

## A Central Limit Theorem for the Population Process of Genetic Algorithms

Stefan Voget\*

*Institut für Mathematik, Universität Hildesheim,  
Marienburger Platz 22, D-31141 Hildesheim, Germany*

**Abstract.** A genetic algorithm (GA) is a stochastic search and optimization algorithm that works by iterative application of several evolutionary operators on populations of solutions. We introduce a central limit theorem for the population process when population size grows. The theorem approximates a GA by a continuous gaussian process. This leads to a numerical method for the examination of the algorithm. In some simple examples we present applications of the method.

### 1. Introduction

A genetic algorithm (GA) adapts principles of natural evolution to combinatorial optimization problems. The concept was introduced in 1975 in [10]. Since that time GAs have been applied to a lot of problems [5, 8]. A GA works by iterative application of operations such as selection, crossover, and mutation. The aim is to propagate a population of possible solutions such that an optimal, or at least a very good, solution can grow in an evolutionary process.

In the first step the members of the population, called individuals, are reproduced, contributing to the next generation according to their fitness. For maximization tasks the fitness may be the optimization function value. For minimization tasks the fitness is calculated by transformations of the optimization function. Crossover, which is viewed as the most essential factor of GAs, combines individuals by exchanging parts of them. Finally, the mutation mechanism changes random parts of the individuals with small probability. The most inherent feature of GA optimization lies in its attempt to explore the search space in a global fashion by studying the entire population.

Although the algorithm is quite successful in applications, the mathematical theory of the operators selection, mutation, and crossover still seems incomplete. In the past, no sufficient theoretical answers have been found for questions such as the following.

---

\*Electronic mail address: [voget@informatik.uni-hildesheim.de](mailto:voget@informatik.uni-hildesheim.de).

- How should parameters be chosen?
- How do the parameters interconnect?
- What crossover should be used?
- How does the population look after finite many steps?

We introduce an approximation method from the theory of quantitative population genetics of biomathematics. In section 2 we present notations and a mathematical model that is used for the description of the algorithm. This model is often approximated by models with infinite population size [18, 28].

In [9] and [12], a diffusion approximation model originally introduced for biomathematics in [13] is used. They focus on the equilibrium state when time becomes infinite.

We adapt another approach introduced in [19, 20] and employed in [17] for biomathematics. With this approach we are able to give a diffusion model for finite population sizes and a finite number of steps.

The main result is described in section 3. It is a central limit theorem for when the population size tends to infinite. Using this method, the discrete population process can be replaced by a continuous gaussian process. Such processes are widely examined. Some simple applications are given in section 4. The main problem in the application of the method lies in the calculation of the mean function of the gaussian process. The applications given in this paper allow the mean function to be calculated analytically. Numerical and structural results for the general case are presented in another paper [29].

## 2. Basic algorithm formulation

Let  $K \in \mathbb{N}$  and  $\Xi_k$  ( $k = 1, \dots, K$ ) be finite sets. The optimization problem is given with the search space  $\Xi = \prod_{k=1}^K \Xi_k$  and an optimization function  $f : \Xi \rightarrow \mathbb{R}^+$  by

$$\arg \max_{i \in \Xi} f(i).$$

By  $N$  we denote the population size of the GA. A population of size  $N$  is defined by a vector  $x^{(N)} = (x_i^{(N)})_{i \in \Xi}$  with the properties

$$x_i^{(N)} \in \left\{0, \frac{1}{N}, \dots, \frac{N-1}{N}, 1\right\}, \quad \sum_{i \in \Xi} x_i^{(N)} = 1.$$

The set of all populations of size  $N$  is denoted by  $P^{(N)} := \{x^{(N)} \mid x^{(N)} \text{ population of size } N\}$ . Further, we introduce populations of infinite size. A population with infinite size is defined by a vector  $x = (x_i)_{i \in \Xi}$  with the properties

$$x_i \in [0, 1], \quad \sum_{i \in \Xi} x_i = 1.$$

The set of all populations of size infinity is denoted by  $P := \{x \mid x \text{ population}\}$ . Notice that  $P^{(N)} \subset P$  for all  $N \in \mathbb{N}$ . Members of a population are called individuals. The coordinate of the population vector with the index  $i$  denotes the relative proportion of the individual  $i$  in the population  $x$ . A GA is given by iterative application of Steps 1, 2, and 3.

**Step 1:** Selection. Given a population  $x(n)$  in a generation with number  $n \in \mathbb{N}$ , members of the population are selected independently. Individuals with higher fitness are preferred for the evolution process. Individuals with low fitness get a lesser chance for selection. The distribution of a randomly chosen individual  $I^S(n)$  after selection is given by

$$\forall i \in \Xi : P(I^S(n) = i \mid X(n) = x(n)) := \frac{\sigma(f(i)) * x_i(n)}{\sum_{j \in \Xi} \sigma(f(j)) * x_j(n)},$$

with a selection function  $\sigma : \mathbb{R} \rightarrow \mathbb{R}^+$ . The selection function is necessary for a scaling of the fitness function. This kind of selection is usually called proportional selection. Discussion of this and other kinds of selection can be found elsewhere [1, 2, 6, 7, 14, 15, 16].

**Step 2:** Crossover. After the independent selection of two individuals, crossover is applied. With a probability  $p_c$  the two parent vectors are mated for the production of one offspring. With the complementary probability  $1 - p_c$  the individuals are not changed, and one of the two parent vectors is taken as the offspring. A crossover schema is given by a random vector  $U$  with values in  $\{0, 1\}^K$ . By  $\bar{u}$  we denote the bitwise complement of  $u$  and by  $\odot$  and  $\oplus$  we denote the componentwise multiplication and summation. The  $k$ th coordinate of  $U$  assigns the parent from which the  $k$ th coordinate is inherited. The distribution of a randomly produced individual  $I^C(n)$  is defined by

$$\begin{aligned} P(I^C(n) = i \mid I^{S_1}(n) = j, I^{S_2}(n) = k) \\ := \frac{1 - p_c}{2} * (\mathbb{1}_{\{i=j \vee i=k\}}) + p_c * P((j \odot U) \oplus (k \odot \bar{U}) = i). \end{aligned}$$

By  $\mathbb{1}_{\{ \cdot \}}$  we denote the indicator function. Provided that a coordinate of  $U$  takes the value 1, that coordinate of the offspring  $i$  is taken from the first selected individual  $I^{S_1}(n) = j$ . If a coordinate of  $U$  takes the value 0, that coordinate is taken from the other parent vector  $I^{S_2}(n) = k$ . In this model two parent vectors produce one offspring. Often two offspring are produced, but the second offspring depends stochastically on the first. This leads to another, more complicated model.

### Example 1.

1. *One-point crossover.* This kind of crossover divides a parent string into two substrings. The substrings of two parents are combined into one new individual, the offspring. In a mathematical description  $U$  has uniform distribution on the set

$$\{(0, \dots, 0, 1), (0, \dots, 0, 1, 1), (0, \dots, 0, 1, 1, 1), \dots, (1, 1, \dots, 1)\},$$

that is, the right substring is taken from  $I^{S_1}(n)$  and the left substring is taken from  $I^{S_2}(n)$ .

2. *Two-point crossover.* One block of  $I^{S_1}(n)$  is inherited and the remainder of the coordinates are taken from  $I^{S_2}(n)$ , that is,  $U$  is uniformly distributed on the set

$$\begin{aligned} &\{(0, \dots, 0, 0, 1), (0, \dots, 0, 0, 1, 0), \dots, (1, 0, 0, \dots, 0), \\ &\quad (0, \dots, 0, 1, 1), (0, \dots, 0, 1, 1, 0), \dots, (1, 1, 0, \dots, 0), \\ &\quad \vdots, \dots, \vdots \\ &\quad (1, \dots, 1)\}. \end{aligned}$$

3. *Uniform crossover.* Each coordinate is taken independently of the other coordinates with probability  $1/2$  from  $I^{S_1}(n)$ . With  $u \in \{0, 1\}^K$  and  $|u| := \sum_{k=1}^K u_k$ , the counting measure of  $U$  is given by

$$P(U = u) = \frac{1}{2^K}.$$

Discussion and further examples of crossover are given elsewhere [3, 4, 11, 21, 22, 24].

Given a population  $x(n)$ , it is not possible to produce all elements of  $\Xi$  with crossover. Crossover only mixes up the coordinates of individuals. It is not possible to produce coordinate values not present in  $x(n)$ . Therefore, the necessity of another production operator, mutation, is obvious.

**Step 3:** Mutation. The coordinates of an individual  $j$  are altered independently with a small probability  $\mu$ . By  $i_k$  we denote the  $k$ th coordinate of the individual  $i$ . If mutation occurs in the  $k$ th coordinate, the value of the coordinate is changed to one of the  $|\Xi_k| - 1$  remaining values. Each of the values is taken with equal probability. The distribution of a randomly mutated individual  $I^M(n)$  is given by

$$P(I^M(n) = i | I^C(n) = j) = \prod_{k=1}^K (1 - \mu)^{1_{\{i_k = j_k\}}} * \left( \frac{\mu}{|\Xi_k| - 1} \right)^{1_{\{i_k \neq j_k\}}}.$$

If  $i_k$  is not equal to  $j_k$ , mutation has to occur in the  $k$ th coordinate and has to produce the value  $i_k$ . Otherwise the coordinate must not be changed.

**Step 4:** Production of the next generation. The next generation is produced by reiteration of the first three steps.  $N$  individuals are produced with selection, crossover, and mutation. Because the iteration occurs independently, the new population is given by a polynomially distributed random vector  $X^{(N)}(n+1)$ . With the help of the abbreviation  $p_M(n, i) := P(I^M(n) = i | X(n) = x(n))$ , the distribution of the next population is given by

$$P_{N * X^{(N)}(n+1) | X(n) = x(n)} = M \left( N, (p_M(n, i))_{i \in \Xi} \right). \quad (1)$$

### 3. A central limit theorem

The model of section 2 gives an exact description of a simple GA with selection, crossover, and mutation. In this section we present an approximation model in which population size converges to infinity. Three parameters (selection pressure, probability of crossover, and probability of mutation) simultaneously converge to zero. This allows control of the connection between parameters and the population process such that a central limit theorem can be given. Let  $(\epsilon_N)_{N \in \mathbb{N}}$  be a sequence of real values fulfilling

$$\epsilon_N > 0, \quad \epsilon_N \longrightarrow 0 \quad \text{and} \quad N * \epsilon_N \longrightarrow \infty \quad (2)$$

for  $N \rightarrow \infty$ . By dependence of  $\epsilon_N$  the selection function is restricted to

$$\sigma(i) = 1 + \epsilon_N * f(i)$$

for all  $i \in \Xi$ . The probability of mutation from an individual  $i$  to an individual  $j$  is restricted to

$$p_{j \rightarrow i}^M := P(I^M = i \mid I^C = j) = \begin{cases} \epsilon_N * \mu_{j \rightarrow i} & j \neq i \\ 1 - \epsilon_N * \sum_{k \in \Xi, k \neq i} \mu_{i \rightarrow k} & i = j \end{cases},$$

and the probability of crossover is restricted to

$$p_c = \epsilon_N * \hat{p}_c$$

with  $0 \leq \mu_{i \rightarrow j} \leq 1$  and  $0 \leq \hat{p}_c \leq 1$ . The parameters should be chosen such that the probabilities are in the interval  $[0, 1]$  and  $\sigma$  takes values greater than zero. This is fulfilled if low enough values of  $(\epsilon_N)_{N \in \mathbb{N}}$  are chosen. The following two lemmas describe changes of the expectation and the covariance of a population after one step of the GA in the limit  $N \rightarrow \infty$ . The changes of the expectation are described by the following definition.

**Definition 1.** Define the *population mean function*

$$m : \begin{cases} P & \longrightarrow \mathbb{R}^\Xi \\ x & \longmapsto (m_i(x))_{i \in \Xi} \end{cases}$$

by

$$\begin{aligned} m_i(x) &:= x_i * \left( f(i) - \sum_{j \in \Xi} f(j) * x_j \right) \\ &\quad + \hat{p}_c * \sum_{j,k} x_j * x_k * P(I^C = i \mid I^{S_1} = j, I^{S_2} = k) - \hat{p}_c * x_i \\ &\quad + \sum_{j \in \Xi} \mu_{j \rightarrow i} * x_j - x_i. \end{aligned}$$

The population mean function  $m$  consists of three main parts. Each part describes the influence of one isolated operator. The first summand arises from the selection. It consists of a distance of the function value from the expectation. The part of the second line arises from crossover. It describes a distance of the expected distribution of an individual after crossover to the one before crossover. The part of the third line arises from mutation. It describes the changes from before to after mutation. The name of  $m$  is explained by the following lemma.

**Lemma 1.** Let  $n \in \mathbb{N}$ ,  $i \in \Xi$ , and  $(\epsilon_N)_{N \in \mathbb{N}}$  be a sequence of real values fulfilling equation (2)

$$\lim_{N \rightarrow \infty} \frac{1}{\epsilon_N} E \left( X_i^{(N)}(n+1) - X_i^{(N)}(n) \mid X^{(N)}(n) = x(n) \right) = m_i(x(n)).$$

*Proof.* See the appendix. ■

A similar property for the covariance function of the population vector is given by the following definition.

**Definition 2.** Let the *population covariance function*

$$v : \begin{cases} P & \longrightarrow \mathbb{R}^{\Xi \times \Xi} \\ x & \longmapsto v_{i,j}(x) \end{cases}$$

be defined by

$$v_{i,j}(x) := x_i * (\delta_{ij} - x_j),$$

where  $\delta_{ij}$  stands for the Kronecker delta function,

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & \text{else.} \end{cases}$$

With  $v$  we are able to give a limit theorem for the covariance of two individuals in the population vector after one step of the GA for  $N \rightarrow \infty$ .

**Lemma 2.** Let  $n \in \mathbb{N}$  and  $i, j \in \Xi$ ,

$$\lim_{N \rightarrow \infty} N * \text{cov} \left( X_i^{(N)}(n+1), X_j^{(N)}(n+1) \mid X^{(N)}(n) = x(n) \right) = v_{i,j}(x(n)).$$

*Proof.* See the appendix. ■

We also need a measure for the comparison of two stochastic processes.

**Definition 3.** A sequence  $(Y^{(N)}(n))_{n, N \in \mathbb{N}}$  of probability variables is called *asymptotically equally distributed* as a sequence  $(Z^{(N)}(n))_{n, N \in \mathbb{N}}$  (in signs:  $Y^{(N)}(n) \sim Z^{(N)}(n)$ ) of probability variables, if for all bounded continuous real-valued functions  $F$  and all constants  $T \in \mathbb{N}$

$$\max_{n \leq T/\epsilon_N} | E [F(Y^{(N)}(n))] - E [F(Z^{(N)}(n))] | \longrightarrow 0$$

for  $N \rightarrow \infty$ .

Now we formulate the main result, a central limit theorem for the population process defined by a GA.

**Theorem 1.** *Let  $x(0) \in P$  be a population and  $z : \mathbb{R}^+ \rightarrow P$  be the unique solution of the differential equation system*

$$\frac{dz}{dt}(t) = m(z(t)) \quad (3)$$

with the condition  $z(0) = x(0)$ . Let a matrix  $K$  with  $x \in P$  be given by

$$[K(x)]_{i,j} := \frac{\partial m_i}{\partial x(j)}(x).$$

Let  $(\epsilon_N)_{N \in \mathbb{N}}$  be a sequence fulfilling equation (2). With  $n \in \mathbb{N}$  and  $t := \epsilon_N * n$ , the probability vector

$$Z^N(n) := \sqrt{N * \epsilon_N} * [X^{(N)}(n) - z(t)]$$

is asymptotically equally distributed as a sequence of multivariate normally distributed random vectors with mean vector  $\vec{0}$  and the covariance matrix

$$\Sigma(t) := \int_0^t \left\{ \exp \left[ \int_{t'}^t K(z(t'')) dt'' \right] \right\} * v(z(t')) * \left\{ \exp \left[ \int_{t'}^t K(z(t'')) dt'' \right] \right\}^T dt',$$

where  $v$  denotes the population covariance function from Lemma 2. In signs

$$Z^{(N)}(n) \sim N(\vec{0}, \Sigma(n * \epsilon_N)).$$

**Remark:** The theorem says that the stochastic process  $(X^{(N)}(n))_{n \in \mathbb{N}}$  of populations can be substituted asymptotically by a continuous gaussian process with mean function  $z(t)$  and covariance function  $\Sigma(t)$ . The main difficulty in the application of Theorem 1 lies in the calculation of  $z(t)$ . The calculation requires the solution of a differential equation system.

*Proof.* We give only a sketch of the proof here. A complete proof is given in [26].

A central limit theorem for stochastic processes is proved in [20]. The application of the theory in [20] rests on the first three moments of

$$\Delta X^{(N)}(n) := X^{(N)}(n+1) - X^{(N)}(n).$$

Lemmas 1 and 2 prove their conditions on the first two moments. To apply the theory of [20], the following condition on the third moment is left. With

$$e_{3,n}^N := \Delta X^{(N)}(n) - E(\Delta X^{(N)}(n) | X^{(N)}(n))$$

we have to show the existence of a constant  $L$  such that

$$E(\|e_{3,n}^N\|^3)^{1/3} \leq L * \left(\frac{1}{N}\right)^{1/2}.$$

This is done by application of the Jensen inequality in [26]. Now the theorem is a special case of Theorem 3 in [20]. ■

The following corollary shows that the solution  $z(t)$  of the differential equation system (3) is a population for all  $t \in \mathbb{R}^+$ .

**Corollary 1.** Let  $x(0)$  be a population.  $z(t)$  calculated by equation (3) is also a population for all  $t \in \mathbb{R}^+$ .

*Proof.* The change of  $z$  in time is influenced by  $m$ . We show that the sum of the coordinates is not changed by equation (3) for all populations  $x \in P$ :

$$\begin{aligned}
 \frac{d}{dt} \sum_{i \in \Xi} x_i(t) &= \sum_{i \in \Xi} m_i(x) \\
 &= \sum_{i \in \Xi} x_i * (f(i) - \sum_{j \in \Xi} f(j) * x_j) \\
 &\quad + \hat{p}_c * \sum_{i,j,k} x_j * x_k * P(I^C = i | I^{S_1} = j, I^{S_2} = k) - \hat{p}_c * \sum_{i \in \Xi} x_i \\
 &\quad + \sum_{i,j \in \Xi} \mu_{j \rightarrow i} * x_j - \sum_{i \in \Xi} x_i \\
 &= \sum_{i \in \Xi} x_i * f(i) - \sum_{i \in \Xi} x_i \sum_{j \in \Xi} f(j) * x_j \\
 &\quad + \hat{p}_c * \sum_{j,k} x_j * x_k - \hat{p}_c * \sum_{i \in \Xi} x_i \\
 &\quad + \sum_{j \in \Xi} 1 * x_j - \sum_{i \in \Xi} x_i \\
 &= 0.
 \end{aligned}$$

The last equality follows with  $\sum_{i \in \Xi} x_i = 1$ . Therefore, the sum of the coordinates is a constant for all  $t \in \mathbb{R}^+$ . The second condition in the definition of a population is that the coordinates take values in  $[0, 1]$ . Because  $z_i(t)$  is continuous in  $t \in \mathbb{R}^+$ , it suffices to examine the first derivative in the boundaries. Let  $z_i(t) = 0$ . It is easy to see that

$$\begin{aligned}
 m_i(z(t)) &= \hat{p}_c * \sum_{j,k} z_j(t) * z_k(t) * P(I^C(t) = i | I^{S_1}(t) = j, I^{S_2}(t) = k) \\
 &\quad + \sum_{j \in \Xi} \mu_{j \rightarrow i} * z_j(t) \geq 0.
 \end{aligned}$$

Let  $z_j(t) = 1$ . Again, it is easy to see that

$$m_i(z(t)) = \mu_{i \rightarrow i} - 1 \leq 0.$$

Therefore, the value  $z_i(t)$  will never leave the interval. ■

#### 4. Simple applications

In this section we present some simple examples for the application of Theorem 1. For more details refer to [26].



### 4.1 Genetic drift

A GA with finite population size  $N$ , without selection pressure ( $f \equiv 0$ ), without crossover ( $\hat{p}_c = 0$ ), and without mutation ( $\mu_{i \rightarrow j} \equiv 0$ ) reduces to simple drawings with replacement. Under these conditions  $m$  reduces to

$$\forall i \in \Xi : m_i(x) \equiv 0.$$

Let  $x(0) \in P$  be a population. The solution of equation (3) is a constant, that is,

$$\forall t \in \mathbb{R}^+ : z(t) = x(0).$$

Since the differential of  $m$  subject to  $x_j$  is zero, the matrix  $K$  denotes the zero matrix. The exponential function of a zero matrix is the unit matrix. The covariance matrix can be calculated as

$$\Sigma(t) = v(x(0)) * t = [x_i(0) * (\delta_{ij} - x_j(0))]_{i,j} * t.$$

The expectation of the population stays constant. The variance increases linearly in  $t$ .

**Example 2.** Because of the condition  $\sum_{i \in \Xi} x_i(0) = 1$  the functions reduce to one-dimensional functions in a search space with  $|\Xi| = 2$ . Let  $i_1$  and  $i_2$  be the two individuals and let  $t \in \mathbb{R}^+$ :

$$z(t) \equiv x_{i_1}(0) \Rightarrow \Sigma(t) = x_{i_1}(0) * (1 - x_{i_1}(0)) * t.$$

The statement of Theorem 1 reduces to

$$Z^{(N)}(n) \sim N(0, x_{i_1}(0) * (1 - x_{i_1}(0)) * n * \epsilon_N).$$

With the definition of  $Z^{(N)}(n)$  we get

$$P_{X^{(N)}(n) | X^{(N)}(0)=x(0)} \approx N\left(x_{i_1}(0), x_{i_1}(0) * (1 - x_{i_1}(0)) * \frac{n}{N}\right). \quad (4)$$

Notice that the sequence  $(\epsilon_N)_{N \in \mathbb{N}}$  does not appear in the last formula. This is because no dependent parameter arises in the genetic drift model. For the values  $N = 1000$ ,  $n = 10$ , and  $x_{i_1}(0) = \frac{1}{2}$  one gets the approximation

$$P_{X^{(1000)}(10) | X^{(1000)}(0)=\frac{1}{2}} \approx N\left(\frac{1}{2}, \frac{1}{400}\right).$$

Figure 1 shows the approximation of the density of  $1000 * X_{i_1}^{(1000)}(0)$  compared with simulations. The simulated density is based on 100,000 repetitions of drawing with replacement. The graph shows good agreement between the approximated and the simulated densities.

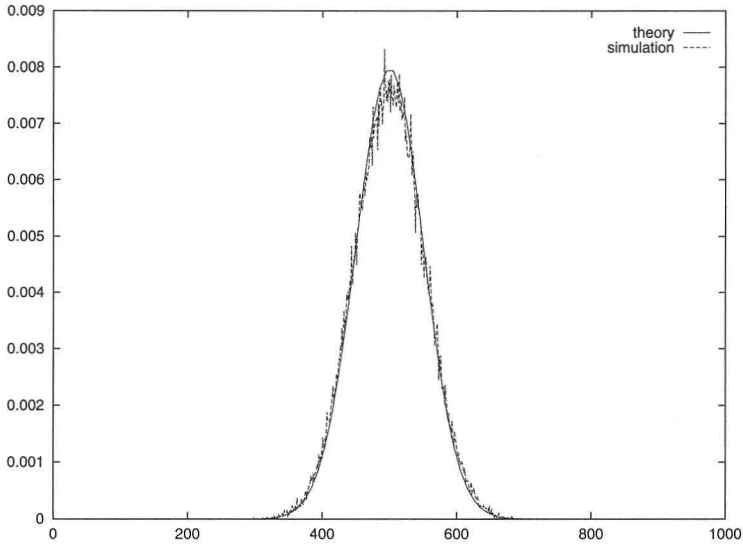


Figure 1: Density of  $1000 * X_{i_1}^{1000}(10)$ . Using (4), the calculated density is compared with results from simulations.

#### 4.2 Search space of dimension $|\Xi| = 2$

In this section we consider a search space with two elements,  $\Xi := \{i_1, i_2\}$ . In this case it makes no sense to consider crossover. Therefore, we will consider a model with only selection and mutation.

Because of the condition  $\sum_{i \in \Xi} z_i(t) = 1$ , the system of differential equations  $\dot{z}_i(t) = m_i(z)$  reduces to a one-dimensional differential equation  $\dot{z} = m(z)$ . This equation is solvable under several adjustments of parameters. We will give one example with positive probability of mutation and with nonzero selection pressure. Further examples for this case are given in [26].

The significance of this application for real-size examples is suggested in the following three interpretations.

*Generalization 1.* First, one may divide the search space into two classes of individuals: the “good” ones and the “bad” ones. The “good” ones may be the ones with a fitness value above a border. This kind of classification is known from the theory of simulated annealing [25].

*Generalization 2.* For a second interpretation one may sign one individual out of the population. This individual is compared with the “rest of the world.” Will the individual’s heritage continue over several generations? Will the individual die because of a low fitness value? The “rest of the world” may be represented by an expected fitness of the other individuals. If the fitness value of the signed individual is higher than the expected fitness, this individual will stay alive; otherwise it will fall into the group that dies out after a few steps.

*Generalization 3.* For a third interpretation consider one special locus in an optimization task of higher problem size. In an optimization task coded in bit-strings, the question arises of whether the 0 or the 1 dies out on this locus (compare the diffusion model of [9]). To answer the question, one individual gets the mean fitness of all individuals with a 0 on this locus. The other individual gets the mean fitness of all individuals with a 1. How fast will one value die out? Or will there be an equilibrium distribution?

Let  $x_{i_1}(0)$  be the relative frequency of individual  $i_1$  in the population  $x(0)$ , and let  $x$  be a real value. The functions of Theorem 1 are calculated with

$$m(x) = \mu_{i_2 \rightarrow i_1} - x * (\mu_{i_2 \rightarrow i_1} + \mu_{i_1 \rightarrow i_2}) + x * (1 - x) * (f(i_1) - f(i_2))$$

and

$$v(x) = x * (1 - x).$$

Let  $z$  be the solution of

$$z(0) = x_{i_1}(0) \quad \text{and} \quad \frac{dz}{dt}(t) = m(z(t)). \quad (5)$$

The function of variance  $\Sigma$  is given by

$$\Sigma(t) = \int_0^t \exp \left( 2 \int_u^t \frac{\partial m(z)}{\partial z}(\xi) d\xi \right) * v(z(u)) du.$$

#### 4.2.1 Selection and mutation in a search space of dimension $|\Xi| = 2$

To motivate the following model we take a population of many different individuals and pick out one individual  $i_1$  of special interest. For the other individuals we are interested only in an expected behavior. These individuals are put into one class. We use  $i_2$  to denote a typical representative of this class. This reduces the model  $\Xi = \{i_1\} \cup (\Xi \setminus \{i_1\})$  to  $\Xi = \{i_1\} \cup \{i_2\}$ .

Let  $i_1$  be of higher fitness than the other. The idea of the GA is that this individual should stay alive over several populations to pass on genes. But the individual may die out because  $i_1$  mutates into another individual, or  $i_1$  is not selected.

On the other hand, the probability for production of this special individual by mutation may be neglected ( $\mu_{i_2 \rightarrow i_1} := 0$ ). Under these conditions the question arises of how the probability of mutation and the pressure of selection influence the quality of the algorithm. The population mean function  $m$  reduces, in this special case, to

$$\begin{aligned} \dot{z}(t) &= m(z(t)) \\ &= z(t) * \mu_{i_1 \rightarrow i_2} + z(t) * (1 - z(t)) * (f(i_1) - f(i_2)) \end{aligned}$$

from which we are able to calculate  $z$ :

$$z(t) = \frac{\mu_{i_1 \rightarrow i_2} - f(i_1) + f(i_2)}{-f(i_1) + f(i_2) + \left( f(i_1) - f(i_2) + (\mu_{i_1 \rightarrow i_2} - f(i_1) + f(i_2)) \frac{1}{x_{i_1}(0)} \right) * e^{(\mu_{i_1 \rightarrow i_2} - f(i_1) + f(i_2)) * t}}$$

$$\Leftrightarrow z(n) = \frac{p^M - \sigma(i_1) + \sigma(i_2)}{-\sigma(i_1) + \sigma(i_2) + \left( \sigma(i_1) - \sigma(i_2) + (p^M - \sigma(i_1) + \sigma(i_2)) \frac{1}{x_{i_1}(0)} \right) * e^{(p^M - \sigma(i_1) + \sigma(i_2)) * n}}$$

with  $p^M := p_{i_1 \rightarrow i_2}^M$  and the condition  $t = n * \epsilon_N$ . It would take too much space to present the solution of  $\Sigma(n)$  here. It is given in detail in [26].

**Example 3.** In Figures 2 and 3, respectively,  $z$  and  $\Sigma$  are given for a population of size  $N = 1000$  after  $n = 10$  steps. The graphs show the values in dependence of the probability of mutation  $p_{i_1 \rightarrow i_2}^M$  (front to back) and the pressure of selection  $Selpr := \sigma(i_1) - \sigma(i_2)$  (right to left). The population  $x(0)$  was chosen with  $x_{i_1}(0) = 0.5$ . The highest variance occurs in a region in

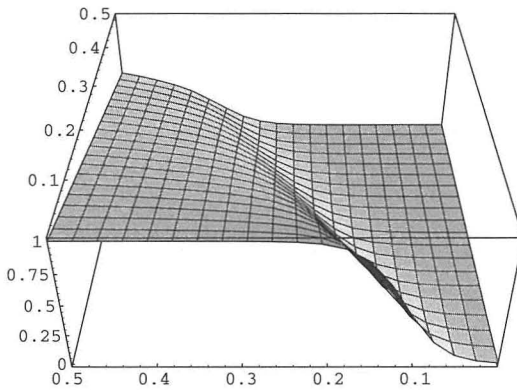


Figure 2:  $z(n)$ . The parameter settings are given in the text. From right to left the pressure of selection increases. From front to back the probability of mutation increases.

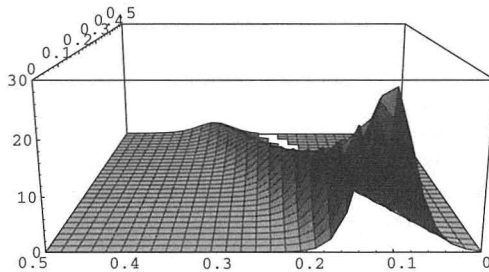


Figure 3:  $\Sigma(n)$ . The parameter settings are given in the text. From right to left the pressure of selection increases. From front to back the probability of mutation increases.

which the change in expectation is high. This region is the interesting one. In the other regions the individual dies out or dominates the population. These two situations are not ideal for an examination of the search space with a GA. Good solutions leave their coordinates to new individuals but do not reproduce themselves.

The variance given by  $\Sigma$  can be interpreted as a kind of uncertainty. If the parameter settings are such that the population process has high variance, the GA produces solutions of different quality in several repetitions. If the population process has low variance, repetitions of the GA produce individuals of the same quality. Because of limited computational resources, in most cases only one repetition of the GA is performed. For a given mean it is better to set the parameters in a region with low variance. The figures show that it is better to set the parameters to high values.

#### 4.3 Crossover in a search space of dimension $|\Xi| = 4$

In this section we consider the simplest case in which crossover makes sense. Let  $\Xi = \{aa, aA, Aa, AA\}$ , a search space of dimension 4. Let  $x \in P$ . For a model with crossover, no pressure of selection ( $\sigma \equiv 1$ ), and no mutation ( $p_{i \rightarrow j}^M = 0$ ),  $m$  reduces with  $\hat{p}_c = 1$  to

$$\forall i \in \Xi: m_i(x) = \sum_{j,k \in \Xi} x_j * x_k * P(I^C = i | I^{S_1} = j, I^{S_2} = k) - x_i.$$

Again, we may reduce the dimension by application of the condition  $\sum_{i \in \Xi} x_i = 1$ . In a search space with four members, the three kinds of crossover introduced in section 2 reduce to two different kinds of crossover. In one-point crossover and uniform crossover, the two coordinates are taken with probability 0.5 from different parents and with probability 0.5 from one parent. In two-point crossover, the two coordinates are taken with certainty from different parents.

**Example 4.** In one-point crossover and uniform crossover the *mean population function* is given by

$$\begin{pmatrix} m_{aa} \\ m_{aA} \\ m_{AA} \end{pmatrix} (x) = \frac{1}{2} * \begin{pmatrix} -x_{aa} + x_{aa} * x_{aa} + x_{aa} * x_{aA} + x_{aa} * x_{Aa} + x_{aA} * x_{Aa} \\ +x_{aa} - x_{aa} * x_{aa} - x_{aa} * x_{aA} - x_{aa} * x_{Aa} - x_{aA} * x_{Aa} \\ +x_{aa} - x_{aa} * x_{aa} - x_{aa} * x_{aA} - x_{aa} * x_{Aa} - x_{aA} * x_{Aa} \end{pmatrix}.$$

With this  $m$  the function  $z$  can be calculated by

$$\begin{aligned} z_{aA}(t) &= -z_{aa}(t) + z_{aA}(0) + z_{aa}(0), \\ z_{Aa}(t) &= -z_{aa}(t) + z_{Aa}(0) + z_{aa}(0), \\ z_{AA}(0) &= 1 - z_{aa}(t) - z_{aA}(t) - z_{Aa}(t), \\ z_{aa}(t) &= [z_{aA}(0) + z_{aa}(0)] * [z_{Aa}(0) + z_{aa}(0)] \\ &\quad + [z_{aa}(0) - [z_{aA}(0) + z_{aa}(0)] * [z_{Aa}(0) + z_{aa}(0)]] * e^{-t/2}. \end{aligned}$$

The first three equations give  $z$  for the individuals in dependence of the individual  $aa$ . In the expression for  $z_{aa}(t)$ , the dependence of the initial population  $z_{aa}(0)$  decreases exponentially. In the limit  $t \rightarrow \infty$ , the remaining term may be compared with an independent distribution of the two coordinates. The marginal distribution is the one of the initial population. If we take as  $x(0)$  the uniform distribution on the four points, the covariance matrix  $\Sigma$  is given by

$$\Sigma(t) = \frac{1}{16} \begin{pmatrix} [1 + 2 * t - e^{-t}] & [-1 + e^{-t}] & [-1 + e^{-t}] \\ [-1 + e^{-t}] & [1 + 2 * t - e^{-t}] & [1 - 2 * t - e^{-t}] \\ [-1 + e^{-t}] & [1 - 2 * t - e^{-t}] & [1 + 2 * t - e^{-t}] \end{pmatrix}.$$

The terms in the main diagonal increase linearly in  $t$ . The covariance between  $aA$  and  $Aa$  decreases linearly in  $t$ .

**Example 5.** For two-point crossover we calculate

$$\begin{aligned} z_{aA}(t) &= -z_{aa}(t) + z_{aA}(0) + z_{aa}(0), \\ z_{Aa}(t) &= -z_{Aa}(t) + z_{Aa}(0) + z_{aa}(0), \\ z_t(AA) &= 1 - z_{aa}(t) - z_{aA}(t) - z_{Aa}(t), \\ z_{aa}(t) &= [z_{aA}(0) + z_{aa}(0)] * [z_{Aa}(0) + z_{aa}(0)] \\ &\quad + [z_{aa}(0) - [z_{aA}(0) + z_{aa}(0)] * [z_{Aa}(0) + z_{aa}(0)]] * e^{-t}. \end{aligned}$$

If we take as  $x(0)$  the uniform distribution on the four points, the covariance matrix  $\Sigma$  is given by

$$\Sigma(t) = \frac{1}{32} \begin{pmatrix} [1 + 4 * t - e^{-2t}] & [-1 + e^{-2t}] & [-1 + e^{-2t}] \\ [-1 + e^{-2t}] & [1 + 4 * t - e^{-2t}] & [1 - 4 * t - e^{-2t}] \\ [-1 + e^{-2t}] & [1 - 4 * t - e^{-2t}] & [1 + 4 * t - e^{-2t}] \end{pmatrix}.$$

In two-point crossover,  $z$  converges faster ( $e^{-t}$ ) than in one-point crossover ( $e^{-t/2}$ ). The multiplier in the covariance matrix is lower than the one for one-point crossover. This shows that in the special situation of only two coordinates, two-point crossover mixes the coordinates more disorderedly. One-point crossover preserves some of the parents. In some applications it may be of benefit to mix the coordinates, and in other applications it may be of benefit to preserve the parents [27].

## 5. Summary and discussion

In this paper we have presented a central limit theorem for the population process of a GA. Some simple examples show how the theorem may be applied. A more detailed description of the method is presented in [26]. These examples show the usefulness of the method for the consideration of dependences between parameters. For further dependences between the probability of mutation and selection pressure, refer to [29].

The theory of GAs is divided into models with infinite and finite population sizes. Most mathematical models deal with infinite populations [18, 23, 28, 30]. Up to now, the strongest results have been given in this area of the theory. But the approximation leads to a deterministic model. The infinite model neglects the stochastic nature of the GA.

The population after a finite number of steps is a random variable. Diffusion models with finite population sizes address this fact. The diffusion models given in [9] and [12] make only a few limitations on the parameters. But the approach is mainly applicable when the number of steps is infinite.

The model presented in this paper describes a useful approximation for a finite population size and a finite number of steps. This is reached at the expense of some limitations on the parameters. Further, Theorem 1 gives a connection of the theory of GAs with the theory of differential equation systems. Therefore, the theorem offers a wide field for research in the future.

## Appendix

*Proof of Lemma 1.* The proof is divided into several steps. Each step considers the influence of one operator.

**Proposition 1.** The changes of the population that are based on selection are given by

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{\epsilon_N} \left( P(I^S(n) = i \mid X^{(N)}(n) = x(n)) - x_i(n) \right) \\ = x_i(n) * \left( f(i) - \sum_{j \in \Xi} f(j) * x_j(n) \right). \end{aligned}$$

*Proof of Proposition 1.*

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{\epsilon_N} \left( P(I^S(n) = i \mid X^{(N)}(n) = x(n)) - x_i(n) \right) \\ = \lim_{N \rightarrow \infty} \frac{1}{\epsilon_N} \left[ \frac{\sigma(i) * x_i(n)}{\sum_{j \in \Xi} \sigma(j) * x_j(n)} - x_i(n) \right] \\ = \lim_{N \rightarrow \infty} \frac{1}{\epsilon_N} \left[ \frac{(1 + \epsilon_N * f(i)) * x_i(n) - x_i(n) \sum_{j \in \Xi} (1 + \epsilon_N * f(j)) * x_j(n)}{\sum_{j \in \Xi} (1 + \epsilon_N * f(j)) * x_j(n)} \right] \\ = \lim_{N \rightarrow \infty} \left[ \frac{f(i) * x_i(n) - x_i(n) \sum_{j \in \Xi} f(j) * x_j(n)}{1 + \epsilon_N * \sum_{j \in \Xi} f(j) * x_j(n)} \right] \\ = f(i) * x_i(n) - x_i(n) \sum_{j \in \Xi} f(j) * x_j(n). \end{aligned}$$

**Proposition 2.** The changes that are based on crossover are given by

$$\lim_{N \rightarrow \infty} \frac{1}{\epsilon_N} \left( P(I^C(n) = i \mid X^{(N)}(n) = x(n)) - x_i(n) \right)$$

$$\begin{aligned}
&= x_i(n) * \left( f(i) - \sum_{j \in \Xi} f(j) * x_j(n) \right) \\
&\quad + \hat{p}_c * \sum_{j,k \in \Xi} x_j(n) * x_k(n) * P(I^C(n) = i \mid I^{S_1}(n) = j, I^{S_2}(n) = k) - \hat{p}_c * x_i(n).
\end{aligned}$$

*Proof of Proposition 2.*

$$\begin{aligned}
&\lim_{N \rightarrow \infty} \frac{1}{\epsilon_N} \left( P(I^C(n) = i \mid X^{(N)}(n) = x(n)) - x_i(n) \right) \\
&= \lim_{N \rightarrow \infty} \frac{1}{\epsilon_N} \left[ \epsilon_N * \hat{p}_c * \sum_{j,k \in \Xi} \frac{\sigma(j) * x_j(n) * \sigma(k) * x_k(n)}{(\sum_{l \in \Xi} \sigma(l) * x_l(n))^2} \right. \\
&\quad * P(I^C(n) = i \mid I^{S_1}(n) = j, I^{S_2}(n) = k) \\
&\quad \left. + (1 - \epsilon_N * \hat{p}_c) * P(I^S(n) = i \mid X^{(N)}(n) = x(n)) - x_i(n) \right] \\
&= \hat{p}_c * \sum_{j,k \in \Xi} x_j(n) * x_k(n) * P(I^C(n) = i \mid I^{S_1}(n) = j, I^{S_2}(n) = k) - \hat{p}_c * x_i(n) \\
&\quad + \lim_{N \rightarrow \infty} \frac{1}{\epsilon_N} \left[ P(I^S(n) = i \mid X^{(N)}(n) = x(n)) - x_i(n) \right] \\
&\stackrel{[1]}{=} x_i(n) * \left( f(i) - \sum_{j \in \Xi} f(j) * x_j(n) \right) \\
&\quad + \hat{p}_c * \sum_{j,k \in \Xi} x_j(n) * x_k(n) * P(I^C(n) = i \mid I^{S_1}(n) = j, I^{S_2}(n) = k) - \hat{p}_c * x_i(n).
\end{aligned}$$

The equality [1] follows from Proposition 1.

*Proof of the lemma.*

$$\begin{aligned}
&\lim_{N \rightarrow \infty} \frac{1}{\epsilon_N} E \left( X_i^{(N)}(n+1) - X_i^{(N)}(n) \mid X^{(N)}(n) = x(n) \right) \\
&\lim_{N \rightarrow \infty} \frac{1}{\epsilon_N} \left[ P(I^M(n) = i \mid X^{(N)}(n) = x(n)) - x_i(n) \right] \\
&\quad \sum_{j \in \Xi, j \neq i} \mu_{j \rightarrow i} \lim_{N \rightarrow \infty} P(I^C(n) = j \mid X^{(N)}(n) = x(n)) \\
&\quad + \lim_{N \rightarrow \infty} \left[ \left( \frac{1}{\epsilon_N} - \sum_{j \in \Xi, j \neq i} \mu_{i \rightarrow j} \right) * P(I^C(n) = i \mid X^{(N)}(n) = x(n)) - \frac{1}{\epsilon_N} * x_i(n) \right] \\
&\quad \sum_{j \in \Xi, j \neq i} \mu_{j \rightarrow i} * x_j(n) - x_i(n) \sum_{j \in \Xi, j \neq i} \mu_{i \rightarrow j} \\
&\quad + \lim_{N \rightarrow \infty} \frac{1}{\epsilon_N} \left[ P(I^C(n) = i \mid X^{(N)}(n) = x(n)) - x_i(n) \right]
\end{aligned}$$

The assertion follows from Proposition 2. ■

*Proof of Lemma 2.* This proof is analogous to the one for Lemma 1. Therefore, we give only a sketch. Because of the polynomial distribution of  $X_n^{(N)}$ ,



the covariance follows with  $i \neq j$ :

$$\begin{aligned}
 & \lim_{N \rightarrow \infty} N * \text{cov}(X_i^{(N)}(n+1), X_j^{(N)}(n+1) | X^{(N)}(n) = x(n)) \\
 &= \lim_{N \rightarrow \infty} \frac{1}{N} \text{cov}(N * X_i^{(N)}(n+1), N * X_j^{(N)}(n+1) | X^{(N)}(n) = x(n)) \\
 &= - \lim_{N \rightarrow \infty} P(I^M(n) = i | X^{(N)}(n) = x(n)) * P(I^M(n) = j | X^{(N)}(n) = x(n)) \\
 &= - \lim_{N \rightarrow \infty} P(I^C(n) = i | X^{(N)}(n) = x(n)) * P(I^C(n) = j | X^{(N)}(n) = x(n)) \\
 &\quad - \lim_{N \rightarrow \infty} P(I^S(n) = i | X^{(N)}(n) = x(n)) * P(I^S(n) = j | X^{(N)}(n) = x(n)) \\
 &\quad - x_i(n) * x_j(n).
 \end{aligned}$$

The proof for the terms in the diagonal is analogous with

$$\begin{aligned}
 & \lim_{N \rightarrow \infty} N * V(X_i^{(N)}(n+1) | X^{(N)}(n) = x(n)) \\
 &= \lim_{N \rightarrow \infty} P(I^M(n) = i | X^{(N)}(n) = x(n)) \\
 &\quad * (1 - P(I^M(n) = j | X^{(N)}(n) = x(n))) . \blacksquare
 \end{aligned}$$

## References

- [1] C. A. Ankenbrandt, "An Extension to the Theory of Convergence and a Proof of the Time Complexity of Genetic Algorithms," in *Foundations of Genetic Algorithms* edited by G. J. E. Rawlins (Morgan Kaufmann, San Mateo, CA, 1991).
- [2] J. E. Baker, "Adaptive Selection Methods for Genetic Algorithms," in *Proceedings of the International Conference on Genetic Algorithms* edited by J. J. Grefenstette (Lawrence Erlbaum, Hillsdale, NJ, 1985).
- [3] L. B. Booker, "Recombination Distributions for Genetic Algorithms," *Foundations of Genetic Algorithms 2* edited by L. D. Whitley (Morgan Kaufmann, San Mateo, CA, 1993).
- [4] Y. Davidor, "Analogous Crossover," in *Proceedings of the International Conference on Genetic Algorithms* edited by D. Schaffer (Morgan Kaufmann, San Francisco, CA, 1989).
- [5] L. Davis, *Handbook of Genetic Algorithms*, (Van Nostrand Reinhold, New York, 1991).
- [6] D. E. Goldberg, "A Note on Boltzmann Tournament Selection for Genetic Algorithms and Population-oriented Simulated Annealing," *Complex Systems*, 4 (1990) 445–460.
- [7] D. E. Goldberg and K. Deb, "A Comparative Analysis of Selection Schemes used in Genetic Algorithms," in *Foundations of Genetic Algorithms* edited by G. J. E. Rawlins (Morgan Kaufmann, San Mateo, CA, 1991).

- [8] D. E. Goldberg and P. Segrest, "Finite Markov Chain Analysis of Genetic Algorithms," in *Proceedings of the International Conference on Genetic Algorithms* edited by J. J. Grefenstette (Lawrence Erlbaum, Hillsdale, NJ, 1987).
- [9] I. Harvey, "The Puzzle of the Persistent Question Marks: A Case Study of Genetic Drift," in *Proceedings of the International Conference on Genetic Algorithms* edited by S. Forrest (Morgan Kaufmann, San Francisco, CA, 1993).
- [10] J. H. Holland, *Adaptation in Natural and Artificial Systems* (University of Michigan Press, Ann Arbor, 1975).
- [11] K. A. de Jong, "An Analysis of the Interacting Roles of Population Size and Crossover in Genetic Algorithms," in *Parallel Problem Solving from Nature* edited by H. P. Schwefel and R. Maenner (Springer, New York, 1990).
- [12] H. Kargupta, "Drift, Diffusion, and Boltzmann Distribution in a Simple Genetic Algorithm," in *Proceedings of the Workshop on Physics and Computation* edited by D. Matzke (IEEE Computer Society Press, New York, 1992).
- [13] M. Kimura, "Diffusion Models in Population Genetics," *Journal of Applied Probability*, **1** (1964) 177–232.
- [14] S. W. Mahfoud, "Finite Markov Chain Models of an Alternative Selection Strategy for the Genetic Algorithm," *Illigal Report*, (Illinois Genetic Algorithms Laboratory, 91007, October 1991).
- [15] S. W. Mahfoud, "An Analysis of Boltzmann Tournament Selection," *Illigal Report*, (Illinois Genetic Algorithms Laboratory, 94007, March 1994).
- [16] M. de Maza and B. Tidor, "An Analysis of Selection Procedures with Particular Attention Paid to Proportional and Boltzmann Selection," in *Proceedings of the International Conference on Genetic Algorithms* edited by S. Forrest (Morgan Kaufmann, San Francisco, CA, 1993).
- [17] T. Nagylaki, "The Gaussian Approximation for Random Genetic Drift," in *Evolutionary Processes and Theory* edited by S. Karlin and E. Nevo (Academic Press, New York, 1986).
- [18] A. E. Nix and M. D. Vose, "Modeling Genetic Algorithms with Markov Chains," *Ann. Math. Artificial Intelligence*, **5** (1992) 79–88.
- [19] M. F. Norman, "A Central Limit Theorem for Markov Processes that Move by Small Steps," *Ann. Probability* **2** (1974) 1065–1074.
- [20] M. F. Norman, "Approximation of Stochastic Processes by Gaussian Diffusions, and Applications to Wright-Fisher Genetic Models," *SIAM Journal of Applied Mathematics*, **29** (1975) 225–242.
- [21] P. W. Poon and J. N. Carter, "Genetic Algorithm Crossover Operators for Ordering Applications," *Computer Operations Research*, **22** (1995) 135–147.

- [22] W. M. Spears and K. A. de Jong, "An Analysis of Multipoint Crossover," in *Foundations of Genetic Algorithms* edited by L. D. Whitley (Morgan Kaufmann, San Mateo, CA, 1993).
- [23] M. Srinivas and L. M. Patnaik, "Binomially Distributed Populations for Modelling Gas," in *Proceedings of the International Conference on Genetic Algorithms* edited by S. Forrest (Morgan Kaufmann, San Francisco, CA, 1993).
- [24] G. Syswerda, "Uniform Crossover in Genetic Algorithms," in *Proceedings of the International Conference on Genetic Algorithms* edited by D. Schaffer (Morgan Kaufmann, San Francisco, CA, 1989).
- [25] M. T. Tran, "Simulated-Annealing: Konstruktion und Analyse einer Auswahl stochastischer Optimierungsverfahren mithilfe inhomogener Markoff-Ketten." Thesis, Universität Hildesheim, 1995. Cuvillier Verlag, Göttingen.
- [26] S. Voget, "Aspekte genetischer Optimierungsalgorithmen: mathematische Modellierung und Einsatz in der Fahrplanerstellung." Thesis, Universität Hildesheim, 1995.
- [27] S. Voget, "Speeding up of a Genetic Algorithm for the Optimization of Periodic Networks," *Hildesheimer Informatik Berichte*, **21** (1995) 1–19.
- [28] S. Voget, "Theoretical Analysis of Genetic Algorithms with Infinite Population Size," *Hildesheimer Informatik Berichte*, **31** (1995) 1–19.
- [29] S. Voget, "Gaussian Diffusion in a Simple Genetic Algorithm," in *Proceedings of the Conference on Parallel Problem Solving from Nature*, 1996, in press.
- [30] D. Whitley, "An Executable Model of a Simple Genetic Algorithm," in *Foundations of Genetic Algorithms* edited by L. D. Whitley (Morgan Kaufmann, San Mateo, CA, 1992).