

## Effects of Population Size on Performance of Genetic Algorithms and Roles of Crossover and Mutation

Yu-an Zhang

Makoto Sakamoto

Hiroshi Furutani

Faculty of Engineering

University of Miyazaki

Miyazaki City 889-2192

### Abstract

One of the most important parameters in the application of GAs is the population size  $N$ . In many cases, the choice of  $N$  determines the quality of obtained solutions. The study of GAs with finite population size requires the stochastic treatment of evolution. In this study, we examined effects of genetic fluctuations on the performance of GA calculations. We considered the roles of crossover and mutation by using the stochastic schema theory within the framework of the Wright-Fisher model of Markov processes. The failure probability of obtaining the optimum solution was investigated experimentally and theoretically.

### 1 Introduction

This study treats the influence of finite population size on the performance of genetic algorithms (GAs). We focused on the effect of genetic drift due to the random sampling of individuals in selection. When we apply a GA to a given problem, we choose the population size  $N$  intuitively or by experience. This is because we do not have any reliable theory of choosing  $N$ . If we choose a small  $N$  to cut the cost of calculations, there appear several problems caused by genetic drift. The main part of the effect of genetic drift may disappear by averaging repeated trials. However there are several cases in which its effect remains finite even after averaging. An example of such cases is a GA on the multiplicative landscape with weak selection [1, 2]. When one uses a small  $N$ , there is a high risk of poor performance caused by the undesirable effect of genetic drift. The stochastic analysis of GAs using finite  $N$  is far more complicated than the deterministic ones assuming infinitely large  $N$ . One of the most fundamental models is the Markov chain theory which includes selection, mutation and crossover. However, it is in general difficult to obtain an analytical expression of Markov process in closed form. Furthermore

the dimension of the transition matrix increases exponentially with string length  $\ell$  and population size  $N$ . This fact means that numerical simulations are impossible when we use realistic values of  $\ell$  and  $N$ . In population genetics, researchers also encountered this type of difficulties in treating the evolution of a finite population by Markov chain model [3]. They found another approach by the application of diffusion equations. Fisher treated the simple case of no selection by the heat diffusion equation. We applied stochastic approaches to the present problem, and studied the evolution of first order schemata in the GA on the multiplicative landscape with finite  $N$ . In this report, we considered the evolution of the GA on the multiplicative landscape to study the influence of genetic drift on the performance of calculations. This study includes the crossover and mutation, and effects of these operators are analyzed by comparing the theory and numerical experiments.

### 2 Models and Methods

#### 2.1 Representations

In this study, individuals in a population are represented by binary strings of the fixed length  $\ell$ . There are  $n = 2^\ell$  genotypes, and the  $i$ th genotype is given by the binary representation of the integer  $i$ ,

$$i = \langle i(\ell), \dots, i(1) \rangle \quad (0 \leq i \leq n - 1),$$

where  $i(k)$  is 0 or 1.

The population size  $N$  is fixed throughout calculations, and

$$N = \sum_{i=0}^{n-1} N_i(t),$$

where  $N_i(t)$  is the frequency of the  $i$ th genotype at generation  $t$ . The relative frequency  $x_i(t)$  is defined

by

$$x_i(t) = N_i(t)/N.$$

If a population is in linkage equilibrium, the distribution of individuals depends only on the frequencies of the first order schemata [4]. Therefore, the relative frequency  $x_i$  is represented by

$$x_i = \prod_{k=1}^{\ell} h_{i(k)}, \quad (1)$$

where  $h_{i(k)}$  is the frequency of the first order schema corresponding to  $i(k)$ . This decomposition enables us to treat the population by using the schema theory of exact form [5].

## 2.2 Deterministic Model

We use the fitness proportionate selection, and the fitness  $f_i$  is the multiplicative function of each bit

$$f_i = \prod_{k=1}^{\ell} \{1 + s i(k)\}, \quad (2)$$

where  $s$  is a parameter of selection strength. Within the framework of the infinite population model of multiplicative landscape, the assumption of linkage equilibrium holds at all generations if the initial state is at linkage equilibrium [6]. Thus the evolution process of the first order schema  $i(k) = 1$  is given by the following difference equation

$$h_1(t+1) = \frac{(1+s)h_1(t)}{1+s h_1(t)}, \quad (3)$$

and the average fitness  $\bar{f}$  is also given by  $h_1$

$$\bar{f}(t) = \prod_{k=1}^{\ell} \{1 + s h_{1(k)}\}, \quad (4)$$

denoting  $i(k) = 1$  by  $1(k)$ .

## 2.3 Wright-Fisher Model

The Wright-Fisher model treats chromosomes having one locus and two alleles, corresponding to the GA of  $\ell = 1$  with genotypes  $i \in \{0, 1\}$ . The number of the first genotype 1 takes the values of  $N_1 = \{0, 1, \dots, N\}$ , and that of the genotype 0 is given by  $N_0 = N - N_1$ . The fitness values of genotypes are

$$f_{i(k)} = \begin{cases} 1 & (i(k) = 0) \\ 1 + s & (i(k) = 1). \end{cases} \quad (5)$$

We analyze selection processes by taking into account the effect of random sampling, and consider the process of choosing offspring randomly from the population in proportion with their fitness values. If there are  $N_1 = i$  copies of the genotype 1 at the current generation  $t$ , the probability  $P(j|i)$  of  $N_1$  taking the value of  $j$  at the next generation  $t+1$  is given by the binomial distribution

$$P(j|i) = \binom{N}{j} a^j (1-a)^{N-j}, \quad (6)$$

where  $a$  is given by substituting the right hand side of equation (3) with

$$h_1(t) \rightarrow \frac{i}{N}.$$

Thus we have

$$a = \frac{(1+s)i}{N+si}. \quad (7)$$

To take into account the effect of mutation with mutation rate  $p_m$ , we replace  $a$  by  $b$  in equation (6)

$$b = (1-2p_m)a + p_m \quad (8)$$

## 2.4 Diffusion Equations

The Wright-Fisher model can be approximated by diffusion equations if  $N$  is not too small. The diffusion equation derived from equation (6) is

$$\frac{\partial \psi(p, t)}{\partial t} = \frac{V(p)}{2} \frac{\partial^2 \psi(p, t)}{\partial p^2} + M(p) \frac{\partial \psi(p, t)}{\partial p}, \quad (9)$$

where  $p$  is the initial value of the relative frequency  $y = i/N$  at time  $t = 0$ , and

$$V(p) = \frac{p(1-p)}{N}, \quad M(p) = sp(1-p).$$

We define the ultimate fixation probability by

$$u(p) = \lim_{t \rightarrow \infty} \psi(p, t). \quad (10)$$

Using equation (9), we have

$$\frac{d^2 u(p)}{dp^2} + \frac{2M(p)}{V(p)} \frac{du(p)}{dp} = 0. \quad (11)$$

The boundary condition is

$$u(0) = 0, \quad u(1) = 1.$$

Since  $2M(p)/V(p) = 2Ns$ , we have

$$u(p) = \frac{1 - \exp(-2Nsp)}{1 - \exp(-2Ns)}. \quad (12)$$

We define the failure probability  $F$  that there is no optimum solution in the stationary state. This probability is approximately given by  $u(p)$

$$F = 1 - u(p)^\ell \quad (13)$$

### 3 Results

We performed numerical calculations of GA with roulette wheel selection on the multiplicative landscape. The results were compared with the deterministic and stochastic models of the first order schema. We compared results of

1. selection with  $s = 0.01$
2. selection with  $s = 0.4$ .

Crossover was done with the uniform crossover of crossover rate  $p_c$ . We used mutation rates  $p_m$  of 0 and 0.001 and compared their results. The string length was  $\ell = 8$ . The initial value of the first order schema was  $h_1 = 1/2$ . The calculations were performed repeatedly, and results were averaged over 1000 runs.

Figure 1 shows the failure probabilities  $F$  with the weak selection of  $s = 0.01$ . The solid line represents the prediction of diffusion model given by equation (13). We can observe the effect of crossover is very large, and crossover greatly improve the performance of GA calculations. We find that even the very small value of  $p_c = 0.01$  has noticeable effect of reducing  $F$ .

Figure 2 also shows the results of  $s = 0.01$  with mutation rates of  $p_m = 0.001$  and 0. We notice that mutation also works as the operator reducing  $F$ .

Figure 3 shows the failure probabilities  $F$  with the strong selection of  $s = 0.4$ . In this case, we have to use more large crossover rate  $p_c = 0.2$  to observe the effect of crossover. When we used the small value of  $p_c = 0.01$ , we cannot observe the effect of crossover.

Figure 4 is the results of  $s = 0.4$  with mutation rates of  $p_m = 0.001$  and 0. This figure also shows the effect of mutation in reducing  $F$ .

### 4 Summary

The study of GA on the multiplicative landscape demonstrated effects of crossover on the performance of GA calculations. Since the failure probability  $F$  has the strong  $N$  dependence, we have to choose carefully the optimum population size  $N$  in the case of weak selection  $s = 0.01$ . In the case of strong selection  $s = 0.4$ , there are notable effects of genetic drift and influence of  $N$  when  $N$  is very small. We also found that mutation has effects of reducing the failure probabilities. It was demonstrated that the diffusion approximation can explain the behavior of GA on the multiplicative landscape.

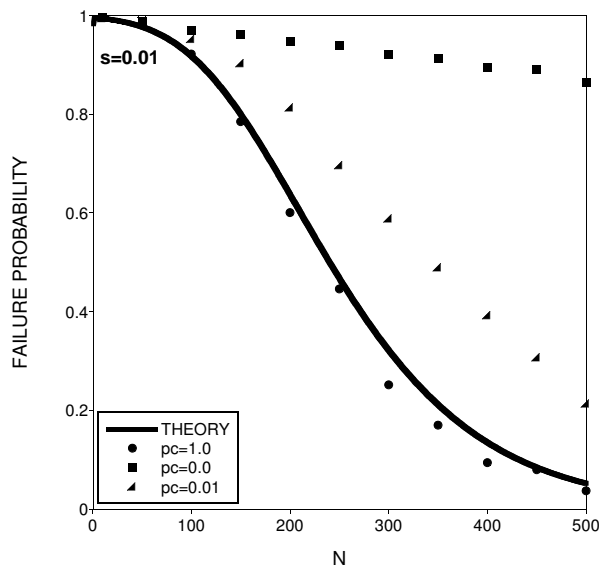


Figure 1: Failure probability  $F$ . With selection strength  $s = 0.01$ , and crossover rates  $p_c = 1.0, 0.01$ , and 0.0.

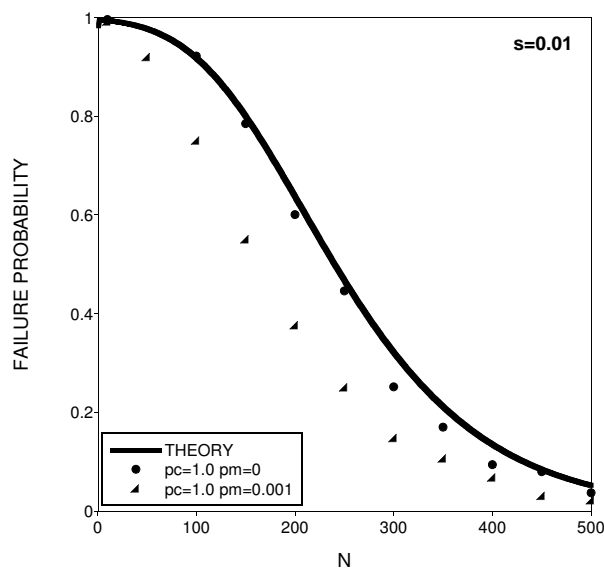


Figure 2: Failure probability  $F$  with  $s = 0.01$  and  $p_c = 1$ . Mutation rates are  $p_m = 0.001$  and 0.

## References

- [1] H. Furutani, S. Katayama, M. Sakamoto and M. Ito: Stochastic Analysis of Schema Distribution in a Multiplicative Landscape, *Artificial Life and Robotics*, **11**, (2007) 101–104
- [2] H. Furutani, T. Fujimaru, Y. Zhang and M. Sakamoto: Effects of Population Size on Computational Performance of Genetic Algorithm on Multiplicative Landscape, *Proceedings of Third International Conference on Natural Computation*, Vol.3, 488–493 (2007).
- [3] J. W. J. Ewens: *Mathematical Population Genetics. I. Theoretical Introduction*, Second Edition. Springer-Verlag, New York (2004)
- [4] H. Asoh and H. Mühlenbein: On the Mean Convergence Time of Evolutionary Algorithms without Selection and Mutation. *Parallel Problem Solving from Nature, Lecture Notes in Computer Science*, **866**, Springer-Verlag, New York, (1994) 88–97
- [5] H. Furutani: Schema Analysis of OneMax Problem, *Foundations of Genetic Algorithms 7*, Morgan Kaufmann, San Francisco, (2003), 9–26
- [6] H. Furutani: Schema Analysis of Genetic Algorithms on Multiplicative Landscape, *Proceedings of the Simulated Evolution and Learning Conference, SEAL'02*, (2002) 230–235

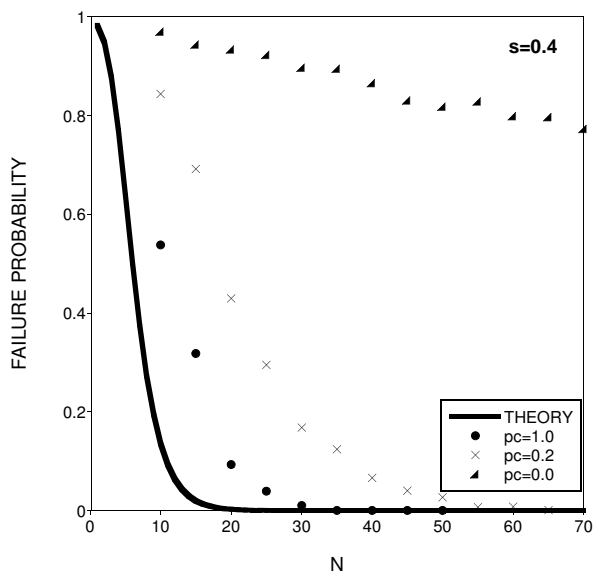


Figure 3: Failure probability  $F$  with  $s = 0.4$ . Crossover rates are  $p_c = 1, 0.2$  and  $0$ .

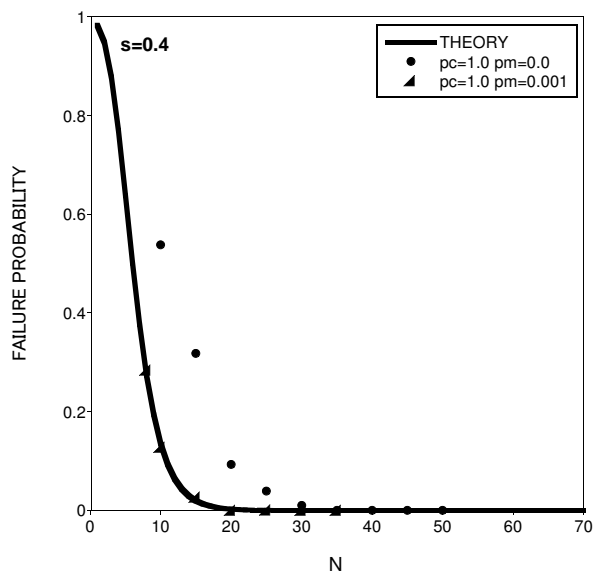


Figure 4: Failure probability  $F$  with  $s = 0.4$  and  $p_c = 1$ . Mutation rate are  $p_m = 0.001$  and  $0$ .