Theoretical Analysis of Selection Operator in Genetic Algorithms

Tatsuya Okabe

Abstract—Genetic algorithms (GAs) have gathered much attention by researchers and practitioners because of successful results in many real-world optimization problems [4], [5], [6]. However, to our knowledge, the working mechanism of GAs, in particular dynamics of GAs, is still unclear and should be investigated. Since we believe that the theoretical analysis of dynamics will help us not only to understand the working mechanism of GAs but also to develop more efficient algorithms, we proposed the theoretical framework to analyze the dynamics of GAs in [14]. Since we formulated only crossover and mutation in [14], we formulate selection operator in this paper as an extension of our previous paper. Based on the proposed theoretical framework, this paper analyzes the dynamics of selection and derives several theorems.

I. INTRODUCTION

Evolutionary algorithms (EAs) have been recognized as powerful tools for solving optimization problems for a couple of decades. In the literature, many algorithms have been introduced and many applications are successfully optimized by EAs [4], [5], [6]. Despite of many success of EAs, theoretical work for EAs seems to be fewer than empirical work. Recently, many researchers have started to concentrate on the theoretical work for EAs. Examples are studies on the accuracy of the approximation of the Pareto front, on the convergence properties, on the diversity of individuals in a population, and on the model of EAs, e.g. [3], [25], [29], [22]. However, the dynamics of EAs during optimization and the role of each genetic operator are still unclear. In our opinion, the analysis of dynamics of EAs is very helpful not only to understand working mechanism of EAs [11] but also to improve performance of EAs and to propose a new algorithm [12] because the final solution from an optimizer is the total result of the dynamics of EAs. The analysis of the dynamics of EAs is completely equal to the one of the role of each genetic operator. Since EAs are consist of several genetic operators, the dynamics of EAs might be divided into several dynamics of genetic operators. Thus, in this paper, we will try to understand the role of genetic operators first and derive several theories. Particularly, the theoretical approach for the dynamics of EAs will be carried out.

In [14], we proposed a theoretical framework of crossover and mutation for genetic algorithms (GAs) to analyze the dynamics. As an extension of the framework proposed in [14], we formulate selection operator to analyze the dynamics of selection in this paper. We will formulate a proportional selection, truncation selections \((\mu, \lambda)\) and \((\mu + \lambda)\). After formulating proportional selection, \((\mu, \lambda)\) selection and \((\mu + \lambda)\) selection, the role of selection operator will be analyzed theoretically and some theories are derived.

The rest of this paper is structured as follows. Section 2 explains the related work in the literature and the motivation of this paper. In Section 3, we summarize the theoretical framework of crossover and mutation previously proposed in [14]. Section 4 newly proposes the framework of proportional selection and truncation selections. Based on the theoretical framework proposed, the dynamics of selection are discussed and some theories are derived in Section 5. Finally, the conclusions of this paper are summarized in Section 6.

II. RELATED WORK AND MOTIVATION OF THIS PAPER

A. Analysis of Dynamics using Cumulants

Several work for analysis of dynamics can be found in the literature, e.g. [17], [21], [3], [16], [32]. In these papers, the distribution of population is not directly treat, but the representative values of the distribution, i.e. cumulants, are used to analyze the dynamics of EAs. The usage of representative values might lead the easy treatment of population.

Prügel-Bennett and Shapiro have shown how to derive a set of equations using cumulants for describing dynamics of a GA [17]. They have shown the influence of operators on the first four cumulants. Prügel-Bennett has also discussed selection and ranking [18], two point crossover [19] and recombination [20] with cumulants. Rogers and Prügel-Bennett have researched the roulette wheel and stochastic universal sampling [23]. They have also analyzed the generational selection and the steady state selection in [24]. Van Nimwegen et al. have analyzed the dynamics of the Royal Road GA with cumulants [27]. They introduced GA dynamics as a flow in the fitness space (objective space). To confirm the theoretical results, Kallel [9] and Matsumura et al. [10] have observed the dynamics of EAs with cumulants empirically. Beyer has also investigated the dynamics without selection, in particular for evolution strategy [2].

Although the cumulants are easier to be treat than the population distribution itself, it is clear that much information will be lost.

B. Modeling of EAs

As the way to treat population directly, Goldberg proposed the model of GAs with two-bit strings for solving the minimum deceptive problems [7]. Vose extended this model to an arbitrary number of strings [28]. In these models (theories), the information of the population distribution is stored in the probability vector where each component indicates the probability of a certain chromosome.

The usage of the probability vector implicitly results in an infinite population size because of its definition [28]. To
tackle the infinite population size, Suzuki explained how to model the GA with Markov chain [26]. By Markov chain, Rudolph has analyzed the convergence of canonical GA [25]. Since an infinite population size is not realistic, some researches for finite size population can be found in [17], [18]. The detailed review can be seen in [1], [30], [31].

C. Motivation of This Paper

Although the existing theories can analyze the dynamics of EAs, the strong merit of these theories is the analysis of the final solutions. This is caused by the fact that the theories are optimized for the analysis of the final solutions. However, our target in this paper is the observation of the dynamics of EAs. Therefore, this paper proposes a new theoretical framework optimized for the analysis of the dynamics.

III. THEORETICAL FRAMEWORK OF CROSSOVER AND MUTATION

In this section, we roughly summarize the theoretical framework of crossover and mutation operators in GAs we proposed previously. For more detailed explanation of equations, see [13], [14].

A. One-point Crossover

Let’s assume the l-locus binary chromosome. Thus, each chromosome, Z, holds Z ∈ {0, 1}^l. Now, we define an index ζ, called chromosome index, and the ratio of individuals with Z in the population P(Z) or P(ζ), called occurrence probability, as follows:

Definition 1 (Chromosome Index ζ): A chromosome index of Z is defined by ζ = \sum_{i=1}^{l} B(Z, i) \times 2^{-i-1}. Here, B(Z, i) is the i-th allele in the chromosome Z. Using a chromosome index, we can name a certain chromosome in the simple way. For example, the chromosome of Z = [00001010] can be written as the chromosome of ζ = 10.

Definition 2 (Occurrence Probability P(Z) = P(ζ)): Let’s consider the chromosome Z (or ζ). An occurrence probability, P(Z) = P(ζ), is defined by P(Z) = P(ζ) = n_{Z}/n_{pop}. Here, n_{Z} and n_{pop} mean the number of individuals with the chromosome Z and the total number of individuals in a population, respectively. The occurrence probability should hold \sum_{ζ} P(ζ) = 1.

To formulate one-point crossover operator, the model in Fig. 1 will be considered. Generally, the crossover operator is carried out by exchanging chromosomes in two parents and generates two offspring, see Fig 1. This model can be divided into two sub-models. One is that the Parent 1 generates the Offspring 1 with support by the Parent 2 and the other is that the Parent 2 generates the Offspring 2 with support by the Parent 1. In this paper, we simply consider the first sub-model because the result obtained by a sub-model leads the same results by two sub-models from stochastic point of view. We assume that two parents are selected randomly from a population.

Now, we will try to generate a certain chromosome Z = (###...#) by crossover operator. Here, the symbol of # means a given allele. To generate the chromosome Z, we have two possibilities. One is that the chromosome Z is selected as the Parent 1, and no crossover occurs. The other is the following parents, Z_{l-i}^{L} and Z_{i}^{F}, are selected and crossover occurs at a proper crossover point indicating ‘|’ (after i-th loci) in Eq. (1).

\[
Z_{l-i}^{L} = \{##...#|*...*\} \\
Z_{i}^{F} = \{*...*|##...#\}
\]

Here, the symbol of ‘*’ is so-called don’t care symbol [7]. The subscript i (or l − i) indicates the number of don’t care symbols. The superscript L (or F) means that the last (or first) parts of chromosome are don’t care symbols.

For a certain building block such Z_{i}^{L} or Z_{i}^{F}, an index, called chromosome index for building block, is also defined as follows:

Definition 3 (Chromosome Index for Building Block ζ_{i}^{L}): A chromosome index for Z_{i}^{L} = {##...#*...*} (The last i alleles are the don’t care symbols) is defined as ζ_{i}^{L} = \sum_{j=1}^{l-i} B(Z, j) \times 2^{-j-1}.

Definition 4 (Chromosome Index for Building Block ζ_{i}^{F}): A chromosome index for Z_{i}^{F} = {*...*##...#} (The first i alleles are the don’t care symbols) is defined as ζ_{i}^{F} = \sum_{j=i+1}^{l} B(Z, j) \times 2^{-i-1}.

Using the occurrence probability of ζ_{i}^{F} and ζ_{i}^{L}, denote P_{e}^{L}(ζ_{i}^{L}) and P_{e}^{F}(ζ_{i}^{F}) respectively, the probability to generate the chromosome Z by crossover operator, denote ξ(Z) called generative probability by crossover, can be calculated as follows:

\[
\xi(Z) = P(Z) \times (1 - P_{e}) + \sum_{k=1}^{l-1} P_{e}^{L}(ζ_{L}^{k}) \times P_{e}^{F}(ζ_{F}^{k}) \times P_{e} \times \frac{1}{l-1}.
\]

Here, P_{e} is a crossover rate. The first term means that the parent with the chromosome Z is selected and no crossover occurs (1 − P_{e}). The second term means that two parents of ζ_{L}^{k} and ζ_{F}^{k} are selected and crossover occurs (P_{e} \times 1/(l − 1)). Since there are several possibilities to generate Z by crossover, all possibilities are summed up. Using Eq. (2), generative probabilities of all possible chromosomes can be calculated.
B. Bit-flipping Mutation

In canonical GA, mutation is conducted after the crossover. Since Eq. (2) gives the generative probability by one-point crossover, the mutation operator should function to this generative probability.

If the Hamming distance \( d \) between two chromosomes \( Z \) and \( Z' \) is given by \( |Z \oplus Z'| \), the probability that the chromosome \( Z' \) becomes \( Z \), denote \( \Xi(Z' \rightarrow Z) \), can be written as:

\[
\Xi(Z' \rightarrow Z) = (1 - P_m)^{d} P_m^{d},
\]

here, \( P_m \) is the mutation rate. Since the Hamming distance indicates the number of different alleles between \( Z \) and \( Z' \), the mutation should occur at \( |Z \oplus Z'| \) alleles and no mutation should occur at \( (l - |Z \oplus Z'|) \) alleles. Since all chromosomes have a chance to become \( Z \), the total generative probability of \( Z \) by mutation, \( \Xi(Z) \), can be calculated as:

\[
\Xi(Z) = \sum_{Z'} (1 - P_m)^{d} P_m^{d} \xi(Z').
\]

This equation can be also written as:

\[
\Xi = \mathcal{M} \xi,
\]

Here, the matrix \( \mathcal{M} \) is called mutation matrix. The vectors of \( \Xi \) and \( \xi \) are constructed by \( \Xi(Z) \) and \( \xi(Z) \) for all \( Z \).

IV. THEORETICAL FRAMEWORK OF SELECTION

In [14], only crossover and mutation operators are formulated. In this section, we newly formulate selection operator to understand the role of selection and to build up the theoretical framework to analyze the dynamics of GAs.

A. Proportional Selection

In this section, proportional selection [7] will be modeled. Assume a probability of offspring distribution to be \( \Xi \). During the proportional selection, a selection probability is calculated according to its fitness. Now, let us assume a fitness of each chromosome to be \( f \), here \( \zeta \) is a chromosome index. Thus, the probability of the next parents after the proportional selection, denote \( \bar{P} \), can be calculated as:

\[
\bar{P} = \frac{1}{\text{tr}(\text{diag}(S\Xi))} S\Xi,
\]

here, \( S \) is given by:

\[
S = \begin{pmatrix}
  f_0 & 0 & \cdots & 0 \\
  0 & f_1 & \ddots & \vdots \\
  \vdots & \ddots & \ddots & 0 \\
  0 & \cdots & 0 & f_{2^n-1}
\end{pmatrix}.
\]

B. \((\mu, \lambda)\) Selection

As a truncation selection, \((\mu, \lambda)\) selection will be modeled. Without loss of generality, the minimization task is assumed. In the \((\mu, \lambda)\) selection, the best \( \mu \) offspring will be selected from \( \lambda \) offspring as the next parents. Assume the offspring probabilities and the next parental probabilities to be \( \zeta^{(t)}(f) \) and \( \rho^{(t+1)}(f) \). Here, the superscript, e.g. \((t)\), is a generational index and \( f \) is the value of an objective. To avoid confusion, the generational index will be added according to necessity. It is worth noticing that since the selection will be carried out in the fitness space, the both probabilities are defined in the fitness space. If the mapping between the parameter space and the fitness space is one-to-one mapping, then the following equation should be satisfied:

\[
\zeta^{(t)}(f) = \Xi^{(t)}(\zeta), \quad \rho^{(t+1)}(f) = \theta^{(t+1)}(f),
\]

here, the fitness value of the chromosome \( \zeta \) corresponds to \( f \). To simplify the further discussion, one-to-one mapping is assumed. We will discuss non-one-to-one mapping in the next paragraph.

The cases of one-to-one and non-one-to-one mapping are shown in Fig. 2. The upper figure is the example of one-to-one mapping and the lower one is the non-one-to-one mapping. If we assume an one-to-one mapping, we can use the probability \( \Xi(\zeta) \) defined in the parameter space without any modification in the fitness space. If non-one-to-one mapping is assumed, a certain summation shown in the low-right figure is necessary.

![Fig. 2. Examples of one-to-one mapping and non-one-to-one mapping.](image-url)
from the second class, it is not important to observe each individual in the first class or the second class. Simply, we should observe the total number of individuals in the first class or the second class.

Let us assume the numbers of offspring in the first class, in the second class and in the third class to be \((\mu - j), i\) and \((\lambda - \mu + i + j)\), respectively. It is worth saying that the value \(j\) can vary from \((\mu - \lambda)\) to \(\mu\). Here, \(\mu - \lambda \leq 0\). Due to the simple calculation later, we assign these numbers. See Fig. 3.

Since the total number of individuals is \(\lambda\) and the total number of selected offspring is \(\mu\), the following constraints should be satisfied:

\[
\begin{align*}
\mu & \leq \lambda, \quad 0 \leq \mu - j \leq \lambda, \quad 0 \leq i \leq \lambda, \\
0 & \leq \mu - \lambda - i \leq j.
\end{align*}
\]

(9)

Now we will take the probability into account that more than one offspring in the second class are selected from the first class and there is no possibility to select more than one offspring with \(f'\). In order to select more than one offspring with \(f'\), the following constraints should be satisfied:

\[
\begin{align*}
j & \geq 1, \quad i \geq 1.
\end{align*}
\]

(10)

This means that more than one individuals in the second class are necessary to select more than one individuals from the second class. Additionally, the number of individuals in the first class should be less than \(\mu\). Otherwise, all \(\mu\) individuals will be selected from the first class and there is no possibility to select individuals in the second class.

By coupling Eq. (9) and Eq. (10), one can get the following constraints:

\[
\begin{align*}
\mu & \leq \lambda, \quad 1 \leq i \leq \lambda, \quad 1 \leq j \leq \mu, \quad \mu - \lambda - i \leq j.
\end{align*}
\]

(11)

The third and forth constraints in Eq. (11) can be merged into:

\[
\max(1, \mu - \lambda + i) \leq j \leq \mu.
\]

(12)

Under the above situation, the numbers of selected offspring after the \((\mu, \lambda)\) selection for the class 1, \(\min(i, j)\) for the class 2 and \((j - \min(i, j))\) for the class 3.

The probability to occur the situation in Fig. 3, denote \(P(f', i, j)\), is given by:

\[
P(f', i, j) = \left( \frac{\lambda}{i} \right) \left( \frac{\lambda - i}{\mu - j} \right) \left( \sum_{f' < f} \varsigma(f') \right)^{\mu - j} \left( \sum_{f' > f} \varsigma(f') \right)^{\lambda - \mu - i + j}.
\]

(13)

Since the expected number of individuals from the class 2 is \(\min(i, j)\) under the above situation, the expected number of individuals from the class 2, denote \(E[N(f')]\), can be calculated as follows:

\[
E[N(f')] = \sum_{i=1}^{\mu} \sum_{j=0}^{\min(i, j)} \left[ \min(i, j) P(f', i, j) \right],
\]

(14)

here,

\[
a = \max(1, \mu - \lambda + i).
\]

(15)

Since the value \(i\) and \(j\) should satisfy Eq. (11) and Eq. (12), the ranges of summation are restricted as shown in Eq. (14). Since probability can be calculated by \(E[N(f')]\) divided by \(\mu\), the next parental probabilities can be calculated as:

\[
\begin{align*}
g^{(t+1)}(f') &= \frac{1}{\mu} \sum_{i=1}^{\lambda} \sum_{j=\max(1, \mu - \lambda + i)}^{\mu} \left[ \min(i, j) \right] \\
& \times \left( \frac{\lambda}{i} \right) \left( \frac{\lambda - i}{\mu - j} \right) \left( \sum_{f' < f} \varsigma(f') \right)^{\mu - j} \\
& \times \left( \sum_{f' > f} \varsigma(f') \right)^{\lambda - \mu + i + j}.
\end{align*}
\]

(16)

\[
\begin{align*}
&\text{with this equation, the distribution of selected population by } (\mu, \lambda) \text{ can be calculated. The implications of this equation will be also discussed in Section 5.}
\end{align*}
\]

C. \((\mu + \lambda)\) Selection

Based on the previous section, \((\mu + \lambda)\) selection will be formulated in this section. In \((\mu + \lambda)\) Selection, the combined population (\(\mu\) parents and \(\lambda\) offspring) is firstly generated. Thereafter, the best \(\mu\) individuals are selected from the combined population. In the \((\mu + \lambda)\) selection, the relation of \(\mu \leq \lambda\) in the \((\mu, \lambda)\) is not necessary. The number of parents and offspring in three classes are assumed as Fig. 4.

The meanings of three classes, i.e. the classes 1, 2 and 3, are the same with Fig. 3.

Here, the following constraints should be satisfied:

\[
\begin{align*}
\mu & \geq 1, \quad \lambda \geq 0, \quad 0 \leq \mu - j \leq \mu, \\
0 & \leq i \leq \mu, \quad 0 \leq j - i \leq \mu, \quad 0 \leq \lambda - m \leq \lambda, \\
0 & \leq k \leq \lambda, \quad 0 \leq m - k \leq \lambda.
\end{align*}
\]

(17)

One can simplify the above constraints as:

\[
\begin{align*}
\mu & \geq 1, \quad \lambda \geq 0, \quad 0 \leq i \leq \mu, \\
0 & \leq j \leq \mu, \quad 0 \leq k \leq \lambda, \quad k \leq m \leq \lambda.
\end{align*}
\]

(18)

To pick up, at least one individual from the class 2 by \((\mu + \lambda)\) selection, the following conditions should be satisfied:

\[
\begin{align*}
0 & \leq \mu - j \leq \mu - 1, \quad 0 \leq \lambda - m \leq \mu - 1, \\
\mu - j + \lambda - m & \leq \mu - 1, \quad i + k \geq 1.
\end{align*}
\]

(19)
\[ \mu \]

\begin{tabular}{ccc}
Class 1 & Class 2 & Class 3 \\
\hline
Parent & $\mu-j$ & $i$ & $j-i$ \\
Offspring & $\lambda-m$ & $k$ & $m-k$ \\
\hline
\end{tabular}

Fig. 4. $(\mu + \lambda)$ selection. $\mu$ parents and $\lambda$ offspring are classified into three classes.

Here, the first three constraints indicate that the number of individuals in the class 1 in both of parents and offspring is, at most, $\mu - i$. Otherwise, all $\mu$ individuals will be selected from the class 1. The last one indicates that the number of individuals in the class 2 is, at least, 1. By merging Eq. (18) and Eq. (19), one can get the following constraints:

\[
0 \leq i \leq \mu, \quad \text{max}(0, 1 - i) \leq k \leq \lambda, \\
\text{max}(\lambda, \lambda - j + 1) \leq m \leq \lambda.
\]  

(20)

Under the assumption of Fig. 4, the number of individuals to be selected in the class 2 can be calculated as:

\[
\min\left(\mu - (\mu - j) - (\lambda - m), i + k\right) = \min\left(j + m - \lambda, i + k\right).
\]  

(21)

The first part, i.e. $(j + m - \lambda)$, indicates the possible slots to select individuals from the class 2. The last part, i.e. $(i + k)$, indicates the number of individuals in the class 2.

The probability of the occurrence of Fig. 4, denote $P(f', i, j, k, m)$, is given by:

\[
P(f', i, j, k, m) = \binom{\mu}{i} \binom{\mu - i}{\mu - j} \binom{\lambda}{k} \\
\times \left(\lambda - k\right) \left(\sum_{f < f'} g^{(i)}(f)^j \sum_{f < f'} c^{(i)}(f)^k\right) \\
\times \left(\sum_{f < f'} g^{(i)}(f)^j \sum_{f < f'} c^{(i)}(f)^k\right)^{m-k}.
\]  

(22)

Thus, the expected number of individuals from the class 2 in the next parents, denote $E[N(f')]$, is given by:

\[
E[N(f')] = \mu \sum_{i=0}^{\mu} \sum_{j=0}^{\mu} \sum_{k=0}^{\lambda} \sum_{m=0}^{\lambda} \left[ \min(j + m - \lambda, i + k) \times P(f', i, j, k, m) \right].
\]  

(23)

\[ \]  

Here, $b_1 = \max(1, i)$, $b_2 = \max(0, 1 - i)$ and $b_3 = \max(k, \lambda - j + 1)$. Since the probability can be calculated by $E[N(f')]$ divided by $\mu$, the next parental probabilities can be calculated as:

\[
\phi^{(t+1)}(f') = \frac{1}{\mu} \sum_{i=0}^{\mu} \sum_{j=0}^{\mu} \sum_{k=0}^{\lambda} \sum_{m=0}^{\lambda} \left[ \min(j + m - \lambda, i + k) \left(\frac{\mu - i}{\mu - j}\right) \right. \\
\times \left(\frac{\lambda}{k}\right) \left(\frac{\lambda - k}{\lambda - m}\right) \sum_{f < f'} \phi^{(i)}(f)^j \right. \\
\left. \times \left(\sum_{f < f'} \phi^{(i)}(f)^j \right)^{m-k} \times \left(\sum_{f < f'} c^{(i)}(f)^k\right)^{m-k} \right],
\]  

(24)

\[ \]  

This equation lets the distribution of selected population be calculable. The implications of this equation will be discussed in [15] due to page limitation.

V. Simulation Results and Discussions

Using obtained equations of proportional selection and truncation selection operators in the last section, we analyze the dynamics of selection operator in this section. Due to the page limitation and the similarities of $(\mu, \lambda)$ and $(\mu + \lambda)$ selections, the results of proportional selection and $(\mu, \lambda)$ selection are introduced. For simplicity, a maximization problem for proportional selection and a minimization problem for truncation selection are considered.

A. Proportional Selection

To observe the role of the proportional selection, no crossover and no mutation are used. After giving the uniform probability to the occurrence probability vector, i.e. $P(Z) = 1/2^l$, $\forall Z$, the transition of probability vector will be calculated with Eq. (6). For the calculation, $l = 10$ is used. As a fitness value $f(Z)$, the following equation is used: $f(Z) = \zeta_Z + 1$, where $\zeta_Z$ is the chromosome index of $Z$. For examples, $f([0000000000]) = 1$ and $f([1111111111]) = 1024$.

The result is shown in Fig. 5. We can see that the probability of the best chromosome increases linearly. In more details, the probabilities of relatively better individuals increase linearly and the probabilities of relatively worse individuals decrease in the generation $t$. Thereafter, in the better individuals in the generation $t$, the same procedures are repeated in the generation $(t+1)$. Finally, only the probability of the best individual achieves to one.
To check the theoretical results, an empirical experiment is carried out. Ten thousand individuals and one hundred individuals are tested. The results are shown in Fig. 6. If sufficiently large number of individuals is used, there is no difference between the empirical result and the theoretical result, see Fig. 6 (a). However, if small number of individuals is used, see Fig. 6 (b), there is a certain difference between the empirical result and the theoretical result. We call this difference as finite population effect caused by the usage of occurrence probability which implicitly assumes infinite population. Since the selection is strongly influenced by the distribution of other offspring, this difference appears if the number of individuals is small.

Now, we will see the details of Fig. 6 (b). The empirical probabilities of individuals with better fitness values are smaller than the theoretical probabilities. Correspondingly, the other empirical probabilities are bigger than the theoretical probabilities because the sum of probabilities should be 1. This situation can be explained as follows: If infinite population can be assumed, the occurrence probabilities of the best individuals in the initial population should be non-zero. Under this situation, the probabilities can follow the theoretical results. However, it is impossible to assume infinite population in a real calculation. Thus, we have to assume finite population. If finite population is assumed, it is possible that there is no best individuals in the initial population. In this case, the probabilities of the best individuals should be zero. By averaging several runs, the probabilities of the best individuals are smaller than the theoretical ones.

**B. (μ,λ) Selection**

With Eq. (16), the transition of each probability is observed. To observe the effect of (μ, λ), no crossover and no mutation are assumed. Under this assumption, the parental distribution can be assigned to the offspring distribution as: $c^{(t)}(f) = g^{(t)}(f)$. In the first generation, we assume all components in the occurrence probability vector to be same, i.e. $1/2^l$.

Here, (9, 10) is investigated. The result is shown in Fig. 7 for 10 bits. Here, the fitness is assumed to be $f(Z) = \zeta_Z$ and the minimization task is used.

Fig. 7 indicates that the probabilities of the better individuals increase exponentially but the probabilities of the worse individuals decrease exponentially.

To check the relationship between $\mu$ and $\lambda$, several combinations are tested. As one of examples, the result of (μ, 10) is shown in Fig. 8.

This figure indicates that the gradient of the probabilities of the better individuals is big if $\mu$ is relatively smaller than $\lambda$. Oppositely, the gradient is small if $\mu$ is relatively similar with $\lambda$.

Now, curves of the best chromosome whose fitness is $f_{opt}$
Theorem 1:
Under (μ, λ) selection without crossover nor mutation, the probability of the best chromosome, denote $P(f_{opt})$, in early generations can be approximated by:

$$P(f_{opt}) \sim \left(\frac{\lambda}{\mu}\right)^t$$

where $t$ is a generational index.

Proof:
Eq. (16) leads the following equation:

$$\frac{\varrho(t+1)(f_{opt})}{\varrho(t)(f_{opt})} = \frac{\lambda}{\mu} \left(1 - \varrho(t)(f_{opt})\right)^{\lambda-1} + \ldots + \varrho(t)(f_{opt})^{\lambda-1}.$$  

When $\varrho(t)(f_{opt}) \approx 0$, the above equation becomes nearly $\lambda/\mu$. Oppositely, $\varrho(t)(f_{opt})$ approaches 1, the equation becomes nearly 1. (Q.E.D.)

With this equation, one can easily derive the necessary generations that the probability of the best individual achieves 1. If we initially generate several individuals randomly, each chromosome has the probability of $1/2^\mu$. Thus, the probability of the best individual will follow the equation of:

$$P(f_{opt}) \approx \frac{1}{2^\mu} \left(\frac{\lambda}{\mu}\right)^t,$$  

here, $t$ is a generation index. Thus, one can obtain the following relationship:

$$\frac{1}{2^\mu} \left(\frac{\lambda}{\mu}\right)^t = 1 \iff t = l \cdot \frac{\ln 2}{\ln \lambda - \ln \mu}. \quad (26)$$

Since Eq. (25) is valid during small $P(f_{opt})$ and $P(f_{opt})$ will be saturated before achieving 1, it is reasonable to replace ‘=’ in Eq. (26) by ‘≥’. Thus, one can obtain the following theorem:

Theorem 2:
Under (μ, λ) selection without crossover nor mutation, the necessary generations $t$ that the probability of the best chromosome $P(f_{opt})$ converges to 1 can be calculated as:

$$t \geq l \cdot \frac{\ln 2}{\ln \lambda - \ln \mu}. \quad (27)$$

here, $l$ is the number of bits. This time is called takeover time [8].

To check the previous theoretical results, empirical investigations are carried out. For (9000, 10000) and (90, 100), the results are shown in Fig. 9. In the figures, the averaged probability over 10000 runs are shown. To avoid complicated figures, only representative results are drawn.

In Fig. 9 (a) and (b), the trend in early generations shows very similar with theoretical results. However, in later generations, the trend seems to be different from the theoretical ones, in particular Fig. 9 (b).

In Fig. 9 (b), the probability of the best chromosome converges a certain value ($\neq 1$). This situation can be explained as follows: Since the number of individuals is restricted, there are some runs where the best individual does not exist. In these cases, the probability of the better individuals, which are not the best but really exist, converges one. By averaging the probabilities over 10000 runs, the saturation observed in Fig. 9 (b) appears.
If we use the big number of individuals like Fig. 9 (a), the trend shows better agreement with the theoretical result during the whole generation. However, the finite population effect can be still visible in later generations.

VI. CONCLUSIONS

In this paper, we newly built up theoretical framework of proportional selection, $\mu, \lambda$ selection and $(\mu + \lambda)$ as an extension of our previous paper [14] which proposed theoretical framework of crossover and mutation. Although the finite population effect should be considered to observe the dynamics of single objective genetic algorithms quantitatively, our theoretical framework can show good agreement with an empirical result qualitatively. The main conclusions of this paper are as follows:

1. Oppositely, the the probabilities of other individuals decreases because the sum of probabilities should be 1. As more details, the dynamics of selection operator is as follows: the probabilities of the better individuals in a current population increases linearly/exponentially; the probabilities of the worst individuals in a current population decreases; in the next generation, the same flow will be repeated for the selected population.

Under proportional selection, selection pressure strongly relates to the gradient of the curve for the best individual.

The finite population effect on selection operator is severe. Although the theoretical probability of the best individual converges to 1 and others to 0 under proportional selection or truncation selection, several probabilities of the better individuals converge to certain values. This is caused by the finite sampling.

The probability of the best individual, $P(f_{\text{opt}})$, under $(\mu, \lambda)$ selection can be approximated by: $P(f_{\text{opt}}) \sim (\lambda/\mu)^t$, where $t$ is a generational index. As indicated in the proof of Theorem 1, this approximation is valid before $P(f_{\text{opt}}) \approx 1$. The so-called takeover time can be calculated as: $t \geq 1 \cdot \ln(2)/(\ln(\lambda) - \ln(\mu))$ for $(\mu, \lambda)$ selection.

ACKNOWLEDGMENT

The authors would like to thank Dr. Yaochu Jin and Dr. Bernhard Sendhoff in Honda Research Institute Europe GmbH for their fruitful discussions. For improving this paper, we also thank the reviewers of this paper for their valuable comments.

REFERENCES