On the Dynamics of EAs without Selection

Hans-Georg Beyer

University of Dortmund, Department of Computer Science XI
D-44221 Dortmund, Germany
beyer@zappa.cs.uni-dortmund.de

Abstract

This paper investigates the dynamics of evolutionary algorithms (EAs) without fitness based selection (constant fitness). Such algorithms exhibit a behavior similar to the MISR effect (mutation-induced speciation by recombination) which has been found in the analysis of $(\mu/\mu_D, \lambda)$ evolution strategies. It will be shown that this behavior can be observed in a variety of EAs, not only in unrestricted search spaces, but also in binary GAs. The quantification of this effect is done by introducing the expected population variance σ_P^2. The evolution of σ_P^2 over the time g is analytically calculated for both unrestricted and binary search spaces. The theoretical predictions are compared with experiments. The genetic drift phenomenon and the diffusion effect are derived from the general σ_P^2 formulae, and it will be shown that MISR is a *finite* population size *sampling* effect which cannot be observed in infinite populations.

1 Introduction

The question of how GAs, or more generally – evolutionary algorithms (EAs), do work is still waiting for a (hopefully) definite system of answers. Even though some intuitively appealing working principles have been proposed in the GA field, i.e. the schema theorem, the implicit parallelism (Holland, 1975), and the building block hypothesis (Goldberg, 1989), there is a continuously growing population of scientists who cast doubt on the correctness and/or foundation and/or utility of those principle, especially when function optimization is considered (Grefenstette, 1991; De Jong, 1993; Rechenberg, 1994; Rudolph, 1994). Furthermore, the question arises whether these principles must necessarily define the only way of thinking about and understanding the working of a GA. That is, is there room for a set of alternative working principles? If the whole EA-class is considered, there is indeed an alternative system of working principles proposed in (Beyer, 1995a, 1997):

1. the *evolutionary progress principle* (EPP),
2. the *genetic repair* (GR) hypothesis,
3. the *mutation-induced speciation by recombination* (MISR) principle.

However, this alternative set of principles has been derived from the analysis of evolution strategies (ESs). Its applicability to all EAs is somewhat *ex pede Herculem*, because the question as to the relevance on GAs in binary search spaces is still open. This paper is related to the third principle, the MISR. Before going into the details, it seems expedient to give a short explanation of EPP, GR, and MISR.

The EPP can be regarded as the *governing* principle of evolutionary optimization/melioration. It states on the qualitative level that the evolutionary progress observed from generation at time g to $g+1$ is a result of two opposite tendencies: the *progress gain* and the *progress loss*.

Progress, considered as a quantity, can have different definitions. It can be the expected fitness change from g to $g+1$, known as the so-called *quality gain* $\overline{Q}$ (for infinite population models also known as "selection response" (Mühlenbein & Schlierkamp-Voosen, 1994)). Or one may consider the change of the population in the search space measuring the approach toward the optimum, known as the *progress rate* φ. It is important to note that "progress" is always considered as a *local* quantity, where "local" refers to *local in time*, but *by no means local in search space*. Analyzing EAs using the EPP paradigm thus means, to extract the gain- and loss-producing forces induced by the genetic operators when applied to the population. It is to remark as an important observation that the selection operator mainly influences the gain part, whereas the recombination operator mainly influences the loss part leaving the gain part almost unchanged. This leads us to the benefit of recombination.

The genetic repair (GR) hypothesis gives an alternative explanation why recombination *can* provide a performance gain over an EA using selection and (perhaps) mutation alone. Genetic repair describes an effect of *extracting similarities* from the *selected* offspring. The selected offspring have certain "traits" or "building blocks" which make them superior to the rest of the offspring. Furthermore, they have "traits" or "building blocks" that deteriorate the offspring's fitness. As has been shown for intermediate multirecombination (Beyer, 1995b), these deteriorative parts in the genome of the selected individuals are nearly statistically independent of each other, whereas the fitness improving parts are correlated by the selection process. Applying the intermediate, i.e. averaging, recombination dampens the statistically independent parts. That is, the deteriorative parts undergo a statistical error correction – the GR-effect. Those parts exhibiting the same tendencies/correlations are more or less conserved by the averaging, thus, recombination performs similarity extraction.

The GR-effect does not necessarily imply a performance gain over a simple mutation-selection EA. That depends on the properties of the fitness landscape; and furthermore, the offspring population must be sufficiently spread over the search space in order to have offspring genomes with both large gain and loss parts. Getting such a sufficiently spread offspring population, i.e. a large population variance in the search space, can be achieved in at least three ways:

a) by random initialization of the parental population at $g = 0$ over the whole search space (only applicable for low cardinality alphabets and in restricted search spaces).

This technique is the preferred one for binary GAs. In its extreme form, the population size is chosen in such a way that the GA can even work without physical mutations, i.e. the mutation rate (bit-flipping rate) is $p_m = 0$ (Goldberg, 1989). However, there is a certain probability of loosing alleles due to genetic drift (see Section 3.3 and 4.3).

b) for intermediate (μ/μ_I)-multirecombination in ESs, the mutation operator with a sufficiently large mutation strength is absolutely needed, because the whole parent population is mapped onto the center of mass by the (μ/μ_I)-recombination (see Beyer, 1995b). The offspring variance in the search space is obtained by physical mutations only. These are applied to the center of mass parent. Therefore, the population variance (in the search space) is equal to the physical mutation strength.

c) by producing population variance through dominant recombination. Actually, the principle behind this way of producing population variance – the *mutation-induced speciation by recombination* (MISR) principle – can be observed in its pure form by considering an EA *without selection.* This paper is devoted to the analysis of the MISR phenomenon.

The MISR-principle states that the physical mutations (e.g. additive mutations, bit-flips) are transformed into genetic variety by dominant recombination. That is, the result of the repeated application of mutation and (dominant) recombination produces a more or less stable population distribution in the search space. This population resembles a *species*: the individuals are "crowded" around an "imaginary wild-type" (known from biology):

MISR-Principle

DOMINANT RECOMBINATION & PHYSICAL MUTATIONS $\Rightarrow$ "SPECIES"

The variance of the population with respect to the "wild-type" is a function of the physical mutation parameter (mutation strength and mutation rate, respectively) and the *effective* population size μ (i.e., those individuals of the population which take part in the reproduction process). There seems to be an "invisible confinement" keeping the individuals together *without any selection pressure.* This at first glance astonishing observation, firstly made by the analysis of the dominant μ/μ_D-multirecombination ES in real-valued search spaces under very special conditions (Beyer, 1995b), seems to be of more general interest, because it is independent of the EPP and the GR-hypothesis. It has been argued in (Beyer, 1997) that it should hold for a larger class of EAs. However, a proof was still pending. This paper gives the proofs, showing that the MISR-principle is indeed a universal phenomenon in mutation-recombination-algorithms with random sampling based recombination operators (for their definition, see below) acting on *finite* populations. Furthermore, it shows that standard GAs exhibit the same variance dynamics; however, in those cases it is *not* caused by the recombination, but by the probabilistic selection. Thus, MISR gets another interpretation: *mutation-induced speciation by random-sampling.*

The rest of this paper is organized as follows. The next section introduces the dominant recombination and the EAs to be investigated. Furthermore, the variance measures are defined. Section 3 is devoted to the analysis of the variance dynamics in unrestricted search spaces. Section 4 addresses the binary search space and Section 5 gives a short summary.

2 How to Measure and Simulate the MISR-Effect

2.1 The Population Variance

The MISR-principle makes a *qualitative* statement on the distribution of an evolving population in the search space. In order to quantify the MISR-effect, one has to define certain measures of diversity which allow for an evaluation of the population dispersion in the search space. An natural measure of dispersion is the population variance. Since the MISR-effect can be observed and analyzed in its purest form by investigating EAs without fitness based selection, statistical independence of the different genome positions in the individuals can be assumed. This will simplify the analysis considerably, because under this condition it suffices to investigate the evolution of just one gene position. That is, we will consider only one component from the individual's genome.

The individual's genome may be denoted by $\vec{x}$ and $\vec{b}$, respectively, depending on the search space considered. Here, $\vec{x}$ is a state from an unrestricted search space such as $\mathbb{R}^N$ or $\mathbb{Z}^N$, and $\vec{b} \in \mathbb{B}^\ell$, where ℓ is the dimension of the binary search space. A single component from the genome will be denoted as x_m and b_m, respectively, indicating that this component belongs to the individual with number m. The individual index m runs from 1 to μ, where μ is the *effective* population size. Note, in evolution strategies (ESs) one usually has the parental population size μ and an intermediate offspring population of size λ. Even in this case, the effective population size is μ, as we will see in the next subsection.

Let us consider a population of μ individuals at a fixed gene position. The population consists of μ random variates x_m. Its *average value* $\langle x \rangle$ is therefore

$$\langle x \rangle := \frac{1}{\mu} \sum_{m=1}^{\mu} x_m. \tag{1}$$

This $\langle x \rangle$ might be referred to as "wild-type gene". All individuals are more or less crowded around this average type. Their deviation from the wild-type, i.e. $x_m - \langle x \rangle$, may serve as dispersion measure when taken to the square and averaged over all μ individuals. Actually, this is just the definition of the variance $\mathrm{Var}\{x\} := \langle (x - \langle x \rangle)^2 \rangle$. It will be called *population variance*

$$\mathrm{Var}\{x\} := \frac{1}{\mu} \sum_{m=1}^{\mu} (x_m - \langle x \rangle)^2. \tag{2}$$

It is really important to realize that $\mathrm{Var}\{x\}$ is a *random variate*, because the population members are random variates having probability density functions which change over the time/generations g. In order to characterize the spread of the population we need to define the *expected population variance*, denoted by $\sigma^2[x]$

$$\sigma^2[x] := \mathrm{E}[\mathrm{Var}\{x\}] = \overline{\mathrm{Var}\{x\}}, \qquad \text{NB:} \quad \mathrm{E}[y] \equiv \overline{y}, \tag{3}$$

where the expectation is taken over the N-dimensional state space and the overline symbol is only a short form for the expectation functional.

The definitions (1), (2), and (3) can be immediately transferred to the binary alphabet case. One just has to replace x by b leading to

$$\mathrm{Var}\{b\} := \frac{1}{\mu} \sum_{m=1}^{\mu} (b_m - \langle b \rangle)^2, \qquad \langle b \rangle := \frac{1}{\mu} \sum_{m=1}^{\mu} b_m \tag{4}$$

and $$\sigma^2[b] := \overline{\mathrm{Var}\{b\}}. \tag{5}$$

The calculation of (2), (3) and (4), (5) will be the main goal in this paper. Before starting the derivation, we first have to introduce the EAs to be analyzed.

2.2 EAs without Fitness Based Selection

As already mentioned, the MISR-effect can be studied much easier when fitness based selection is switched off. Figure 1 shows the data flow in $(\mu/\mu_D, \lambda)$ dominant multiparent recombination ESs (Rechenberg, 1994; Beyer, 1995b). One single gene position of the individuals is considered and displayed as numbered boxes. On the upper side of the "Generation Loop", there is the parental population consisting of μ genes having an expected population variance σ_P^2 at generation g. The index "$_P$" stands for parent. Then, dominant multiparent recombination is applied to the parental population. The result is labeled by the subscript "$_R$" indicating recombinants. One might wonder why further gene positions are not displayed in this picture. The reason is that the effect of recombination at the single gene level is just a *random sampling*: each recombinant $\boxed{1}_R$ is simply obtained by randomly choosing one of the μ parents

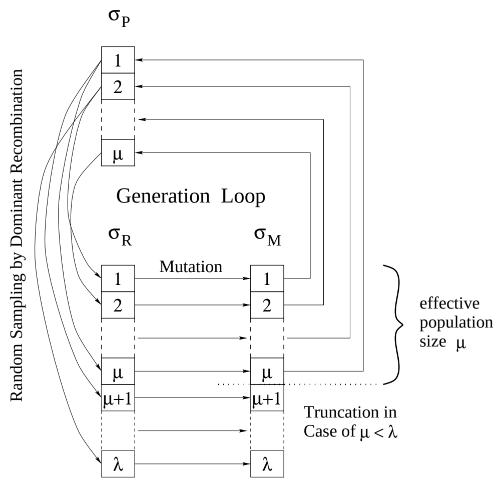


Figure 1: The $(\mu/\mu_D, \lambda)$-ES with generation gap comprising dominant recombination, mutation, and reproduction. It does *not* contain a selection operator. Note, only one gene position is displayed. Although it is trivial fact, it seems important to be remarked that the notion "selection" refers to a selective preference of certain *individuals* as to their *fitness*. That is, selection is on the individual's level; it always picks the genetic information of an individual as a whole. In contrast to that, the sampling process by dominant recombination is a random process which uniformly picks up information from the gene pool destroying the "genetic integrity" of the individuals.

$$\boxed{1}_R := \boxed{\mathrm{Random}\{1, \ldots, \mu\}}_P. \tag{6}$$

That is, in contrast to standard crossover techniques in GAs, only one gene is produced by the recombination technique (6) in each step.

After the recombination step, mutation is applied to the recombinant individual in Fig. 1. Its concrete form will be given in the Subsections 3.1.2 and 4.1.2, respectively. For the time being it is only to remark that this process is applied to all individuals in a one-by-one fashion. That is, the individuals keep their individuality, the genes may be changed; however, there is *no* re-sampling. This has been indicated by straight lines leading from the recombinant to the mutant.

In order to close the generation loop, the population must be eventually reduced in case of $\lambda > \mu$. This process is usually attained by fitness based selection which is not considered in this paper. Instead, the individuals $\mu+1, \ldots, \lambda$ are simply dropped. If we ask for the population variance or its expectation σ_M^2 at this point, then it becomes clear that σ_M^2 is the σ_P^2 of the new generation and the loop is closed. Tracing back to the recombinant, we see further that σ_M^2 is constituted by a sampling process which takes μ random samples from the parental population. That is, even in case of $\lambda > \mu$ and selection, it suffices to consider only those individuals that really take part in the reproduction process. In other words, we have an *effective population size* of μ.

2.3 Special Cases

The full-blown $(\mu/\mu_D, \lambda)$-ES without selection as given by Fig. 1 contains two special cases displayed in Fig. 2. In the left picture mutation has been switched off. The ES performs

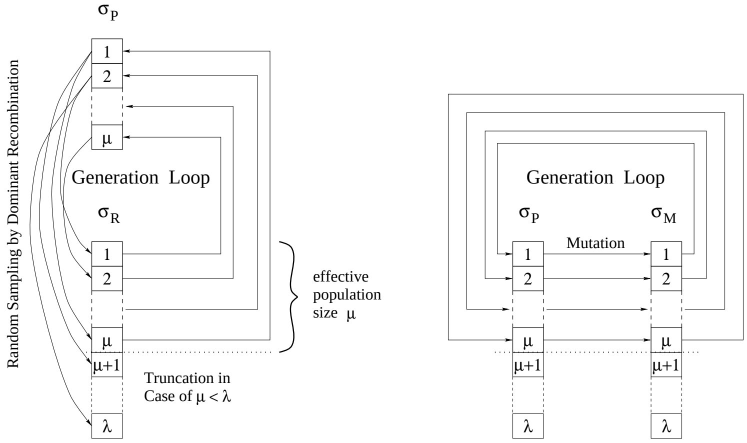


Figure 2: Special cases of the $(\mu/\mu_D, \lambda)$-ES without fitness based selection. Left picture: generation loop performing dominant recombination only. Right picture: generation loop with mutation only.

dominant multiparent recombination only. As we will see later on, this special case exhibits a behavior called *random genetic drift*, known from quantitative genetics and also subject to investigations in GAs (see below).

The right picture of Fig. 2 displays the case without recombination. That is, there is *no* random sampling. Only mutations are applied to the genes. The individuals keep their individuality. As the analysis will show, the long time behavior is diametrally opposite to the "recombination only" case.

2.4 Relations to GA Models

The model in Fig. 1 can be given an alternative interpretation making it equivalent to models investigated in GA theory. In ES theory, selection is in general regarded as a non-uniform selection process which is related to the offspring's/parent's fitness. However, one might as well interpret the recombinative sampling in Fig. 1 as some kind of *selection* which transfers the genetic information from the parents to the offspring.[1] Since the standard recombination operators one-point, n-point (Goldberg, 1989), and uniform crossover (Syswerda, 1989) pick

[1]The author is grateful to Reviewer #8 for pointing him to this important fact.

up two parents and produce two offspring (this is in contrast to dominant multiparent recombination and other standard ES recombination techiques), the allele frequency at a single bit position is necessarily conserved. That is, standard crossover in GAs does not influence the single bit variance dynamics (note, fitness based selection is switched off). Thus, the model in Fig. 1 becomes equivalent to that one investigated in Goldberg and Segrest (1987) by (numerical) Markov chain analysis. From this point of view, the models in Fig. 1 and Fig. 2 can be differentiated with respect to the way how selection is performed. In Fig. 1 and Fig. 2, left picture, we have selection by *random sampling with replacement*, whereas the right picture in Fig. 2 can be regarded as a *deterministic* selection method. The left picture in Fig. 2 is the standard genetic drift model firstly investigated by De Jong (1975). It has received new attention in a paper by Asoh and Mühlenbein (1994).

It should be mentioned that the analysis to be presented here is not restricted to these two selection types (i.e., random sampling with replacement and deterministic selection). The author succeeded in the analytical calculation of the variance dynamics of "random sampling without replacement" (to be published) which has been numerically investigated by Schaffer et al. in this FOGA proceedings (see the respective paper).

3 The MISR-Effect in Unrestricted Search Spaces

This section is devoted to the MISR-effect in unrestricted search spaces, e.g. the $\mathbb{R}^N$ or the $\mathbb{Z}^N$. Section 3.1 contains the derivations, 3.2 the comparisons with experiments, and in 3.3, the asymptotic behavior of the population will be investigated.

3.1 On the Derivation of the Population Variance

The derivation will be performed in three steps. First, the effect of random sampling on σ_P^2 is investigated. This yields the expected population variance of the recombinant population σ_R^2. In the next step the influence of the mutation operator on σ_R^2 will be determined and in the third step, the dynamics of σ_P^2 will be calculated from the evolution equation.

3.1.1 The Effect of Random Sampling

In order to calculate the expected population variance, we start from the definition of the population variance (2) keeping in mind that the μ individual genes are μ random variates

$$\begin{aligned}
\mathrm{Var}\{x\} &= \frac{1}{\mu}\sum_{m=1}^{\mu}(x_m - \langle x\rangle)^2 \;=\; \frac{1}{\mu}\sum_{m=1}^{\mu}\left(x_m^2 - 2x_m\langle x\rangle + \langle x\rangle^2\right) \\
&= \frac{1}{\mu}\sum_{m=1}^{\mu}x_m^2 - 2\langle x\rangle\frac{1}{\mu}\sum_{m=1}^{\mu}x_m + \frac{1}{\mu}\sum_{m=1}^{\mu}\langle x\rangle^2 \;=\; \frac{1}{\mu}\sum_{m=1}^{\mu}x_m^2 - 2\langle x\rangle\cdot\langle x\rangle + \langle x\rangle^2 \\
&= \frac{1}{\mu}\sum_{m=1}^{\mu}x_m^2 - \left(\frac{1}{\mu}\sum_{m=1}^{\mu}x_m\right)^2 \;=\; \frac{1}{\mu}\sum_{m=1}^{\mu}x_m^2 - \frac{1}{\mu^2}\sum_{j=1}^{\mu}\sum_{k=1}^{\mu}x_j x_k. \qquad (7)
\end{aligned}$$

The last line has been obtained by taking definition (1) into account. Now, the double sum is decomposed in such a way that quadratic terms can be collected in the first sum, i.e., the

$j = k$ diagonal elements are removed from the double sum and are put into the first sum

$$\begin{aligned}
\mathrm{Var}\{x\} &= \frac{1}{\mu}\sum_{m=1}^{\mu} x_m^2 - \frac{1}{\mu^2}\sum_{m=1}^{\mu} x_m^2 - \frac{1}{\mu^2}\sum_{j=1}^{\mu}\sum_{k\neq j}^{\mu} x_j x_k \\
&= \frac{\mu-1}{\mu^2}\sum_{m=1}^{\mu} x_m^2 - \frac{1}{\mu^2}\sum_{j=1}^{\mu}\sum_{k\neq j}^{\mu} x_j x_k. \qquad (8)
\end{aligned}$$

Since the $x_j x_k$ products are symmetrical with respect to index exchange and the diagonal elements $j = k$ are excluded in the double sum, Eq. (8) can be further simplified to

$$\mathrm{Var}\{x\} = \frac{\mu-1}{\mu}\frac{1}{\mu}\sum_{m=1}^{\mu} x_m^2 - \frac{2}{\mu^2}\sum_{j=2}^{\mu}\sum_{k=1}^{j-1} x_j x_k. \qquad (9)$$

Now, the expected population variance (3) can be calculated

$$\sigma_R^2[x] = \frac{\mu-1}{\mu}\frac{1}{\mu}\sum_{m=1}^{\mu} \mathrm{E}[x_m^2] - \frac{2}{\mu^2}\sum_{j=2}^{\mu}\sum_{k=1}^{j-1} \mathrm{E}[x_j x_k]. \qquad (10)$$

The random variates x_m, x_j, x_k all belong to the same population. Since there is no fitness based selection, they all have the same probability density function and the random variates are pairwise statistically independent of each other. Therefore, one can write $\mathrm{E}[x_m^2] = \mathrm{E}[x^2] = \overline{x^2}$ and $\mathrm{E}[x_j x_k] = \mathrm{E}[x_j]\mathrm{E}[x_k] = (\mathrm{E}[x])^2 = \overline{x}^2$, where $\overline{x}$ and $\overline{x^2}$ are the first and second moment of the parental population distribution. One obtains from (10)

$$\sigma_R^2[x] = \frac{\mu-1}{\mu}\frac{1}{\mu}\overline{x^2}\sum_{m=1}^{\mu} 1 - \frac{2}{\mu^2}\overline{x}^2\sum_{j=2}^{\mu}\sum_{k=1}^{j-1} 1. \qquad (11)$$

The double sum in (11) can be easily calculated: The sum $\sum_{k=1}^{j-1} = j-1$; it follows

$$\sum_{j=2}^{\mu}\sum_{k=1}^{j-1} 1 = \sum_{j=2}^{\mu}(j-1) = \sum_{j=1}^{\mu-1} j = (\mu-1)\frac{\mu}{2}.$$

Using these results, Eq. (11) can be written as

$$\sigma_R^2[x] = \frac{\mu-1}{\mu}\left[\overline{x^2} - \overline{x}^2\right] = \frac{\mu-1}{\mu}\,\sigma_P^2[x], \qquad (12)$$

where the definition of the variance of a random variate x has been used. As one can see, random sampling in *finite* populations reduces the expected population variance.

3.1.2 The Effect of Mutation

Let us consider a single individual from the recombinant pool in Fig. 1. Its genetic information may be described by the random variate x_R (the subscript "$_R$" stands for the recombinant population). The mutation operator in unrestricted search spaces works usually by adding a random number z. That is, the mutant x_M (the subscript "$_M$" stands for mutant) is obtained by $x_M := x_R + z$. Note, this is in contrast to restricted or low cardinality search spaces (for the binary search space, see Section 4.1.2). The calculation of the x_M-variance is an easy task. From the variance definition one gets

$$\sigma_M^2[x] = \mathrm{E}\left[(x_M - \mathrm{E}[x_M])^2\right] = \overline{(x_M - \overline{x_M})^2}. \tag{13}$$

After the substitution of x_M by $x_R + z$ it follows

$$\begin{aligned} \sigma_M^2[x] &= \overline{\left(x_R + z - \overline{x_R + z}\right)^2} = \overline{\left((x_R - \overline{x_R}) + (z - \overline{z})\right)^2} \\ &= \overline{(x_R - \overline{x_R})^2} + 2\overline{(x_R - \overline{x_R})\,(z - \overline{z})} + \overline{(z - \overline{z})^2}. \end{aligned} \tag{14}$$

Because of the independence of random sampling and mutation, the random variates $(x_R - \overline{x_R})$ and $(z - \overline{z})$ are statistically independent of each other. Therefore, one has $\overline{(x_R - \overline{x_R})(z - \overline{z})} = \overline{(x_R - \overline{x_R})}\ \overline{(z - \overline{z})} = 0 \cdot 0 = 0$ and σ_M^2 becomes

$$\sigma_M^2[x] = \sigma_R^2[x] + \sigma_Z^2, \qquad \text{with} \quad \sigma_Z^2 = \sigma^2[z] = \overline{(z - \overline{z})}, \tag{15}$$

where σ_Z^2 is the variance of the mutations.

3.1.3 The Dynamics of Random Sampling & Mutation

The single effects of random sampling and mutation given by (12) and (15) can be put together. Looking at Fig. 1, one can read that the expected population variance of the parents at generation g, i.e. σ_P^2, is the result of the successive application of random sampling and mutation on the parental distribution at generation $g-1$

$$\sigma_P^{2(g)} = \sigma_Z^2 + \left(1 - \frac{1}{\mu}\right)\sigma_P^{2(g-1)}. \tag{16}$$

Here, the brackets covering the random variates have been omitted just for simplicity reasons. Equation (16) constitutes an iterative scheme of type

$$y^{(g)} = a + b y^{(g-1)} \tag{17}$$

leading to

$$y^{(g)} = a \sum_{j=0}^{g-1} b^j + b^g y^{(0)}. \tag{18}$$

The sum in (18) is a series of geometric progression which has a closed solution

$$y^{(g)} = a\frac{1 - b^g}{1 - b} + b^g y^{(0)}. \tag{19}$$

Identifying a and b in (16) and inserting these in (19) yields finally

$$\boxed{\sigma_P^{2(g)} = \mu\sigma_Z^2\left[1 - \left(1 - \frac{1}{\mu}\right)^g\right] + \left(1 - \frac{1}{\mu}\right)^g \sigma_P^{2(0)}.} \tag{20}$$

This equation describes the time evolution of the expected parental population variance starting from an initial $\sigma_P^{2(0)}$. It was published in (Beyer, 1997) without a derivation. A first derivation can be found in Beyer (1996, in German), however, this early result was obtained under strong restrictions, such as $x \in \mathbb{R}$, z was Gaussian distributed with zero mean, and dominant multirecombination. The derivation presented here shows that the effects to be discussed in the next subsection are not the result of very special assumptions made on the kind of crossover, mutation, or search space chosen.

It is to be remarked that the special case "Mutation Only" (left picture in Fig. 2) is not directly described by Eq. (20), though it can be derived from (20) as we will see in Section 3.3. However, the dynamics can be easily derived from (15) after substituting $\sigma_R^2[x]$ by $\sigma_P^2[x]$, because there is no random sampling. This leads to

$$\sigma_P^{2(g)} := \sigma_P^{2(g-1)} + \sigma_Z^2. \tag{21}$$

Compared with (16) and (17), one reads $b = 1$. This gives with (18) $y^{(g)} = ag + y^{(0)}$ and finally

$$\boxed{\text{"Mutation Only":} \qquad \sigma_P^{2(g)} = g\sigma_Z^2 + \sigma_P^{2(0)}.} \tag{22}$$

3.2 Comparison with Experiments and Discussion

The predictions made by the dynamical Eq. (20) can be verified by simulations. All experiments are performed in N-dimensional continuous search spaces and Gaussian mutations were used. An (effective) population size of $\mu = 300$ has been chosen in order to have the opportunity of displaying the evolution of the population in the search space. For each generation time g, the *empirical* variance is measured as

$$\mathrm{Var}_e\{x\} := \frac{1}{\mu-1}\sum_{m=1}^{\mu}(x_m - \langle x\rangle)^2 \qquad \text{with} \qquad \langle x\rangle := \frac{1}{\mu}\sum_{m=1}^{\mu} x_m. \tag{23}$$

Since Var_e exhibits large fluctuations, it is averaged over the individual's N gene positions,[2] and furthermore, over a number of G independent evolution runs. Figure 3 shows the simulation results of various different experimental scenarios to be discussed below. The results were obtained for $N = 30$, $G = 50$, i.e. each data point has been obtained as the average over $30 \times 50 = 1500$ variance calculations. As one can see, the simulation results are close to the theoretical curves. Let us discuss the observations in detail.

The "Mutation Only" case – the parental genes are deterministically transfered to the offspring pool (i.e., there is no random sampling, cf. the right picture in Fig. 2) – is displayed in the upper picture of Fig. 3. As initial condition, $\sigma_P^{(0)} = 0$ has been chosen, i.e. all parents were concentrated at one point in the search space, $\vec{x} = (0, 0, \ldots, 0)$. The mutation strength is $\sigma_Z = 1$. One observes a straight line. That is, the expected variance of the population increases as a linear function of time g, just as predicted by Eq. (22). Figure 4 shows this behavior in the search space for two dimensions. As one can see, the case "Mutation Only" produces a population which disperses successively in the (unrestricted) search space.

The "No Mutation" case, $\sigma_Z^2 = 0$, where the parental population is varied by random sampling only (cf. the left picture in Fig. 2), is displayed in the right picture of Fig. 3. As recombination operator the one-point crossover has been used with crossover probability $p_c = 1$. In order to observe an effect of the random sampling, there must be an initial

[2] This is allowed because the N gene positions in the individuals are independent of each other.

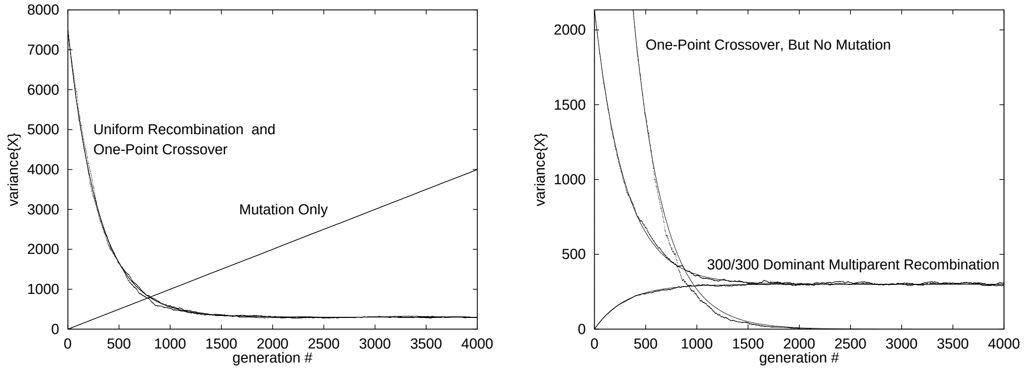
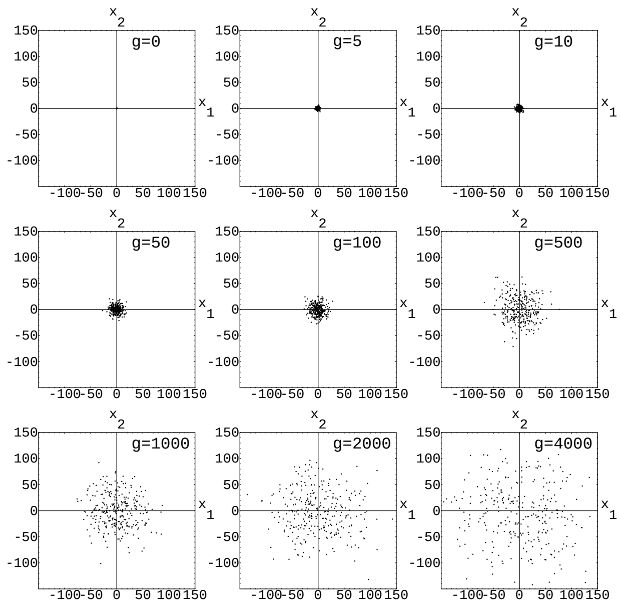


Figure 3: On the time evolution of the average of $\mathrm{Var}_e\{x\}$. The simulation results are close to the theoretical curves obtained from Eq. (20) and Eq. (22), respectively. The left picture shows experiments with uniform and one-point crossover including random sampling as well as an example for an EA with "Mutation Only". The right picture shows simulations with $300/300_D$ dominant multirecombination starting from different initial conditions. Furthermore, a one-point crossover experiment without mutation is displayed.



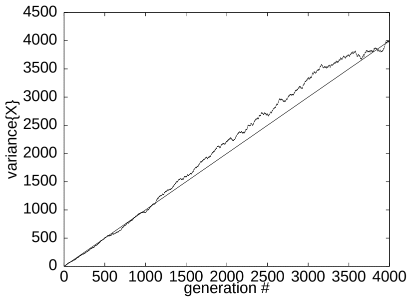
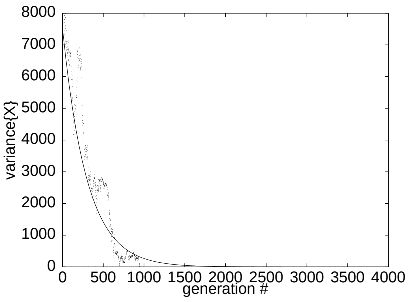


Figure 4: On the effect of "Mutation Only" (random sampling is switched off). The left pictures show the evolution of two coordinates of the population. The picture above the caption displays the actually measured $\mathrm{Var}_e\{x\}$ (no averaging) exhibiting fluctuations around the theoretically expected $\sigma_P^{2(g)}$ displayed as a line.

population distribution. For the experiment each gene position of the $\mu = 300$ individuals was randomly generated from a uniform distribution with $x \in [-150, 150]$, leading to a $\sigma_P^{2(0)} = 300^2/12 = 7500$ (NB: the variance of a uniform variate $x \in [x_l, x_h]$ is given by $\sigma^2[x] = (x_h - x_l)^2/12$). The evolution of the population can be observed in Fig. 5. The population is successively shrinking and is finally fixed at a random position. This phenomenon resembles very much an effect from population genetics (Christiansen & Feldman, 1986), known as *random genetic drift*. This effect will be quantified in Section 3.3.



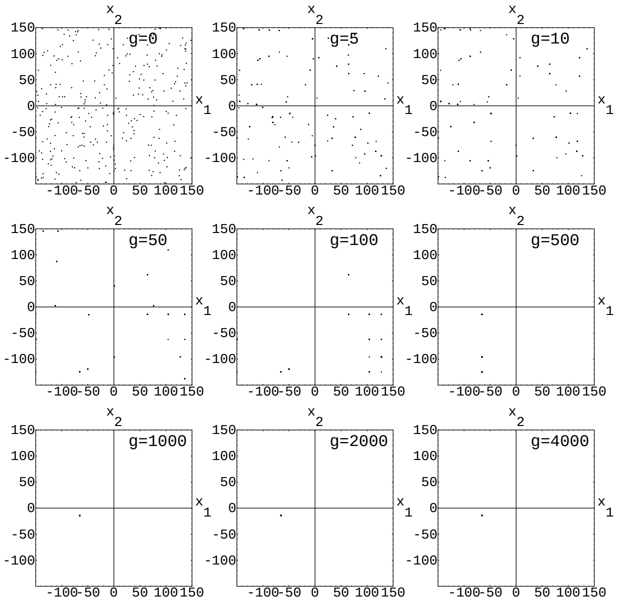
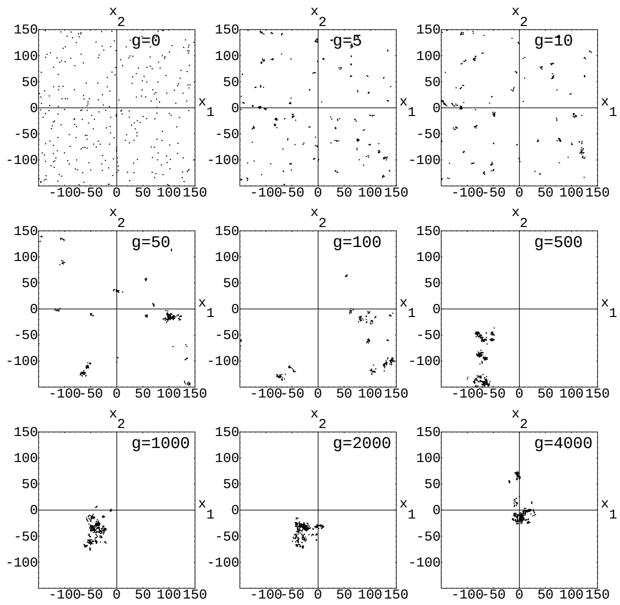


Figure 5: On the effect of random sampling (mutation is switched off). As recombination operator the one-point crossover was used with $p_c = 1$. One observes the collapse of the population into a (random) point. This is the random genetic drift effect. For further explanations, see also Fig. 4.

Let us come back to the left picture in Fig 3. There are two experiments which use mutations, $\sigma_Z^2 = 1$, *and* recombination, $p_c = 1$. Both the one-point crossover and the uniform crossover simulations start from the same initial population which is uniformly distributed over $[-150, 150]$. One observes the same average behavior. One can see – as expected – the standard recombination operators do not have any influence on the dynamics of the single gene variance dynamics. However, most astonishingly, the initially observed shrinking of the population converges for sufficiently large g to a steady state (expected) population variance which is even larger than the variance of the physical mutations ($\sigma_Z^2 = 1$). The analysis (cf. Section 3.3) will show that $\sigma_P^2 \xrightarrow{g \to \infty} \mu \sigma_Z^2$. It is illuminating to watch this behavior in the state space displayed in Fig. 6. The population first shrinks and some kind of clustering occurs, but finally, the population is catched in a cloud. This cloud has a certain stability: From time to time there are random bursts driving some individuals away from the cloud; however, on average they are driven back to the cloud. The cloud moves random-like through the search space.

In order to further investigate this interesting effect two more examples using $300/300_D$ dominant multirecombination, as defined by (6), are considered. The right picture in Fig. 3 displays the average behavior of this strategy for two different initial conditions and fixed mutation strength $\sigma_Z^2 = 1$. Due to space restrictions, plots of the population distribution must be omitted here. The similarities, however, are striking. In all "mutation *and* random sampling" cases one observes the evolution of a stable population cloud, i.e. the individuals are crowded around a center – the average $\langle x \rangle$ – that may be interpreted as wild-type. The effect resembles very much the creation of a *species*. Note, usually one expects speciation as a result of fitness based selection; however, as we can see, the effect of speciation[3] takes

[3]Note, using the wording "random speciation" would be a bit more precise, because the center of mass $\langle x \rangle$ of the population performs a random walk in the search space. From this point of view, fitness based *selection* aims at giving $\langle x \rangle$ a direction in the search space.



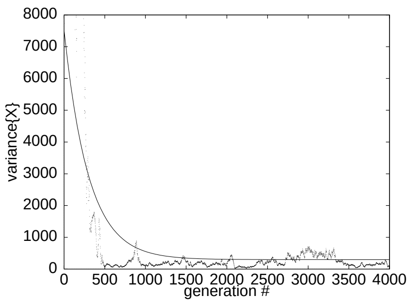


Figure 6: On the effect of random sampling *and* mutation, $\sigma_Z^2 = 1$. The initially large extended population successively shrinks forming a more or less stable cloud which randomly moves through the state space (one-point crossover with $p_c = 1$ used).

always place in EAs as long as both mutation and random sampling are applied. Since this effect is universal, i.e. it does not depend on the search space, the recombination type, and mutation type used, it was proposed in (Beyer, 1995a) as a general working principle of EAs – the MISR-principle (mutation-induced speciation by recombination).[4] In Section 4 it will be shown that it can also be quantified for binary search spaces.

3.3 Asymptotic Behavior

It is the aim of this section to analytically derive the empirically observed behavior from the dynamical Eq. (20). Three scenarios will be discussed:

a) $\sigma_Z^2 > 0$, $\mu < \infty \Longrightarrow$ MISR-effect,

b) $\sigma_Z^2 = 0$, $\mu < \infty \Longrightarrow$ random genetic drift effect,

c) $\sigma_Z^2 > 0$, $\mu \to \infty \Longrightarrow$ diffusion effect.

The special case a) leads to the MISR-effect for sufficient large g. If one investigates the influence of g in Eq. (20) one sees two effects. First, the initial parental distribution with variance $\sigma_P^{2(0)}$ looses more and more influence and fades away for $g \to \infty$. Secondly, the effect of the physical mutations with strength σ_Z^2 is amplified. For $g \to \infty$ it increases up to $\mu\sigma_Z^2$. Thus, one observes a "species" with constant expected population variance $\sigma_\infty^2 := \lim_{g\to\infty} \sigma_P^{2(g)}$

$$\boxed{\text{MISR-effect:} \qquad \sigma_P^2 \xrightarrow{g\to\infty} \mu\sigma_Z^2 \ =: \ \sigma_\infty^2.} \tag{24}$$

[4]For that time, it was conjectured that the MISR effect is related to the recombinative sampling (dominant recombination). However, as we see now, this effect can be observed in *all* EAs with random sampling. Therefore, the "R" in MISR should better stand for "random-sampling".

The time scale at which this speciation takes place can be easily estimated. To this end, let us consider the special case $\sigma_P^{2(0)} = 0$. Convergence up to a relative value $c := \sigma_P^{2(g_c)}/\mu\sigma_Z^2$ gives, inserted into (20), $c\mu\sigma_Z^2 = \mu\sigma_Z^2\left[1 - \left(1 - \frac{1}{\mu}\right)^{g_c}\right]$. Resolved for g_c, one obtains

$$g_c = \frac{\ln(1-c)}{\ln(1-1/\mu)}. \tag{25}$$

For sufficiently large μ the $\ln(1+x) \approx x$ approximation can be applied. Thus, one gets the

$$\text{MISR-CONVERGENCE TIME:} \qquad g_c \approx \mu \ln\left(\frac{1}{1-c}\right), \tag{26}$$

i.e., the convergence time is (nearly) proportional to the effective population size. For the general case $\sigma_P^{2(0)} = k\,\mu\sigma_Z^2$ one obtains a similar convergence time formula

$$g_c \approx \mu \ln\left(\frac{|k-1|}{1-c}\right). \tag{27}$$

The case b) of vanishing mutation strength $\sigma_Z^2 = 0$ leads immediately from (20) to

$$\sigma_P^{2(g)} = \left(1 - \frac{1}{\mu}\right)^g \sigma_P^{2(0)}. \tag{28}$$

Obviously, $g \to \infty$ leads to $\sigma_P^2 = 0$. The population has lost its variance and is concentrated in one point $x = \langle x \rangle$. One observes gene fixation. Resolving (28) for g and approximating the logarithm gives an estimate for the convergence time similar to (26)

$$\text{GENETIC DRIFT CONVERGENCE TIME:} \qquad g \approx 2\mu \ln\left(\frac{\sigma_P^{(0)}}{\sigma_P^{(g)}}\right). \tag{29}$$

Again, this time depends linearly on the effective population size μ.

The case c) is a somewhat special one. The second term in (20) yields $\sigma_P^{2(0)}$ for $\mu \to \infty$, keeping $g < \infty$ fixed. The first term needs further treatment. In order to calculate $\lim_{g\to\infty} \mu\left[1 - \left(1 - \frac{1}{\mu}\right)^g\right]$, $\left(1 - \frac{1}{\mu}\right)^g$ is expanded by the binomial theorem. One obtains

$$\begin{aligned}
\mu\left[1 - \left(1 - \frac{1}{\mu}\right)^g\right] &= \mu\left[1 - \sum_{k=0}^{g}\binom{g}{k}\left(-\frac{1}{\mu}\right)^k\right] \\
&= \mu\left[1 - \left(1 - \frac{g}{\mu} + \sum_{k=2}^{g}\binom{g}{k}\left(-\frac{1}{\mu}\right)^k\right)\right] \\
&= \mu\left[\frac{g}{\mu} - \frac{1}{\mu^2}\sum_{j=0}^{g-2}\binom{g}{j+2}\left(-\frac{1}{\mu}\right)^j\right] \\
&= g - \frac{1}{\mu}\sum_{j=0}^{g-2}\binom{g}{j+2}\left(-\frac{1}{\mu}\right)^j.
\end{aligned} \tag{30}$$

Taking the limit $\mu \to \infty$ for $g < \infty$ yields g and (20) becomes

$$\mu \to \infty : \qquad \sigma_P^{2(g)} = g\sigma_Z^2 + \sigma_P^{2(0)}. \tag{31}$$

However, this is equivalent to the "Mutation Only" case governed by Eq. (22). This is remarkable because it shows that the MISR-effect can be observed in *finite* populations only. It cannot be derived from the analysis of infinite populations. Infinite populations can only exhibit a dispersive population behavior. That is, the effect of mutation is random-walk-like. Indeed, for $\sigma_P^{2(0)} = 0$ we have the $\sqrt{g}$-law for the standard deviation of Brownian particle motion

$$\textsc{Brownian Random Motion:} \qquad \sigma_P^{(g)} = \sqrt{g}\, \sigma_Z. \tag{32}$$

4 The MISR-Effect in Binary GAs

This section is organized in the same manner as Section 3. The first part contains the derivation, 4.2 compares with experiments, and in 4.3 the asymptotic behavior will be investigated.

4.1 On the Derivation of the Population Variance

4.1.1 The Effect of Random Sampling

Although random sampling can be treated by the approach of Section 3.1.1, it will be given a second derivation which takes the peculiarities of the binary search space into account. The population variance is defined by Eq. (4). Using the same calculation as in (7), one obtains

$$\mathrm{Var}\{b\} = \frac{1}{\mu} \sum_{j=1}^{\mu} b_j^2 - \left(\frac{1}{\mu} \sum_{j=1}^{\mu} b_j \right)^2. \tag{33}$$

Since $b_j \in \{0, 1\}$, the sums in (33) are just counting the number of "1". Let n_P be the number of ones in the parental population then the expected population variances, defined by Eq. (5), are given as

$$\sigma_P^2 = \mathrm{E}\left[\frac{n_P}{\mu} - \left(\frac{n_P}{\mu} \right)^2 \right] \qquad \text{and} \qquad \sigma_R^2 = \mathrm{E}\left[\frac{n_R}{\mu} - \left(\frac{n_R}{\mu} \right)^2 \right]. \tag{34}$$

Here, n_P and n_R are random variates $n_P, n_R \in \{0, \ldots, \mu\}$. By random sampling the n_P variate is transformed into n_R. The transition probability is given by the matrix

$$\Pr(n_P \to n_R) = \binom{\mu}{n_R} \left(\frac{n_P}{\mu} \right)^{n_R} \left(1 - \frac{n_P}{\mu} \right)^{\mu - n_R} =: P(n_R | n_P), \tag{35}$$

where n_P/μ is the probability of finding a "1" in the parental population. Equation (35) is the binomial distribution which gives the probability of getting n_R "1" when μ independent samplings are performed.

In order to calculate σ_R^2 in (34) the conditional expectations $\mathrm{E}[n_R|n_P]$ and $\mathrm{E}[n_R^2|n_P]$ are to be determined:

$$\mathrm{E}[n_R|n_P] = \sum_{n_R=0}^{\mu} n_R P(n_R|n_P) \qquad \text{and} \qquad \mathrm{E}[n_R^2|n_P] = \sum_{n_R=0}^{\mu} n_R^2 P(n_R|n_P). \tag{36}$$

This can be easily done by considering the binomial theorem $(a+b)^\mu \equiv \sum_{j=0}^{\mu} \binom{\mu}{j} a^j b^{\mu-j}$. If one calculates the first and second order derivatives $\frac{\partial}{\partial a}$ and $\frac{\partial^2}{\partial a^2}$ one obtains the identities

$$\mu(a+b)^{\mu-1} \equiv \sum_{j=0}^{\mu} j \binom{\mu}{j} a^{j-1} b^{\mu-j} \quad \text{and} \quad \mu(\mu-1)(a+b)^{\mu-2} \equiv \sum_{j=0}^{\mu} j(j-1) \binom{\mu}{j} a^{j-2} b^{\mu-j}.$$

After multiplication with a and a^2, respectively, and setting $b = 1-a$ it follows

$$\mu a = \sum_{j=0}^{\mu} j \binom{\mu}{j} a^j (1-a)^{\mu-j} \qquad \text{and} \qquad \mu(\mu-1)a^2 = \sum_{j=0}^{\mu} j(j-1) \binom{\mu}{j} a^j (1-a)^{\mu-j}. \tag{37}$$

If the left equation is added to the right one, one further obtains

$$\mu(\mu-1)a^2 + \mu a = \sum_{j=0}^{\mu} j^2 \binom{\mu}{j} a^j (1-a)^{\mu-j}. \tag{38}$$

Now, we can identify $a = n_P/\mu$ and $j = n_R$ by considering (36) and (35). Thus, we get

$$\mathrm{E}[n_R|n_P] = n_P \qquad \text{and} \qquad \mathrm{E}[n_R^2|n_P] = n_P + \frac{\mu-1}{\mu} n_P^2. \tag{39}$$

After inserting these results into the right equation of (34) one gets

$$\sigma_R^2 \;=\; \mathrm{E}\left[\frac{n_P}{\mu} - \frac{n_P}{\mu^2} - \frac{\mu-1}{\mu}\frac{n_P^2}{\mu^2}\right] \;=\; \mathrm{E}\left[\frac{\mu-1}{\mu}\frac{n_P}{\mu} - \frac{\mu-1}{\mu}\left(\frac{n_P}{\mu}\right)^2\right].$$

The comparison with the left equation in (34) finally yields

$$\sigma_R^2 \;=\; \frac{\mu-1}{\mu}\,\mathrm{E}\left[\frac{n_P}{\mu} - \left(\frac{n_P}{\mu}\right)^2\right] \;=\; \frac{\mu-1}{\mu}\,\sigma_P^2. \tag{40}$$

This result is in agreement with (12) obtained by an alternative approach.

4.1.2 The Effect of Mutation

Unlike the random sampling, the effect of mutation in restricted search spaces is usually different to unrestricted search spaces with respect to the variance evolution of the population. Mutation in restricted search spaces cannot be treated independently from the individual state. That is, the correlation term in (14) does not vanish and Eq. (15) does not hold. Therefore, one has to consider the state transition matrix of the mutation process.

In binary search spaces mutation is usually performed by flipping each bit with the mutation rate p_m. This bit-flipping is a random process that affects each gene with equal probability and there is no correlation between different genes. This simplifies the analysis considerably, because it suffices to consider the state change of a single gene. Since there are only two possible states “0” and “1” before and after the application of the mutation operator, the transition matrix simply reads

$$P(b_M|b_R) \;=\; \begin{pmatrix} \Pr(\text{“0”} \to \text{“0”}) & \Pr(\text{“1”} \to \text{“0”}) \\ \Pr(\text{“0”} \to \text{“1”}) & \Pr(\text{“1”} \to \text{“1”}) \end{pmatrix} \;=\; \begin{pmatrix} 1-p_m & p_m \\ p_m & 1-p_m \end{pmatrix}. \qquad (41)$$

Here, b_R stands for state of a single gene after recombination and b_M stands for state after the application of the mutation operator. Let us consider the probabilities of being “1”

$$P_M := \Pr(b_M = 1) \qquad \text{and} \qquad P_R := \Pr(b_R = 1), \qquad (42)$$

again, “$_M$” stands for state after the mutation and “$_R$” before the mutation. The transition matrix (41) transforms “$_R$” to “$_M$”:

$$\begin{pmatrix} 1-P_M \\ P_M \end{pmatrix} \;=\; \begin{pmatrix} 1-p_m & p_m \\ p_m & 1-p_m \end{pmatrix} \begin{pmatrix} 1-P_R \\ P_R \end{pmatrix}. \qquad (43)$$

Thus, one gets

$$\begin{aligned} P_M &= p_m\,(1-P_R) + (1-p_m)\,P_R \\ P_M &= p_m + (1-2p_m)\,P_R. \end{aligned} \qquad (44)$$

Now, the variance σ_M^2 can be calculated.
Since $\sigma^2[b] = \mathrm{E}[b^2] - (\mathrm{E}[b])^2$, and $\mathrm{E}[b^k] = \sum_{x=0}^{1} x^k \Pr(b=x) = 0^k \Pr(b=0) + 1^k \Pr(b=1)$, one obtains

$$\sigma_M^2 = P_M - P_M^2 \qquad \text{and} \qquad \sigma_R^2 = P_R - P_R^2. \qquad (45)$$

The insertion of (44) yields

$$\begin{aligned} \sigma_M^2 &= p_m + (1-2p_m)\,P_R - [p_m + (1-2p_m)\,P_R]^2 \\ \sigma_M^2 &= [p_m + (1-2p_m)\,P_R]\,[(1-p_m) + (1-2p_m)\,P_R] \\ \sigma_M^2 &= p_m\,(1-p_m) - p_m\,(1-2p_m)\,P_R + (1-2p_m)(1-p_m)\,P_R - (1-2p_m)^2\,P_R^2 \\ \sigma_M^2 &= p_m\,(1-p_m) - (1-2p_m)^2(P_R - P_R^2). \end{aligned} \qquad (46)$$

For $(1-2p_m)^2$ one can write $(1-4p_m+4p_m^2) = 1-4p_m\,(1-p_m)$ in (46); furthermore,

$$\sigma_Z^2 := p_m\,(1-p_m) \qquad (47)$$

is introduced as the *variance of the mutation operator*. With (45) this finally leads to

$$\sigma_M^2 = \sigma_Z^2 + (1-4\sigma_Z^2)\sigma_R^2. \qquad (48)$$

Note, this formula is different to the unrestricted search space case (15).

4.1.3 The Dynamics of Random Sampling & Mutation

In order to get the evolution equation for the expected population variance, Eq. (40) is inserted into (48). With $\sigma_P^2 = \sigma_P^{2(g-1)}$ and $\sigma_M^2 = \sigma_P^{2(g)}$ one obtains

$$\sigma_P^{2(g)} = \sigma_Z^2 + \left(1-4\sigma_Z^2\right)\left(1-\frac{1}{\mu}\right)\sigma_P^{2(g-1)}. \qquad (49)$$

This equation is of type (17) leading to the closed solution (19). Identifying $a = \sigma_Z^2$ and further $b = (1-4\sigma_Z^2)\left(1-\frac{1}{\mu}\right)$ one gets

$$\sigma_P^{2(g)} \quad = \quad \sigma_Z^2 \frac{1-(1-4\sigma_Z^2)^g\left(1-\frac{1}{\mu}\right)^g}{1-(1-4\sigma_Z^2)\left(1-\frac{1}{\mu}\right)} + (1-4\sigma_Z^2)^g\left(1-\frac{1}{\mu}\right)^g \sigma_P^{2(0)}. \tag{50}$$

This can be alternatively written as

$$\boxed{\sigma_P^{2(g)} \quad = \quad \frac{\mu\sigma_Z^2}{1+(\mu-1)4\sigma_Z^2}\left[1-(1-4\sigma_Z^2)^g\left(1-\frac{1}{\mu}\right)^g\right] + (1-4\sigma_Z^2)^g\left(1-\frac{1}{\mu}\right)^g \sigma_P^{2(0)}.} \tag{51}$$

The case “Mutation Only” can be derived from Eq. (48) setting $\sigma_R^2 = \sigma_P^{2(g-1)}$ and $\sigma_M^2 = \sigma_P^{2(g)}$

$$\sigma_P^{2(g)} = \sigma_Z^2 + (1-4\sigma_Z^2)\,\sigma_P^{2(g-1)}. \tag{52}$$

If one compares (49) and (52) then one sees that $\sigma_P^{2(g)}$ can be obtained from (50) by setting $\left(1-\frac{1}{\mu}\right) = 1$.

$$\boxed{\text{“Mutation Only”:} \qquad \sigma_P^{2(g)} \; = \; \frac{1}{4}\left[1-(1-4\sigma_Z^2)^g\right] + (1-4\sigma_Z^2)^g\,\sigma_P^{2(0)}.} \tag{53}$$

4.2 Comparison with Experiments and Discussion

The predictions of Eq. (51) and (53) have been extensively tested for a large set of GA variants. As examples, four recombination variants will be presented: (a) one-point crossover with $p_c = 0.6$, (b) one-point crossover with $p_c = 1$, (c) uniform crossover with $p_c = 1$, and (d) dominant multiparent recombination, Eq. (6), $p_c = 1$. As population size $\mu = 40$ was chosen. Each individual has a string length $\ell = 100$. The empirical variance $\mathrm{Var}_e\{b\}$ was calculated according to (23) using b_m instead of x_m. Then, this $\mathrm{Var}_e\{b\}$ is averaged over all ℓ string positions and over a number of $G = 50$ independent evolution runs, i.e. for each generation $\ell \times G = 5000$ Var_e values were calculated and averaged.

Unlike the $x \in \mathbb{R}$ continuous gene case, the 2D-visualization of the population in the binary search space is not very illuminating (there are only 4 points). Therefore, only the expected population variance will be depicted in Fig. 7 and 8.

Figure 7 shows simulations which use the one-point crossover (a) with $p_c = 0.6$, and (b) with $p_c = 1$. Figure 8 displays the results for the uniform crossover and the $40/40_D$-dominant multiparent recombination both performed with $p_c = 1$. As one can see, the simulations are very close to the theoretical curves obtained from Eq. (51) and (53). Since all four pictures in Fig. 7 and 8 exhibit nearly the same behavior, it suffices to discuss one of them.

For the “Mutation Only” case a mutation rate $p_m = 0.02$ was used and the population was initialized such that all genes were set to “0”. Since all gene positions are the same, there is no variance in the population $\sigma_P^{2(0)} = 0$. The evolution of this special case (cf. the right picture of Fig. 2) is governed by Eq. (53). The expected population variance increases monotonously with g. However, unlike the unrestricted search space case (cf. Fig. 4),

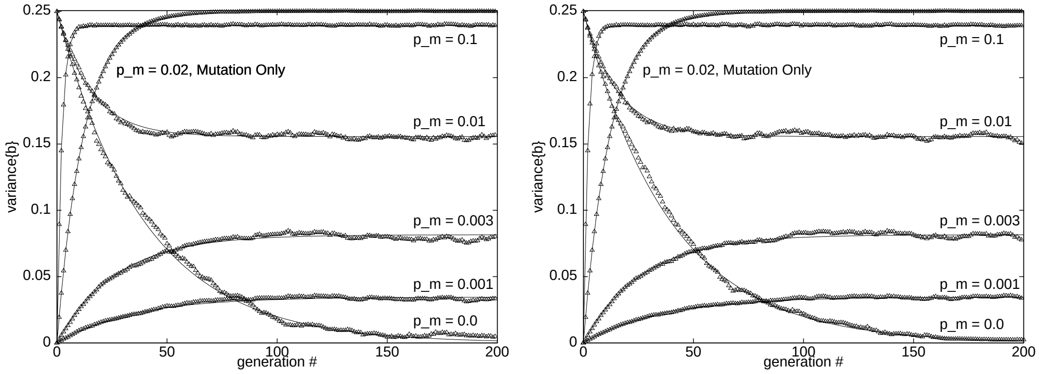


Figure 7: On the time evolution of the average of $\mathrm{Var}_e\{b\}$ for GAs with one-point crossover. Left picture: $p_c = 0.6$. Right picture: $p_c = 1$. The curves are plots of Eq. (51) and (53), respectively, and the "$\triangle$" are the simulation results.

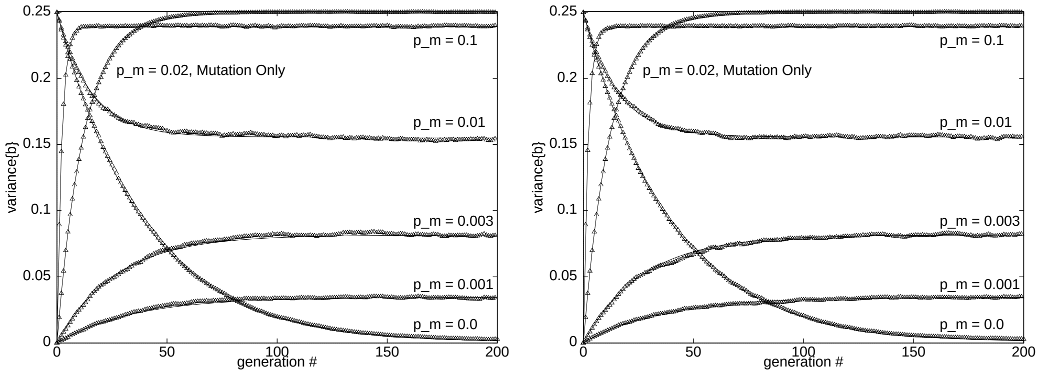


Figure 8: On the time evolution of the average of $\mathrm{Var}_e\{b\}$. Left picture: GA with uniform crossover, $p_c = 1$. Right picture: $40/40_D$-dominant multiparent recombination, $p_c = 1$. The curves are plots of Eq. (51) and (53), respectively, and the "$\triangle$" are the simulation results.

the variance is bounded by 1/4. This variance corresponds to maximal uncertainty in the population:

$$\sigma_P^2 \; = \; \Pr(b_P = 1)\,(1 - \Pr(b_P = 1)) \to \max \quad \Longleftrightarrow \quad \Pr(b_P = 1) = \frac{1}{2}.$$

It is therefore somewhat analogous to the unrestricted search space case where the population disperses in the state space.

The opposite extreme case is given by "$p_m = 0.0$", i.e. the mutation operator is switched off. Initially all genes were randomly initialized $\Pr(b = 1) = 1/2 \Rightarrow \sigma_P^{2(0)} = 1/4$. For $g \to \infty$, the expected population variance becomes

$$\lim_{g \to \infty} \sigma_P^{2(g)} = 0 \qquad \Longleftrightarrow \qquad \forall m: \;\; b_m - \langle b \rangle = 0.$$

That is, random gene fixation has taken place. As in the case of unrestricted search spaces, we observe the *random genetic drift* phenomenon.

The rest of the examples are of the type "Mutation *and* Random Sampling". One observes a good agreement between the simulations and the theoretical curves given by Eq. (51). Obviously, there is some kind of synergism which produces a population with a constant expected population variance. The steady state value does neither depend on the recombination type used nor on the initial conditions. And starting from $\sigma_P^{2(0)} = 0$, one sees that this population distribution is the result of the MISR-effect: physical mutations with mutation rate p_m (NB: $\sigma_Z^2 = p_m(1-p_m)$) are transformed by the random sampling into genetic variance. It should be emphasized once more, one usually expects the speciation effect as a result of fitness based selection; however, as we see the building of a species does *not* necessarily require selection pressure.

4.3 Asymptotic Behavior

The asymptotic behavior of GAs without selection in binary search spaces is similar to that of the unrestricted search space case discussed in Section 3.3. Therefore we will proceed in the same manner.

The MISR-effect is observed for $\sigma_Z^2 > 0$, $\mu < \infty$, and $g \to \infty$. The steady state population variance $\sigma_\infty^2 := \lim_{g\to\infty} \sigma_P^{2(g)}$ is easily obtained from (51)

$$\sigma_\infty^2 = \frac{\mu\sigma_Z^2}{1+4(\mu-1)\sigma_Z^2}, \qquad \text{with} \qquad \sigma_Z^2 = p_m(1-p_m). \tag{54}$$

As one can see, the extension of the population increases with the mutation strength σ_Z of the bit-flipping operator. If $\mu\sigma_Z^2 \ll 1$ then the denominator in (54) can be expanded, this leads to

$$\sigma_\infty^2 \approx \mu\sigma_Z^2\,\left[1-4(\mu-1)\sigma_Z^2\right] \;=\; \mu\sigma_Z^2 + \ldots. \tag{55}$$

That is, for small $\mu\sigma_Z^2$ and therefore for sufficiently small mutations rates p_m, this formula becomes equivalent to the unrestricted search space formula (24).

The convergence time g_d for reaching a certain vicinity of σ_∞^2 can be calculated. Let $\sigma_P^{2(g_d)} := \sigma_\infty^2(1\pm d)$, where the "+" sign holds for the approach from initial values $\sigma_P^{2(0)} > \sigma_\infty^2$ and the "$-$" sign holds for $\sigma_P^{2(0)} < \sigma_\infty^2$. The d value defines the relative deviation from the steady state σ_∞^2. After inserting $\sigma_P^{2(g_d)}$ in (51) and taking (54) into account, one gets

$$\begin{aligned}
(1\pm d)\,\sigma_\infty^2 &= \sigma_\infty^2 + \left(\sigma_P^{2(0)} - \sigma_\infty^2\right)\left[\left(1-4\sigma_Z^2\right)\left(1-\frac{1}{\mu}\right)\right]^{g_d} \\
\ln\frac{\pm d\sigma_\infty^2}{\sigma_P^{2(0)}-\sigma_\infty^2} &= g_d\,\ln\left[\left(1-4\sigma_Z^2\right)\left(1-\frac{1}{\mu}\right)\right] \\
g_d &= \ln\left|\frac{d\sigma_\infty^2}{\sigma_P^{2(0)}-\sigma_\infty^2}\right| \Big/ \ln\left[\left(1-4\sigma_Z^2\right)\left(1-\frac{1}{\mu}\right)\right].
\end{aligned} \tag{56}$$

There is no simple approximation for $g_d = f(\mu, p_m)$, this is in contrast to the case of unrestricted search spaces (cf. Eq. (27)).

The special case "$p_m = 0$" (only random sampling) leads to $\sigma_Z^2 = 0$; therefore Eq. (51) reduces to (28). This is the random genetic drift scenario. The convergence time (29) is nearly a linear function of the population size. It is interesting to notice that in an early work of Goldberg and Segrest (1987) the same linear dependency has been found by numerical investigation of the first passage time of the corresponding Markov chain model. In contrast to that early result, there is now a simple analytical derivation.

The special case "$\mu \to \infty$" immediately leads from (51) to the "Mutation Only" Eq. (53). Its steady state value $\sigma^2_\infty = 1/4$, i.e. for $g \to \infty$, the population is maximally spread over the state space, there is maximum uncertainty and the MISR-effect cannot be observed. Equation (56) can be used for the estimation of the "divergence time" when $\left(1 - \frac{1}{\mu}\right)$ is set to 1 and $\sigma^2_\infty = 1/4$. One sees, *in*finite population models can exhibit a behavior which does *not* reflect the behavior of a *finite* population.

5 Summary

This paper was devoted to the analytical analysis of the variance dynamics of EAs without fitness based selection operating on finite populations. It was the aim of this paper to demonstrate that certain qualitative aspects, such as the random genetic drift and the MISR effect can be observed in any EA. The main message of MISR is that an EA, comprising (at least) a mutation operator and a random sampling operator, transforms (physical) mutations into (larger) population variance. The result looks like a species crowded around an average type ("wild-type") which randomly moves through the state space. If there is selection pressure in the EA then this randomly moving species gets a direction leading to fitter regions of the search space.

The investigations have shown with respect to the MISR, there is not such a big difference between ESs and binary GAs. However, not all possible EA variants are covered by the analysis presented: Steady state EAs and EAs using higher cardinality alphabets, as e.g. genetic programming (GP), remain to be investigated.

Furthermore, although the recombination type used does not influence the MISR-effect, it does not mean that there will not be significant differences when the performance of EAs is evaluated with respect to fitness optimization. Selection introduces some kind of linkage or correlations between different gene positions. This is not covered by the MISR-effect. Therefore, additional principles are needed to explain the working of EAs, such as the *evolutionary progress principle* (EPP) and the *genetic repair* (GR) hypothesis, shortly explained in the introductory section of this paper. Proving their relevance/significance is a task for future research.

6 Acknowledgments

The author is grateful to A. Irfan Oyman for helpful comments. The final version of this paper was very much influenced by comments of Reviewer #8 and by discussions with Ken De Jong, Alex Rogers, and Dave Schaffer. The author is Heisenberg fellow of the DFG under grant Be 1578/4-1.

References

Asoh, H., & Mühlenbein, H. (1994). On the Mean Convergence Time of Evolutionary Algorithms without Selection and Mutation. In Y. Davidor, R. Männer, & H.-P. Schwefel (Eds.), *Parallel Problem Solving from Nature, 3* (pp. 88–97). Heidelberg: Springer-Verlag.

Beyer, H.-G. (1995a). *How GAs do* NOT *Work – Understanding GAs without Schemata and Building Blocks* (Technical Report No. SYS-2/95). University of Dortmund: Department of Computer Science.

Beyer, H.-G. (1995b). Toward a Theory of Evolution Strategies: On the Benefit of Sex – the $(\mu/\mu, \lambda)$-Theory. *Evolutionary Computation*, *3*(1), 81–111.

Beyer, H.-G. (1996). *Zur Analyse der Evolutionsstrategien.* University of Dortmund: Habilitationsschrift.

Beyer, H.-G. (1997). An Alternative Explanation for the Manner in which Genetic Algorithms Operate. *BioSystems*, *41*, 1–15.

Christiansen, F. B., & Feldman, M. W. (1986). *Population Genetics.* Palo Alto, CA: Blackwell Scientific Publications.

De Jong, K. (1975). *An analysis of the behavior of a class of genetic adaptive systems.* Unpublished doctoral dissertation, University of Michigan.

De Jong, K. (1993). Genetic algorithms are not function optimizers. In L. D. Whitley (Ed.), *Foundations of Genetic Algorithms* (pp. 5–17). San Mateo, CA: Morgan Kaufmann.

Goldberg, D., & Segrest, P. (1987). Finite markov chain analysis of genetic algorithms. In J. Grefenstette (Ed.), *Genetic Algorithms and Their Applications: Proc. of the Second Int'l Conference on Genetic Algorithms* (pp. 1–8). Hillsdale, NJ.

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

Grefenstette, J. J. (1991). Conditions for Implicit Parallelism. In G. J. E. Rawlins (Ed.), *Foundations of Genetic Algorithms, 1* (pp. 252–261). San Mateo, CA: Morgan Kaufmann.

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

Mühlenbein, H., & Schlierkamp-Voosen, D. (1994). The science of breeding and its application to the breeder genetic algorithm BGA. *Evolutionary Computation*, *1*(4), 335–360.

Rechenberg, I. (1994). *Evolutionsstrategie '94.* Stuttgart: Frommann–Holzboog Verlag.

Rudolph, G. (1994). Convergence Properties of Canonical Genetic Algorithms. *IEEE Transaction on Neural Networks*, *5*(1), 96–101.

Syswerda, G. (1989). Uniform Crossover in Genetic Algorithms. In J. D. Schaffer (Ed.), *Proc. 3rd Int'l Conf. on Genetic Algorithms* (pp. 2–9). San Mateo, CA: Morgan Kaufmann.