^2 + \sum_{\alpha \neq \beta} n_{\alpha} n_{\beta}. \quad (6)$$

Averaging over all possible selections gives

$$P^2 = P E[n^2] + P(P-1) E[n_{\alpha} n_{\beta}] \quad (7)$$

and thus

$$E[n_{\alpha} n_{\beta}] = \frac{P - E[n^2]}{P-1}. \quad (8)$$

Substituting this expression into (5) gives

$$E[\kappa'_2] = \frac{P - E[n^2]}{P-1} \left[\left(\frac{1}{P} - \frac{1}{P^2} \right) \sum_{\alpha=1}^P F_{\alpha}^2 - \frac{1}{P^2} \sum_{\alpha \neq \beta} F_{\alpha} F_{\beta} \right]. \quad (9)$$

The term within the square brackets is simply the fitness variance of the initial population given in (2) and thus

$$E[\kappa'_2] = \frac{P - E[n^2]}{P-1} \kappa_2. \quad (10)$$

We can find the change in population fitness variance for any selection scheme simply by calculating $E[n^2]$ —the expected square of the number of times any population member is selected. This is related to the variance in the number of times any member is selected— $V[n]$. As $V[n] = E[n^2] - E[n]^2$, we can rewrite (10) in these terms

$$E[\kappa'_2] = \left(1 - \frac{V[n]}{P-1} \right) \kappa_2. \quad (11)$$

This expression is the basis for the results derived in this paper. It describes the change in population fitness variance due to selection, genetic drift, in terms of the variance in the number of times any individual is selected.

III. RESULTS

The change in the population fitness variance due to selection, genetic drift, is dependent only on the variance of the number of times any individual population member is selected— $V[n]$. If we select each population member once and only once then $V[n] = 0$ and our expression in (11) is equal to one. As expected, we see no change in population variance—indeed, the population has not changed.

To compare each selection scheme we need only calculate $V[n]$. To allow direct comparison between traditional generational selection we normalize the results to one generation—we apply steady-state selection P times and selection with generation gap G , $1/G$ times. We define the ratio r as the change in variance after one generation

$$r = \frac{E[\kappa'_2]}{\kappa_2}. \quad (12)$$

This gives a very simple picture of the change in genetic drift for differing selection schemes. We present the calculations in more detail in the next section but give the results here. While the first expression for generational selection is exact, the other expressions are approximations that are accurate to terms in $1/P$.

$$\text{Generational: } r = 1 - \frac{1}{P}$$

$$\text{Steady State: } r \approx 1 - \frac{2}{P}$$

$$\text{Generation Gap } G: r \approx 1 - \frac{2-G}{P}$$

$$\text{CHC Algorithm: } r \approx 1 - \frac{1}{2P}.$$

The rate of genetic drift in generational selection is well known as the result of sampling P times with replacement from a finite population.

The rate of genetic drift in steady-state selection is twice that of generational selection. This result has previously been shown by the authors [13] in an analysis of steady-state selection using Boltzmann selection. Varying the generation gap produces a smooth progression between these two limits. The simple model of the CHC algorithm shows half the genetic drift of the generational selection scheme. This is in agreement with the empirical observation and numerical Markov chain comparison by Schaffer *et al.* [2].

Fig. 1 shows a comparison of these analytical results with simulation data. A population of 100 was initially drawn from a normal distribution ($\kappa_2 = 1$) and selection repeatedly performed. The plot shows the decreasing population fitness variance for five different selection schemes—steady-state selection (SSGA), generation gap $G = 0.2$, generation gap $G = 0.5$, traditional generational selection (GA), and CHC-style selection (CHC). Simulation data were averaged over 10000 runs.

IV. PERFORMING THE CALCULATIONS

To calculate $V[n]$ for each selection scheme is an exercise in probability. We use two results from standard probability

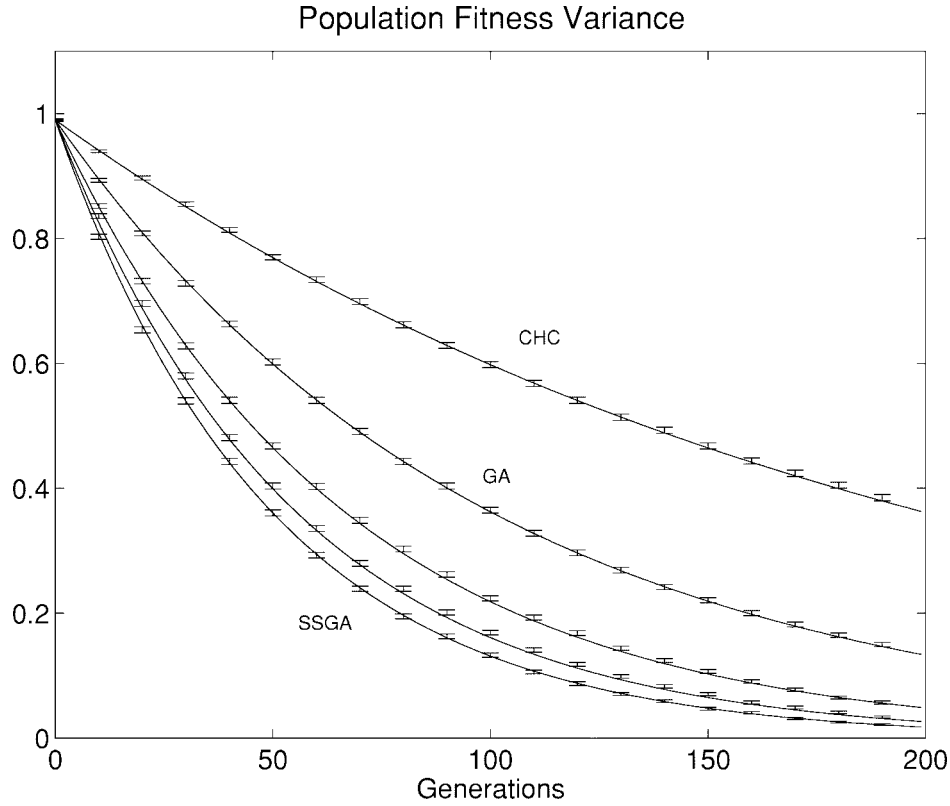


Fig. 1. Population fitness variance for five different selection schemes. Solid lines are analytical results and error bars are simulation results averaged over 10000 runs. Curves presented are steady-state (SSGA), generation gap $G = 0.2$, generation gap $G = 0.5$, generational (GA), and a simple model of the CHC algorithm (CHC). Population size is 100.

theory regarding binomial and hypergeometric distributions [14].

Selecting from a population with replacement gives rise to a binomial distribution $B(N, p)$ where we select N times with probability of success p . In this case, the expected number of times any individual is selected and its variance are given by

$$E[n] = Np \quad V[n] = Np(1 - p).$$

When we are selecting without replacement, the result is a hypergeometric distribution $H(M, m, N)$. Here M is the size of the population, N is the number of times we select, and m is the number of copies of each individual in the initial population. This gives the known result

$$E[n] = \frac{Nm}{M} \quad V[n] = \frac{Nm(M - N)(M - m)}{M^3 - M^2}.$$

In each case, we calculate $E[n]$ to check that population size is conserved, as expected, and then use $V[n]$ in (11) to give the expected change in population fitness variance and thus the rate of genetic drift.

A. Generational Selection

In a generational selection scheme under random sampling, we are drawing P members from a population with replacement. This gives rise to a binomial distribution, $B(P, 1/P)$ and thus

$$E[n] = 1 \\ V[n] = 1 - 1/P.$$

As required, $E[n] = 1$ and we can thus substitute $V[n]$ directly into (11) to give

$$E[\kappa'_2] = \left(1 - \frac{1}{P}\right)\kappa_2. \quad (13)$$

Using the definition of r in (12)

$$r = 1 - \frac{1}{P}. \quad (14)$$

B. Steady-State Selection

In the steady-state genetic algorithm we select one member at random, replicate it, and replace another random member with the copy in each time step.

We can calculate this by dividing the population into two. We draw one member with replacement into subpopulation A and then draw $P - 1$ members without replacement into subpopulation B . We then combine these two to form the next population. For subpopulation A we have a binomial distribution $B(1, 1/P)$ and hence

$$E[n_A] = 1/P \\ V[n_A] = (P - 1)/P^2.$$

For subpopulation B we have a hypergeometric distribution $H(P, 1, P - 1)$ and hence

$$E[n_B] = 1 - 1/P \\ V[n_B] = (P - 1)/P^2.$$

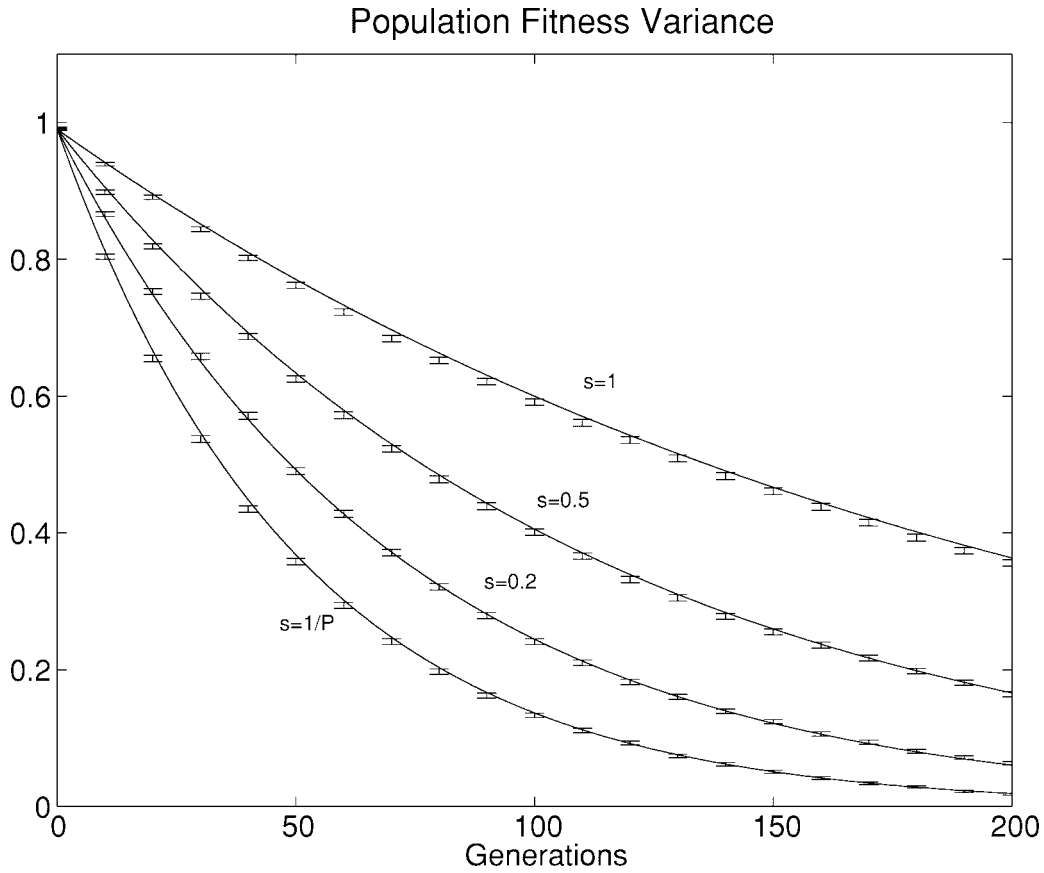


Fig. 2. Population fitness variance for $(\mu + s\mu)$ selection for varying s . Solid lines are analytical results and error bars are simulation results averaged over 10000 runs. Population size is 100.

Since the two populations are independent, we can simply sum for the final population

$$\begin{aligned} E[n] &= E[n_A] + E[n_B] = 1 \\ V[n] &= V[n_A] + V[n_B] = 2(P-1)/P^2. \end{aligned}$$

As required, $E[n] = 1$ and we can thus substitute $V[n]$ directly into (11) to give

$$E[\kappa'_2] = \left(1 - \frac{2}{P^2}\right) \kappa_2. \quad (15)$$

It is often more convenient to compare P of these selections to one generational selection so using the definition of r as the change after one generation

$$r = \left(1 - \frac{2}{P^2}\right)^P \approx 1 - \frac{2}{P}. \quad (16)$$

It is clear that the rate of genetic drift is twice that of the generational case.

C. Varying Generation Gap

To generalize between these two cases we use the concept of generation gap (G). We select GP members with replacement from the original population and delete GP members at random to make room.

Again we can consider two subpopulations. We draw GP members with replacement from the original population into

subpopulation A and then draw $P(1-G)$ members without replacement into subpopulation B .

For subpopulation A we have a binomial distribution $B(GP, 1/P)$ and hence

$$\begin{aligned} E[n_A] &= G \\ V[n_A] &= G(1-1/P). \end{aligned}$$

For subpopulation B we have a hypergeometric distribution $H(P, 1, P-GP)$ and hence

$$\begin{aligned} E[n_B] &= 1-G \\ V[n_B] &= G-G^2. \end{aligned}$$

Again, we simply sum these for the final population

$$\begin{aligned} E[n] &= 1 \\ V[n] &= 2G - G^2 - G/P. \end{aligned}$$

As required, $E[n] = 1$ and we can thus substitute $V[n]$ directly into (11) to give

$$E[\kappa'_2] = \left(1 - \frac{2G - G^2 - G/P}{P-1}\right) \kappa_2. \quad (17)$$

To compare this to one generation we apply the selection operator $1/G$ times. Thus approximating to first-order terms in $1/P$ we get

$$r = \left(1 - \frac{2G - G^2 - G/P}{P-1}\right)^{1/G} \approx 1 - \frac{2-G}{P}. \quad (18)$$

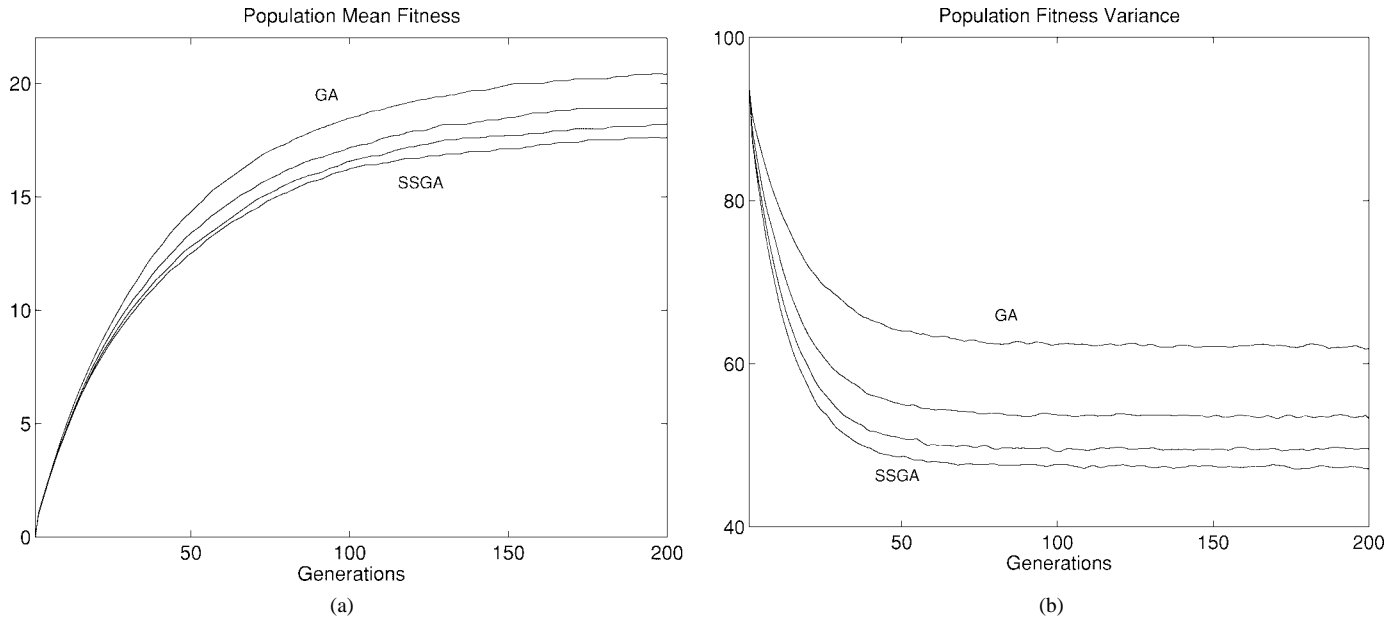


Fig. 3. Population mean fitness and variance for four different selection schemes. Simulation results are averaged over 10 000 runs and the error bars are the thickness of the lines. Curves presented are (in order) steady-state (SSGA), generation gap $G = 0.2$, generation gap $G = 0.5$, and generational (GA).

Thus there is a gradual transition between the two rates of genetic drift as generation gap changes.

D. CHC Algorithm and Evolution Strategies

Eshelman's CHC algorithm uses another nontraditional form of selection whereby crossover is performed among the initial population and then selection is performed without replacement from the combined population of parents and offspring.

A simple model of this used by Schaffer *et al.* [2] in a numerical genetic drift comparison is to duplicate each member of the population and then draw P members from the population of $2P$ without replacement. In terms of evolution strategies this is $(\mu + \lambda)$ selection with $\lambda = \mu$.

This selection gives rise to a hypergeometric distribution $H(2P, 2, P)$ where we select P times from an initial population of $2P$ which consists of two copies of each individual

$$\begin{aligned} E[n] &= 1 \\ V[n] &= (P-1)/(2P-1). \end{aligned}$$

As required $E[n] = 1$ and we can thus substitute $V[n]$ directly into (11) to give

$$E[\kappa'_2] = \left(1 - \frac{1}{2P-1}\right) \kappa_2. \quad (19)$$

As we draw P members from the population, we can compare this directly to the generational case and simply make a first-order approximation,

$$r \approx 1 - \frac{1}{2P}. \quad (20)$$

Thus genetic drift in this model of CHC selection is at half the rate of that of the traditional generational algorithm.

While we have only considered the case here equivalent to CHC selection, the technique presented is immediately

applicable to other evolution strategy selection schemes. When λ is a whole number multiple of μ , the above approach gives the correct expression. However the more common and more interesting case where λ is some fraction of μ is more complicated due to the need to average over the population.

We consider a $(\mu + \lambda)$ evolution strategy where $\mu = P$ and $\lambda = sP$ where s is some fraction, $0 \leq s \leq 1$. When we apply selection, we are selecting from two subpopulations, one consisting of $P(1-s)$ individuals and the other of size $2sP$ containing sP pairs. If n_1 is the number of individuals and n_2 the number of pairs in the final population, the variance in the number of times any population member is selected can be shown to be simply

$$V[n] = \frac{2n_2}{P} \quad (21)$$

as $PE[n] = n_1 + 2n_2$, $PE[n^2] = n_1 + 4n_2$, $E[n] = 1$, and $V[n] = E[n^2] - E[n]^2$. If we draw X times without replacement from the subpopulation of pairs, the number of pairs drawn and thus the number of pairs in the final population is given by

$$n_2 = \frac{X^2 - X}{2(2sP - 1)}. \quad (22)$$

Substituting (22) into (21) and averaging over X gives

$$V[n] = \frac{E[X^2] - E[X]^2}{P(2sP - 1)}. \quad (23)$$

The expectations of X are described by the hypergeometric distribution $H(P(1+s), 2sP, P)$, as we are drawing P times without replacement from a population of $P(1+s)$. Using $V[X] = E[X^2] - E[X]^2$ and the standard results for the hypergeometric distribution given earlier, gives the result

$$V[n] = \frac{2s(P-1)}{(1+s)[P(1+s)-1]}. \quad (24)$$

As before, we can substitute $V[n]$ directly into (11) and normalize the expression by applying the selection $1/s$ times to give the final rate of genetic drift

$$r = \left(1 - \frac{2s}{(1+s)[P(1+s)-1]} \right)^{1/s} \approx 1 - \frac{2}{(1+s)^2 P}. \quad (25)$$

The rate of genetic drift covers the same range as that seen for the genetic algorithm selection schemes. Fig. 2 shows a plot of these analytical result against simulation data. Four different values of s are considered and the population size is again 100.

V. DISCUSSION

Analyzing genetic drift in terms of the change in population fitness variance allows exact analytical expressions to be derived for any selection scheme. From these expressions we can make some comparisons of the effect that genetic drift has on the convergence of a GA under varying generation gap. If we consider a GA using a small population and weak selection, these effects will be most pronounced.

Fig. 3 shows the population fitness mean and variance for steady-state, generational, and varying generation gap ($G = 0.2$ and 0.5) implementations of GA on the ONEMAX problem. All use a population size of 50 with probabilistic tournament selection ($s = 0.1$), string length 96, point mutation rate $1/96$, and uniform crossover. CHC is not included in the comparison as the other features of the algorithm lead to more significant differences than genetic drift alone.

Selection pressure is the same in each case as evidenced by the identical initial gradients of the mean fitness curves. As variance decreases through selection, the change in mean fitness decreases. For the steady-state GA, variance decreases fastest due to the higher rate of genetic drift and thus the mean fitness evolves to a lower final value.

These results illustrate how genetic drift can influence the convergence of a GA. It is not always detrimental, however. In another paper analyzing steady-state Boltzmann selection [13], the authors show that in the weak selection limit, rescaling the parameters of a steady-state GA enables it to reproduce the dynamics of a generational GA but at half the computational cost. Definitive statements about the performance of different

selection schemes are difficult to make. However, it is clear that genetic drift is another factor, alongside more commonly understood factors such as selection pressure, which affects the convergence of the GA and can be controlled by the choice of selection scheme.

ACKNOWLEDGMENT

The authors wish to thank the reviewers for highlighting the use of the hypergeometric probability distribution and suggestions on notation. Both contributions improve the clarity of the calculations.

REFERENCES

- [1] K. A. DeJong, "An analysis of the behavior of a class of genetic adaptive systems," Ph.D. dissertation, Univ. Michigan, Ann Arbor, 1975.
- [2] J. Schaffer, M. Mani, L. Eshelman, and K. Mathias, "The effect of incest prevention on genetic drift," in *Foundations of Genetic Algorithms 5*, W. Banzhaf and C. Reeves, Eds. San Francisco, CA: Morgan Kaufmann, 1998, in press.
- [3] P. A. P. Moran, "Random processes in genetics," *Proc. Cambridge Philos. Soc.*, vol. 54, pp. 60–71, 1958.
- [4] M. Kimura, "Diffusion models in population genetics," *Applied Probability* (Methuen's Review Series in Applied Probability), 1964.
- [5] W. J. Ewens, *Mathematical Population Genetics*. Berlin, Germany: Springer-Verlag, 1979.
- [6] A. Prügel-Bennett and J. L. Shapiro, "An analysis of genetic algorithms using statistical mechanics," *Phys. Rev. Lett.*, vol. 72, no. 9, pp. 1305–1309, 1994.
- [7] ———, "The dynamics of a genetic algorithm for simple random Ising systems," *Phys. D*, vol. 104, pp. 75–114, 1997.
- [8] A. Prügel-Bennett, "Modeling evolving populations," *J. Theor. Biol.*, vol. 185, pp. 81–95, 1997.
- [9] M. Rattray, "Modeling the dynamics of genetic algorithms using statistical mechanics," Ph.D. dissertation, Manchester Univ., Manchester, U.K., 1996.
- [10] H. Mühlenbein, "Genetic algorithms," in *Local Search in Combinatorial Optimization*, E. Aarts and J. K. Lenstra, Eds. New York: Wiley, 1997, pp. 137–171.
- [11] K. A. DeJong and J. Sarma, "Generation gaps revisited," in *Foundations of Genetic Algorithms 2*, L. D. Whitley, Ed. San Mateo, CA: Morgan Kaufmann, 1993, pp. 19–28.
- [12] L. Eshelman, "The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination," in *Foundations of Genetic Algorithms*, G. Rawlins, Ed. San Mateo, CA: Morgan Kaufmann, 1991, pp. 265–283.
- [13] A. Rogers and A. Prügel-Bennett, "Modeling the dynamics of steady-state genetic algorithms," in *Foundations of Genetic Algorithms 5*, W. Banzhaf and C. Reeves, Eds. San Francisco, CA: Morgan Kaufmann, 1998, in press.
- [14] P. A. P. Moran, *An Introduction to Probability Theory*. Oxford, U.K.: Clarendon, 1968.