From recombination of genes to the estimation of distributions II. Continuous parameters

H. Mühlenbein[1] and J. Bendisch[1] and H.–M. Voigt[2]*

[1] GMD – Forschungszentrum Informationstechnik, 53754 Sankt Augustin , Germany
[2] GFaI, Rudower Chaussee 5, 12484 Berlin, Germany

Abstract. The Breeder Genetic Algorithm (BGA) is based on the equation for the response to selection. In order to use this equation for prediction, the variance of the fitness of the population has to be estimated. For the usual sexual recombination this can be difficult. In this paper the new points (offspring) are generated from distributions, a uniform distribution and a distribution generated by univariate marginal distributions. For a class of unimodal fitness functions the performance of the BGA is analytically computed. The results are compared to gene recombination methods. The uniform distribution is approximately generated by line recombination; recombination methods acting independently on each gene approximate the second distribution.

1 Introduction

The Breeder Genetic Algorithm (BGA) is based on the classical science of livestock breeding. The central part of this theory is the equation for the response to selection

$$R(t) = b(t) \cdot I \cdot \sigma(t) \tag{1}$$

Here R denotes the response, which is defined as the difference between the mean fitness of the population at generation $t+1$ and t, $b(t)$ is the *realized heritability*, I is the *selection intensity* and σ is the standard deviation of the fitness [3]. If $b(t)$ and $\sigma(t)$ can be estimated, the equation can be used for predicting the mean fitness of the population after selection. In livestock breeding many methods have been developed to estimate the heritability [3], estimating the variance is still an open question [5].

Looking closer to the equation for the response to selection, we see that the equation does not use a microscopic genetic model at all. It needs the distribution of the fitness. In this paper we directly use distributions to generate new points (offspring). This facilitates the performance analysis for the algorithm. We compute the performance of the BGA for two different distributions. Comparisons are made with popular sexual recombination methods.

* HMV is also with the Technical University of Berlin

2 The uniform distribution BGA

We will use the following class of functions for a performance analysis

$$f_{n,k}(x) = \sum_{i=1}^{n} {x_i}^k \quad k = 1, 2, \ldots \tag{2}$$

These functions are to be minimized in

$$S_{n,k} = \{x | \sum_{i=1}^{n} x_i^k \leq 1, \quad 0 \leq x_i\}.$$

The unique minimization point is $x^* = 0$ with $f^*_{n,k} = 0$. The conceptual Uniform Distribution Breeder Genetic Algorithm (UDBGA) is defined as follows:

UDBGA

- **STEP0:** Set $t = 1$. Generate $N \gg 0$ points randomly in $S_{n,k}$.
- **STEP1:** Select the best $100 \cdot \rho\%$ points ($0 < \rho < 1$, truncation selection).
- **STEP2:** Generate N new points uniformly within the convex hull defined by the selected points. Set $t = t + 1$.
- **STEP3:** If not terminated, go to STEP1.

The above algorithm is conceptual, because the convex hull generated by the selected points maybe difficult to determine. For the above class of test functions, it is easy to see that truncation selection generates convex hulls $S_{n,k,\rho}(t)$, which are given by

$$S_{n,k,\rho}(t) = \{x | \sum_{i=1}^{n} x_i^k \leq d^k_{n,\rho}(t), \quad 0 \leq x_i\} \tag{3}$$

For the analysis the value of $d_{n,\rho}(t)$ is needed. The following lemma gives the result.

Lemma 1: *The boundary value of the domain $S_{n,k,\rho}(t)$ is given by*

$$d_{n,\rho}(t) = \rho^{(t-1)/n} \tag{4}$$

Proof: We start with $d_{n,\rho}(1) = 1$. The next values are determined from the recursion $vol(S_{n,k,\rho}(t+1)) = \rho \cdot vol(S_{n,k,\rho}(t))$. Computing the volume of $S_{n,k,\rho}(t)$ in form of $vol(S_{n,k,\rho}(t)) = \gamma(n, k) \cdot d^n_{n,\rho}(t)$ gives the result. □

Note that $d_{n,\rho}(t)$ is independent of k. We are now able to formulate the main theorem.

Theorem 1. *The expected average fitness E_t of a population evolving according to UDBGA is given by*

$$E_t(f_{n,k}) = \frac{n}{n+k} d^k_{n,\rho}(t) \tag{5}$$

The variance V_t is given by

$$V_t(f_{n,k}) = \frac{n \cdot k^2}{(n+k)^2(n+2k)} d_{n,\rho}^{2k}(t) \tag{6}$$

Proof: For the proof $n-$dimensional integrals have to be computed. We used the following method, based on geometric arguments. Let $\varphi_k = f_{n,k}(x)$. With density

$$\theta_{n,\rho,t}(\varphi) = n \cdot d_{n,\rho}^{-n}(t) \cdot \varphi^{n-1}.$$

the expected fitness is obtained from

$$E_t(f_{n,k}) = \int_0^{d_{n,\rho}(t)} \varphi_k \theta_{n,\rho,t}(\varphi) d\varphi.$$

The variance is computed by the same method. □

In order to use the equation for the response to selection for prediction the selection intensity I and the coefficient of variation CV is needed. The selection intensity for an arbitrary fitness function f is defined as

$$I(f) = \frac{E_t(f) - E_t(f_\rho)}{V^{\frac{1}{2}}(f)},$$

where $E_t(f_\rho)$ denotes the expected mean of the fitness of the selected points at generation t. For our conceptual algorithm we obviously have a heritability of one, i.e. $E_t(f_\rho) = E_{t+1}(f)$. The coefficient of variation is defined for $f \geq 0$ as

$$CV(f) = \frac{V^{\frac{1}{2}}(f)}{E(f)}$$

From theorem 1 follows:

Theorem 2. *For UDBGA the selection intensity I is independent of t and given by*

$$I_{n,k,\rho} = \left(1 - \sqrt[n]{\rho^k}\right) \sqrt{\frac{n(n+2k)}{k^2}}. \tag{7}$$

The coefficient of variation CV is independent of ρ and given by

$$CV_{n,k} = \sqrt{\frac{k^2}{n(n+2k)}}. \tag{8}$$

Corollary: *In the limit we have*

$$\lim_{n \to \infty} I_{n,k,\rho} = \ln \frac{1}{\rho} \qquad 0 < \rho \leq 1 \tag{9}$$

Note that for *evolution strategies* exactly the same limit has been obtained for the progress coefficient $c_{1,\lambda}$ giving the maximum rate of progress ([1]).

Next we show how to use the above results in order to compute the computational complexity and the convergence rate of the algorithm.

3 Computational complexity and convergence

We recall the following definition from mathematical optimization theory.
Definition: *An optimization algorithm converges linearly if*

$$|f(x_{m+1}) - f^*| \leq c|f(x_m) - f^*| \quad 0 < c < 1 \tag{10}$$

For convenience we assume $f^* = 0$.

Theorem 3. *If the coefficient of variation $CV(f)$, the realized heritability b and the selection intensity I are independent of t, and $0 < b \cdot I \cdot CV(f) < 1$, the breeder genetic algorithm converges linearly*

$$E_{t+1}(f) = (1 - b \cdot I \cdot CV(f))\, E_t(f) \tag{11}$$

The proof follows directly from the equation for the response to selection.

Corollary:*For the previous class of test functions $f_{n,k}$ and truncation selection with parameter ρ UDBGA converges linearly in number of generations t*

$$E_{t+1}(f_{n,k}) = \sqrt[n]{\rho^k} E_t(f_{n,k}) \tag{12}$$

For unimodal functions it is obvious that the convergence speed increases with decreasing ρ, i.e severe selection is the best.

In order to evaluate the computational complexity, we introduce the notion of $\epsilon-$computational complexity.

Definition: *The $\epsilon-$computational complexity is defined as the number of steps t_ϵ needed to reduce the initial approximation error by a given fraction $\epsilon > 0$.*

Theorem 4. *For the class of test functions $f_{n,k}$ and truncation selection with parameter ρ UDBGA has a computational complexity of $O(n)$ in number of generations t_ϵ*

$$t_\epsilon(n, k, \rho) = \frac{n}{k} \frac{\ln \epsilon}{\ln \rho} \tag{13}$$

Proof: From Equations (4) and 5 we obtain

$$E_{t+1}(f_{n,k}) = \left(\sqrt[n]{\rho^k}\right) E_t(f_{n,k}) = \left(\sqrt[n]{\rho^k}\right)^t E_1(f_{n,k})$$

This gives

$$\frac{E_{t+1}(f_{n,k})}{E_1(f_{n,k})} = \left(\sqrt[n]{\rho^k}\right)^t = \epsilon$$

Taking the logarithm gives equation 13. □

The above results are given in number of generations t, not number of function evaluations. For the analysis it is assumed that the population is very large, so that the points are uniformly distributed in the statistical sense. In order to minimize the number of functions evaluations, the size of the population has to be as small as possible. This question leads to the difficult problem of the *minimal population size*. For binary genes this problem has been investigated in [4].

4 Line recombination

The conceptual UDBGA is based on the following principle. Generate initial points according to a uniform distribution. Do selection. Generate uniformly distributed new points in the domain implicitly defined by the selected points. The crucial questions for an implementation of this algorithm are: How to generate points uniformly in a given domain? How to identify the domain to be tested? It turns out that it is surprisingly easy to generate uniform distributed vectors in a convex polytope. This new recombination method will be investigated in a forthcoming paper. Here we will make a comparison to a more classical genetic algorithm with sexual recombination.

For convex domains, one of the easiest ways to create new points from two given points x and y is the linear combination

$$z = \alpha x + (1-\alpha)y \quad x,y \in R^n \tag{14}$$

These recombinations schemes are known as line recombination in genetic algorithms (see for instance [2]). For a certain number of generations, these recombination methods will generate points which are approximately uniformly distributed. Usually α is taken to be uniformly distributed in $[0,1]$, or even fixed at 0.5. Both methods have severe shortcomings because they reduce the variance. This will be shown with the following lemma.

Lemma 2 *Let x and y be independent random vectors from the same distribution with existing variance $V(x) = V(y)$. Let the stochastic real number α be independent from x and y and arbitrarily distributed with expectation $E(\alpha) = \bar{\alpha}$ and variance $V(\alpha) = v_\alpha$, then*

$$V(\alpha x + (1-\alpha)y) = (1 - 2\bar{\alpha} + 2\bar{\alpha}^2 + 2v_\alpha) \cdot V(x) \tag{15}$$

The proof has to be omitted. We use the above lemma to compute the variance of the fitness of line recombination methods. The lemma can be only applied for the linear function $f_{n,1}(x)$, for $k > 1$ higher moments have to be computed.

Theorem 5. *For $f_{n,1}(x) = \sum_{i=1}^n x_i$ and $z = \alpha x + (1-\alpha)y$ we have*

$$V(f_{n,1}(z)) = (1 - 2\bar{\alpha} + 2\bar{\alpha}^2 + 2v_\alpha) \cdot V(f_{n,1}(x)) \tag{16}$$

Example 1: For $\alpha \equiv 0.5$ we have $\bar{\alpha} = 0.5$ and $v_\alpha = 0$. Therefore the well known result $V(f_{n,1}(z)) = 0.5V(f_{n,1}(x))$ is obtained.

Example 2: For α uniformly distributed in $[0,1]$ we have $\bar{\alpha} = 0.5$ and $v_\alpha = 1/12$. Therefore $V(f_{n,1}(z)) = \frac{2}{3}V(f_{n,1}(x))$. The variance is severely reduced.

Example 3: Extended line recombination with α uniformly distributed in $[-d, 1+d]$ and $d \approx 3/8$ gives $V(f_{n,1}(z)) \approx V(f_{n,1}(x))$.

Example 4: For α derived from the bimodal fuzzy distribution defined in [7] we have $\bar{\alpha} = 0.5$ and $v_\alpha = 7/24$. Therefore $V(f_{n,1}(z)) = \frac{13}{12}V(f_{n,1}(x))$.

Example 5: For α uniformly distributed in $[-0.25, 0.25]$ and $[-0.75, 1.25]$ (two plateau distribution) we have $\bar{\alpha} = 0.5$ and $v_\alpha = 13/48$. This gives $V(f_{n,1}(z)) =$

$\frac{25}{24}V(f_{n,1}(x))$.

Both in example 1 and example 2 we have a reduction of the variance. This can lead to premature convergence. The following table shows some numerical results. The initial points were uniformly placed in the cube $[0,1]^n$. For the initial population we have $E_0(f_{n,1}) = n/2$ and $V_0(f_{n,1}) = n/12$.

	D(1)			D(2)			D(3)		
t	$\bar{f}$	σ	CV	$\bar{f}$	σ	CV	$\bar{f}$	σ	CV
0	1.011	4.048	0.40	0.991	4.013	0.40	1.001	4.044	0.40
1	0.679	1.767	0.26	0.657	2.131	0.32	0.677	2.380	0.35
2	0.529	0.863	0.16	0.485	1.272	0.26	0.488	1.662	0.34
3	0.464	0.467	0.10	0.382	0.746	0.19	0.355	1.347	0.38
4	0.427	0.330	0.08	0.325	0.493	0.15	0.247	1.001	0.40
5	0.403	0.235	0.06	0.286	0.351	0.12	0.159	0.583	0.36
6	0.388	0.174	0.05	0.260	0.285	0.11	0.110	0.386	0.35
7	0.378	0.195	0.05	0.210	0.160	0.08	0.077	0.273	0.35

Table 1. Three line recombination methods ($k = 1$,$n = 2$,$\rho = 0.5$)

In Table 1 the results for three distributions of α are shown: (D1) $\alpha \equiv 0.5$, (D2) α uniform in $[0,1]$, and (D3) α uniform in $[-0.25, 0.25], [0.75, 1.25]$. The predicted CV from UDBGA is $CV_{2,1} = 0.35$. This value is fairly accurately reproduced by extended line recombination (D3). The other two line recombination methods ((D1),(D2)) reduce the variance too much, leading to premature convergence. For higher dimensions this problem gets also important for extended line recombination (D3).

	D(3)			D(3)		
t	$\bar{f}$	σ	CV	$\bar{f}$	σ	CV
0	2.00	0.5694	0.28	0.6710	0.4237	0.63
1	0.80	0.1744	0.21	0.3177	0.1909	0.60
2	0.36	0.0636	0.18	0.1538	0.0935	0.61
3	0.18	0.0352	0.19	0.0733	0.0465	0.63
4	0.11	0.0065	0.06	0.0338	0.0212	0.62
5	0.10	0.0042	0.04	0.0160	0.0086	0.53

Table 2. Two results for (D3) ($k = 1$, $n = 4$,$\rho = 1/24$; $k = 2$,$n = 2$,$\rho = 0.5$)

Table 2 gives some results for two other examples. In the first example the predicted CV from UDBGA is $CV = 0.20$. Here $CV \approx 0.20$ is achieved for only three generations, then it gets much smaller leading to premature convergence of the algorithm. The reason is the severe selection. A much larger population is needed to circumvent premature convergence. The second example shows that the UDBGA approximation can also be used for $k = 2$. The coefficient of variation for UDBGA is $CV = \sqrt{1/3} = 0.58$. This value is fairly accurately reproduced by extended line recombination.

We have done more simulations, all of which confirm our theoretical results. For large values of k the difference between UDBGA and line recombination based on (D3) gets larger. This shows that extended line recombination does not create uniformly distributed points in higher dimensions.

5 The univariate marginal distribution BGA

The uniform distribution does not take the fitness distribution into account. The search strategy of this algorithm is fairly conservative, i.e it does not exploit areas of good fitness values. In this section we will analyze a distribution which uses the fitness distribution. This distribution generates more points in areas where good fitness values are expected. The simplest way is to generate the new points according to the univariate marginal distributions of the selected points. The analysis of this algorithm is more difficult. We restrict the analysis to $k = 1$ and $k = 2$. The Univariate Marginal Distribution Breeder Genetic Algorithm (UMDBGA) is similar to the UDBGA, only the new points are generated from a different distribution.

UMDBGA

- **STEP0:** Set $t = 0$. Generate $N \gg 0$ points randomly within the unit cube $C_n(0)$.
- **STEP1:** Select the best $100 \cdot \rho\%$ points ($0 < \rho < 1$, truncation selection).
- **STEP2:** Generate N new points x according to $p(x) = \prod_{i=1}^{n} \theta(x_{i,\rho})$. Here $\theta(x_{i,\rho})$ denotes the univariate marginal distribution of variable $x_{i,\rho}$ in the domain of selected points.
- **STEP3:** If not terminated, go to STEP1

For the given fitness function, the domain defined by selection is $S_{n,k,\rho}(t)$ as before, but UMDBGA will generate new points non uniformly within a cube $C_{n,\rho}(t)$ with side length $d_{n,\rho}(t)$. Selection gives domain $S_{n,k,\rho}(t+1)$. Unfortunately, the determination of the marginal distributions within this domain turns out to be difficult, because the points are not uniformly distributed in the cube $C_{n,\rho}(t)$. We will use an approximation by assuming that we have a uniform distribution within the cube. For $k = 1$ we will use the density

$$\theta_{n,\rho,t}(x_{i,\rho}) = \frac{n}{d_{n,\rho}^{n}(t)} \left(d_{n,\rho}(t) - x_{i,\rho}\right)^{n-1}. \tag{17}$$

Theorem 6. *The expected average fitness of a population evolving according to UMDBGA with the marginal distribution θ and fitness function $f_{n,1}$is given by*

$$E_t(f_{n,1}) = \frac{n}{n+1} d_{n,\rho}(t). \tag{18}$$

The variance is given by

$$V_t(f_{n,1}) = \frac{n^2}{(n+1)^2(n+2)} d_{n,\rho}^2(t). \tag{19}$$

Proof: Let $d = d_{n,\rho}(t)$. We have

$$E_t(f_{n,1}) = n \cdot E_t(f_{1,1}) = n \int_0^d \frac{n}{d^n}(d-\xi)^{n-1}\xi d\xi = \frac{n^2 d^{n+1}}{n(n+1)d^n}$$

Similarly the variance is obtained from

$$V_t(f_{n,1}) = n \int_0^d \theta_{n,\rho,t}(\xi)\xi^2 d\xi - E_t^2(f_{n,1}). \qquad \square$$

Corollary: For UMDBGA with marginal distribution (17) the coefficient of variation is

$$CV_{n,1} = \sqrt{\frac{1}{n+2}} \tag{20}$$

The computation of the side length $d_{n,\rho}$ seems to be difficult for the general case. We just give the solution for $n = 2$ and $\rho = 0.5$.
Lemma 3 *For dimension $n = 2$ and $\rho = 0.5$ the side length of the domain $C_{2,0.5}(t)$ is given by*

$$d_{2,0.5}(t+1) = 0.642 \cdot d_{2,0.5}(t) \quad d_{2,0.5}(1) = 1 \tag{21}$$

Proof: Let $d_{2,\rho}(t) = d(t)$. With truncation selection parameter ρ the side length $d(t+1)$ is obtained from

$$\int_0^{d(t+1)} \int_0^{d(t+1)} \frac{2}{d(t)^2}(d(t)-x_1)\frac{2}{d(t)^2}(d(t)-x_2)dx_1 dx_2 = \rho$$

This leads to the equation

$$\eta^4 - 8\eta^3 + 12\eta^2 - 6\rho = 0 \quad \eta = \frac{d(t+1)}{d(t)}$$

For $\rho = 0.5$ the result is obtained. $\square$

Univariate marginal distributions are approximated by volume oriented sexual recombination schemes as used in genetic algorithms. In this case the recombination schemes discussed in the previous section are applied *independently* for each variable. We will show numerical results only for discrete recombination [2].

t	$\bar{f}$	σ	CV	$\bar{f}$ Theory	σ Theory
0	1.000	0.406	0.40	1.000	0.401
1	0.662	0.278	0.41	0.666	0.333
2	0.430	0.196	0.45	0.428	0.214
3	0.274	0.128	0.46	0.275	0.138
4	0.166	0.078	0.47	0.176	0.088
5	0.101	0.047	0.46	0.113	0.057
6	0.063	0.024	0.38	0.073	0.037

Table 3. Discrete recombination $(n = 2, k = 1, \rho = 0.5)$

In table 3 simulation results are displayed for $n = 2$ and $\rho = 0.5$. In this case UMDBGA has $CV_{2,1} = 0.5$. Discrete recombination has a coefficient of variation, which is about 10% less. Note that the average fitness of the genetic recombination BGA and UMDBGA are very similar for the first generations. Then discrete recombination will of course lead to premature convergence.

UMDBGA can be also analyzed for other values of k. But the mathematical difficulties are tremendous. We just outline the case $k = 2$. The selected domain is now part of a hyper sphere. If the points would be uniformly distributed then the marginal distribution would be

$$\theta_{n,\rho,t}(x_{i,\rho}) = c\left(\sqrt{d^2_{n,\rho}(t) - x^2_{i,\rho}}\right)^{n-1} \tag{22}$$

c can be computed from the volume of the $(n-1)$-dimensional hyper-sphere. As for $k = 1$, we make the assumption that the marginal distribution is given by equation (22). The proof of the following theorem is based on sophisticated integration of n-dimensional integrals and has to be omitted.

Theorem 7. *The expected average fitness of a population evolving according to UMDBGA with marginal distribution (22) is given for odd $n \geq 3$ and $k = 2$ by*

$$E_t(f_{n,2}) = \frac{n}{n+2} d^2_n(t) \tag{23}$$

The variance is given by

$$V_t(f_{n,2}) = \frac{2n(n+1)}{(n+2)^2(n+4)} d^4_n(t) \tag{24}$$

Corollary: Under the assumptions of the above theorem the coefficient of variation of UMDBGA is given by

$$CV_{n,2} = \sqrt{\frac{2(n+1)}{n(n+4)}} \tag{25}$$

Note that for large n we have $CV_{n,2} = \sqrt{2/n}$. This value was already empirically determined for gene pool recombination with fuzzy recombination ([6]).

This shows that a UMDBGA accurately describes the behavior of a BGA with volume oriented genetic recombination.

6 Conclusion and outlook

We have shown that the performance of genetic algorithms is mainly governed by distributions. In this paper we have investigated algorithms based on two simple distributions. UDBGA using a uniform distribution is explorative, UMDBGA using marginal distributions is more exploitative. But using univariate marginal distributions is only the first step. For more complex fitness functions, like general quadratic forms, covariances and principal component analysis has to be used [6].

In order to predict the performance of classical genetic algorithms with sexual recombination, the distributions have to be estimated which are generated by recombination. We have shown that extended line recombination approximately generates uniformly distributed points; and volume oriented recombination methods approximate univariate marginal distributions. Now the questions arise: Why not directly using distributions instead of sexual recombination? Why not using multivariate distributions for fitness functions where the variables are correlated?

Acknowledgment: The authors thank Gerd Paaß, who proved lemma 2.

References

1. Hans-Georg Beyer. Toward a theory of evolution strategies: Some asymptotical results from the $(1, \lambda)$-theory. *Evolutionary Computation*, 1:165–188, 1994.
2. H. Mühlenbein and D. Schlierkamp-Voosen. Predictive Models for the Breeder Genetic Algorithm I. Continuous Parameter Optimization. *Evolutionary Computation*, 1:25–49, 1993.
3. H. Mühlenbein and D. Schlierkamp-Voosen. The science of breeding and its application to the breeder genetic algorithm. *Evolutionary Computation*, 1:335–360, 1994.
4. H. Mühlenbein and D. Schlierkamp-Voosen. The theory of breeding and the breeder genetic algorithm. In J. Stender, E. Hillebrand, and J. Kingdon, editors, *Genetic Algorithms in Optimisation, Simulation and Modelling*, Frontiers in Artificial Intelligence Applications, pages 27–64, Amsterdam, 1994. IOS Press.
5. T. Naglyaki. *Introduction to Theoretical Population Genetics.* Springer, Berlin, 1992.
6. H.-M. Voigt and H. Mühlenbein. Gene Pool Recombination and the Utilization of Covariances for the Breeder Genetic Algorithm. In Z. Michalewicz, editor, *Proc. of the 2nd IEEE International Conference on Evolutionary Computation*, pages 172–177, New York, 1995. IEEE Press.
7. H.-M. Voigt, H. Mühlenbein, and D. Cvetković. Fuzzy recombination for the continuous breeder genetic algorithm. In L. J. Eshelman, editor, *Proc. of the Sixth Int. Conf. on Genetic Algorithms*, pages 104–112, San Francisco, 1995. Morgan Kaufmann.

This article was processed using the LaTeX macro package with LLNCS style