The Response to Selection Equation for Skew Fitness Distributions[1]

Hans–Michael Voigt[2]
GFaI
12484 Berlin, Germany
voigt@fb10.tu-berlin.de

Heinz Mühlenbein
GMD
53754 Sankt Augustin, Germany
muehlenbein@gmd.de

Dirk Schlierkamp–Voosen
GMD
53754 Sankt Augustin, Germany
dirk@gmd.de

Abstract— The equation for the response to selection is a powerful analysis and modeling tool for genetic algorithms. In this paper we extend the classical analysis which is restricted to a normal distribution to skew fitness distributions. We show that for a small number of variables the Gamma distribution fits the distribution of the fitness values better than a normal distribution. We compute the selection intensities for the Gamma distribution. It is shown that with these values the prediction for the mean fitness of the population is very accurate. Finally we show that multi-modal functions may lead to fitness distributions having several modal values.

I. Introduction

Let an optimization problem be given on a domain $G \subset R^n$

$$f^* = f(x^*) = \min_{x \in G} f(x), \quad G \subset R^n. \qquad (1)$$

We make no assumptions concerning the convexity and differentiability of the function $f(x)$. For the minimization a number of algorithms have been proposed. In this paper we apply the Breeder Genetic Algorithm BGA [6] to obtain approximations. The BGA uses a continuous representation and a uniform fuzzy gene pool recombination [19]. We will investigate the behavior of the optimization process by the equation for the *response to selection*, both empirically and theoretically . The equation states that the progress of the average fitness of the population is proportional to the *selection intensity*, the *heritability* and the *standard deviation* of the fitness. In [18] we have shown that this equation leads to a design principle for operators. An efficient recombination operator should maximize the progress of the average fitness for a number of generations. This means that it should *maximize the product of the heritability and the standard deviation.* In this paper we will investigate the selection intensity. Previously the selection intensity was computed under the assumption of a normal distribution of the fitness values of a population. We will show that this assumption need not be valid even for very simple fitness functions. If the fitness distribution is skew then this has a significant influence on the modeling of the convergence behavior of the BGA given the same heritability and phenotypic variance.

Skew fitness distributions have already been investigated by Karl Pearson as early as 1894. There are three memoirs devoted to skew distributions [10] [11] [12]. Based on a large amount of data like measurements of Naples crabs he stated:

In a series of memoirs to the Royal Society I have endeavored to show that the Gaussian–Laplace normal distribution is very far from being a general law of frequency distribution either for errors of observation or for the distribution of deviations from type such occur in organic populations.

The outline of the paper is as follows. First we recall the response to selection equation and some of its features. Then we show the evidence of skew fitness distributions based on simulation experiments. In the following section we give a general outline for the determination of selection intensities for large and small population sizes even for skew fitness distributions. Then comparison of a quantitative prediction model with simulation results is made. Finally we show that multi-modal fitness functions may lead to fitness distributions having itself several modes.

II. Response to selection

In this section we summarize our previous theory [6; 7] in more detail. Let $\overline{f(t)}$ be the average fitness of the population at generation t. Then the response to selection is is defined as

$$R(t) = \overline{f(t+1)} - \overline{f(t)}. \qquad (2)$$

The amount of selection is measured by the selection differential

$$S(t) = \overline{f_s(t)} - \overline{f(t)}, \qquad (3)$$

where $\overline{f_s(t)}$ is the average fitness of the selected parents. The equation for the response to selection relates R and S by

$$R(t) = b(t) \cdot S(t). \qquad (4)$$

$b(t)$ is called the *realized heritability.* For many fitness functions and selection schemes the selection differential can be expressed as a function of the phenotypic standard deviation σ. For truncation selection (selecting the $T \cdot N$ best individuals from a population of size N) one obtains

$$S(t)/\sigma(t) = I. \qquad (5)$$

I is called the *selection intensity.* Finally, the famous equation for the response to selection is obtained [4]:

$$R(t) = I \cdot b(t) \cdot \sigma(t). \qquad (6)$$

These equations have been considered for normally distributed fitness values [4]. But they are valid for a large range of distributions, not just for a normal distribution. The response depends on the selection intensity, the realized heritability, and the standard deviation of the fitness distribution. In order to use the above equation for prediction, one has to estimate $b(t)$ and σ and the selection intensity I induced by the fitness distribution. The equation also gives a design criterion for genetic operators – to *maximize the product of heritability and standard deviation* [18]. In this paper we will analyze the influence of the type of the fitness distribution on the selection intensity I.

III. Empirical Fitness Distributions

Throughout this section, the analysis on skew fitness distributions will be done for the sphere and declining plane model

$$f_{sphere}(x) = \sum_{i=1}^{n} x_i^2 \quad \text{and} \quad f_{plane}(x) = \sum_{i=1}^{n} x_i. \qquad (7)$$
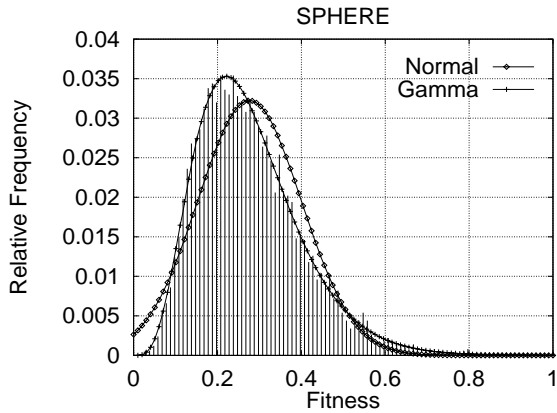

Fig. 1. Fitness distribution of the sphere model, $n = 10$, Generation 10

They are interesting model functions because the sphere has a unique fixed point as the optimal value whereas the declining plane, assuming a finite approximation, has a manifold of optimal solutions. Thus, for the sphere the phenotypic variance has to decrease in approaching the optimum. For the declining plane the phenotypic variance has to increase to gain sufficient convergence velocity. These model functions were extensively used in developing adaptation rules for normally distributed mutations in Evolution Strategies. For recent publications see [14] [15].

In [19] we distinguished between different phases of the optimization process for the sphere model. In the beginning we observed an adaptation of the phenotypic standard deviation σ_P followed by a phase with a constant coefficient of variation $\sigma_P / \bar{f}$.

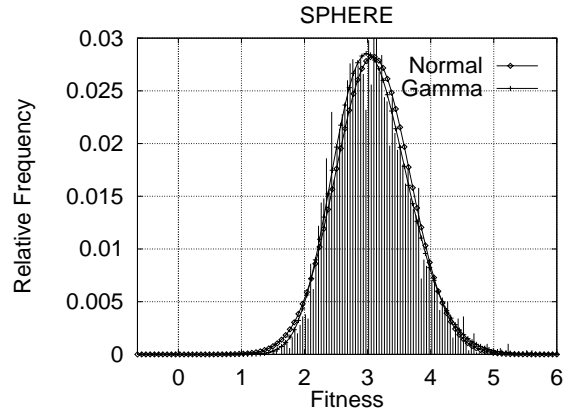Figures 1 through 4 refer to the phase of already adapted standard deviations.

Fig. 2. Fitness distribution of the sphere model, $n = 80$, Generation 20
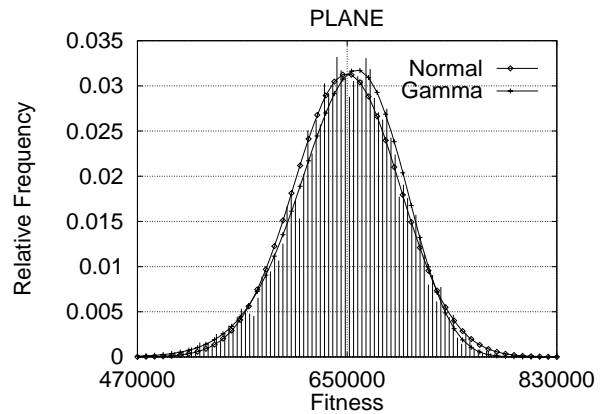

Fig. 3. Fitness distribution of the plane model, $n = 10$, Generation 65

From the histogram plots it can be seen at once that the fitness distributions are by no way symmetric but are more or less skewed. For the sphere model the skewness depends on the number of variables n. The smaller n the more skewed is the distribution. For very high n the distribution can be approximated by a normal distribution. The distribution is skewed to the right such that for a minimiza-
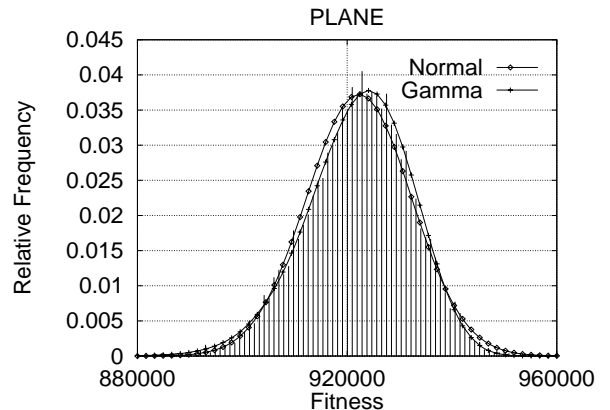

Fig. 4. Fitness distribution of the plane model, $n = 80$, Generation 50

tion problem as considered here the selection intensity will be below the selection intensity for a normal distribution. For the plane model the skewness does not seem to depend on n. But the distribution is skewed to the left which leads to a higher selection intensity than that for a normal distribution. For the plane model the skewness is only very slight compared to that for the sphere model with small n. For all distributions a normal distribution approximation is shown.

In [19] we proposed to use a Gamma distribution to have a closer resemblance of the actual distribution.

We consider the shifted Gamma distribution with $p > 0$, $f \geq f_u$, and $\theta > 0$ having the probability density function

$$p(f) = \frac{\theta^p (f - f_u)^{p-1}}{\Gamma(p)} \cdot e^{-\theta(f - f_u)} \qquad (8)$$

with the expectation $\bar{f}$ and variance σ_P^2

$$\bar{f} = \frac{1}{\theta} \cdot p + f_u \quad \text{and} \quad \sigma_P^2 = \frac{1}{\theta^2} \cdot p. \qquad (9)$$

Given the expectation and variance we have

$$f_u = \bar{f} - \sigma_P \cdot \sqrt{p} \quad \text{and} \quad \theta = \sqrt{p} \cdot \frac{1}{\sigma_P}. \qquad (10)$$

Having the expectation and variance computed from the actual data this relation was used to fit the parameter p to the actual data. For the sphere model we got in every case the best fit for $p = n/2$. This is shown in Figures 1 and 2. It is striking that the Gamma distribution fits the distribution of the actual data so well.

IV. TRUNCATION SELECTION

Truncation selection is a very common technique in breeding [5]. Therefore, in quantitative genetics this selection type has been extensively analyzed for large and small populations (e.g. [4; 13]) having normally distributed metric characters to be selected. Hence, the selection intensities for a normal distribution are tabulated ([4]: large populations Table A, small populations Table B, detailed tables for both cases can be found in [2]). The method for computing I is e.g. given in [13]. Recently, these results have been used to characterize convergence models for tournament- and (μ, λ)-selection for normal distributions [1]. In general, it cannot be assumed that fitness distributions are normal. In the next two subsections we present methods for computing the selection intensities for general unimodal fitness distributions and apply them to the Gamma distribution encountered in the empirical observations.

Let us be given a unimodal fitness distribution with the probability density function (p.d.f.) $p(f)$ having the expectation $\bar{f}$ and the variance σ_P^2. Then the standardized and normalized p.d.f. is given by $p(z)$ with

$$z = \frac{f - \bar{f}}{\sigma_P} \qquad (11)$$

with $\bar{z} = 0$ and $\sigma_z = 1$. This p.d.f. will be used subsequently.

A. Truncation selection in large populations

Let us assume a very large population $N \to \infty$. The proportion of selected parents T for a minimization problem is given by

$$T = \int_{-\infty}^{z_T} p(z) dz \qquad (12)$$

z_T is called the *truncation threshold*. Then, the expectation of the selected parents $\bar{z}_s$ is given by

$$\bar{z}_s = \frac{1}{T} \cdot \int_{-\infty}^{z_T} z \cdot p(z) dz \qquad (13)$$

As the selection intensity we use $I = |z_s|$. Using the normalized and standardized Gamma distribution from Equation (8) with $f_u = -\sqrt{p}$ and $\theta = \sqrt{p}$ one gets for the sphere model with $p = n/2$ the selection intensities shown in Figure 5. It is remarkable that I depends on the problem size n. The smaller the problem size n the larger is the deviation of the selection intensity compared to the selection intensity for a normal distribution.
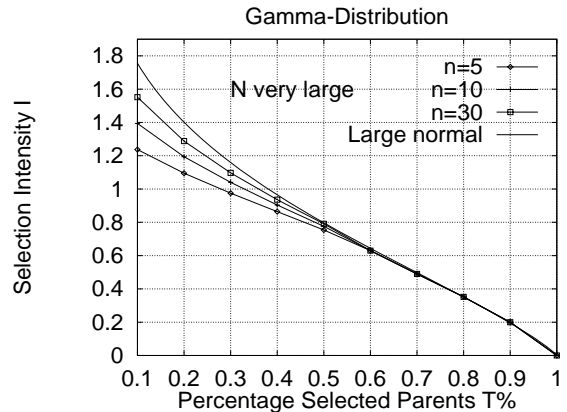

Fig. 5. Selection intensity for the Gamma distribution for different values of n for a large population size $N \to \infty$

B. Truncation selection in small populations

For small populations we have to use order statistics [13] [3] to compute the selection intensities. This is well known in quantitative genetics and has been recently introduced for the study of evolutionary algorithms [1] for normal distributions.

Let us assume a population size N. The standardized and normalized fitness values of the population should have a p.d.f. $p(z)$ and should be arranged in ascending order

$$z_1 \leq z_2 \leq ... \leq z_{N-1} \leq z_N. \qquad (14)$$

$P(z)$ should denote the cumulative density function (c.d.f.) of $p(z)$. Then the p.d.f. of the rth order statistic, $r = 1, 2, ..., N$, is given by [3]

$$p_r(z) = \frac{P(z)^{r-1} \cdot (1 - P(z))^{N-r}}{B(r, N - r + 1)} \cdot p(z). \qquad (15)$$

$B(a, b)$ is Euler's Beta function. The expectation is given by

$$\bar{z}_r = \int_{-\infty}^{\infty} z \cdot p_r(z) dz \qquad (16)$$

Let us assume that N_s parents having for a minimization problem the smallest fitness values are selected we get the average fitness of the selected parents by

$$\bar{z}_s = \frac{1}{N_s} \sum_{r=1}^{N_s} \bar{z}_r \qquad (17)$$

and finally $I = |\bar{z}_s|$. Obviously, the proportion of selected parents is given by $T = N_S / N$.
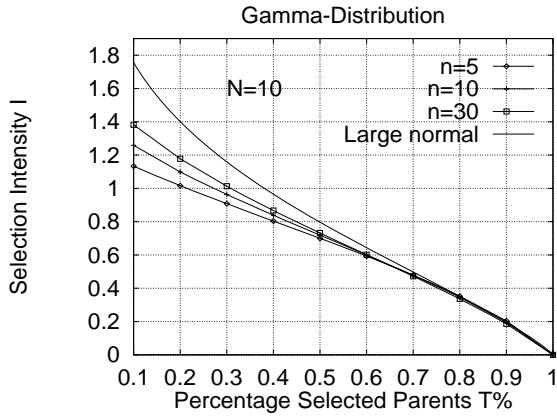

Fig. 6. Selection intensity for the Gamma distribution for different values of n for a small population size $N = 10$

For the sphere function we computed the selection intensities using a Gamma distribution with $p = n/2$ as for the case of a large population. This is shown in Figure 6 for a population size of $N = 10$. The selection intensities depend just like for a large population size on the problem size n. The selection intensities are slightly smaller then those for a large population and especially for small n far from the selection intensity for a large population with a normal fitness distribution function.

V. Quantitative prediction models

In [19] we approximated the response to selection equation by a differential equation introducing a coefficient c_f dependent on the function to be optimized. Better results can be obtained by using the original recurrence relation if it is possible to solve it. No additional coefficient has to be introduced. For a minimization problem the response to selection equation then reads

$$\overline{f(t+1)} - \overline{f(t)} = -b_n \cdot I \cdot CV_n \cdot (\overline{f(t)} - f^*) \qquad (18)$$

where CV_n is the coefficient of variation of the fitness distribution. In [19] we observed for the sphere model for the phase of adapted standard deviation σ_P a constant coefficient of variation CV_n and an average constant realized

heritability b_n. In this case the recurrence relation can be solved

$$\overline{f(t)} = (\overline{f(0)} - f^*) \cdot (1 - b_n \cdot I \cdot CV_n)^t + f^* \qquad (19)$$

For the Gamma distribution with $p = n/2$ we have a coefficient of variation $CV = \sqrt{2/n}$. $\overline{f(0)}$ is the average fitness at generation $t = 0$.
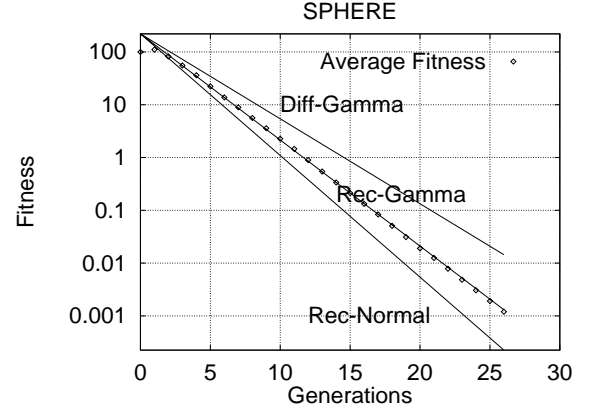

Fig. 7. Prediction models for the sphere function with $n = 20$ and a measured average realized heritability $\bar{b} = 0.93$;**Rec-Gamma**: recurrence equation and selection intensity computed from the Gamma distribution, **Rec-Normal**: recurrence equation and selection intensity computed from the normal distribution, **Diff-Gamma**: differential equation [19] and selection intensity computed from the Gamma distribution; the dotted graph shows the average fitness values from simulation

We are now able to give an estimate for the number of generations until convergence gen^*. Let us assume that the average initial fitness $\overline{f(0)}$ is given and that the standard deviation σ_P is adapted such that we have a constant coefficient of variation CV_n. The population size N is assumed to be sufficiently high. Solving Equation 19 the number of generations to reach a desired approximation of $\overline{f(gen^*)} - f^* = \epsilon$ is given by

$$\epsilon = (\overline{f(0)} - f^*) \cdot (1 - b_n \cdot I \cdot CV_n)^{gen^*} \qquad (20)$$

with $0 < (1 - b_n \cdot I \cdot CV_n) < 1$ assumed. Taking the logarithm finally yields

$$gen^*(I, n, \epsilon, \overline{f(0)}, b_n) \geq \qquad (21)$$
$$-\ln\left(\frac{|f^* - \overline{f(0)}|}{\epsilon}\right) \Big/ \ln\left(1 - b_n \cdot I \cdot \sqrt{\frac{2}{n}}\right)$$

where the equality holds if there is no adaptation phase necessary. Otherwise the inequality holds. In Figure 7 there are depicted four different graphs. The dotted graph represents the average fitness from simulation. The graph labeled *Rec-Gamma* which approximates the observed data best was obtained from Equation 19 by using the selection intensity from the Gamma distribution. For the graph labeled *Rec-Normal* the selection intensity from a normal distribution was used. It displays a much to high convergence speed and is not in accordance with the data. The graph

labeled *Diff-Gamma* shows an approximation using the solution of the differential equation and a selection intensity from the Gamma distribution. The convergence speed is estimated to low.

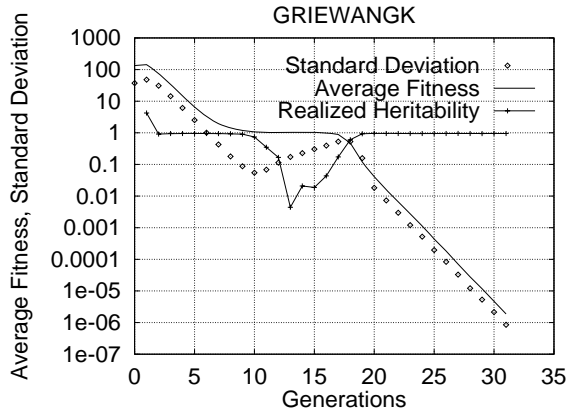## VI. MULTI-MODAL FITNESS FUNCTIONS

Fig. 8. Average fitness $\bar{f}(t)$, phenotypic standard deviation $\sigma_P(t)$, and realized heritability $b(t)$ for Griewangk's function with $n = 10$ and $T = 0.1$
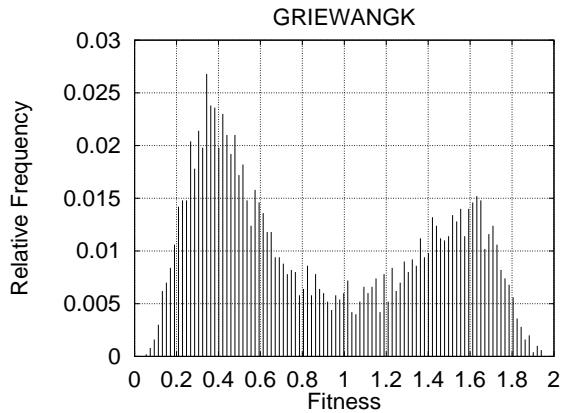

Fig. 9. Fitness distribution at generation 17 for Griewangk's function with $n = 10$ and $T = 0.1$

It should be noted that the prediction may be more complicated because of the fitness distribution may change. An example is the highly multi-modal Griewangk function [6; 17]. Figure 8 shows that the convergence speed at the beginning and at the end of the search is exactly like the convergence speed for the sphere. There is a plateau where the speed of convergence is reduced. Between generation 10 and 18 the average fitness of the population remains constant. The realized heritability gets very small. After generation 20 the heritability returns to a value of almost 1. The reason for this behavior lies in the structure of this function. On a broad scale it looks like a sphere. If the attractor region of the infimum is found, the function again looks like a sphere. In-between these two regions it oscillates like a sine function, and the convergence slows down because the heritability gets very small. The plateau region is very small for strong selection $T = 0.1$.

The fitness distribution at generation 17 is shown in Figure 9 because at this time the standard deviation has already increased to the highest level after a low at generation 10. The distribution is bimodal characterizing the transition from the second best optimum to the global optimum. Therefore we don't have a homogeneous statistical model but the overlay of two distributions. The analysis of the implication of statistical mixtures on selection has to be postponed.

VII. CONCLUSIONS

The equation for the response to selection has been shown to be a powerful method for designing, analyzing and predicting genetic algorithms. In a first approximation one may assume normal distributed fitness values. In this paper we have shown that the prediction gets more accurate if a skewed distribution like the Gamma distribution is used. Highly multi-modal functions may lead to totally different fitness distributions. It remains to be investigated if the concept of selection intensity can be extended to deal also with some of these cases.

REFERENCES

[1] T. Bäck. Generalized Convergence Models for Tournament- and (μ, λ)-Selection In Larry J. Eshelman, editor, *Proc. Sixth Int. Conf. on Genetic Algorithms*, 2-8. Morgan Kaufmann Publishers, San Francisco, 1995

[2] W.A. Becker. *Manual of Quantitative Genetics* Academic Enterprises, Pullman, 1964

[3] H.A. David. *Order Statistics* Wiley, New York, 1981

[4] D.S. Falconer. *Introduction to Quantitative Genetics*. Longman, London, 1981

[5] J.L. Lush. *Animal Breeding Plans, 3rd Ed.* Iowa State University Press, Ames, 1945

[6] H. Mühlenbein and D. Schlierkamp-Voosen. Predictive Models for the Breeder Genetic Algorithm I. Continuous Parameter Optimization. *Evolutionary Computation*, 1:25–49, 1993

[7] H. Mühlenbein and D. Schlierkamp-Voosen. The science of breeding and its application to the breeder genetic algorithm. *Evolutionary Computation*, 1:335–360, 1994

[8] H. Mühlenbein and H.-M. Voigt. Gene pool recombination in Genetic Algorithms. In I.H. Osman and J.P. Kelly, editors, *Proc. of the Metaheuristics Conference.* Kluwer Academic Publishers, Norwell, USA, 1995

[9] H.N. Nagaraja. Some finite sample results for the selection differential. *Ann. Inst. Statist. Math.*, 33:437–448, 1981

[10] K. Pearson. Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material. *Phil. Trans. of the Royal Society of London*, series A, vol. 186, 343–414, 1895

[11] K. Pearson. Contributions to the Mathematical Theory of Evolution. X. Supplement to a Memoir on Skew

Variation. *Phil. Trans. of the Royal Society of London*, series A, vol. 197, 443–459, 1901

[12] K. Pearson. Contributions to the Mathematical Theory of Evolution. XIX. Second Supplement to a Memoir on Skew Variation. *Phil. Trans. of the Royal Society of London*, series A, vol. 216, 429–457, 1916

[13] D. Rasch *Einführung in die mathematische Statistik* VEB Deutscher Verlag der Wissenschaften, Berlin, 1976

[14] I. Rechenberg. *Evolutionsstrategie '94* Frommann-Holzboog, Stuttgart, 1994

[15] H.-P. Schwefel. *Evolution and Optimum Seeking.* John-Wiley, New York, 1995

[16] H.-M. Voigt. Fuzzy evolutionary algorithms. Technical report TR-92-038, *International Computer Science Institute* Berkeley, 1992

[17] H.-M. Voigt Soft genetic operators in evolutionary algorithms. In W. Banzhaf and F. H. Eeckman, editors, *Evolution and Biocomputation: Computational Models of Evolution*, Lecture Notes in Computer Science 899, pages 123–141. Springer-Verlag, Berlin 1995

[18] H.-M. Voigt, H. Mühlenbein, and D. Cvetcovic. Fuzzy recombination for the Breeder Genetic Algorithm. In Larry J. Eshelman, editor, *Proc. Sixth Int. Conf. on Genetic Algorithms*, 104-113. Morgan Kaufmann Publishers, San Francisco, 1995

[19] H.-M. Voigt and H. Mühlenbein. Gene Pool Recombination and Utilization of Covariances for the Breeder Genetic Algorithm. In *Proc. 1995 IEEE Int. Conf. on Evolutionary Computing*, 172–177, Perth 1995, IEEE Press