A Revision of the Schema Theorem

Rainer Menke
Univ. Dortmund, FB Informatik, LS II, 44221 Dortmund, Germany
menke@ls2.cs.uni-dortmund.de

December 4, 1997

Abstract

Due to an approximation error the schema theorem implies a wrong estimate for the frequency of instances of a schema. In this article an example is given for which the schema theorem gives a wrong estimate. Based on a modeling which allows a mathematical analysis of genetic algorithms, the schema theorem is revised and a corrected estimate is shown.

1 Introduction

Today evolutionary algorithms and especially genetic algorithms are known to be efficient in many problem domains. But up to now a gap between theoretical foundations and empirical studies about their behavior and quality exists. This is displeasing since they are based on a solid grounding of mathematical theory which had been suppressed as their practical application indicates their benefit on various problem domains. Even if this gap cannot be filled within the next years it seems to be essential to build a solid grounding on which newer and better results can be achieved. In this article the schema theorem and the derived understandings will be revised due to the fact that an assumption is used which cannot be verified.

This article is divided into two parts. The first is dealing with the mathematical modeling of genetic algorithms. For this in subsection 2.1 the used notations and abbreviations are introduced. In subsection 2.2 the components of genetic algorithms are modeled which are genetic objects, genetic operators, populations, selection operators, and replacement strategies. This modeling allows a mathematical analysis of genetic algorithms and is based on the definitions given in [Men96] which is the base document of this article. The restriction of the used sets to be finite in the modeling is not necessary and can be avoided with a greater formal and technical effort. We use this modeling in subsection 2.3 to describe the standard genetic algorithm. Section 3 is dealing with the theoretical foundations of standard genetic algorithms. First in 3.1 the schema theorem, the building block hypothesis, and the explanation of the implicit parallelism are cited to understand the difference to the revised results which are presented in 3.3. By using a simple example we show in 3.2 that the estimate for the schema theorem is wrong. In 3.4 we discuss the results and their implications for the behavior of genetic algorithms. In Section 4 we discuss an alternative genetic algorithm, which is also assumed to be a modeling for which the schema theorem is valid. First we give in 4.1 an example that the estimate of the schema theorem is not correct. In 4.2 we show a correct estimate for the expected frequency of instance of schemata.

2 Modeling of genetic algorithms

2.1 Used symbols

Even if only elementary definitions of probability theory are used it is assumed that the reader is familiar with probability theory. With $E(X)$ we denote the expected value of X, $\wp(X)$ is the power set of X, and $\overline{X}$ is used as an abbreviation for $\mathbb{G} \setminus X$ if $X \subset \mathbb{G}$ where the set $\mathbb{G}$ will be introduced later. As the indicator whether X is true we use the symbol $\mathbb{1}_{(X)}$. Additionally we remember that for $\omega \in \Omega$ a transition matrix $K : \Omega \times \wp(\Omega') \to [0,1]$ can be identified with a random variable of the probability space over Ω' because this identification will be used frequently in this article.

With $\mathcal{F} : \mathbb{G} \to (0,1]$ as a fitness function, $\mathcal{P}_i(x)$ as the frequency of the set x in the i-th population $\mathcal{P}_i$, and

$$\mathcal{F}(\mathcal{P}_t, x) := \frac{\sum_{g \in x} \mathcal{F}(g)\mathcal{P}_t(g)}{\mathcal{P}_t(x)}$$

as the average fitness of elements of x in the population $\mathcal{P}_t$ we will use

$$N_t(x) := \frac{\mathcal{P}_t(x)}{\mathcal{P}_t(\mathbb{G})}$$

as an abbreviation for the relative frequency of elements of x in the population $\mathcal{P}_t$,

$$\mathcal{F}_t(x) := \frac{\mathcal{F}(\mathcal{P}_t, x)}{\mathcal{F}(\mathcal{P}_t, \mathbb{G})}$$

as the relative average fitness of elements of x in the population $\mathcal{P}_t$, and

$$E_{t+1}(x) := \frac{E(\mathcal{P}_{t+1}(x))}{E(\mathcal{P}_{t+1}(\mathbb{G}))}$$

for the expected relative frequency of elements of x in the population $\mathcal{P}_{t+1}$.

2.2 Modeling the components

A genetic object will be interpreted as a random variable $\mathcal{G}$ of a finite probability space over $\mathbb{G} = \{0,1\}^M$, where 0 and 1 will be called alleles. The hyperplanes of $\mathbb{G}$ are called schemata and an element of a schema is one of its instances. A schema S can be interpreted as an element of the set $\{0,1,*\}^M$ where $*$ will be called gene. The positions on which a schema has alleles will be called defining positions. The order of a schema is the number of defining positions and the length is the distance between the rightmost and leftmost defining position. Additionally a fitness function will be assumed to be a function

$$\mathcal{F} : \mathbb{G} \to (0,1]$$

over the search space $\mathbb{G}$.

A genetic operator $\mathcal{O}$ which creates one genetic object from two given will be modeled as a transition matrix

$$\mathcal{O} : \mathbb{G}^2 \times \wp(\mathbb{G}) \to [0,1].$$

If the probability measure μ of a genetic object $\mathcal{G}$ fulfils $\mu(g) = \mathcal{O}(\mathcal{G}_f, \mathcal{G}_m, g)$ for all $g \in \mathbb{G}$ then we denote with

$$\mathcal{G} := \mathcal{O}(\mathcal{G}_f, \mathcal{G}_m)$$

the resultant genetic object $\mathcal{G}$ created from the genetic objects $\mathcal{G}_f$ and $\mathcal{G}_m$. Additionally the minimum probability of a genetic object $\mathcal{G} = \mathcal{O}(\mathcal{G}_f, \mathcal{G}_m)$ to be an element of the set γ'' if $\mathcal{G}_f$ is an element of γ and $\mathcal{G}_m$ of γ' will be called minimum creation probability for the genetic operator $\mathcal{O}$ (for γ, γ' and γ'') and will be denoted with $\mathcal{O}^{\gamma''}_{\gamma,\gamma'}$.

The mutation ξ_{P_M} with the parameter P_M applied on a genetic object $\mathcal{G} = (g_1, \ldots, g_M)$ will result in a genetic object $\mathcal{G}_c = (g'_1, \ldots, g'_M)$ where the probability that g'_i differs from g_i will be P_M for all $i \in \{1, \ldots, M\}$. As a result it can be shown that for any schema S the mutation ξ_{P_M} will have a creation probability of

$$\xi_{P_M}{}^S_S = (1 - P_M)^{o(S)}. \qquad (1)$$

When the one-point-crossing-over χ is applied on two genetic objects $\mathcal{G}_f = (g_1, \ldots, g_M)$ and $\mathcal{G}_m = (g'_1, \ldots, g'_M)$ it will generate offsprings $(g_1, \ldots, g_i, g'_{i+1}, \ldots, g'_M)$ and $(g'_1, \ldots, g'_i, g_{i+1}, \ldots, g_M)$ where i is selected uniformly from $\{1, \ldots, M-1\}$. "To incorporate crossing-over directly [...] one of the resultant structures is discarded"([Hol92] pp.98). Since there are several alternatives for discarding one genetic object we decide that each resultant will be discarded with a probability of 0.5. Based on this modeling the minimum creation probabilities are given by

$$\begin{aligned} \chi^S_{S,\overline{S}} &= \chi^S_{\overline{S},S} = \frac{M - 1 - l(S)}{2(M-1)}, \\ \chi^S_{\overline{S},\overline{S}} &= 0 \text{ and } \chi^S_{S,S} = 1. \end{aligned} \qquad (2)$$

Because it is elementary to show that the equalities $\chi^S_{S,S} = 1$ and $\chi^S_{\overline{S},\overline{S}} = 0$ are valid only the remaining equality is proven. Let us assume that the two genetic objects $\mathcal{G}_f$ and $\mathcal{G}_m$ are given and that their offsprings z and z' are generated as $z := (g_1, \ldots, g_i, g'_{i+1}, \ldots, g'_M)$ and $z' := (g'_1, \ldots, g'_i, g_{i+1}, \ldots, g_M)$. If $\mathcal{G}_f$ is an instance of schema S and $\mathcal{G}_m$ is not one then z will be an instance of S if $i \geq R(S)$. In an analogous way it can be noted that z'

will be an instance of S if $i < L(S)$. By summing up the cases when either z or z' is an instance of schema S we easily see that only $M - R(S) + L(S) - 1 = M - l(S) - 1$ of the $2(M-1)$ possible offsprings can be an instance of S. The last genetic operator we use is the competition $\odot_{P_C}$ with parameter P_C whose resultant $\mathcal{G}_c = \odot_{P_C}(\mathcal{G}_f, \mathcal{G}_m)$ is with a probability of P_C the genetic object $\mathcal{G}_f$ and with the remaining probability $\mathcal{G}_m$. It is trivial to see that

$$\odot_{P_C}{}^{S}_{S,\overline{S}} \;=\; 1 - \odot_{P_C}{}^{S}_{\overline{S},S} = P_C \tag{3}$$

is valid for the minimum creation probability.

Like genetic objects we will interpret a population as a random variable $\mathcal{P}$ of a probability space over $\mathbb{P}$ which is a set of measurable functions from $\mathbb{G}$ to $\mathbb{N}$. Except subsection 3.1 we assume $\mathbb{P}$ to be finite to simplify and reduce the needed effort of modeling genetic algorithms. With the assumption that $\mathbb{P}$ is a finite set we define selection operators as transition matrices

$$\mathcal{S} : \mathbb{P} \times \wp(\mathbb{G}) \to [0,1]$$

which fulfil for $p \in \mathbb{P}$ the restriction $\mathcal{S}(p, \{g\}) = 0$ for all $g \in \mathbb{G}$ with $p(g) = 0$. This restriction is made due to the fact that only genetic objects can be selected which exist in the population. If the probability measure μ of a genetic object $\mathcal{G}$ fulfils $\mu(g) = \mathcal{S}(\mathcal{P}, g)$ for all $g \in \mathbb{G}$ then the selection of a genetic object from the population $\mathcal{P}$ will be denoted with

$$\mathcal{G} := \mathcal{S}(\mathcal{P}).$$

Based on [Hol92] the proportional selection is given by

$$\mathcal{S}_P(\mathcal{P}_t, G) \;:=\; N_t(G)\mathcal{F}_t(G) \tag{4}$$

and the uniform selection by

$$\mathcal{S}_U(\mathcal{P}_t, G) := N_t(G) \tag{5}$$

for all subsets G of $\mathbb{G}$. Also we define replacement strategies as transition matrices

$$\mathcal{E} : \mathbb{P}^2 \times \wp(\mathbb{P}) \to [0,1]$$

which fulfil $\mathcal{E}(p, p', p'') = 0$ for $p, p', p'' \in \mathbb{P}$ if there exists a $g \in \mathbb{G}$ with $p''(g) > p(g) + p'(g)$. We make this restriction to avoid that genetic objects are introduced into the resultant population p'' which do not exist in the given populations p and p'. If the probability measure μ of a population $\mathcal{P}''$ fulfils $\mu(p) = \mathcal{E}(\mathcal{P}, \mathcal{P}', p)$ for all $p \in \mathbb{P}$ then the replacement of the populations $\mathcal{P}$ and $\mathcal{P}'$ will be denoted by

$$\mathcal{P}'' := \mathcal{E}(\mathcal{P}, \mathcal{P}').$$

The replacement strategy which will be used for the standard genetic algorithm is the generational replacement. It is given through

$$\mathcal{E}_G(\mathcal{P}, \mathcal{P}', P) := \mathbb{1}_{(\mathcal{P}' \in P)},$$

so the resultant of the replacement of the populations $\mathcal{P}$ and $\mathcal{P}'$ is always the population $\mathcal{P}'$.

The last and easiest component we use is needed for building populations from genetic objects. We denote with

$$\mathcal{P} = \left\{ \bigotimes_{i=1,\ldots,n} \mathcal{G}_i \right\}$$

the population, which consists of the n genetic objects $\mathcal{G}_1, \ldots, \mathcal{G}_n$.

2.3 Modeling the algorithm

We will now model the standard genetic algorithms by the above formalism. This shows that our formalism is easy to apply. We choose the standard genetic algorithms for our example because the results we are about to derive pertain to them.

It has been noted already that standard genetic algorithms use the generational replacement. Therefore the population

$$\mathcal{P}_{t+1} := \mathcal{E}_G(\mathcal{P}_t, \mathcal{P}'_t) = \mathcal{P}'_t$$

at time $t+1$ is the generational replacement of the current population $\mathcal{P}_t$ and a temporary population

$$\mathcal{P}'_t := \left\{ \bigotimes_{i=1,\ldots,n} \mathcal{G}_{(t,i)} \right\}$$

which consists of n independent identically distributed genetic objects $\mathcal{G}_{(t,1)}, \ldots, \mathcal{G}_{(t,n)}$. To generate the genetic object

$$\mathcal{G}_{(t,i)} := \xi_{P_M}\left(\mathcal{G}'_{(t,i)}\right)$$

we mutate a temporary genetic object $\mathcal{G}'_{(t,i)}$ with a mutation parameter P_M. The genetic object

$$\mathcal{G}'_{(t,i)} := \odot_{P_C}\left(\mathcal{G}''_{(t,i)}, \mathcal{G}'''_{(t,i)}\right)$$

is the competition with parameter P_C of the genetic object

$$\mathcal{G}'''_{(t,i)} = \mathcal{S}_P\left(\mathcal{P}_t\right)$$

which we select with the proportional selection from the current population $\mathcal{P}_t$ and

$$\mathcal{G}''_{(t,i)} = \chi\left(\mathcal{S}_P\left(\mathcal{P}_t\right), \mathcal{S}_U\left(\mathcal{P}_t\right)\right) \tag{6}$$

which is the resultant of the crossing-over of genetic objects selected with the proportional selection and uniform selection from the population $\mathcal{P}_t$.

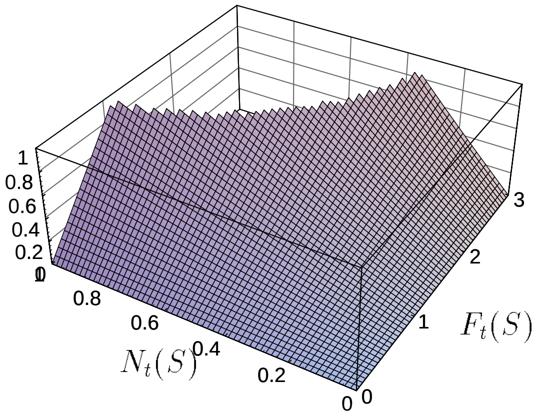

Figure 1: Estimate (7) in dependency to $\mathcal{F}_t(S)$ and $N_t(S)$

3 Theoretical foundations

3.1 Standard genetic algorithms

Genetic algorithms are based on reproductive plans for which Theorem 6.2.3 in [Hol92] – which is known as the schema theorem – gives a lower bound for the frequency of instances of schemata. For the proof of the theorem the assumption that "no more than a proportion $\frac{(1-N_t(S))l(S)P_C}{M-1}$ of the modified offspring of S can be expected to be instances of schemata other than S"([Hol92] pp.102) is used. If this assumption is correct the given estimate

$$N_{t+1}(S) \;\geq\; (1-P_M)^{o(S)}\, N_t(S)\mathcal{F}_t(S)\left(1-\frac{P_C l(S)}{M-1}(1-N_t(S))\right) \tag{7}$$

– which is valid for infinite populations only – can be generalized to finite populations in the way that the expected frequency for instances of a schema S in the successor population $\mathcal{P}_{t+1}$ would be

$$E_{t+1}(S) \;\geq\; (1-P_M)^{o(S)}\, N_t(S)\mathcal{F}_t(S)\left(1-\frac{P_C l(S)}{M-1}(1-N_t(S))\right). \tag{8}$$

The importance of estimates (7) and (8) is not the opportunity to calculate a lower bound for the frequency of instances for certain schemata. It is more important that from the estimate (7) understandings are derived which try to describe, explain and predict the optimization process of genetic algorithms.

One explanation is the building block hypothesis which says that "just as a child creates magnificent fortress through the arrangement of simple blocks, so does a genetic algorithm seek near optimal performance through the juxtaposition of short, low-order, high performance schemata, or building blocks"([Gol89], p.41). Another explanation is the implicit parallelism, which is an $O(n^3)$ estimate meaning "that despite the processing something of only n structures each generation, a genetic algorithm processing something like n^3 schemata in parallel with no special bookkeeping or memory other than the population itself"([Gol89],p.40).

3.2 Falsification for the schema theorem

As a motivation for the revision of the schema theorem we show that the estimate given by the schema theorem can be incorrect for certain populations.

Proposition 1: It exists a finite population $\mathcal{P}_t \in \mathbb{P}$ so that the expected frequency of instances $E_{t+1}(S)$ of a schema S in the successor population $\mathcal{P}_{t+1}$ fulfills

$$E_{t+1}(S) \;<\; (1-P_M)^{o(S)}\, N_t(S)\mathcal{F}_t(S)\left(1-\frac{P_C\, l(S)}{M-1}(1-N_t(S))\right).$$

Proof: Let us assume that we have a population $\mathcal{P}_t$ consisting of the genetic object $g=(0,\ldots,0)$ and $n-1$ copies of the genetic object $g'=(1,\ldots,1)$. We assume that the fitness function fulfills $\mathcal{F}(g)=1$ and $\mathcal{F}(g')=\epsilon>0$ with $\epsilon<1$. It follows that

$$\mathcal{F}_t(g)=\frac{\mathcal{F}(\mathcal{P}_t,g)}{\mathcal{F}(\mathcal{P}_t,\mathbb{G})}=\frac{n}{(n-1)\epsilon+1}.$$

If we regard the schema $S=(*,0,\ldots,0)$ we easily achieve that $l(S)=M-2$ and $o(S)=M-1$. With the assumption of $P_M=0$ and $P_C=1$ estimate (8) can be simplified to

$$\begin{aligned}
E_{t+1}(S) &\geq (1-P_M)^{o(S)}\, N_t(S)\mathcal{F}_t(S)\left(1-\frac{P_C\, l(S)}{M-1}(1-N_t(S))\right)\\
&= \frac{1}{n}\frac{n}{(n-1)\epsilon+1}\left(1-\frac{M-2}{M-1}\frac{n-1}{n}\right)=\frac{2}{n}\frac{1}{(n-1)\epsilon+1},
\end{aligned}$$

if we choose $M=n$ additionally.
Using $\chi(g,g,g)=1-\chi(g',g',g)=1$ and $\chi(g,g',g)=\chi(g',g,g)=\frac{1}{2(n-1)}$ we are able to show that

$$\begin{aligned}
E_{t+1}(S) &= \sum_{a\in\mathbb{G}}\left(\sum_{b\in\mathbb{G}}(\chi(a,b,S)\mathcal{S}_P(\mathcal{P}_t,a)\mathcal{S}_U(\mathcal{P}_t,b))\right)\\
&= \mathcal{S}_P(\mathcal{P}_t,g)\mathcal{S}_U(\mathcal{P}_t,g)\chi(g,g,g)+\mathcal{S}_P(\mathcal{P}_t,g')\mathcal{S}_U(\mathcal{P}_t,g')\chi(g',g',g)\\
&\quad+\mathcal{S}_P(\mathcal{P}_t,g')\mathcal{S}_U(\mathcal{P}_t,g)\chi(g',g,g)+\mathcal{S}_P(\mathcal{P}_t,g)\mathcal{S}_U(\mathcal{P}_t,g')\chi(g,g',g)\\
&= N_t(g)\mathcal{F}_t(g)N_t(g)\\
&\quad+(1-N_t(g)\mathcal{F}_t(g))N_t(g)\frac{1}{2(n-1)}+N_t(g)\mathcal{F}_t(g)(1-N_t(g))\frac{1}{2(n-1)}\\
&= \frac{1}{n}\frac{1}{(n-1)\epsilon+1}
\end{aligned}$$

$$
\begin{aligned}
&+\left(1-\frac{1}{(n-1)\epsilon+1}\right)\frac{1}{n}\frac{1}{2(n-1)}+\frac{1}{(n-1)\epsilon+1}\left(1-\frac{1}{n}\right)\frac{1}{2(n-1)}\\
=\;& \frac{2}{2n}\frac{1}{(n-1)\epsilon+1}+\frac{\epsilon}{(n-1)\epsilon+1}\frac{1}{2n}+\frac{1}{(n-1)\epsilon+1}\frac{1}{2n}\\
=\;& \frac{3+\epsilon}{2n}\frac{1}{(n-1)\epsilon+1}<\frac{2}{n}\frac{1}{(n-1)\epsilon+1},
\end{aligned}
$$

because the population $\mathcal{P}_t$ consists only of the genetic objects g' and $n-1$ copies of g'. □

3.3 Revised theoretical foundations

Based on the mathematical formalism introduced in Section 2 we will derive a estimate for the standard genetic algorithm denoted in Subsection 2.3. To do so we first show the following lemma:

Lemma 2: The expected frequency for a genetic object $\mathcal{G}=\mathcal{O}(\mathcal{S}(p),\mathcal{S}'(p))$ to be element of a set G will be at least

$$
\begin{aligned}
E^*(\mathbb{1}_{(\mathcal{G}\in G)}) \;:=\;& \mathcal{O}^{G}_{\overline{G},\overline{G}}+\mathcal{S}(p,G)\left(\mathcal{O}^{G}_{G,\overline{G}}-\mathcal{O}^{G}_{\overline{G},\overline{G}}\right)+\mathcal{S}'(p,G)\left(\mathcal{O}^{G}_{\overline{G},G}-\mathcal{O}^{G}_{\overline{G},\overline{G}}\right)\\
&+\mathcal{S}(p,G)\mathcal{S}'(p,G)\left(\mathcal{O}^{G}_{G,G}-\mathcal{O}^{G}_{\overline{G},\overline{G}}\right)+\mathcal{S}(p,G)\mathcal{S}'(p,G)\left(\mathcal{O}^{G}_{\overline{G},\overline{G}}-\mathcal{O}^{G}_{\overline{G},G}\right)
\end{aligned}
\tag{9}
$$

where $\mathcal{S}$ and $\mathcal{S}'$ denote selection operators and $\mathcal{O}$ a genetic operator.

Proof: A proof is given in the following way:

$$
\begin{aligned}
E(\mathbb{1}_{(\mathcal{G}\in G)}) \;=\;& \sum_{a\in\mathbb{G}}\left(\sum_{b\in\mathbb{G}}\left(\mathcal{O}(a,b,G)\mathcal{S}(p,a)\mathcal{S}'(p,b)\right)\right)\\
\geq\;& \mathcal{S}(p,G)\mathcal{S}'(p,G)\mathcal{O}^{G}_{G,G}+(1-\mathcal{S}(p,G))(1-\mathcal{S}'(p,G))\mathcal{O}^{G}_{\overline{G},\overline{G}}\\
&+(1-\mathcal{S}(p,G))\mathcal{S}'(p,G)\mathcal{O}^{G}_{\overline{G},G}+\mathcal{S}(p,G)(1-\mathcal{S}'(p,G))\mathcal{O}^{G}_{G,\overline{G}}\\
=:\;& E^*(\mathbb{1}_{(\mathcal{G}\in G)}),
\end{aligned}
$$

because $\mathcal{O}^{\gamma''}_{\gamma,\gamma'}$ denotes the minimum probability of creating an element of γ'' with the genetic operator $\mathcal{O}$ if γ and γ'' are given. □

Because we now have a simple estimate for one genetic object, we use it to achieve an estimate for the successor population:

Theorem 3: The expected frequency of instances of a schema S in the successor population for the standard genetic algorithm as described in subsection 2.3 is

$$
\begin{aligned}
E_{t+1}(S) \;\geq\;& (1-P_M)^{o(S)}\,N_t(S)\left(P_C\frac{M-1-l(S)}{2(M-1)}+\mathcal{F}_t(S)\chi_t(S)\right)\\
=\;& (1-P_M)^{o(S)}\,N_t(S)\mathcal{F}_t(S)\left(1-\frac{P_C l(S)}{M-1}(1-N_t(S))\right)\\
+\;& (1-P_M)^{o(S)}\,N_t(S)P_C\frac{M-1-l(S)}{2(M-1)}(1-\mathcal{F}_t(S))
\end{aligned}
\tag{10}
$$

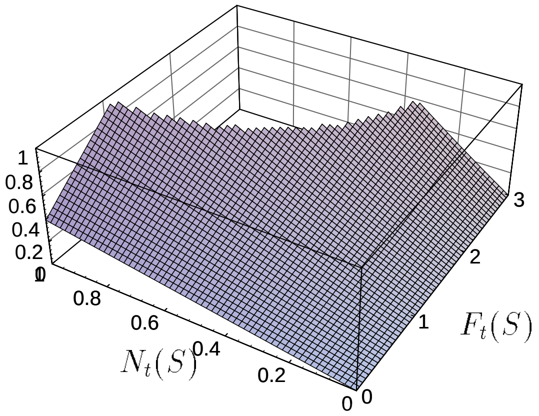

Figure 2: Dependency of estimate (10) to $\mathcal{F}_t(S)$ and $N_t(S)$

with $\chi_t(S) := 1 - P_C \frac{M-1+l(S)(1-2N_t(S))}{2(M-1)}$.

Proof: If the successor population $\mathcal{P}_{t+1}$ is generated from independent identically distributed genetic objects $\mathcal{G}_1, \ldots, \mathcal{G}_n$ it follows that

$$E_{t+1}(S) = \frac{E(\mathcal{P}_{t+1}(S))}{E(\mathcal{P}_{t+1}(\mathbb{G}))} = \frac{\sum_{i=1}^{n} E(\mathbb{1}_{(\mathcal{G}_i \in S)})}{n} = E(\mathbb{1}_{(\mathcal{G}_1 \in S)}) \geq E^*(\mathbb{1}_{(\mathcal{G}_1 \in S)})$$

so it remains to show the estimate for the genetic object

$$\begin{aligned} \mathcal{G} &= \xi_{P_M}(\mathcal{G}') \\ \text{with } \mathcal{G}' &= \odot_{P_C}(\mathcal{G}'', \mathcal{G}'''), \\ \mathcal{G}'' &= \chi\left(\mathcal{S}_P(\mathcal{P}_t), \mathcal{S}_U(\mathcal{P}_t)\right), \\ \text{and } \mathcal{G}''' &= \mathcal{S}_P(\mathcal{P}_t). \end{aligned}$$

For estimating $E(\mathbb{1}_{(\mathcal{G}'' \in S)})$ we use the lemma 2, the estimates (2) and the estimates for the uniform (5) and proportional selection operators (4) and achieve

$$\begin{aligned} E(\mathbb{1}_{(\mathcal{G}'' \in S)}) &= \chi^{S}_{\overline{S},\overline{S}} + \mathcal{S}_P(\mathcal{P}_t, S)\left(\chi^{S}_{S,\overline{S}} - \chi^{S}_{\overline{S},\overline{S}}\right) + \mathcal{S}_U(\mathcal{P}_t, S)\left(\chi^{S}_{\overline{S},S} - \chi^{S}_{\overline{S},\overline{S}}\right) \\ &\quad + \mathcal{S}_P(\mathcal{P}_t, S)\mathcal{S}_U(\mathcal{P}_t, S)\left(\chi^{S}_{S,S} - \chi^{S}_{S,\overline{S}}\right) + \mathcal{S}_P(\mathcal{P}_t, S)\mathcal{S}_U(\mathcal{P}_t, S)\left(\chi^{S}_{\overline{S},\overline{S}} - \chi^{S}_{\overline{S},S}\right) \\ &= (N_t(S))^2 \mathcal{F}_t(S)\left(1 - \frac{M-1-l(S)}{M-1}\right) + N_t(S)\left(\mathcal{F}_t(S)+1\right)\frac{M-1-l(S)}{2(M-1)} \end{aligned}$$

and for the genetic object $\mathcal{G}'''$ we see

$$E(\mathbb{1}_{(\mathcal{G}''' \in S)}) = \mathcal{S}_P(\mathcal{P}_t, S) = N_t(S)\mathcal{F}_t(S).$$

Using these results and the estimate (3) we are able to show that

$$E(\mathbb{1}_{(\mathcal{G}' \in S)}) = P_C E(\mathbb{1}_{(\mathcal{G}'' \in S)}) + (1-P_C)\, E(\mathbb{1}_{(\mathcal{G}''' \in S)}) = N_t(S)\left(P_C \chi^{S}_{S,\overline{S}} + \mathcal{F}_t(S)\chi_t(S)\right)$$

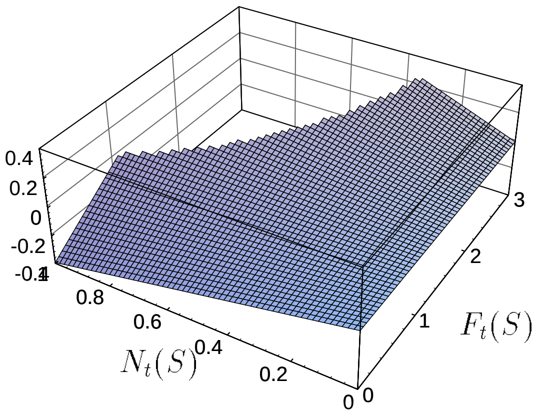

Figure 3: $(1-P_M)^{o(S)} N_t(S) P_C \frac{M-1-l(S)}{2(M-1)} (1-\mathcal{F}_t(S))$ in dependency to $\mathcal{F}_t(S)$ and $N_t(S)$

is valid and using the equality (1) we easily see

$$E(1\!\!1_{(\mathcal{G}\in S)}) \;=\; E\left(1\!\!1_{(\mathcal{G}\in S\wedge\mathcal{G}'\in S)} + 1\!\!1_{\left(\mathcal{G}\in S\wedge\mathcal{G}'\in\overline{S}\right)}\right) \geq \xi_{P_M}{}^{S}_{S} E\left(1\!\!1_{(\mathcal{G}'\in S)}\right) = (1-P_M)^{o(S)}\, E\left(1\!\!1_{(\mathcal{G}'\in S)}\right)$$

which completes the proof. □

For the calculation of the lower bounds of (8) and (10) in the Figures 1 and 2 the constants $(1-P_M)^{o(S)} = P_C = 0.95$, $M = 40$, $l(S) = 5$ and the elementary but important equality

$$\mathcal{F}(\mathcal{P}_t, \mathbb{G}) \;=\; \mathcal{F}(\mathcal{P}_t, G) N_t(G) + \mathcal{F}(\mathcal{P}_t, \overline{G}) N_t(\overline{G}) \tag{11}$$

are used. In the Figures 3 and 4 the difference between the estimates is computed and visualized to emphasize the understanding.

If $\mathcal{F}_t(S)$ is smaller than 1 the estimate (10) will be greater than (8). This will not imply problems to the analysis of genetic algorithms because the estimate of the schema theorem denotes a lower bound for the expected frequency for instances of schemata. Nevertheless the graph of (10) describes the effects of the one-point-crossing-over and mutation in genetic algorithms. First there is for almost all schemata a positive probability of being explored in the successor population even if the average fitness of their instances is nearly 0. Second the frequency of instances will be decreasing for the schemata with an average fitness beyond the average fitness of the population.

But in the case when $\mathcal{F}_t(S)$ is greater than 1 the corrected estimate (10) will be smaller than (8). This leads to problems in the analysis and understanding of genetic algorithms due to the fact that schemata with above average fitness will not have the attraction in the optimization process of genetic algorithms as proposed in the literature. The difference will not only be the asymptotic limits when $\mathcal{F}_t(S)$ diverges against infinty. It is more important that the understandings based on the schema theorem are influenced by this behavior. So up to now the author is not able to prove either the implicit parallelism of order $O(n^3)$ as observed in [Gol89] and [BD93] or the building block hypothesis whose proofs use the incorrect calculation of $\chi^S_{S,\overline{S}}$ and $\chi^S_{\overline{S},S}$.

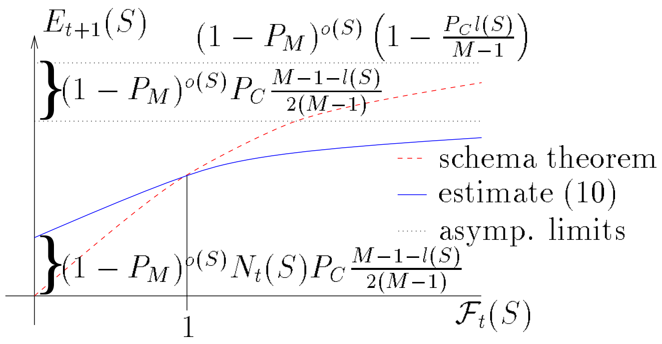

Figure 4: Visualisation of the dependency between $E_{t+1}(S)$ and $\mathcal{F}_t(S)$

Even if it is unclear if these understandings can be proven it is the author's opinion that they point in the correct direction for the understanding of genetic algorithms. But to be usable it is needed that some assumptions are made and that they are interpreted with the correct abstraction.

For the building block hypothesis it has to be remarked that the building blocks neither are visible to the genetic algorithm so that they can be used in the optimization process nor can they be separated from the applied genetic operators and selection operators. Another argument against the building block hypothesis raises from the fact that the schemata which can be combined to form better schemata do not have to exist. So the building block hypothesis "isn't incorrect, if taken only as an empirical observation from the optimization process. But if the hypothesis is regarded as a rational for how the GA processes solutions then cause and effect are reversed"([Bey97] p.3). Additionally it has to be remarked that the building block hypothesis has a trapdoor due to three elementary facts:

- Schemata with an order of $o(s)$ are descriptions of hyperplanes with $2^{M-o(s)}$ elements. But the more elements a hyperplane has the closer its (static) average fitness will be to 1. Regarding this with the dynamics of genetic algorithms it is clear that the average fitness of a schema of a population will converge faster to 1 the lower its order is. But a schema being usable for the building block hypothesis should have an average fitness greater than 1 and due to this its order must be great.

- Even schemata with great length can have low order so that the length of a schema cannot be as important for the building block hypothesis as its order can be. Although like the order the length has to be great to describe schemata with above average fitness because only in this case the corresponding hyperplanes will have few elements.

- The equality (11) implies that even the average fitness of schemata which are described by the building block hypothesis changes and this will influence their importance for the building block hypothesis.

The misunderstanding with the so-called implicit parallelism is based on the fact that every genetic object is an instance of 2^M schemata and that there exist 3^M schemata. "Leaving aside that the n^3-counting argument is questionable, implicit parallelism is certainly not unique to GAs. The counting argument is mainly based upon the number of schemata owned by an individual string of length M"([Bey97] p.3). Also the processing of n^3 schemata can be misunderstood that a genetic algorithm is more effective the greater its populations are. But there is not only the upper limit of 3^M schemata which can be processed although there is no knowledge if genetic algorithms are able to reduce the number of schemata being processed dramatically when the genetic algorithm proceeds. This would be a more sophisticated indicator for an efficient working of genetic algorithms on schemata.

3.4 Explanation for the behavior of genetic algorithms

It is the author's opinion that the optimization process of genetic algorithms is based on antagonistic effects and aspects and that the quality of genetic algorithms depends on how well these can be combined. In a more general way it could be said that "the behavior of an evolving population is governed by a complicated interplay of a few major genetic and population forces"([vNCM97] p.8) and the way they are allowed to interact with each other. These forces are the genetic drift, the selection and genetic operators, and last but not least the replacement strategies.

Even if we are not able to describe the effects of these interacting forces for all genetic algorithms in detail we can do so for standard genetic algorithms. For them the generational replacement is one of the weakest forces because this replacement strategy is unable to reject any genetic object which will be created or selected from the current population. For the selection operators we have to separate the proportional from the uniform selection due to their different effects. The proportional selection increases the frequence of fitter genetic objects. Therefore the population will converge to a homogeneous population which consists of the best genetic object that has been found. In contrast to this, the uniform selection increases the frequence of frequent genetic objects such that the current population will converge to a homogeneous population which consists of the most frequent genetic object in the current population. But only if the currently best genetic object is also the currently most frequent genetic object the two selection operators will yield the same homogenous population.

The mutation and crossing-over are forces that drive genetic mixing between genetic objects. The effect of the crossing-over of the parents selected by proportional and uniform selection operators respectively will be that the created offspring will have the opportunity of having above average fitness and frequent alleles. In contrast to this the mutation which is applied on proportionally selected genetic objects only will increase the exploration of regions with above average fitness. These effects are static behaviors but are also depending on the crossing-over-rate P_C. In short term it can be noted that the crossing-over is a mechanism for merging different qualities and implies that the population will converge to homogeneity and in contrast to this the mutation is a mechanism for establishing inhomogeneity and implies that the population keeps exploring the search space.

The genetic drift is a result of sampling fluctuations induced by the finite size of the population. It is the sum of not well understood random fluctuations which will be stronger when the other forces – especially the replacement strategy – are weaker and therefore the population will tend to premature convergence.

4 Variant of genetic algorithms

Because there are several articles in which it is assumed that offsprings are generated from two proportionally selected parents, we regard this special case. For these genetic algorithms we compare the estimate of the schema theorem with the expected frequency for instances of a schema and apply lemma 2 for estimating an lower bound. Before going further, we first modify the genetic algorithms described in subsection 2.3 to our purpose. As explained before everything remains unchanged except (6) which has to be replaced by

$$\mathcal{G}''_{(t,i)} \quad = \quad \chi\left(\mathcal{S}_P\left(\mathcal{P}_t\right), \mathcal{S}_P\left(\mathcal{P}_t\right)\right). \tag{12}$$

4.1 Estimate of the schema theorem

Like in subsection 3.2 we first want to show, that the schema theorem gives wrong estimates for the expected frequency of instances of schemata in certain cases. For this we use the following proposition.

Proposition 4: It exists a finite population $\mathcal{P}_t \in \mathbb{P}$ so that the expected frequency of instances $E_{t+1}(S)$ of a schema S in the successor population $\mathcal{P}_{t+1}$ of the previous described genetic algorithm fulfills

$$E_{t+1}(S) \quad < \quad (1-P_M)^{o(S)}\, N_t(S)\mathcal{F}_t(S)\left(1-\frac{P_C\, l(S)}{M-1}\,(1-N_t(S))\right).$$

Proof: Let us assume that we have a population $\mathcal{P}_t$ consisting of 100 genetic objects; 80 copies of the genetic object $g = (0,\ldots,0)$ and 20 copies of the genetic object $g' = (1,\ldots,1)$. We assume that the fitness function fulfills $\mathcal{F}(g) = 1/6$ and $\mathcal{F}(g') = 1$. It follows that

$$\mathcal{F}_t(g) = \frac{\mathcal{F}(\mathcal{P}_t, g)}{\mathcal{F}(\mathcal{P}_t, \mathbb{G})} = \frac{1}{2}.$$

If we regard the schema $S = (*,0,\ldots,0)$ we easily achieve that $l(S) = M-2$ and $o(S) = M-1$. With the assumption of $P_M = 0$ and $P_C = 1$ estimate (8) can be simplified to

$$\begin{aligned} E_{t+1}(S) \quad &\geq \quad (1-P_M)^{o(S)}\, N_t(S)\mathcal{F}_t(S)\left(1-\frac{P_C\, l(S)}{M-1}\,(1-N_t(S))\right) \\ &= \quad \frac{4}{10}\left(1-\frac{M-2}{M-1}\frac{2}{10}\right) = \frac{4}{10}\frac{8M+6}{10(M-1)} = \frac{32M+24}{100(M-1)}. \end{aligned}$$

Using $\chi(g,g,g) = 1 - \chi(g',g',g) = 1$ and $\chi(g,g',g) = \chi(g',g,g) = \frac{1}{2(M-1)}$ we are able to show that

$$\begin{aligned}
E_{t+1}(S) &= \sum_{a\in\mathbb{G}}\left(\sum_{b\in\mathbb{G}}(\chi(a,b,S)\mathcal{S}_P(\mathcal{P}_t,a)\mathcal{S}_P(\mathcal{P}_t,b))\right)\\
&= \mathcal{S}_P(\mathcal{P}_t,g)\mathcal{S}_P(\mathcal{P}_t,g)\chi(g,g,g) + \mathcal{S}_P(\mathcal{P}_t,g')\mathcal{S}_P(\mathcal{P}_t,g')\chi(g',g',g)\\
&\quad +\mathcal{S}_P(\mathcal{P}_t,g')\mathcal{S}_P(\mathcal{P}_t,g)\chi(g',g,g) + \mathcal{S}_P(\mathcal{P}_t,g)\mathcal{S}_P(\mathcal{P}_t,g')\chi(g,g',g)\\
&= N_t(g)\mathcal{F}_t(g)N_t(g)\mathcal{F}_t(g) + 2(1-N_t(g)\mathcal{F}_t(g))N_t(g)\mathcal{F}_t(g)\frac{1}{2(M-1)}\\
&= \frac{16}{100} + \frac{24}{100(M-1)} = \frac{16M+8}{100(M-1)} < \frac{32M+24}{100(M-1)}
\end{aligned}$$

because of the definition of the population $\mathcal{P}_t$ and the assumption $M > 1$. □

4.2 Revised estimate

As for the genetic algorithm described in subsection 2.3 we show a lower bound for the expected frequency for instances of a schema.

Theorem 5: The expected frequency of instances of a schema S in the successor population for the genetic algorithm assumed in this subsection is

$$\begin{aligned}
E_{t+1}(S) &\geq (1-P_M)^{o(S)}\, N_t(S)\mathcal{F}_t(S)\left(1 - \frac{P_C\, l(S)}{M-1}(1 - N_t(S)\mathcal{F}_t(S))\right) \qquad (13)\\
&= (1-P_M)^{o(S)}\, N_t(S)\mathcal{F}_t(S)\left(1 - \frac{P_C\, l(S)}{M-1}(1 - N_t(S))\right)\\
&+ (1-P_M)^{o(S)}\, N_t(S)^2\mathcal{F}_t(S)\frac{P_C\, l(S)}{M-1}(\mathcal{F}_t(S) - 1)
\end{aligned}$$

Proof: The proof of this theorem is mostly identical to the proof of theorem 3. If the successor population $\mathcal{P}_{t+1}$ is generated from independent identically distributed genetic objects $\mathcal{G}_1, \ldots, \mathcal{G}_n$ it follows that $E_{t+1}(S) \geq E^*(\mathbb{1}_{(\mathcal{G}_1\in S)})$. So it remains to show the estimate for the genetic object

$$\begin{aligned}
\mathcal{G} &= \xi_{P_M}(\mathcal{G}')\\
\text{with } \mathcal{G}' &= \odot_{P_C}(\mathcal{G}'', \mathcal{G}'''),\\
\mathcal{G}'' &= \chi(\mathcal{S}_P(\mathcal{P}_t), \mathcal{S}_P(\mathcal{P}_t)),\\
\text{and } \mathcal{G}''' &= \mathcal{S}_P(\mathcal{P}_t).
\end{aligned}$$

For estimating $E(\mathbb{1}_{(\mathcal{G}''\in S)})$ we use the lemma 2, the estimates (2) and the estimates for the proportional selection operators (4) and achieve

$$E(\mathbb{1}_{(\mathcal{G}''\in S)}) = \chi^S_{\overline{S},\overline{S}} + \mathcal{S}_P(\mathcal{P}_t, S)\left(\chi^S_{S,\overline{S}} - \chi^S_{\overline{S},\overline{S}}\right) + \mathcal{S}_P(\mathcal{P}_t, S)\left(\chi^S_{\overline{S},S} - \chi^S_{\overline{S},\overline{S}}\right)$$

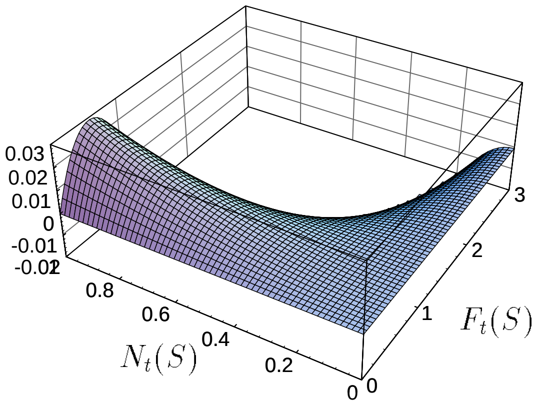

Figure 5: $(1-P_M)^{o(S)}\, N_t(S)^2\mathcal{F}_t(S)\frac{P_C l(S)}{M-1}\,(\mathcal{F}_t(S)-1)$ in dependency to $\mathcal{F}_t(S)$ and $N_t(S)$

$$
\begin{aligned}
& +\mathcal{S}_P(\mathcal{P}_t,S)\mathcal{S}_P(\mathcal{P}_t,S)\left(\chi^{S}_{S,S}-\chi^{S}_{S,\overline{S}}\right)+\mathcal{S}_P(\mathcal{P}_t,S)\mathcal{S}_P(\mathcal{P}_t,S)\left(\chi^{S}_{\overline{S},\overline{S}}-\chi^{S}_{\overline{S},S}\right) \\
= \;& N_t(S)^2\mathcal{F}_t(S)^2\left(1-2\frac{M-1-l(s)}{2(M-1)}\right)+N_t(S)\mathcal{F}_t(S)\left(2\frac{M-1-l(s)}{2(M-1)}\right) \\
= \;& N_t(S)\mathcal{F}_t(S)\left(\frac{M-1-l(S)}{M-1}+N_t(S)\mathcal{F}_t(S)\frac{l(S)}{M-1}\right) \\
= \;& N_t(S)\mathcal{F}_t(S)\left(1-\frac{l(S)(1-N_t(S)\mathcal{F}_t(S))}{M-1}\right)
\end{aligned}
$$

and for the genetic object $\mathcal{G}'''$ we know $E(\mathbb{1}_{(\mathcal{G}'''\in S)})=N_t(S)\mathcal{F}_t(S)$. Using these results and the estimate (3) we are able to show that

$$
\begin{aligned}
E(\mathbb{1}_{(\mathcal{G}'\in S)}) &= P_C E(\mathbb{1}_{(\mathcal{G}''\in S)})+(1-P_C)\,E(\mathbb{1}_{(\mathcal{G}'''\in S)}) \\
&= N_t(S)\mathcal{F}_t(S)\left(1-\frac{P_C l(S)(1-N_t(S)\mathcal{F}_t(S))}{M-1}\right)
\end{aligned}
$$

is valid and using the equality (1) we easily see

$$
\begin{aligned}
E(\mathbb{1}_{(\mathcal{G}\in S)}) &= E\left(\mathbb{1}_{(\mathcal{G}\in S\wedge\mathcal{G}'\in S)}+\mathbb{1}_{\left(\mathcal{G}\in S\wedge\mathcal{G}'\in\overline{S}\right)}\right)\geq \xi_{P_M}{}^{S}_{S}E\left(\mathbb{1}_{(\mathcal{G}'\in S)}\right) \\
&= (1-P_M)^{o(S)}\,N_t(S)\mathcal{F}_t(S)\left(1-\frac{P_C l(S)(1-N_t(S)\mathcal{F}_t(S))}{M-1}\right)
\end{aligned}
$$

which completes the proof. □

For the visualisation of the difference between the estimate 8 of the schema theorem and the estimate (13) in the Figure 5 the constants $(1-P_M)^{o(S)}=P_C=0.95$, $M=40$, $l(S)=5$ and the equality (11) are used. To emphase their different behaviour a schematic visualisation is given in Figure 6 additionally.

Even if for these algorithms the schema theorem is only greater than the estimate (13) if $\mathcal{F}_t(S)$ is smaller than 1 it's the authors opinion, that the criticism given in subsection

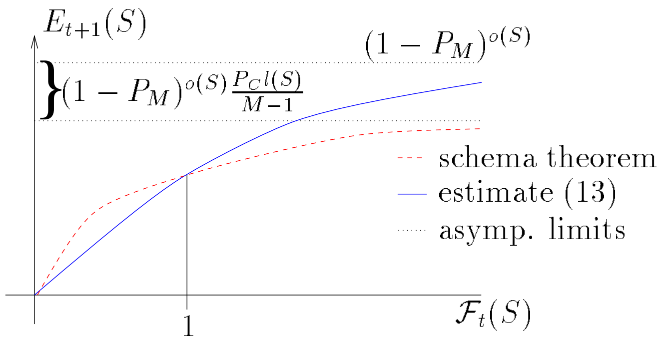

Figure 6: Visualisation of the dependency between $E_{t+1}(S)$ and $\mathcal{F}_t(S)$

3.3 are correct. This is mainly based on the fact that the building block hypothesis is a static description for a dynamic effect and that the implicit parallelism is a description for an effect which occur in each parallel (search-)algorithm.

5 Conclusions

By using equality (9) it is not hard to show that a greater lower bound than (10) exists although an analog estimate to equality (9) could be used for calculating an upper bound for the frequency of instances of a schema. But as a matter of fact the exacter the estimate is the complexer the estimate would be. Due to this we evade exacter estimates and present a simpler estimate pointing into the right direction for the revision of the schema theorem and the understandings of its derived statements. Additionally it is possible to calculate with equality (9) the frequency of instances of schemata when other genetic algorithms are used. But this is out of the scope of this article. The same is valid for the case when genetic operators are used which generate more than one offspring or use more than two parent individuals or when selection operators are used which select more than one genetic object at one time. Even if the presented modeling could easily be extended to these cases the modeling is omitted in this article but is grasped in [Men96].

6 Acknowledgments

I would not have been able to write this article if I had not got the help and support I received. So I want to thank the department of computer science of the University Dortmund especially Dr. Hans-Georg Beyer, Stefan Droste, Thomas Jansen, Dr. Günter Rudolph, Prof. Dr.-Ing. Hans-Paul Schwefel, and Prof. Dr. Ingo Wegener for encouraging me to write this article and last but not least for the discussions we had.

References

[BD93] A. Bertoni and M. Dorigo. Implicit parallelism in genetic algorithms. *Artificial Intelligence*, 61(2):307–314, 1993.

[Bey97] H.-G. Beyer. An alternative explanation for the manner in which genetic algorithms operate. *BioSystems*, 41:1–15, 1997.

[Gol89] D.E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley, 1989.

[Hol92] J. H. Holland. *Adaption in Natural and Artificial Systems.* MIT Press, 1992.

[Men96] R. Menke. Genetische Algorithmen mit Anwendung zur frequenzausschöpfenden Kanalzuweisung im Mobilfunknetz. Master's thesis, Westfälische Wilhelmsuniversität Münster, March 1996. Available by anonymous ftp as menke96.ps.gz in /pub/GA/papers at ftp://lumpi.informatik.uni-dortmund.de.

[vNCM97] E. van Nimwegen, J.P. Crutchfield, and M. Mitchell. Statistical dynamics of the royal road genetic algorithm. Technical Report 97-04-035, Santa Fe Institute, 1997.