# A Method for Estimating Mean First-passage Time in Genetic Algorithms

**Hiroshi Furutani**
*Department of Information Science,*
*Kyoto University of Education,*
*Fushimi-ku, Kyoto, 612-8522 Japan*

This paper presents a simple method for estimating the mean time at which the best solution, or optimum allele, appears for the first time in the search process of genetic algorithms. A set of equations modified from Eigen's evolution equation is used for calculating the mean first-passage time. The essential point of the method is that the optimum allele is considered as an absorbing state. It is shown that the theory can generally reproduce numerical experiments on three types of fitness landscapes.

## 1. Introduction

One of the most important and serious problems in genetic algorithms (GAs) is how to predict the performance of GAs when a task is specified. For solving this problem, it is essential to understand the mechanism of search processes in GAs. Holland's schema theorem provides a good description of the dynamics of GAs [1], and has been used most widely in theoretical analysis. This theorem can predict the increase of the mean fitness of a subpopulation from one generation to the next [2]. However, it seems too simple to make quantitative analysis of the performance of GAs. It is in general very difficult to predict the hardness of a problem by using this method alone.

For evaluating the performance of GAs, one of the most important measures is a first-passage time $T_a$ at which the best solution appears for the first time in a population. In this paper, we present a simple method for estimating the mean first-passage time $\overline{T_a}$ in GAs by extending the method used in [3], which was based on the deterministic theory of molecular evolution developed by Eigen and coworkers [4, 5]. In a previous paper [3], we presented the theory of GA dynamics by applying Eigen's model of evolution. The time dependent behavior of a population is given by a set of ordinary differential equations which includes the effect of selection and mutation. The effect of crossover was neglected. We found that the solution for the equations could reproduce well the relative frequencies of alleles in GA calculations on three types of fitness landscapes.

Several methods of calculating the first-passage time in stochastic processes have been reported [6]. There are, for example, the absorbing boundary, adjoint equation and renewal approaches, and so on. Among them, the absorbing boundary method is the most popular and has been used in a variety of applications [7, 8]. In this paper, a method for estimating $\overline{T_a}$ is obtained simply by introducing artificially an absorbing state in Eigen's evolution system. In our method, the absorbing state corresponds to the optimum solution which we want to find. We employed this absorbing boundary condition in order to avoid counting the probability flux, which reaches the optimum solution more than twice. That is, we have to remove a part of probability flux from the system when it arrives at the optimum. If we do not remove that part of probability, it will reach the optimum later again and brings longer $T_a$ [6].

De Jong, Spears, and Gordon proposed a GA theory to compute mean first-passage times within the framework of the Markov chain [9]. They used the Markov model of GAs developed by Nix and Vose [10], and their theory includes selection, mutation, and crossover. The Nix and Vose model can calculate exact state transition probabilities, and thus provides a method for estimating $\overline{T_a}$ on the firm foundation of stochastic process theory. However, there is a serious shortcoming inherent in their theory, which prevents its practical application to real world GAs. The defect of their approach is that the number of states $M$ in the Markov process increases exponentially as the size of population and string length increase. For example, the number $M$ becomes more than two million when population size is 6 and string length is 5. Since the theory involves an $M \times M$ matrix, the analysis of the Markov processes by means of such a rigorous theory would be practically impossible even for a small system like this.

This suggests that it is inevitable for us to introduce some approximations into a stochastic theory when we perform GA analysis in more realistic situations. Shapiro, Prügel-Bennet, and Rattray made another approach for describing GA dynamics by the use of statistical mechanics [11–13]. Their formalism takes into account only a few macroscopic statistical parameters and the remaining microscopic degrees of freedoms are averaged out. This approximation makes it possible for their model to handle general classes of problems in GAs.

Flyvbjerg and Lautrup proposed another method treating the temporal aspect of evolution by means of the adaptive walk model [14]. In their method, evolution is modeled as a stepwise optimization process of a population, and the genetic variability of the population is neglected. They derived analytical expressions for the average length and duration of adaptive walks in a rugged landscape.

The method proposed here uses the deterministic theory of evolution introduced by Eigen [4, 5], which includes quite a few parameters.

Therefore the method is tractable for us even if the system size is large. The other side of this fact means that we can estimate only the average of a stochastic variable. A more complete description would require developing the stochastic theory of random variable $T_a$. Especially, the variance of $T_a$ is another important measure for evaluating the performance of GAs. We hope that the present work serves as a first step toward the comprehensive understanding of first-passage times in GAs.

## 2. Eigen's evolution model

For describing the GA process, we have used a set of difference equations based on the discrete time model of evolution. The method is essentially the same one given by Eigen for describing the evolutionary process of self-replicative macromolecules such as DNA or RNA, in the framework of the continuous time model which uses a set of differential equations [4, 5].

The application of Eigen's continuous time model of evolution to GAs was presented in [3]. In Eigen's theory, the method for solving the differential equations has already been developed by many groups (e.g., [15–17]). However, since most GA applications use generational GAs, it may be better to use the discrete time model for simulating real GAs as discussed in [18]. Therefore we have adopted the discrete time model of GA evolution represented by a set of difference equations. In the following, we give a short description of the problem and describe the method for solving the difference equations.

### 2.1 Discrete time model of evolution

We studied the behavior of a population composed of fixed length binary strings. The evolutionary process of GAs is treated within the framework of a one-locus multiple-allele model. Hence there are $n = 2^l$ alleles represented by binary strings of length $l$, $\mathbf{S}_i \quad (i = 0, \ldots, n-1)$. As an evolutionary procedure, we used a simple genetic algorithm (SGA) described in Goldberg's textbook [2]. The SGA is a generational GA, in which we utilize two nonoverlapping populations at each generation with the birth of offspring and the entire replacement of parents.

We take into account the effect of selection and mutation, while neglecting crossover. The role of crossover in first-passage time problems will be discussed in forthcoming papers. As a selection scheme, we applied proportionate reproduction, and used the roulette wheel selection in numerical experiments.

The size of the population is $N$ which is assumed to be fixed throughout the GA evolution, and the number of the $i$th allele is $N_i$ $(i = 0, \ldots, n-1)$ with $N = N_0 + \cdots + N_{n-1}$. Since we wanted to develop an infinite population model for GA evolution, we used the relative

frequency $x_i(t)$ at generation $t$ instead of $N_i$,

$$x_i(t) = N_i/N \qquad (i = 0, \ldots, n-1).$$

For $x_i(t)$, the condition of $N_0 + \cdots + N_{n-1} = N$ is represented by

$$x_0(t) + \cdots + x_{n-1}(t) = 1.$$

We will sometimes use the vector representation of $x_i(t)$;

$$\boldsymbol{x}(t) = (x_0(t), x_1(t), \ldots, x_{n-1}(t))^T,$$

where $T$ stands for transpose.

In proportionate selection, the relative frequencies at generation $t+1$ are given in terms of the relative frequencies at generation $t$ as

$$x_i(t+1) = \frac{f_i}{\bar{f}(t)} x_i(t) \qquad (i = 0, \ldots, n-1), \tag{2.1}$$

$$\bar{f}(t) = \sum_{i=0}^{n-1} f_i x_i(t). \tag{2.2}$$

Here $f_i$ is the fitness of string $\mathbf{S}_i$ and $\bar{f}(t)$ stands for the mean fitness of the population at generation $t$. We assume that the fitness values $f_i$ are all positive. It is not difficult to find the solution for the equations as given in [18]:

$$x_i(t) = \frac{f_i^t x_i(0)}{\sum_j f_j^t x_j(0)} \qquad (i = 0, \ldots, n-1).$$

In the mutation process, frequencies at the next generation $t+1$ are given by

$$x_i(t+1) = \sum_{j=0}^{n-1} M_{ij} x_j(t) \qquad (i = 0, \ldots, n-1), \tag{2.3}$$

where $M_{ij}$ is an element of the mutation matrix $M$ representing the rate of transition $\mathbf{S}_j \rightarrow \mathbf{S}_i$ by mutation. The mutation matrix $M$ can be described as

$$M_{ij} = (1-p)^{l-d(i,j)} p^{d(i,j)}, \tag{2.4}$$

where $d(i,j)$ is the Hamming distance between strings $\mathbf{S}_i$ and $\mathbf{S}_j$, and we assume that the strings are reproduced with mutation rate $p$ per bit per generation. The following relation is important for solving the Eigen equation

$$\sum_{i=0}^{n-1} M_{ij} = 1, \tag{2.5}$$

and will be used in section 2.2.

The equations of evolution with selection and mutation are obtained by combining equations (2.1) and (2.3)

$$x_i(t+1) = \frac{1}{\bar{f}(t)} \sum_{j=0}^{n-1} M_{ij} f_j x_j(t)$$
$$= \frac{1}{\bar{f}(t)} \sum_{j=0}^{n-1} A_{ij} x_j(t) \qquad (i = 0, \ldots, n-1). \tag{2.6}$$

Here $A_{ij} = M_{ij} f_j$ is an element of the selection-mutation matrix $A$. We call this set of difference equations the discrete time Eigen equation.

### 2.2 Solution for the Eigen equation

We can calculate the frequencies $\boldsymbol{x}(t)$ by applying the discrete time Eigen equation (2.6) $t$ times to the initial distribution $\boldsymbol{x}(0)$. However, there is also an analytical method for solving this set of difference equations. In the following, the procedure given by Nimwegen, Crutchfield, and Mitchell is presented [19].

New variables are introduced

$$\boldsymbol{y}(t) = (y_0(t), \ldots, y_{n-1}(t))^T,$$

which satisfy a system of difference equations

$$y_i(t+1) = \sum_{j=0}^{n-1} A_{ij} y_j(t) \qquad (i = 0, \ldots, n-1), \tag{2.7}$$

assuming the initial distribution $\boldsymbol{y}(0) = \boldsymbol{x}(0)$.

The solution of this system has the form

$$\boldsymbol{y}(t) = A^t \boldsymbol{y}(0) = A^t \boldsymbol{x}(0). \tag{2.8}$$

We can obtain the explicit form of this solution by solving an eigenvalue equation $A\boldsymbol{v} = \alpha \boldsymbol{v}$. It is shown in [3] that the eigenvalues of the matrix $A$ are all real. Furthermore, remembering the assumption that all fitness values $f_i$ are positive, and using the Perron–Frobenius theorem, we can verify that the eigenvalue of the largest magnitude is nondegenerate and positive from the fact that all matrix elements of $A$ are positive when the mutation rate is nonzero. Rumschitzki gives detailed discussions on the properties of the spectrum of selection-mutation matrix $A$ in [17].

We can choose $n$ linearly independent eigenvectors of $A$ with their eigenvalues $\alpha_i$

$$A\boldsymbol{v}^{(i)} = \alpha_i \boldsymbol{v}^{(i)} \qquad (i = 0, \ldots, n-1).$$

For convenience of representation, we choose the normalization of $\boldsymbol{v}^{(i)}$ as

$$\sum_{j=0}^{n-1} v_j^{(i)} = 1 \qquad (i = 0, \ldots, n-1). \tag{2.9}$$

We can construct a solution for the modified equation (2.7) as

$$\boldsymbol{y}(t) = \sum_{i=0}^{n-1} b_i {\alpha_i}^t \boldsymbol{v}^{(i)}, \tag{2.10}$$

where the constants $\{b_i\}$ are determined from the initial condition

$$\boldsymbol{y}(0) = \boldsymbol{x}(0) = \sum_{i=0}^{n-1} b_i \boldsymbol{v}^{(i)}.$$

To calculate $\{b_i\}$, we define the transformation matrix $G$ in terms of the eigenvectors $\boldsymbol{v}^{(j)}$

$$G_{ij} \equiv v_i^{(j)} \qquad (i, j = 0, \ldots, n-1). \tag{2.11}$$

Using the inverse matrix of $G$, the constants $\{b_i\}$ are given by

$$b_i = \sum_{j=0}^{n-1} G_{ij}^{-1} x_j(0) \qquad (i = 0, \ldots, n-1). \tag{2.12}$$

Finally, equation (2.6) is solved. We can show that the solution $\boldsymbol{x}(t)$ is given in terms of $\boldsymbol{y}(t)$

$$x_i(t) = \frac{y_i(t)}{\sum_j y_j(t)} \qquad (i = 0, \ldots, n-1). \tag{2.13}$$

This equation can be proven inductively. First there is a relation

$$\begin{aligned} \frac{y_i(1)}{\sum_j y_j(1)} &= \frac{\sum_j A_{ij} y_j(0)}{\sum_{j,k} A_{jk} y_k(0)} = \frac{\sum_j A_{ij} x_j(0)}{\sum_k f_k x_k(0)} \\ &= \frac{1}{f(0)} \sum_j A_{ij} x_j(0) = x_i(1). \end{aligned}$$

Equation (2.5) is used to show

$$\sum_j A_{jk} = \sum_j M_{jk} f_k = f_k.$$

Second, equation (2.13) is assumed, and then

$$\begin{aligned} \frac{y_i(t+1)}{\sum_j y_j(t+1)} &= \frac{\sum_j A_{ij} y_j(t)}{\sum_{j,k} A_{jk} y_k(t)} = \frac{\sum_j A_{ij} y_j(t)}{\sum_k f_k y_k(t)} \\ &= \frac{\sum_j A_{ij} x_j(t)}{\sum_k f_k x_k(t)} = \frac{1}{f(t)} \sum_j A_{ij} x_j(t) = x_i(t+1) \end{aligned}$$

is demonstrated. Thus we can prove that equation (2.13) also holds at $t+1$.

## 3. Mean first-passage time in genetic algorithms

By generalizing equation (2.6), we obtained a simple method to evaluate the mean number of generations until the optimum allele is introduced in the population of size $N$. The essential point of this method is the assumption that the optimum allele behaves as an absorbing state [6].

Let the allele $\mathbf{S}_0$ be the best solution for a given problem, or equivalently $f_0 > f_i \quad (i \neq 0)$. It is easy to extend the method to landscapes where more than one allele has the maximum fitness value. We consider the situation that, when the optimum allele $\mathbf{S}_0$ appears in a GA calculation, it is removed from the population. By assuming this, we can treat the optimum allele $\mathbf{S}_0$ as an absorbing state. Therefore, in the discrete time model, the evolution equation (2.6) is modified to

$$x_0(t+1) = 1 - \sum_{j=1}^{n-1} x_j(t+1), \tag{3.1}$$

$$x_i(t+1) = \frac{1}{\bar{g}(t)} \sum_{j=1}^{n-1} A_{ij} x_j(t) \qquad (i \neq 0), \tag{3.2}$$

where $\bar{g}(t)$ stands for the average fitness of the nonoptimum alleles

$$\bar{g}(t) = \frac{\sum_{j=1}^{n-1} f_j x_j(t)}{\sum_{j=1}^{n-1} x_j(t)} = \frac{\sum_{j=1}^{n-1} f_j x_j(t)}{1 - x_0(t)}.$$

The calculation starts under the normalization condition

$$\sum_{j=0}^{n-1} x_j(0) = 1,$$

which will be satisfied at all generations.

The equations for the nonoptimum alleles (3.2) can be obtained by setting $x_0(t) = 0$ in equation (2.6). At first glance, the system of equations (3.2) seems to be independent of $x_0(t)$. However, it is to be noted that the reduction of $x_i(t)$ through mutation $x_i(t) - M_{ii}x_i(t) = \{1 - (1-p)^l\}x_i(t)$ includes a component that decays into the absorbing state $s_0$.

From equation (2.5), we find

$$\sum_{i=1}^{n-1} A_{ij} = \sum_{i=1}^{n-1} M_{ij} f_j = \sum_{i=0}^{n-1} M_{ij} f_j - M_{0j} f_j = f_j - M_{0j} f_j,$$

and from equation (3.2), we have

$$\sum_{i=1}^{n-1} x_i(t+1) = \frac{1}{\bar{g}(t)} \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} A_{ij}x_j(t) = \frac{1}{\bar{g}(t)} \sum_{j=1}^{n-1} x_j(t) \sum_{i=1}^{n-1} A_{ij}$$
$$= \frac{1}{\bar{g}(t)} \sum_{j=1}^{n-1} x_j(t) f_j(1 - M_{0j}) = 1 - x_0(t) - \frac{1}{\bar{g}(t)} \sum_{j=1}^{n-1} M_{0j} f_j x_j(t).$$

Then an alternative expression for $x_0(t+1)$ is obtained from equation (3.1)

$$x_0(t+1) = x_0(t) + \frac{1}{\bar{g}(t)} \sum_{j=1}^{n-1} M_{0j} f_j x_j(t).$$

Since all $f_i$ are positive as we have assumed, the increment of $x_0(t)$ at one generation step is

$$\Delta x_0(t) \equiv x_0(t+1) - x_0(t) = \frac{1}{\bar{g}(t)} \sum_{j=1}^{n-1} M_{0j} f_j x_j(t) > 0.$$

Therefore the frequency $x_0(t)$ is a monotonically increasing function of time $t$.

When the frequency of the optimum allele $x_0(t)$ becomes some fixed value $P$, what is the distribution of the number of the optimum allele $N_0$? The discussion in the Appendix answers that $N_0$ has a binomial distribution with the mean $NP$. We define an estimate of $\overline{T_a}$ as a function of the population size $N$

$$\widehat{T_a}(N) = \min_{i \geq 0} \left\{ x_0(i) \geq \frac{1}{N} \right\} = \min_{i \geq 0} \{ N_0(i) \geq 1 \}, \tag{3.3}$$

where $N_0(t) = x_0(t)N$. This definition implies that we choose $P = 1/N$. Hence at $t = \widehat{T_a}(N)$, the expected value of $N_0$ is 1. In other words, equation (3.3) means that we use the estimate of the first-passage time at which $N_0$ becomes 1 for estimating $\overline{T_a}$. Therefore this definition of $\widehat{T_a}(N)$ is not a direct estimate for $\overline{T_a}$. However, as we will show in section 4, $\widehat{T_a}(N)$ works quite well in many cases of GA applications.

It is to be noted that $\widehat{T_a}(N)$ takes an integer value while the value $\overline{T_a}$ is in general a real number. Therefore this method essentially has an error within a maximum of one generation. However one may not care about this degree of error in the real situation of GA applications.

## 4. Numerical experiments

In this section the results of numerical experiments in mean first-passage times of GAs on three types of fitness landscapes are reported. The main

aim of the experiments was to study how well the estimate $\widehat{T_a}(N)$ defined by equation (3.3) can quantitatively predict mean first-passage times of GAs. We examine the flat, multiplicative, and deceptive landscapes.

### 4.1 Flat landscape

The first example is the flat landscape represented by the constant fitness function $f_i \equiv 1$. On this landscape, the discrete time Eigen equation reduces to the equation for the system without selection as given by equation (2.3) for the mutation process

$$x_i(t+1) = \frac{1}{\bar{f}(t)} \sum_{j=0}^{n-1} M_{ij} f_j x_j(t)$$
$$= \sum_{j=0}^{n-1} M_{ij} x_j(t) \qquad (i = 0, \ldots, n-1),$$

since $\bar{f}(t) = \sum_{i=0}^{n-1} f_i x_i(t) = \sum x_i(t) = 1$. Although the selection operator with a constant fitness function, which is equivalent to random sampling, seems to have no effect on the distribution of a population, the selection (sampling) process actually causes genetic drift on it. The effects of random sampling are discussed in the following.

Figures 1 and 2 show the mean first-passage time $\overline{T_a}$ as a function of the population size $N$ in GAs on the flat landscape with the mutation

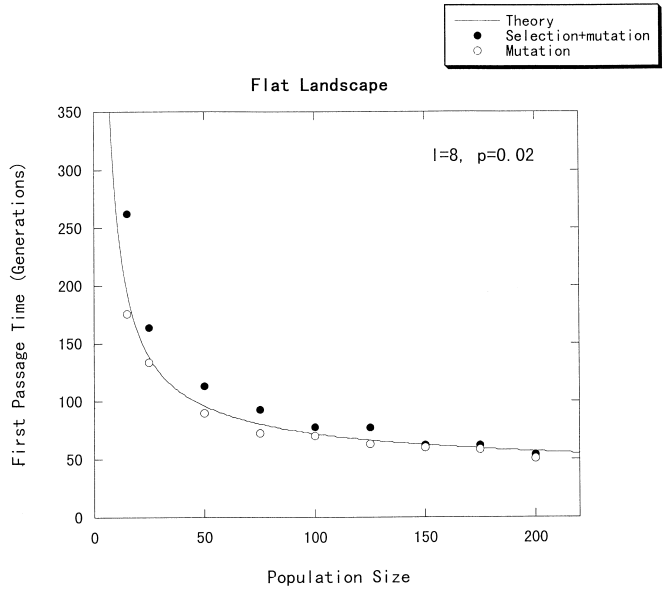



**Figure 1.** The mean first-passage time in GAs on the flat landscape with $p = 0.02$.

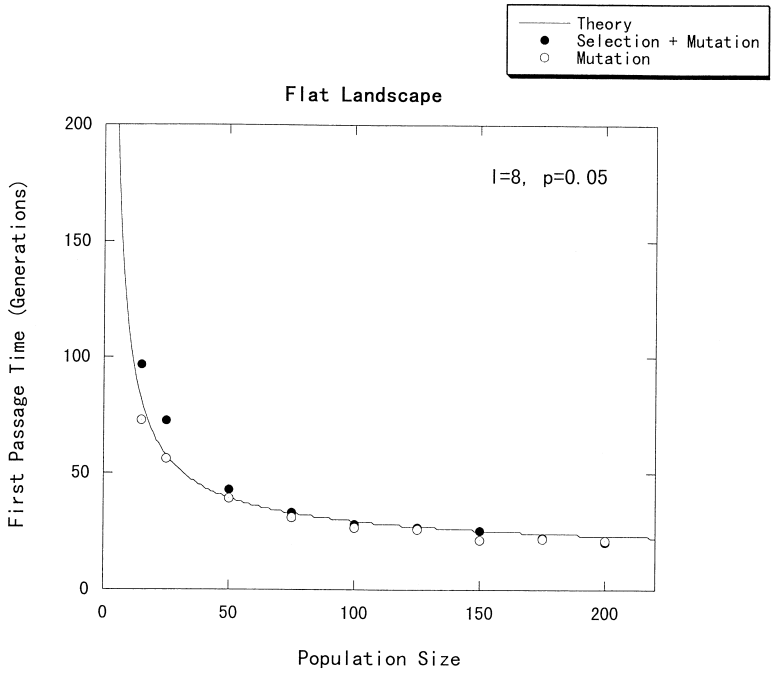



**Figure 2.** The mean first-passage time in GAs on the flat landscape with $p = 0.05$.

rate $p = 0.02$ and 0.05, respectively. The length of the strings is $l = 8$. In these calculations, we used the evolutionary process of SGA [2] described in section 3. In both figures, the initial condition is such that all members of the population start from one string $x_{n-1}(0) = (1, 1, \ldots, 1)$; that is, bits of the string are all 1. We assumed that the optimum allele is the string $(0, 0, \ldots, 0)$, which actually is not the best solution in this case. The closed circles correspond to the mean values of $T_a$ in each 100 runs of the GA calculations on the flat landscape. The open circles represent the mean values of each 100 runs of the GA with mutation alone. Since one of the purposes of the experiment was to study the effects of sampling on $T_a$, we included these calculations for comparison. The mutation-only calculations were performed in the framework of the SGA [2] while omitting the selection procedure. The solid lines in these figures show the theoretical prediction $\widehat{T_a}(N)$ defined by equation (3.3). As reported earlier, the infinite population model predicted the same result for the GA on the flat landscape plus mutation and for the GA with mutation alone.

The theory agrees almost completely with experiments on the flat landscape in the region of large population size $N$. At smaller $N$, however, some discrepancies between theory and experiment were observed.

For example, at the point of $N = 15$ in Figure 1, a large amount of discrepancy can be seen between them. The experimental $\overline{T_a}$ was 253.3 while the theory predicted $\widehat{T_a}(N) = 195$. It can also be seen by comparing both figures that experiments with strong mutation showed better agreement with the theoretical prediction than those with weak mutation.

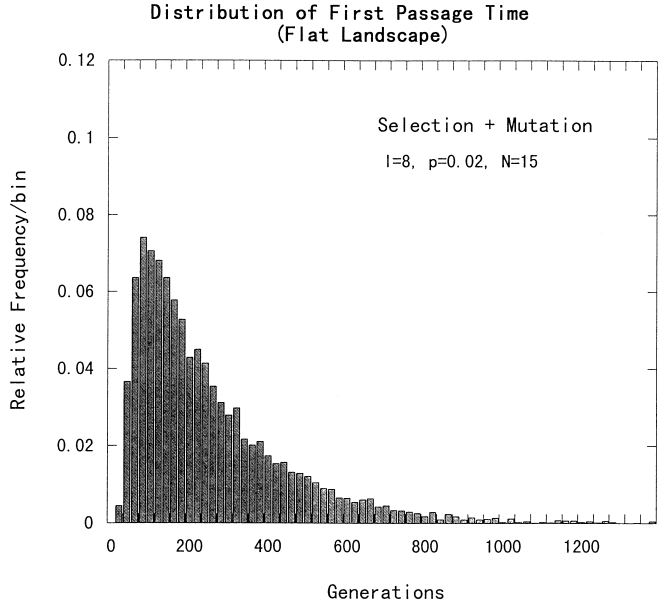
The results of GA experiments on the flat landscape and those without sampling show some differences in the region of small $N$. In general the results of GA without sampling can be reproduced well by $\widehat{T_a}$, while the experimental values of $\overline{T_a}$ with sampling plus mutation are larger than the theoretical estimation at small $N$. This fact suggests that random sampling has the effect of increasing $\overline{T_a}$.

For comparison, we have shown the distributions of $T_a$ in the flat landscape and in the absence of sampling in Figure 3. The parameters in these calculations are $p = 0.02$ and $N = 15$, which give maximum differences between $\overline{T_a}$ in Figures 1 and 2. We carried out 10,000 GA runs repeatedly in each case. The inclusion of sampling causes longer $\overline{T_a}$ and a broader distribution of $T_a$ in Figure 3(a) than 3(b). We obtained $\overline{T_a} = 253.3$ and the standard deviation of $T_a$ $\sigma = 203.3$ in Figure 3(a) while $\overline{T_a} = 186.9$ and $\sigma = 135.6$ in Figure 3(b).

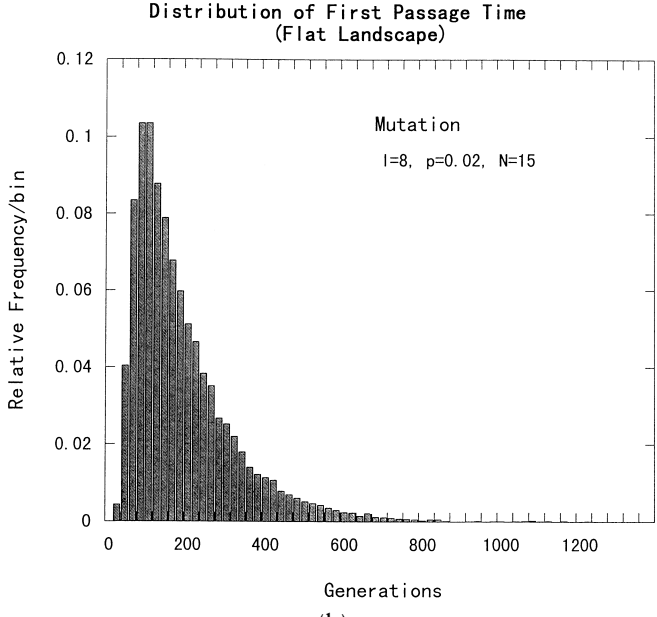
Intuitively, random sampling on the flat landscape seems to have no effect on $T_a$. However, it is well known in population genetics that sampling has a side effect on the population distribution. For example, Derrida and Peliti give a discussion on the evolution of a finite population in the flat landscape [20], see also [21]. Since the infinite population model used here neglects the stochastic effects of random sampling, it seems quite natural that the theory underestimates $\overline{T_a}$ when sampling is present.

To show the effect of random sampling, we performed GA experiments without the absorbing boundary condition, which means that a calculation does not stop even if the optimum allele appears in the population. Figure 4 illustrates the time dependent behavior of the mean and variance of the Hamming distance from the optimum allele $(0, 0, \ldots, 0)$. In this figure it can be seen that while the sampling has a negligible effect on the behavior of the mean value, it drastically reduces the variance of the Hamming distance when the population size $N$ is small. The reason for this change may be understood from the following consideration. Since a parent of each individual can be chosen at random from the previous generation in the sampling process, there is a probability $1/N$ that two individuals have the same parent [20, 21].

On the other hand, in the mutation-only process, each individual evolves independently. Therefore the variance of the Hamming distance becomes small when sampling is present. Also note in Figure 4 that this effect of sampling is small when the population has large $N$. The




(a)




(b)

**Figure 3.** The distribution of the first-passage time $T_a$ in GA on the flat landscape with $p = 0.02$ and $N = 15$; (a) GA with random sampling, (b) mutation-only process. The bin size is 20 generations.

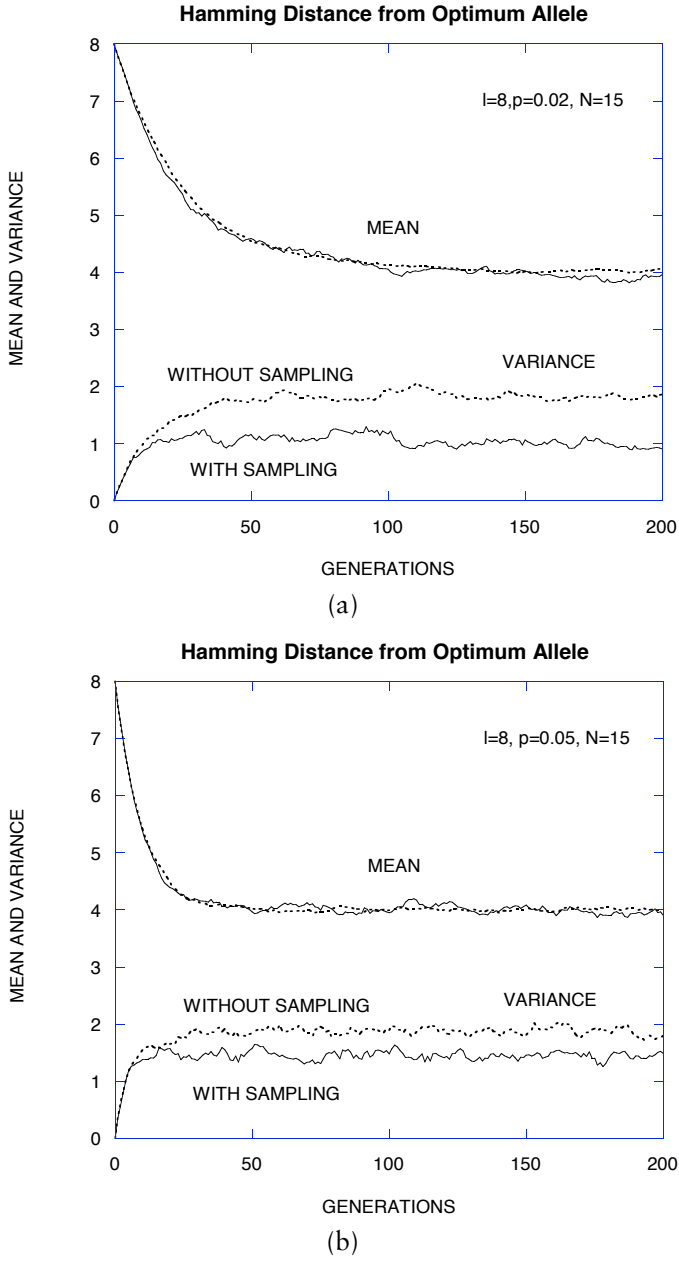
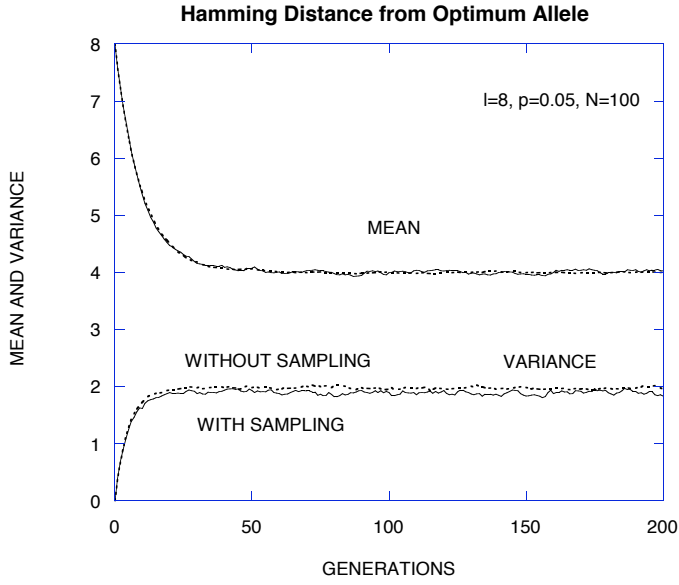



**Figure 4.** The mean and variance of the Hamming distance from the optimum allele in the GA on the flat landscape with $l = 8$; (a) $p = 0.02$ and $N = 15$, with sampling (solid lines) and without sampling (dotted lines), (b) $p = 0.05$ and $N = 15$, and (c) $p = 0.05$ and $N = 100$. The data were obtained by taking the averages of the mean and variance of the population over 100 runs.




(c)

**Figure 4.** (*continued*).

larger variance of the Hamming distance means that the members in the mutation-only process can reach the optimum allele more rapidly than those in the GA process with sampling. The present study suggests that it is important to take into account the variation within a population.

### 4.2 Multiplicative landscape

The second example we give is the multiplicative landscape defined as

$$f_i = (1 - s)^{|i|} \qquad (i = 0, \ldots, n - 1),$$

where $s$ is a parameter representing the strength of selection, and $|i| \equiv d(0, i)$ is the Hamming distance between the optimum allele $\mathbf{S}_0$ and an allele $\mathbf{S}_i$. Here we use the standard binary representation of nonnegative integer $i$. The time dependent behavior of allele frequencies on this landscape without an absorbing state has been studied previously in [3], and it was demonstrated that Eigen's model can reproduce the results of GA calculations of frequencies $x_i(t)$ well.

Figure 5 shows the distribution of $T_a$ in GA on the multiplicative landscape. The GA experiment was carried out 10,000 times. The initial distribution of strings is the same used in section 4.1. By comparing this figure with Figures 3 and 4, we can observe a drastic shortening of $T_a$ ($\overline{T_a} = 42.2$) and realize the remarkable effect of the selection operator. It can also be seen in this figure that the distribution peak becomes very sharp. This property is desirable for the practical application of GAs.

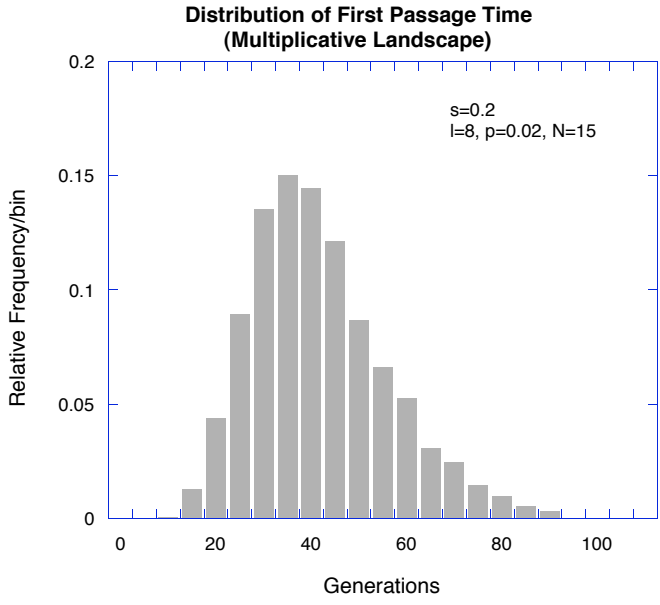



**Figure 5.** The distribution of the first-passage time $T_a$ on the multiplicative landscape with $l = 8$, $p = 0.02$, $N = 15$, and the selection strength $s = 0.2$. Bin size is 5 generations.

Figures 6 and 7 demonstrate the comparison between the theory and experiments with parameters $l = 8$, $s = 0.2$, and $p = 0.02$ and 0.05, respectively. The dotted lines are theoretical estimations on the flat landscape, which are added for the comparison of the two landscapes. As mentioned previously, the result with a strong mutation calculation in Figure 7 shows better agreement with the theory than that of the weak mutation in Figure 6. The theoretical prediction in general underestimates experimental $\overline{T_a}$ in both figures. Also note that the agreement of the theory and experiments is good in large $N$, while some discrepancies are observed in small $N$. This tendency is just like the one observed in section 4.1.

### 4.3 Deceptive landscape

There have been many reports on the deceptiveness of fitness functions [2, 22]. Deceptiveness is usually defined as the structure of a fitness landscape which misleads the evolutionary path by indicating the wrong direction to the selection operator. As an example of deceptive landscapes, we studied a class of fitness functions discussed in [22].

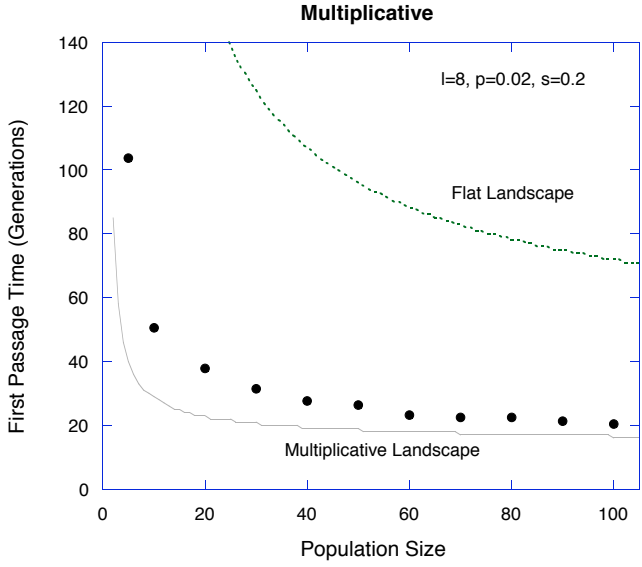



**Figure 6.** The mean first-passage time in GAs on the multiplicative landscape with $p = 0.02$. The closed circles represent the experimental $\overline{T_a}$, and the solid and dotted lines represent the theoretical estimation $\widehat{T}_a$. Calculations were carried out 100 times at each point.

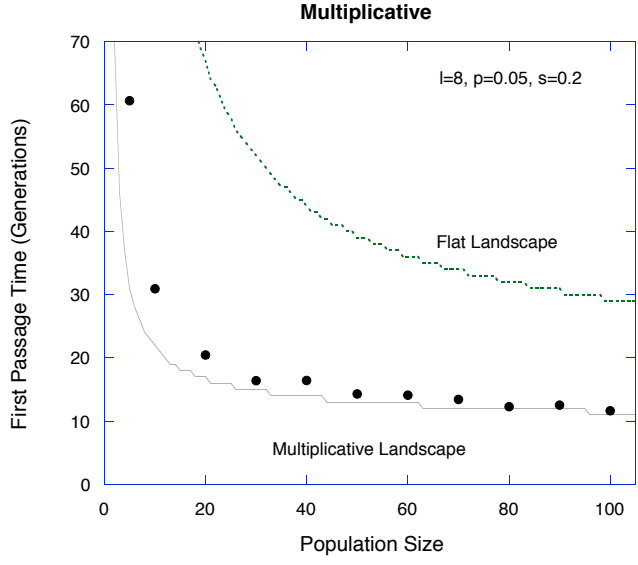



**Figure 7.** The mean first-passage time in GAs on the multiplicative landscape as in Figure 6 with $p = 0.05$.

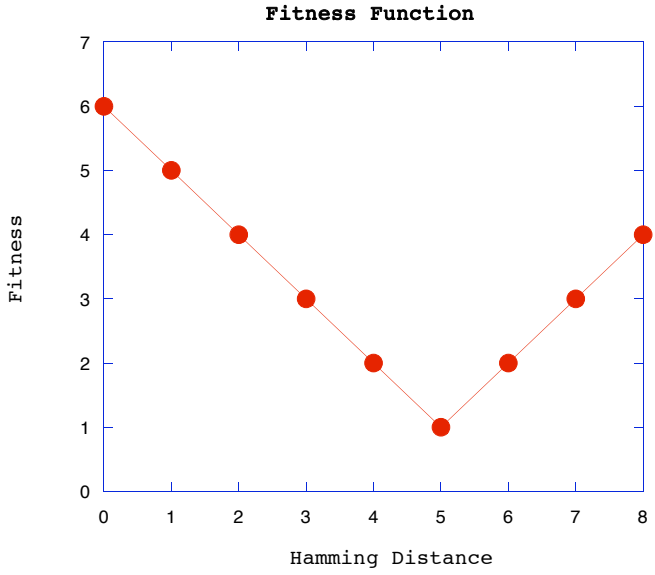



**Figure 8.** The fitness of the deceptive landscape as a function of the Hamming distance $|i|$.

The result of using the fitness function of the following form, referred to as a trap function in [22], defined in terms of $|i| = d(0, i)$ is

$$f_i = \begin{cases} 6 - |i| & (|i| \leq 5) \\ |i| - 4 & (|i| > 5). \end{cases}$$

Figure 8 illustrates the shape of this landscape.

The initial condition is the same as that used previously. That is, the calculation starts from the initial point $|i| = 8$ and the population moves to the direction of the target point $|i| = 0$. Therefore GA operators, especially mutation in this case, have to drive some members of the population to the intermediate point $|i| = 5$ against the unfavorable slope in the fitness function.

One can easily expect that $\overline{T_a}$ of this example would be rather longer than $\overline{T_a}$ of easy problems like the GA on the multiplicative landscape. Figures 9 and 10 illustrate that this is really true. Since, in many GA runs, the calculations did not produce the optimum string within the appropriate computational time when we used the mutation rates of $p = 0.02$ and 0.05, the mutation rate $p$ is increased to 0.07 in Figure 9 and 0.10 in Figure 10, respectively.

The theoretical estimate $\widehat{T_a}$ for this problem deviated from experimental values when the weaker mutation rate of 0.07 was used. Furthermore, in small $N$ regions, the theory underestimated the experimental $\overline{T_a}$. These are the same tendencies observed in other landscapes.

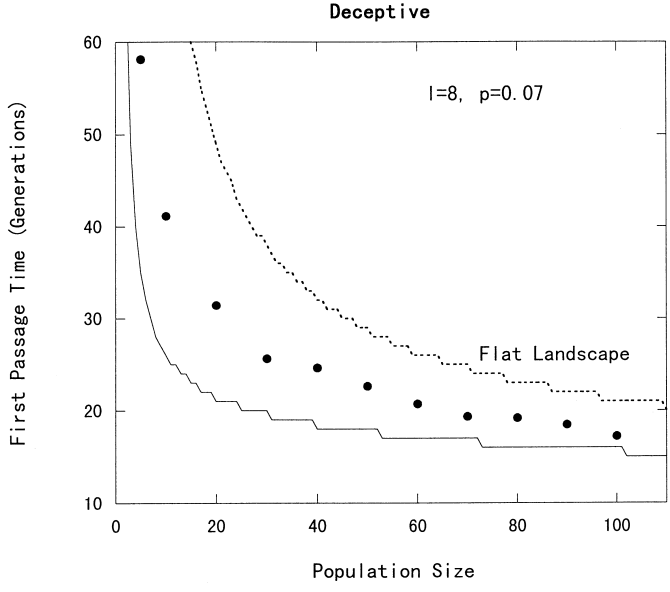



**Figure 9.** The mean first-passage time in GAs on the deceptive landscape as in Figure 6 with $p = 0.07$.

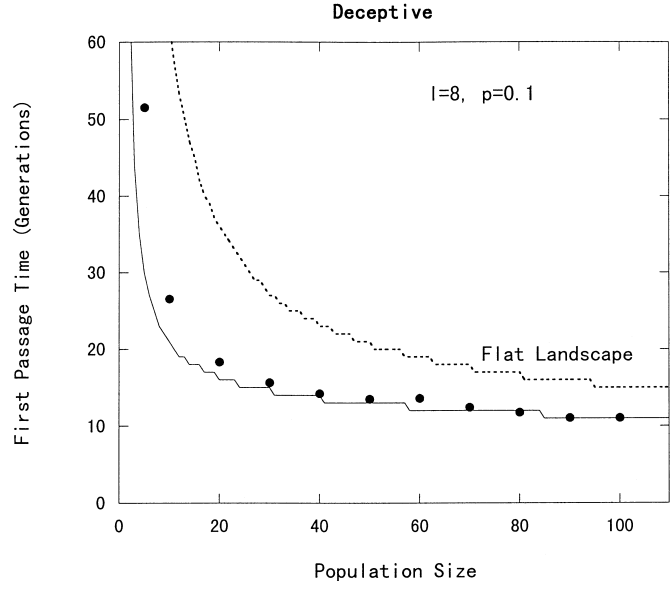



**Figure 10.** The mean first-passage time in GAs on the deceptive landscape as in Figure 6 with $p = 0.10$.

| Landscape | $p$ | $N$ | $\overline{T_a}$ | $\sigma$ | Median | $\widehat{T_a}$ |
|---|---|---|---|---|---|---|
| Flat | 0.02 | 15 | 253.3 | 203.3 | 193 | 195 |
| | | 200 | 58.7 | 23.0 | 55 | 57 |
| | 0.05 | 15 | 92.1 | 72.2 | 70 | 82 |
| | | 200 | 22.4 | 8.1 | 21 | 23 |
| Multiplicative | 0.02 | 15 | 42.2 | 15.3 | 40 | 25 |
| | | 200 | 17.6 | 3.5 | 18 | 15 |
| | 0.05 | 15 | 24.0 | 10.1 | 22 | 18 |
| | | 200 | 10.3 | 2.1 | 10 | 10 |
| Deceptive | 0.07 | 15 | 34.6 | 17.4 | 31 | 23 |
| | | 200 | 14.5 | 3.9 | 14 | 14 |
| | 0.10 | 15 | 21.5 | 9.4 | 20 | 18 |
| | | 200 | 9.5 | 2.3 | 9 | 10 |

**Table 1**. Statistical quantities of $T_a$. Experimental values $\overline{T_a}$, $\sigma$ (standard deviation) and Med. (median) of $T_a$ are obtained in 10,000 runs.

### 4.4 Other properties of $T_a$

In this section we briefly discuss some other statistical properties of $T_a$. Table 1 lists $\overline{T_a}$ for the three landscapes together with the standard deviation and median of $T_a$ obtained in 10,000 GA runs. The theoretical estimation $\widehat{T_a}$ is also included for comparison.

As we have pointed out, the agreement between the theory and experiment becomes poor when both $p$ and $N$ are small. In such a case, the median also deviates from the mean and the distribution has a quite asymmetrical shape. For example, the large $T_a$ region has a very long tail in Figure 3. Since the theoretical $\widehat{T_a}$ is actually the estimate for the first-passage time of $N_0 = 1$, the estimation may deteriorate when the distribution is not symmetrical. On the contrary, when both $p$ and $N$ are large, the theoretical estimation agrees quite well with the experiment and the mean and median of $T_a$ are almost identical. As an example, the distribution of $T_a$ for the multiplicative landscape with $p = 0.05$ and $N = 200$ are shown in Figure 11. One can observe the almost symmetrical shape of the distribution in this figure. In this case $\overline{T_a}$, the median of $T_a$ and $\widehat{T_a}$, almost coincide with each other.

## 5. Discussion

This paper has presented a simple method for estimating the mean first-passage time $\overline{T_a}$ in GAs. The method is based on the theory of macromolecular evolution developed by Eigen and coworkers [4, 5]. There are several reports on the applications of Eigen's theory to genetics and population genetics (e.g., [23–26]). Since the GA theory has an

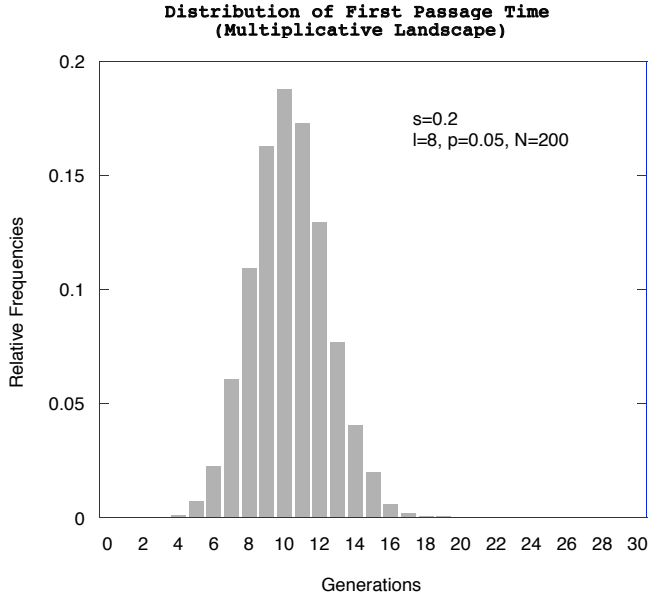



**Figure 11.** The distribution of the first-passage time $T_a$ on the multiplicative landscape with $l = 8$, $p = 0.05$, $N = 200$, and the selection strength $s = 0.2$.

aspect of one field of genetics, we thought that the theory of Eigen might be also useful in GAs [3]. In practice, we have applied Eigen's theory to GAs for describing the time dependence of allele frequencies. We showed that the system of ordinary differential equations with the selection-mutation matrix $A$ reproduced the GA experiments with large population size excellently.

Since GAs are a class of stochastic algorithms, it is necessary to develop a stochastic theory if we want to describe GA processes precisely. Several methods based on the finite population model for analyzing GAs [9, 10] have been proposed. These methods use the theory of the Markov chain and can provide a precise description of GA processes. However, as we have stated, these methods are too rigid for us to apply to GAs in realistic environments. On the contrary, the method proposed here is an extension of Eigen's theory which assumes an infinitely large population and does not include stochastic effects caused by random sampling. However, these shortcomings make this method very easy to use and applicable in a wide range of situations.

Although this method estimates $\overline{T_a}$ only indirectly, the GA simulation in general shows good agreement between theory and experiment. The theory performs very well in large $N$ regions for all examined landscapes. This result seems quite natural because it depends on the infinite population model. The agreement is insufficient in the regions of very small $N$

where the diffusion caused by the random sampling may become significant. Another important observation is that GAs with strong mutation evolve faster than those with weak mutation. In this respect, we need further study to draw a definite conclusion.

Many articles in population genetics have a close connection with the evolution of GA processes studied in this work, and some of them treat finite population models. For example, Kimura gave some analytical results on the neutral evolution of a finite population [27], and there are several papers studying evolution in the flat landscape [20, 21]. Since the system in the flat landscape has a very simple structure, we think that there is a possibility to construct more quantitative methods for the first-passage time analysis in the near future.

There are also several papers treating the evolutionary processes on the multiplicative landscape [23, 28–30]. These papers give some analytical formulae representing the allele distribution of evolving populations. For example, Higgs and Woodcock obtained equations for describing the evolution of the moments of the allele distribution [29] and the exact expression for the allele frequencies in the stationary state [30]. Therefore, it is interesting to extend their results to the analysis of the first-passage time problems in a finite population.

On the deceptive landscape, Barton and Rouhani studied the frequency of shifts between alternative equilibrium states in a finite population model [31, 32]. They discuss the transition of the population from lower to higher fitness peaks through the fitness valley which must be overcome by random drift. The situation of their landscape is very similar to the deceptive landscape presented in section 4, and therefore their study may provide valuable insight into the mechanism of gene flow through the valley separating fitness peaks.

It may be true that only a theory which takes into account the random sampling of a population can give the complete description of $T_a$, the probability distribution $P(T_a)$. Even though the present method is deterministic, we expect that this work will become the first step toward the complete theory of the first-passage time problem in GAs.

## Acknowledgement

The author would like to thank the anonymous reviewer for making a number of useful suggestions and for pointing out some important articles.

## Appendix. Distribution of $N_0$

We considered a system of $N \times M$ strings evolving through selection and mutation. Here $M$ was the number of populations, and $N$ is the size of each population. We assumed that $M$ was very large and that the system

consisted of an infinitely large number of strings. Therefore it was reasonable to assume that the evolution of the system could be described by the discrete time Eigen equation which is independent of the population size $N$. If we removed the optimum strings whenever they appeared in the population, the system may follow equations (3.1) and (3.2).

When the relative frequency of the optimum allele $x_0$ becomes $P$, what is the distribution of $N_0$ in the $M$ populations? The answer is given by considering the following example. Let $m_1$ and $m_2 = m - m_1$ be the numbers of red and black balls in an urn, respectively. If we randomly take out $r$ balls from the urn, the probability of having $k$ red balls is given in the form of the hypergeometric distribution [33]

$$q_k = \frac{1}{\binom{m}{r}} \binom{m_1}{k} \binom{m - m_1}{r - k} = \frac{1}{\binom{m}{m_1}} \binom{r}{k} \binom{m - r}{m_1 - k}.$$

Then we have the inequality [33]

$$\binom{r}{k} \left(P - \frac{k}{m}\right)^k \left(Q - \frac{r-k}{m}\right)^{r-k} < q_k < \binom{r}{k} P^k Q^{r-k} \left(1 - \frac{r}{m}\right)^{-r},$$

where $P = m_1/m$ and $Q = 1 - P$. As $m \to \infty$ with fixed $P$, we can derive the probability of the binomial distribution from the inequality

$$q_k = \binom{r}{k} P^k Q^{r-k}.$$

By substituting $m = N \times M$, $m_1 = N \times M \times P$, $r = N$, and $k = N_0$, we obtain the answer to the previous question about the distribution of $N_0$:

$$q_{N_0} = \binom{N}{N_0} P^{N_0} (1 - P)^{N - N_0}. \tag{5.1}$$

The mean of $N_0$ is given by

$$\overline{N_0} = NP. \tag{5.2}$$

We have the probability of the case where there is no optimum string

$$q_0 = (1 - P)^N. \tag{5.3}$$

## References

[1] J. H. Holland, *Adaptation in Natural and Artificial Systems* (University of Michigan Press, Ann Arbor, 1975).

[2] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning* (Addison-Wesley, Reading, MA, 1989).

[3] H. Furutani, "Application of Eigen's Evolution Model to Infinite Population Genetic Algorithms with Selection and Mutation," *Complex Systems*, **10** (1996) 345–366.

[4] M. Eigen, "Self-organization of Matter and the Evolution of Biological Macromolecules," *Die Naturwissenschaften*, **58** (1971) 465–523.

[5] M. Eigen and P. Schuster, "The Hypercycle A: A Principle of Natural Self-organization: Emergence of the Hypercycle," *Die Naturwissenschaften*, **64** (1977) 541–565.

[6] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1997).

[7] E. W. Montroll and K. E. Shuler, "The Application of the Theory of Stochastic Processes to Chemical Kinetics," *Advances in Chemical Physics*, **1** (1958) 361–399.

[8] D. A. McQuarrie, "Stochastic Approach to Chemical Kinetics," *Journal of Applied Probability*, **4** (1967) 413–478.

[9] K. A. De Jong, W. M. Spears, and D.F. Gordon, "Using Markov Chains to Analyze GAFOs," in *Foundations of Genetic Algorithms 3*, edited by L. D. Whitley and M. D. Vose (Morgan Kaufmann, San Francisco, CA, 1995).

[10] A. E. Nix and M. D. Vose, "Modeling Genetic Algorithms with Markov Chain," *Annals of Mathematics and Artificial Intelligence*, **5** (1992) 79–88.

[11] A. Prügel-Bennet and J. L. Shapiro, "The Dynamics of a Genetic Algorithm for Simple Random Ising Systems," *Physica D*, **104** (1997) 75–114.

[12] A. Prügel-Bennet, "Modelling Evolving Populations," *Journal of Theoretical Biology*, **185** (1997) 81–95.

[13] M. Rattray and J. L. Shapiro, "The Dynamics of a Genetic Algorithm for a Simple Learning Problem," *Journal of Physics A: Mathematical and General*, **29** (1996) 7451–7473.

[14] H. Flyvbjerg and B. Lautrup, "Evolution in a Rugged Fitness Landscape," *Physical Review A*, **46** (1992) 6714–6723.

[15] C. J. Thompson and J. L. McBride, "On Eigen's Theory of the Self-organization of Matter and the Evolution of Biological Macromolecules," *Mathematical Biosciences*, **21** (1974) 127–142.

[16] B. L. Jones, R. H. Enns, and S. S. Rangnekar, "On the Theory of Selection of Coupled Macromolecular Systems," *Bulletin of Mathematical Biology*, **38** (1976) 15–28.

[17] D. S. Rumschitzki, "Spectral Properties of Eigen Evolution Matrices," *Journal of Mathematical Biology*, **24** (1987) 667–680.

[18] D. E. Goldberg and K. Deb, "A Comparative Analysis of Selection Schemes used in Genetic Algorithms," in *Foundations of Genetic Algorithms*, edited by G. J. E. Rawlins (Morgan Kaufmann, San Francisco, CA, 1991).

[19] E. van Nimwegen, J. P. Crutchfield, and M. Mitchell, "Statistical Dynamics of the Royal Road Genetic Algorithm," *Santa Fe Institute Working Paper*, 97-04-35 (1997).

[20] B. Derrida and L. Peliti, "Evolution in a Flat Fitness Landscape," *Bulletin of Mathematical Biology*, **53** (1991) 355–382.

[21] P. G. Higgs and B. Derrida, "Genetic Distance and Species Formation in Evolving Populations," *Journal of Molecular Evolution*, **35** (1992) 454–465.

[22] K. Deb and D. E. Goldberg, "Analyzing Deception in Trap Functions," in *Foundations of Genetic Algorithms 2*, edited by L. D. Whitley (Morgan Kaufmann, San Francisco, CA,1993).

[23] P. G. Higgs, "Error Thresholds and Stationary Mutant Distributions in Multi-locus Diploid Genetics Models," *Genetical Research*, **63** (1994) 63–78.

[24] T. Wiehe, E. Baake, and P. Schuster, "Error Propagation in Reproduction of Diploid Organisms: A Case Study on Single Peaked Landscape," *Journal of Theoretical Biology*, **177** (1995) 1–15.

[25] J. Swetina and P. Schuster, "Self-replication with Errors: A Model for Polynucleotide Replication," *Biophysical Chemistry*, **16** (1982) 329–345.

[26] M. Nowak and P. Schuster, "Error Thresholds of Replication in Finite Populations: Mutation Frequencies and the Onset of Muller's Ratchet," *Journal of Theoretical Biology*, **137** (1989) 375–395.

[27] M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, 1983).

[28] J. Haigh, "The Accumulation of Deleterious Genes in A Population—Muller's Ratchet," *Theoretical Population Biology*, **14** (1978) 251–267.

[29] P. G. Higgs and G. Woodcock, "The Accumulation of Mutations in Asexual Populations and the Structure of Genealogical Trees in the Presence of Selection," *Journal of Mathematical Biology*, **33** (1995) 677–702.

[30] G. Woodcock and P. G. Higgs, "Population Evolution on a Multiplicative Single-peaked Fitness Landscape," *Journal of Theoretical Biology*, **179** (1996) 61–73.

[31] N. H. Barton and S. Rouhani, "The Frequency of Shifts between Alternative Equilibria," *Journal of Theoretical Biology*, **125** (1987) 397–418.

[32] N. H. Barton and S. Rouhani, "Adaptation and the 'Shifting Balance'," *Genetical Research (Cambridge)*, **61** (1993) 57–74.

[33] W. Feller, *An Introduction to Probability Theory and its Application*, volume 1 (Wiley, New York, 1957).