The Equation for Response to Selection and Its Use for Prediction

Article *in* Evolutionary Computation · February 1997

DOI: 10.1162/evco.1997.5.3.303 · Source: PubMed

1 author:

Heinz Mühlenbein
Fraunhofer AiS
165 PUBLICATIONS **9,101** CITATIONS

SEE PROFILE

The equation for the response to selection and its use for prediction

H. Mühlenbein
RWCP[1] Theoretical Foundation GMD[2] Laboratory
D-53754 Sankt Augustin
muehlenbein@gmd.de

Abstract
The Breeder Genetic Algorithm (BGA) was designed according to the theories and methods used in the science of livestock breeding. The prediction of a breeding experiment is based on the *response to selection* (RS) equation. This equation relates the change in a population's fitness to the standard deviation of its fitness, as well as to the parameters *selection intensity* and *realized heritability*. In this paper the exact RS equation is derived for proportionate selection given an infinite population in *linkage equilibrium*. In linkage equilibrium the genotype frequencies are the product of the univariate marginal frequencies. The equation contains Fisher's fundamental theorem of natural selection as an approximation. The theorem shows that the response is approximately equal to the quotient of a quantity called *additive genetic variance*, V_A, and the average fitness. We compare Mendelian two-parent recombination with gene-pool recombination, which belongs to a special class of genetic algorithms which we call *univariate marginal distribution algorithms* (UMD) algorithms. UMD algorithms keep the genotypes in linkage equilibrium. For UMD algorithms an exact RS equation is proven which can be used for long term prediction. Empirical and theoretical evidence is provided which indicates that Mendelian two-parent recombination is also mainly exploiting the additive genetic variance. We compute an exact RS equation for binary tournament selection. It shows that the two classical methods for estimating realized heritability, the *regression heritability* and the *heritability in the narrow sense* may give poor estimates. Furthermore realized heritability for binary tournament selection can be very different from that of proportionate selection. The paper ends with a short survey about methods which extend standard genetic algorithms and UMD algorithms by detecting interacting variables in nonlinear fitness functions and using this information to sample new points.

Keywords
response to selection, heritability, selection intensity, linkage equilibrium, Robbins' proportions, Fisher's theorem, univariate marginal distributions, tournament selection

1 Introduction

The Breeder Genetic Algorithm (BGA) (Mühlenbein et al., 1994) was designed according to the methods and theories used in the science of livestock breeding. Before we could implement BGA, we first had to find and understand the most important concepts in the science of breeding. Then we had to transfer these concepts to the domain of breeding artificial populations on a computer. The first step was more difficult than expected. The main reason was that variation in most traits of animals and plants is almost continuous,

[1] Real World Computing Partnership
[2] GMD - Forschungszentrum Informationstechnik

and the classical laws of Mendel do not deal with this case.

The methods for analyzing measurements on continuously varying traits of individuals — and, from these, describing how the traits are inherited and then predicting the performance of an individual's relatives — form the discipline of *quantitative genetics*. This field deals with populations of individuals and describes the properties of traits in terms of their means and their degree of variation in the population. From these measurements, parameters such as *heritability* are derived. The usefulness of this approach has to be judged by its ability to describe and predict observations. Predictive equations in quantitative genetics can be of two kinds: (1) *microscopic*, based on changes of gene frequency at individual loci — which might then, for example, be summed over the loci to derive the changes in the trait — and (2) *macroscopic*, based on the measurement of traits in a population. The latter is the approach of the biometricians, who, for example, might perform a regression of progeny fitness on parent fitness under the assumption that traits are normally distributed in both populations.

Selection poses a major problem in quantitative genetics. The biometric approach assumes that heritability remains constant, but since selection alters gene frequencies, it also alters heritability, and the assumption becomes invalid. In breeding practice, however, heritability usually does in fact remain constant for a number of generations.

Quantitative genetics has had a major influence on modern statistics. We mention only two important contributions. Linear regression was invented by Galton in 1885 as a means of characterizing the inheritance of a trait. The analysis of variance and covariance was invented by Fisher in 1918 to compute the correlation between relatives, using a genetic chance model where large numbers of genes influence a single quantitative trait.

In this paper we investigate how to predict the evolution of an artificial genetic population such as is used by BGA by the classical techniques of livestock breeding. For the analysis we assume an infinite population. Offspring are created by mating and recombination of genes. Mutation is neglected.

The outline of the paper is as follows. First, we define and discuss the basic concepts of the science of breeding: *response to selection, selection intensity,* and *heritability.* These concepts are used to formulate the classical *response to selection* (RS) equation. Because of its importance for breeding it is also called the *breeders' equation.* In the remaining part of the paper we mainly investigate under which conditions this equation is a good approximation for the response. First we discuss selection intensity in depth, showing that it is fairly independent of the fitness distribution. Then we investigate the response for two loci. It turns out that the mathematical analysis of Mendelian two-parent genetic recombination (TPR) is difficult, even for two loci. Selection leads the population away from *linkage equilibrium*, but the difference equations that describe gene-frequency evolution seem impossible to solve unless the population is assumed to remain in linkage equilibrium. In linkage equilibrium genotype frequencies are the product of the univariate marginal frequencies.

We then analyze the problem of linkage disequilibrium. We numerically investigate a theorem of Geiringer (1944). It states that if recombination is applied in a large population without selection, the population will move towards linkage equilibrium. In Section 6 we prove an exact expression for the response to selection. If the genotype frequencies are in linkage equilibrium, the response mainly depends on a value called the population's *additive genetic variance*.

We then investigate algorithms that keep the gene frequencies in linkage equilibrium.

We call these algorithms *univariate marginal distribution* (UMD) algorithms. For UMD algorithms we derive an exact equation for the response which uses univariate marginal frequencies only. We compare in detail TPR and UMD algorithms for two and three loci. Binary tournament selection for UMD algorithms is discussed in Sections 12 and 13.

The question remains open whether the results for UMD algorithms can be extended to two-parent recombination as it is used in standard genetic algorithms. A mathematical solution of this problem could be based on an extension of our Theorem 9. Empirical evidence for the conjecture that TPR is also mainly exploiting the additive genetic variance and not discovering and exploiting higher order gene interactions is summarized in Section 11.

Given the empirical and theoretical evidence that genetic algorithms with TPR have the same limitations as UMD algorithms concerning the optimization of fitness functions with interacting genes, we decided to stop our efforts to compute the exact response for TPR algorithms. It is easier to extend UMD algorithms because there are known statistical techniques which detect gene interactions. Some of these techniques are briefly described in Section 15.

If not otherwise noted, we assume discrete genes, an infinite population and the recombination operator produces one child from two mating parents.

2 Response to selection, heritability and regression

The Breeder Genetic Algorithm (BGA) is based on the classical science of livestock breeding as it was formulated by Falconer (1981) in the 60's. In this section we will describe the major concepts and give also some historical remarks about the researchers who made major contributions.

Let $\bar{f}(t)$ be the average fitness of the population at generation t. The response to selection is defined as:

$$R(t) = \bar{f}(t+1) - \bar{f}(t). \tag{1}$$

The amount of selection is measured by the selection differential, $S(t)$

$$S(t) = \bar{f}_s(t) - \bar{f}(t), \tag{2}$$

where $\bar{f}_s(t)$ is the average fitness of the selected parents. The equation for the response to selection relates R and S:

$$R(t) = b(t) \cdot S(t), \tag{3}$$

where $b(t)$ is called the *realized heritability*. The concept of realized heritability was first introduced by Falconer (1981). The importance of the above equation for quantitative genetics has been recently emphasized by Lynch and Walsh (1997). They just call it the *Breeders' equation*.

While the selection differential S is a convenient and simple measure of selection, it does not really tell much about the *strength* of selection. Therefore breeders introduced the normalized selection differential, the *selection intensity* I. It is defined as:

$$I(t) = \frac{S(t)}{\sigma(t)}, \tag{4}$$

where $\sigma(t) = \sqrt{V(t)}$ denotes the standard deviation and $V(t)$ the variance of the fitness values.

The concept of selection intensity was introduced much earlier by Haldane (1932) to investigate the influence of selection. For large livestock populations breeders mainly use *mass* or *truncation* selection. Here the $\lfloor \tau \cdot N \rfloor$ best individuals are selected from a population of size N, $0 < \tau < 1$. If the fitness $f(t)$ has a normal distribution, the selection intensity I_τ for a given τ can be computed fairly easily (Falconer, 1981). Haldane observed that a high competition, i.e., a very small value of τ, does not lead to a correspondingly large increase of the response. The relation between τ and the response is highly nonlinear. In contrast, the response depends linearly on the intensity of selection. The nonlinearity is hidden in the relation between I_τ and τ (see the curves in Falconer (1981)). The selection intensity is independent of t, it depends on the fitness distribution. This problem will be investigated in the next section.

Using the selection intensity one obtains the equation

$$R(t) = I \cdot b(t) \cdot \sigma(t). \tag{5}$$

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 I, $b(t)$ and $\sigma(t)$. The estimation of $b(t)$ and $\sigma(t)$ is difficult. In this paper we will concentrate on the estimation of $b(t)$.

Falconer (1981) showed that realized heritability can be estimated by the regression coefficient from offspring to mid-parent. The regression coefficient is given by (Rao, 1973)

$$b_{F_o F_{mp}}(t) = \frac{cov(F_{mp}(t), F_o(t))}{var(F_{mp}(t))}. \tag{6}$$

F_o is a variable that represents the fitness of an offspring, F_{mp} is the mean fitness of its two parents, also called mid-parent fitness . The variance of the mid-parent fitness is half the variance of the parent fitness, $var(F_{mp}(t)) = V(t)/2$. The regression coefficient is determined for the whole population, not using selection according to fitness.

If the regression coefficient of selected parents and offspring is identical to the regression coefficient which is obtained without selection, then it can be used as an estimate of realized heritability (Mühlenbein et al., 1994)

$$b(t) = \frac{cov(F_{mp}(t), F_o(t))}{\frac{1}{2}V(t)}. \tag{7}$$

The above assumption turns out to be very strong. For complicated fitness functions it is not satisfied.

Historically, the regression coefficient was introduced much earlier than realized heritability. It was invented at the end of the last century by Galton and Pearson. The regression coefficient is the foundation of the purely macroscopic approach developed by a school that later came to be called the *biometricians*. After the rediscovery of Mendel's law a new school arose, the *Mendelians*. The most famous biometrician Pearson "proved" (Pearson, 1904) that Mendel's laws or any modification of them are useless for predicting the regression coefficient, used very successfully by the biometric approach. This posed a major problem for the Mendelians. The battle between these two schools continued for more than 25 years.

Judged from todays perspective, Fisher in 1918 derived for certain fitness functions the biometric regression coefficient as well as the correlation coefficient between relatives from a genetic chance model which can be considered as an extension of Mendel's model to the case of many genes influencing a single trait. Fisher's paper is considered to be the most important one in quantitative genetics. It has lead in statistics to the analysis of variance. In order to understand the result of Fisher's paper, some definitions are necessary.

Let $\boldsymbol{x} = (x_1, \ldots, x_n)$, $x_i \in \{0, 1, \ldots, L\}$ be a genotype, let $f(\boldsymbol{x})$ be the fitness and $p(\boldsymbol{x}, t)$ be the frequency at generation t. Then the *univariate marginal frequencies* are given by

$$p_i(x_i, t) = \sum_{\boldsymbol{x}|x_i} p(\boldsymbol{x}, t), \tag{8}$$

where the sum is taken over all $\boldsymbol{x}$ with x_i held fixed.

Definition: *The genotype frequencies are in Robbins' proportions (Robbins, 1918) if*

$$p(\boldsymbol{x}, t) = \prod_{i=1}^{n} p_i(x_i, t). \tag{9}$$

Robbins' proportions simply state that the x_i are statistically independent. This is also called *linkage equilibrium* in population genetics.

The following theorem has been proven by Fisher (1918) under strong assumptions. A concise proof can be found in Asoh and Mühlenbein (1994a).

Theorem 1 *Let the gene frequencies be in Robbins' proportions. Then the variance of the population can be decomposed at generation t into*

$$V(t) = V_1(t) + V_2(t) + \cdots + V_n(t). \tag{10}$$

The covariance can be decomposed into

$$cov(F_{mp}(t), F_o(t)) = \frac{1}{2}V_1(t) + \frac{1}{4}V_2(t) + \cdots + \frac{1}{2^n}V_n(t). \tag{11}$$

$V_1(t)$ is called the *additive genetic variance* $V_A(t)$. For gene frequencies in Robbins' proportions a closed expression for V_A can be found

$$V_A(t) = \sum_{i=1}^{n} \sum_{v=0}^{L} p_i(v, t) \left(F_i(v, t)\right)^2 .$$

It will be derived in Section 7. For a precise definition of the *interaction variances* V_j see Asoh and Mühlenbein (1994a).

$V_A(t)$ will be of critical importance for our analysis. It depends on the univariate marginal frequencies, making an interpretation difficult. Fisher later applied his method to general statistical problems. It lead to the development of the analysis of variance (ANOVA) and covariance. In statistical problems there is no evolving population, instead the concept of a representative sample set is used. Therefore it is implicitly assumed that the p_i are constant. This assumption has been also made by Reeves & Wright (1995) who reintroduced ANOVA into the theoretical analysis of genetic algorithms. They used ANOVA to compute the epistasis of some fitness functions. But for a genetic population the ANOVA decomposition can be used only for the initial population which is generated

randomly. Later p_i will change according to the dynamics of the genetic population. This might change the decomposition dramatically.
From the above theorem using Equation 7 we obtain the estimate:

Corollary: *Under the assumptions of Theorem 1 the regression coefficient can be estimated by*

$$b_{F_o F_{mp}}(t) = \frac{\frac{1}{2}V_A(t) + \frac{1}{4}V_2(t) + \cdots \frac{1}{2^n}V_n(t)}{\frac{1}{2}V(t)}. \tag{12}$$

Both the definition and the computation of the interaction variances is difficult. Therefore Fisher's paper was of theoretical value only. It shows that the connection between a genetic microscopic model and macroscopic regression is very complex.

In 1930 Fisher made another important contribution, which he called the *fundamental theorem of natural selection* (FTNS) (Fisher, 1958). Fisher claimed for proportionate selection that

$$R(t) \approx \frac{V_A(t)}{\bar{f}(t)}, \tag{13}$$

where V_A is the additive genetic variance introduced before. It is easy to show that for proportionate selection we have (Mühlenbein et al., 1994)

$$S(t) = \frac{V(t)}{\bar{f}(t)}.$$

Therefore Fisher's theorem can be written as

$$R(t) \approx \frac{V_A(t)}{V(t)} S(t).$$

This suggests another estimate of realized heritability. It is called the *heritability in the narrow sense* denoted as h^2 (Falconer, 1981):

$$b(t) \approx h^2(t) = \frac{V_A(t)}{V(t)}. \tag{14}$$

Now we arrived at the following problem. From Fisher's FTNS follows that, if $V_A = 0$, the realized heritability is zero. But the regression coefficient is as large as 0.5 for $V_A = 0$ and $V_2 = V$. Which estimate is better? This problem will be investigated in Section 8. Breeders use heritability in the narrow sense as the estimate of realized heritability (Falconer, 1981). This is justified because under the assumptions of Theorem 1 the relation

$$h^2(t) \leq b_{F_o F_{mp}}(t) \tag{15}$$

holds. Inserting h^2 into Equation 5 gives the famous response to selection equation used by breeders (Falconer, 1981)

$$R(t) \approx I \cdot h^2(t) \cdot V^{1/2}(t) = I \cdot h(t) V_A^{1/2}(t). \tag{16}$$

The response to selection is approximately equal to the product of selection intensity, heritability in the narrow sense, and the square root of the additive genetic variance.

In genetic algorithms, but also in breeding of livestock the goal is not to optimize the response for one generation, but the cumulative response for T generations. This is given by

$$R^T = \sum_{t=1}^{T} I \cdot h(t) \cdot V_A^{1/2}(t). \tag{17}$$

This is a very short description of the major concepts and results in classical quantitative genetics. The different definitions of heritability are at first confusing. Some of the results mentioned have been only vaguely derived for diploid organisms. In their forthcoming book Lynch and Walsh (1997) derive conditions under which the breeders' equation can be applied. Furthermore they discuss when this equation should not be used from the viewpoint of quantitative genetics.

We believe that researchers in genetic algorithms should be familiar with the basic concepts of quantitative genetics. We have already mentioned the analysis of variance. A second example is the correlation between parents and offspring. It was defined by Galton and Pearson. In 1991 Manderick et al. proposed the correlation between parent and offspring as a measure for comparing genetic operators. But the correlation coefficient and the regression coefficient $b_{F_o F_{mp}}$ are closely related (see Mühlenbein et al., 1994). The RS equation makes it clear that the correlation measure alone is not sufficient to define a good genetic operator. Instead the product of correlation and the variance of the fitness of the offspring has to be taken. This is a mathematical formulation of the exploitation vs. exploration problem. A high correlation means that offspring are very similar to parents (exploitation), a high variance means that offspring might be very different from parents.

In the next Section we investigate the concept of selection intensity in more detail.

3 Selection intensity

In order to compute selection intensity, we will use the notation and the results of order statistics. (For a recent introduction into order statistics, see Arnold et al., 1992.) Order statistics has already been used by Bäck (1995) to compute the selection intensity of truncation selection and tournament selection, but Bäck only investigated normal distributions. A detailed investigation of selection intensity for different selection schemes has been done by Blickle & Thiele (1997). They also assume a normal distribution. In this section we will compute the selection intensity for some well-known discrete and continuous distributions.

Let $X_{1:s} \leq X_{2:s} \leq \cdots \leq X_{s:s}$ denote the order statistics of a random sample of size s (i.e., $X_{i:s}$ is the ith smallest member of the sample set). The sample is drawn from a continuous distribution with probability density function (PDF) $d(x)$, cumulative distribution function (CDF) $D(x)$, mean μ, and variance σ^2. We first will compute the selection intensity I_s for *tournament selection,* then for truncation selection. In tournament selection, only the largest value, $X_{s:s}$, is taken. Therefore we have to compute

$$I_s = \frac{E(X_{s:s}) - \mu}{\sigma}. \tag{18}$$

The expected value of the largest element is given by

$$E(X_{s:s}) = \int_{-\infty}^{+\infty} x d_{s:s}(x) dx, \tag{19}$$

where $d_{s:s}$ is the PDF of $X_{s:s}$, which for continuous distributions was shown by Arnold et al. (1992) to be:

$$d_{s:s}(x) = sD(x)^{s-1}d(x). \tag{20}$$

For a given continuous distribution one can use these equations to compute the selection intensity I_s. In Table 1 we give results for the following PDF's: the normal distribution $N(0,1)$, the uniform distribution $U(0,1)$, the exponential distribution EXP, and several discrete binomial distributions, $B(n,p)$. In the case of the binomial distributions, we have (Arnold et al., 1992)

$$E(X_{s:s}) = \sum_{i=0}^{n-1}(1 - D(i)^s). \tag{21}$$

For tournament sizes of ten or less (i.e., $s \leq 10$), the selection intensity is very similar for all distributions considered, despite the wide range of values for $E(X_{s:s})$. This means that the values obtained using a normal distribution can be used as an approximation. The binomial distribution arises if the fitness function is the discrete *ONEMAX* function of size n (Mühlenbein et al., 1994). The selection intensity for this discrete distribution is surprisingly similar to those for the continuous distributions.

s	E/I_s	$N(0,1)$	$B(15,0.5)$	$B(10,0.5)$	$B(15,0.1)$	$B(10,0.1)$	$U(0,1)$	EXP
2	$E(X)$	0.5641	8.5832	5.8809			0.6666	1.5
	I_s	0.5641	0.5593	0.5571			0.5773	0.5
5	$E(X)$	1.1629	9.7298	6.8114	2.9116	2.1550	0.8333	2.2833
	I_s	1.1629	1.1514	1.1456	1.2149	1.2174	1.1547	1.2833
10	$E(X)$	1.5387	10.4383	7.3817	3.4480	2.6089	0.9090	2.9289
	I_s	1.5387	1.5173	1.5063	1.6765	1.6959	1.4171	1.9289
∞	$E(X)$	∞	15	10	15	10	1.0	∞
	I_s	∞	3.8729	3.1622	11.6189	9.4868	1.7321	∞

Table 1: $E(X_{s:s})$ and selection intensity I_s for tournament selection of size s

For discrete distributions with a small number of states, the selection intensity may be very different from that obtained by assuming a normal distribution. This is shown in the next theorem.

Theorem 2 *For the binomial distribution $B(2,p)$ we have*

$$E(X_{2:2}) - \mu = 2p(1-p)(1-p+p^2) \tag{22}$$

$$I_2(2,p) = \sqrt{2p(1-p)}(1-p+p^2). \tag{23}$$

For the binomial distribution $B(3,p)$ we obtain

$$E(X_{2:2}) - \mu = 3p(1-p)(1-2p+4p^2-4p^3+2p^4) \tag{24}$$

$$I_2(3,p) = \sqrt{3p(1-p)}(1-2p+4p^2-4p^3+2p^4). \tag{25}$$

Proof: We prove the theorem for $B(2,p)$. Obviously $\mu = 2p$. With $D(0) = (1-p)^2$ and $D(1) = 1-p^2$ we obtain from Equation 21 the term $E(X_{2:2}) - \mu$. The selection intensity is obtained by dividing this expression by $\sigma = \sqrt{2p(1-p)}$. The proof for $B(3,p)$ is

similar.
□

For the binomial distribution the selection intensity depends on the parameters n and p. But even for $n = 3$ the value obtained from the normal distribution is approximately valid in the range $0.2 < p < 0.8$, as can be seen in Figure 1. We obtain $I_2(3, 0.5) = 0.536$, whereas the normal distribution gives $I_2 = 0.56$ (see Table 1).

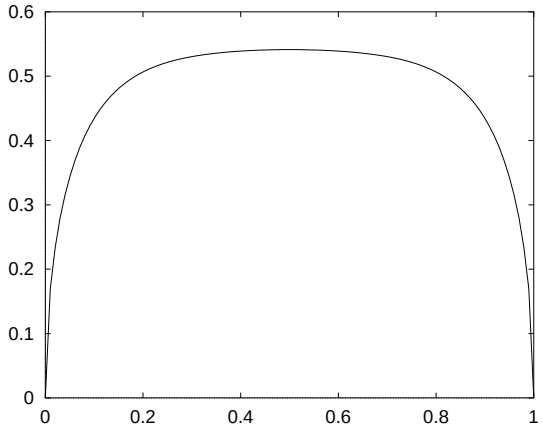

Figure 1: Selection intensity I_2 for $B(3, p)$.

The computation of the selection intensity is more difficult for *truncation selection*. With truncation selection the k best values are selected where k depends on the truncation value τ. If μ is the average of the population, the selection intensity is defined as

$$I(k, N) = \frac{\frac{1}{k}\sum_{i=N-k+1}^{N} X_{i:N} - \mu}{\sigma}, \tag{26}$$

where N now denotes the sample size instead of s in order to be consistent with the notion of population size. The cumulative distribution function CDF of $I(k, N)$ does not have a closed form; however, its mean and variance can be computed from the means, variances, and covariances of the order statistics. Extensive tables are available for the normal distribution (Falconer, 1981). Truncation selection has been widely used in evolution strategies. For normally distributed variates we have $I(k, N) = c_{k/k,N}$, where $c_{k/k,N}$ is called the progress coefficient in evolution strategies (Bäck (1995)).

A closed solution of the expected value of $I(k, N)$ can be obtained if the sample is from an exponential distribution (Nagaraja, 1982):

$$E(I(k, N)) = \sum_{i=k+1}^{N} i^{-1}. \tag{27}$$

If N is large then this can be approximated as:

$$E(I(k, N)) \approx ln(\frac{N}{k}). \tag{28}$$

For arbitrary distributions, even if the sample values are dependent, it can be shown

that (Nagaraja, 1982)

$$E(I(k,N)) \leq \sqrt{\frac{N-k}{k}}. \tag{29}$$

For sufficiently large N, and for $k = \lfloor \tau N \rfloor$, this gives the bound

$$E(I(k,N)) \leq \sqrt{\frac{1-\tau}{\tau}}. \tag{30}$$

In the following table we compare the selection intensity of exponentially distributed with normally distributed fitness values.

Dist.	τ	N=4	N=8	N=16	N=32	$N=\infty$	(30)
EXP	0.5	0.583	0.635	0.663	0.678	0.693	1.0
$N(0,1)$		0.663	0.725	0.760	0.785	0.798	1.0
EXP	0.25		1.218	1.297	1.341	1.387	1.73
$N(0,1)$			1.138	1.201	1.25	1.271	1.73
EXP	0.125			1.881	1.975	2.079	2.65
$N(0,1)$				1.525	1.62	1.645	2.65

Table 2: Selection intensity for exponential and normal fitness distributions. The far-right column is computed from Equation 30.

One can see from Table 2 that the difference between the selection intensities for normal and exponential distributions is roughly 20%. For $\tau = 0.5$ the selection intensity for the normal distribution is higher. This is reversed with more severe selection ($\tau \leq 0.25$).

The upper bound (Equation 30) is between 20% and 60% higher than the intensities obtained from the distributions. Since we are only attempting to approximate the RS equation, it seems reasonable to use the selection intensity derived for the normal distribution as a first approximation of I_τ.

For an accurate prediction the specific distribution has to be taken into account. This was done by Voigt and Mühlenbein (1995), who showed that for continuous unimodal fitness functions, the fitness distribution is better approximated by a gamma distribution than by a normal distribution.

We now turn to the investigation of realized heritability, beginning with an investigation of two-parent genetic recombination as it is normally used in genetic algorithms.

4 Analysis of two-parent recombination for two loci

The difficulty in analyzing two-parent genetic recombination (TPR) will be shown by way of a simple example using proportionate selection and two loci. In this case there are four possible genotypes $\boldsymbol{x}$: $(0,0),(0,1),(1,0)$, and $(1,1)$. We denote their fitness values $f(\boldsymbol{x})$. Let $p(\boldsymbol{x},t)$ be the frequency of genotype $\boldsymbol{x}$ at generation t. For simplicity we restrict the analysis to *uniform crossover* (Syswerda, 1989), an example of two-parent recombination.

Theorem 3 *For proportionate selection and uniform crossover the gene frequencies obey the following difference equation*

$$p(\boldsymbol{x}, t+1) = \frac{f(\boldsymbol{x})}{\bar{f}(t)} p(\boldsymbol{x}, t) + (-1)^{|\boldsymbol{x}|^2+1} \frac{1}{2} \frac{D_s(t)}{\bar{f}(t)^2}. \tag{31}$$

$|\boldsymbol{x}|^2$ *denotes the number of ones in* $\boldsymbol{x}$. $\bar{f}(t) = \sum_{\boldsymbol{x}} p(\boldsymbol{x}, t) f(\boldsymbol{x})$ *is the average fitness of the population; and* $D_s(t)$ *is defined as*

$$D_s(t) = f(0,0) f(1,1) p(0,0,t) p(1,1,t) - f(0,1) f(1,0) p(1,0,t) p(0,1,t) \tag{32}$$

Proof: For proportionate selection the gene frequencies $p^s(\boldsymbol{x}, t)$ after selection are given by

$$p^s(\boldsymbol{x}, t) = \frac{f(\boldsymbol{x})}{\bar{f}(t)} p(\boldsymbol{x}, t).$$

Now we pair randomly between the selected parents and count how often genotype $\boldsymbol{x}$ arises after uniform crossover. Taking $\boldsymbol{x} = (0,0)$ as an example, and computing the probabilities of mating, we obtain

$$\begin{aligned} p(0,0,t+1) &= p^s(0,0,t) \left(p^s(0,0,t) + p^s(0,1,t) + p^s(1,0,t) + \frac{1}{2} p^s(1,1,t) \right) \\ &\quad + \frac{1}{2} p^s(0,1,t) p^s(1,0,t) \end{aligned}$$

Using the fact that $p^s(0,0,t)+p^s(0,1,t)+p^s(1,0,t)+p^s(1,1,t) = 1$ we obtain the theorem for $\boldsymbol{x} = (0,0)$. The remaining equations are obtained in the same manner.
□

Equations 31 are formally identical to those known for diploid organisms in population genetics (Crow & Kimura , 1970), despite the fact that the underlying genetic recombination is different. Uniform crossover can thus be thought of as Mendelian recombination for haploid organisms. Obviously, the same equations are obtained for single point crossover with crossover probability of 0.5. The difference equations have not yet been analytically solved (Nagylaki, 1992).

For the univariate marginal distributions $p_1(1,t) = p(1,0,t) + p(1,1,t)$ and $p_2(1,t) = p(0,1,t) + p(1,1,t)$ we obtain

$$p_1(1,t+1) = \frac{f(1,0) p(1,0,t) + f(1,1) p(1,1,t)}{\bar{f}(t)} \tag{33}$$

$$p_2(1,t+1) = \frac{f(0,1) p(0,1,t) + f(1,1) p(1,1,t)}{\bar{f}(t)}. \tag{34}$$

Here the term D_s has vanished. If the genotypes are in Robbins' proportions then we get

$$p_1(1,t+1) = p_1(1,t) \frac{f(1,0)(1 - p_2(1,t)) + f(1,1) p_1(1,t)}{\bar{f}(t)} \tag{35}$$

$$p_2(1,t+1) = p_2(1,t) \frac{f(0,1)(1 - p_1(1,t)) + f(1,1) p_1(1,t)}{\bar{f}(t)}. \tag{36}$$

We will derive the equations for an arbitrary number of loci n in Section 6. For two loci we can compute an exact expression for realized heritability.

Theorem 4 *The realized heritability $b(t)$ for uniform crossover is given by*

$$b(t) = 1 - \frac{1}{2}(f(0,0) - f(0,1) - f(1,0) + f(1,1))\,\frac{D_s(t)}{\bar{f}(t)V(t)}. \tag{37}$$

Proof: By summation we obtain

$$\begin{aligned} R(t) &= \bar{f}(t+1) - \bar{f}(t) \\ &= \frac{V(t)}{\bar{f}(t)} - \frac{1}{2}(f(0,0) + f(1,1) - f(0,1) - f(1,0))\frac{D_s(t)}{\bar{f}(t)^2} \\ &= \left(1 - \frac{1}{2}(f(0,0) - f(0,1) - f(1,0) + f(1,1))\,\frac{D_s(t)}{\bar{f}(t)V(t)}\right) S(t) \\ &= b(t)S(t). \end{aligned}$$

Here we used the equation $S(t) = V(t)/\bar{f}(t)$ (Mühlenbein et al. 1994).
□

We will return to these equations in Section 8. Note that $D_s(t) = 0$ if $p(0,0,t)p(1,1,t) = p(0,1,t)p(1,0,t)$ and $f(0,0)f(1,1) = f(0,1)f(1,0)$. The first condition is the mathematical definition of *linkage equilibrium* in population genetics. We will soon show that linkage equilibrium is identical to the genotypes being in Robbins' proportions.

Realized heritability is 1 for the additive case $f(0,0) + f(1,1) = f(0,1) + f(1,0)$. Realized heritability is also 1 for the multiplicative case $f(0,0)f(1,1) = f(0,1)f(1,0)$ if the initial population is in linkage equilibrium. But in general, uniform crossover after selection leads to difficult systems of difference equations; the genetic population moves away from linkage equilibrium.

The assumption of linkage equilibrium is not as severe as one might think. The next theorem shows that without selection, the gene frequencies of a population mating randomly will converge to linkage equilibrium. This means that linkage equilibrium can be considered to be the limit distribution of any genetic recombination scheme applied without selection.

Theorem 5 *Let $D(t) = p(0,0,t)p(1,1,t) - p(0,1,t)p(1,0,t)$. If there is no selection then*

$$D(t) = (-1)^{|x|^2}(p(\boldsymbol{x},t) - p_1(x_1,0)p_2(x_2,0)). \tag{38}$$

Furthermore the factor $D(t)$ is halved each generation

$$D(t+1) = \frac{1}{2}D(t). \tag{39}$$

Proof: Without selection the univariate marginal frequencies are independent of t, because for an infinite population a recombination operator based on the Mendelian chance model does not change them. Then from

$$\begin{aligned} p(1,1,t) - p_1(1,0)p_2(1,0) &= p(1,1,t) - (p(1,0,t) + p(1,1,t))(p(0,1,t) + p(1,1,t)) \\ &= p(1,1,t) - p(0,1,t)p(1,0,t) - p(1,1,t)(1 - p(0,0,t)) \end{aligned}$$

we obtain

$$D(t) = p(1,1,t) - p_1(1,0)p_2(1,0).$$

This gives Equation 38 for $\boldsymbol{x} = (1,1)$. The other cases are proven in the same way.

Without selection we have from Equation 31

$$p(\boldsymbol{x}, t+1) = p(\boldsymbol{x}, t) + (-1)^{|\boldsymbol{x}|^2+1}\frac{1}{2}D(t). \tag{40}$$

By computing D(t+1) Equation 39 is obtained.

□

We will use as a measure for the deviation from Robbins' proportions the mean square error $DSQ(t)$

$$DSQ(t) = \sum_{\boldsymbol{x}} \left(p(\boldsymbol{x}, t) - p_1(x_1)p_2(x_2)\right)^2. \tag{41}$$

From the above Theorem we obtain

Corollary: *For two loci we have*

$$DSQ(t+1) = \frac{1}{4}DSQ(t)$$

The genotype frequencies $p(\boldsymbol{x}, t)$ converge to Robbins' proportions for $t \to \infty$.

The limit distribution for an arbitrary number of loci will be investigated in the next section.

5 Recombination without selection

The problem of determining the limit distribution was solved by Geiringer (1944). She considered diploid organisms, where the genes may be linked (in modern genetics terms, the genes may be on the same chromosome). The classical Mendel's laws are valid for genes of different chromosomes — unlinked genes — only.

Theorem 6 (Geiringer) *The limit distribution is the product of the n univariate marginal distributions $p_i(x_i)$, which are derived from $p(\boldsymbol{x}, 0)$, the distribution of gametes in the initial population.*

Note that the limit distribution is independent of the specific recombination method used. The special case, assuming no linkage between genes, was already solved by Tietze (1923) in a very interesting, but rather involved paper. The proof by Geiringer is simpler and shorter, but still very sophisticated.

A rather informal proof of Geiringer's theorem for haploid organism was done by Holland (1992). We will concentrate on the speed of convergence to Robbins' proportions by numerical simulations. For more than 2 loci, the equations for uniform crossover and one-point crossover are different. Uniform crossover should convergence faster, because it mixes the genes much more than one-point crossover. Table 3 gives numerical results for $n = 8$ loci.

It is very difficult to obtain empirical laws from our simulations because of the stochastic fluctuations in a finite population. This is demonstrated by Table 4. There the numerical value for $DSQ(t)$ is shown for different population sizes. In addition $c = DSQ(t+1)/DSQ(t)$ is displayed. From Theorem 5 a factor of $c = 0.25$ is expected.

t	q_0	q_1	q_2	$q_0 - 1/2^n$	$q_1 - 1/2^n$	$q_2 - 1/2^n$
0	0.5	0.0	0.0	0.4961	-3.906E-3	-3.906E-3
1	0.3774	0.0177	0.0	0.3735	1.369E-2	-3.906E-3
2	0.2879	0.0262	0.0012	0.2840	2.229E-2	-2.706E-3
3	0.2225	0.0303	0.0028	0.2186	2.639E-2	-1.106E-3
4	0.1768	0.0314	0.0042	0.1729	2.749E-2	+0.294E-3
5	0.1421	0.0298	0.0050	0.1382	2.589E-2	+1.094E-3
0	0.5	0.0	0.0	0.4961	-3.906E-3	-3.906E-3
1	0.2533	0.0020	0.0023	0.2494	-1.927E-3	-1.646E-3
2	0.0895	0.0097	0.0101	0.0856	+5.834E-3	+6.174E-3
3	0.0323	0.0093	0.0102	0.0283	+5.434E-3	+6.244E-3
4	0.0148	0.0074	0.0072	0.0108	+3.574E-3	+3.254E-3
5	0.0090	0.0057	0.0056	0.0051	-1.794E-3	+1.794E-3

Table 3: Comparison of convergence to Robbins' proportions for $n = 8$ loci, onepoint (upper half) and uniform crossover, $q_0 = p(0,\ldots,0)$, $q_1 = p(0,\ldots,1)$, $q_2 = p(0,\ldots,1,0)$, population size $N = 1000$, averages over 100 runs.

	$N = 1000$		$N = 10000$		$N = 20000$	
t	DSQ	c	DSQ	c	DSQ	c
0	9.00E-2		9.00E-2		9.00E-2	
1	2.28E-2	0.25	2.25E-2	0.25	2.25E-3	0.25
2	6.37E-3	0.28	5.77E-3	0.25	5.59E-3	0.25
3	2.34E-3	0.37	1.55E-3	0.27	1.45E-3	0.26
4	1.62E-3	0.70	4.96E-4	0.32	4.24E-4	0.29
5	1.91E-3	1.03	2.43E-4	0.49	1.67E-4	0.39
8	2.62E-3	1.13	2.25E-4	1.10	1.41E-4	1.10

Table 4: Convergence to linkage equilibrium for n=2 loci

The numerical simulations show that finite populations behave as expected for a short time only. For a population of $N = 1000$ the minimum deviation DSQ_{min} from Robbins' proportions is already achieved after four generations, then DSQ slowly increases due to stochastic fluctuations by genetic drift. Ultimately the population will consist of one genotype only. Genetic drift has been analyzed by Asoh & Mühlenbein (1994b). It will not be considered here.

Table 5 shows the value of c for larger values of n. We see that the reduction factor c decreases for increasing n in generations 2 and 3. We will derive this result analytically. The analysis is valid for $N \ll 2^n$ and $n > 8$. Let q_i denote the frequency of genotype $\boldsymbol{x}$, where i is the integer representation of the binary string $\boldsymbol{x}$.

Theorem 7 *The minimal deviation DSQ_{min} from Robbins' proportions is given for $N \ll 2^n$ and $n > 8$ by*

$$DSQ_{min} = \frac{1}{N} - 2^{-n} \tag{42}$$

	n			
t	4	8	12	16
1	0.25	0.25	0.25	0.25
2	0.23	0.13	0.08	0.07
3	0.25	0.15	0.07	0.03
4	0.26	0.24	0.25	0.24
5	0.31	0.38	0.65	0.84
6	0.49	0.63	0.89	0.95

Table 5: Reduction factor c averaged over 100 runs, $N = 10000$, uniform crossover

Proof: For $N \ll 2^n$ we have at equilibrium about N different genotypes. $2^n - N$ genotypes are not represented in the population. Therefore

$$\begin{aligned} DSQ_{min} &= N(\frac{1}{N} - 2^{-n})^2 + (2^n - N)(0 - 2^{-n})^2 \\ &= \frac{1}{N} - 2^{-n} \end{aligned}$$

□

Next we explain why c is decreasing for generation 2 with increasing n.

Lemma: *For $N \ll 2^n$ and starting with $q_0 = q_{2^n-1} = 0.5$ we have at generation 1*

$$\begin{aligned} q_0 &\approx \frac{1}{4} \\ q_i &\approx 0 \qquad 0 < i < 2^n - 1 \\ q_{2^n-1} &\approx \frac{1}{4} \end{aligned}$$

This gives $DSQ(1) \approx \frac{1}{8}$ and $c(1) = \frac{1}{4}$. In generation 2 we have for $n > 16$

$$\begin{aligned} q_0 &\approx \frac{1}{16} \\ q_{2^n-1} &\approx \frac{1}{16}. \end{aligned}$$

This gives approximately $DSQ(2) \approx \frac{1}{128}$ and $c(2) = \frac{1}{16}$.

Proof: We compute the probability that in generation 1 genotype q_0 is obtained. The genotype appears as a result of mating between q_0 and q_0 and q_0 with q_{2^n-1}. This gives the probability

$$q_0 = \frac{1}{2} * \frac{1}{2} + 2\frac{1}{4}2^{-n} \approx \frac{1}{4}$$

The same probability is obtained for q_{2^n-1}. Furthermore,

$$q_i = 2\frac{1}{4}2^{-n} = 2^{-n-1} \approx 0 \qquad 0 < i < 2^n - 1.$$

A similar computation can be done for the second generation. We obtain for $n > 16$ approximately $q_0 = q_{2^n-1} \approx \frac{1}{16}$.

□

Remark: For the limit $n \to \infty$, $N \ll 2^n$, we conjecture using the same arguments as in the above proof $DSQ(1) = 2^{-3}$, $DSQ(2) = 2^{-7}$, $DSQ(3) = 2^{-13}$, $DSQ(4) = 2^{-25}$ etc. This shows how fast uniform crossover is moving into the direction of Robbins' proportions.

We now summarize the results of the simulation and the theoretical analysis. A finite population normally will not exactly reach Robbins' proportions. It will remain at a minimal distance of DSQ_{min}. The number of steps to reach DSQ_{min} depends mainly on N. It is for $N \ll 2^n$ almost independent of n. In all simulations with $n > 4$ the deviation from Robbins' proportions was less than 1% after 7 generations. This strongly supports to analyze genetic algorithms by assuming Robbins' proportions.

The work of Geiringer was recently rediscovered by Booker (1993). He wrote: "Geiringer's convergence results suggest that the most important difference among recombination operators is the rate at which they converge to equilibrium in the absence of selection." We will later show that it is difficult to extrapolate recombination results without selection to results with selection. The really astonishing result is that *without selection all reasonable two-parent recombination methods converge to the same equilibrium - given by Robbins' proportions.* This means that any given distribution of 2^n genotype frequencies will converge to a distribution defined by n variables only - the univariate marginal distributions.

Therfore we conclude that the theorem supports concentrating the theoretical analysis to gene frequencies being in Robbins' proportions. Given our numerical results we conjecture that all two-parent recombination operators create genotype frequencies which are fluctuations around the trajectory given by Robbin's proportions.

We show in the next section that all two-parent recombination operators give the same univariate gene frequencies, even after one step of selection.

6 Difference equations for univariate marginal frequencies

The difference equations for genotype frequencies soon get complicated if the number of loci increases. They can only be obtained by a computer program. A nice calculus for computing these equations has been developed by Vose and Wright (1994), though these equations are too detailed for our purposes. They describe the evolution of all possible genotypes, which means that for $n = 15$ there are 32768 independent variables!

The following theorem gives the difference equations for the univariate marginal frequencies.

Theorem 8 *For proportionate selection the univariate marginal frequencies are determined by*

$$p_i(v, t+1) = \sum_{\boldsymbol{x}|x_i=v} \frac{p(\boldsymbol{x}, t) f(\boldsymbol{x})}{\bar{f}(t)}. \tag{43}$$

This equation is valid for any *recombination/crossover scheme based on the Mendelian chance model.*

Proof: After selection the univariate marginal frequencies are given by

$$p_i^s(v,t) = \sum_{\boldsymbol{x}|x_i=v} p^s(\boldsymbol{x},t) = \sum_{\boldsymbol{x}|x_i=v} \frac{p(\boldsymbol{x},t)f(\boldsymbol{x})}{\bar{f}(t)}.$$

Now the selected individuals are randomly paired. Since Mendelian recombination does not change the allele frequencies, these operators do not change the univariate marginal frequencies. Therefore

$$p_i(v,t+1) = p_i^s(v,t).$$

□

This result is very important. For $n = 2$ we have already proven it with Equation 33. A number of conclusions can be derived.

Let $H_i(v) = (*,\ldots,*,v,*,\ldots,*)$ be a first-order schema at locus i. This schema includes all strings where the gene at locus i is fixed at v. Then the fitness of the schema at generation t is given by (Holland, 1992):

$$f(H_i(v),t) = \frac{1}{p_i(v,t)} \sum_{\boldsymbol{x}|x_i=v} p(\boldsymbol{x},t)f(\boldsymbol{x}) \tag{44}$$

Our univariate marginal frequency $p_i(v,t)$ is obviously identical to the frequency of schema H_i. From Theorem 8 we obtain:

Corollary (First-order schema theorem): *For a genetic algorithm with proportionate selection using any Mendelian recombination the frequency of first-order schemata changes according to*

$$p_i(v,t+1) = p_i(v,t)\frac{f(H_i(v),t)}{\bar{f}(t).} \tag{45}$$

Note that the above corollary is valid for an infinite population with proportionate selection and recombination. Holland's famous schema theorem (1992) implies for first order schemata (schema defining length of 0)

$$p_i(v,t+1) \geq p_i(v,t)\frac{f(H_i(v),t)}{\bar{f}(t)}.$$

Theorems using univariate marginal distributions are of limited use for prediction. They can only make single-step predictions. The computation of $f(H_i,t)$ and $\bar{f}(t)$ needed for Equation 45 requires all genotype frequencies $p(\boldsymbol{x},t)$. The next corollary directly follows from Theorem 8.

Corollary: *If $p^*(\boldsymbol{x})$ is a fixed point of a genetic algorithm with proportionate selection, then*

$$\bar{f}^* = f^*(H_i(v)) \qquad i = 1,\ldots,n \tag{46}$$

A necessary condition for a fixed point is that the fitness of all first order schemata is equal to the average fitness.

The above condition is necessary, but not sufficient for a fixed point. The following example shows this. Let the fitness function be defined as $f(0,0) = f(1,1) = 0$, $f(1,0) = f(0,1) = 1$. Then for identical genotype frequencies $p(\boldsymbol{x}) = 1/4$ the condition is fulfilled,

but $p(\boldsymbol{x})$ is not a fixed point for uniform crossover. For UMDA counterexamples can be also found.

If the genotype frequencies are in Robbins' proportions, an expression using only univariate marginal distributions can be given. The corollary immediately follows from Theorem 8.

Corollary: *Let the genotype frequencies be in Robbins' proportions. Then for any genetic algorithm with proportionate selection the univariate marginal frequencies obey the difference equation*

$$p_i(v,t+1) = p_i(v,t)\frac{\bar{f}_i(v,t)}{\bar{f}(t)}, \tag{47}$$

where

$$\bar{f}_i(v,t) = \sum_{\boldsymbol{x}|x_i=v} f(\boldsymbol{x}) \prod_{\substack{j=1 \\ j\neq i}}^{n} p_j(x_j,t). \tag{48}$$

The difference equation (47) can also be written in the form

$$p_i(v,t+1) = p_i(v,t) + p_i(v,t)\frac{F_i(v,t)}{\bar{f}(t)}, \tag{49}$$

where

$$F_i(v,t) = \bar{f}_i(v,t) - \bar{f}(t). \tag{50}$$

The expression $F_i(v,t)$ was introduced by Asoh and Mühlenbein (1994a), but was denoted $f_{(i)}(v,t)$. These values minimize

$$\sum_{\boldsymbol{x}} p(\boldsymbol{x},t)\left(f(\boldsymbol{x}) - \bar{f}(t) - \sum_{i=1}^{n} g_i(x_i,t)\right)^2,$$

for varying $g_i(x_i,t)$. $\sum_{i=1}^{n} F_i(x_i,t)$ is the best additive approximation to $f(\boldsymbol{x}) - \bar{f}(t)$. The expressions are used to define the *additive genetic variance* $V_A(t)$, which was introduced briefly in Section 2.

$$V_A(t) = \sum_{i=1}^{n}\sum_{v=0}^{L} p_i(v,t)\left(F_i(v,t)\right)^2. \tag{51}$$

Note that in general a genetic algorithm is not fully described by the n univariate marginal distributions. Even if the genotype frequencies are in Robbins' proportions, they will in general be in linkage disequilibrium after one step of selection.

We are now ready to prove the main theorem, which is related to Fisher's Fundamental Theorem of Natural Selection (Fisher, 1958). It is the exact RS equation for proportionate selection.

Theorem 9 *Let the genotype frequencies be in Robbins' proportions. Then for any genetic algorithm with proportionate selection the response to selection is given by*

$$R(t) = \frac{V_A(t)}{\bar{f}(t)} + \sum_{\boldsymbol{x}} \Delta p(\boldsymbol{x})\left(f(\boldsymbol{x}) - \bar{f}(t) - \sum_{i=1}^{n} F_i(x_i,t)\right) \tag{52}$$

where $\Delta p(\boldsymbol{x}) = p(\boldsymbol{x},t+1) - p(\boldsymbol{x},t)$.

Proof: We have $\sum_{\boldsymbol{x}} \Delta p(\boldsymbol{x}) f(\boldsymbol{x}) = R(t)$ and $\sum_{\boldsymbol{x}} \Delta p(\boldsymbol{x}) = 0$. Let $\Delta p_i(v) = p_i(v, t+1) - p_i(v, t)$. Then

$$\begin{aligned}
\sum_{\boldsymbol{x}} \Delta p(\boldsymbol{x}) \sum_{i=1}^{n} F_i(x_i, t) &= \sum_{i=1}^{n} \sum_{v=0}^{L} F_i(v, t) \sum_{\boldsymbol{x}|x_i=v} \Delta p(\boldsymbol{x}) \\
&= \sum_{i=1}^{n} \sum_{v=0}^{L} F_i(v, t) \Delta p_i(v) \\
&= \sum_{i=1}^{n} \sum_{v=0}^{L} p_i(v, t) F_i^2(v, t) / \bar{f}(t) \\
&= \frac{V_A(t)}{\bar{f}(t)}
\end{aligned}$$

Summing up all terms gives the response equation. We used that for Robbins' proportions $\sum_{\boldsymbol{x}|x_i=v} \Delta p(\boldsymbol{x}) = \Delta p_i(v, t)$. Furthermore Equation 49 was inserted.
□

For a genetic algorithm with two-parent recombination TPR the theorem can be used for one step only, because the genotype frequencies will not remain in Robbins' proportions.

Fisher (1958) stated his theorem as follows: *"The rate of increase in fitness of any organisms at any time is equal to its genetic variance in fitness at that time",* or mathematically $R(t) \approx V_A(t)$. Fisher assumed continuous generations, which leads to differential equations instead of difference equations. For discrete generations the corresponding expression would be $R(t) \approx V_A(t)/\bar{f}(t)$. This is just the first factor of Equation 52. The second term is of second order, because it is a summation of a product of the changes of the genotype frequencies times the error between $f(\boldsymbol{x})$ and its best additive approximation. The theorem indicates that Fisher's theorem is approximately correct.

By neglecting the sum in Equation 52 we obtain:

Corollary: *Let the genotype frequencies be in Robbins' proportions. Then the realized heritability can be estimated by*

$$b(t) \approx h^2(t) = \frac{V_A(t)}{V(t)}. \tag{53}$$

The estimate is valid for any genetic algorithm with proportionate selection.

Proof:

$$R(t) \approx \frac{V_A(t)}{\bar{f}(t)} = \frac{V_A(t)}{V(t)} \frac{V(t)}{\bar{f}(t)} = \frac{V_A(t)}{V(t)} S(t)$$

□

It is known in population genetics that Fisher's theorem is mathematically false. There are even counterexamples with $R(t) < 0$. In these counterexamples the population is not in linkage equilibrium. But the biological interpretation of Fisher's conjecture is still open. Can it be that in nature we only find fitness functions where Fisher's theorem is valid? Second, how important is linkage disequilibrium in natural populations? For a recent discussion of Fisher's theorem in population genetics the interested reader should consult Ewens (1989, 1992, 1995).

A different expression for $R(t)$ has been given by Altenberg (1995). Altenberg derived a general formula for any kind of recombination scheme. But even if linkage equilibrium is assumed, Altenberg's formula is very difficult to apply for a given fitness function.

There have been other approaches to obtain a more precise equation for the response. The most promising approach seems to extend the equation for the response to selection to a set of equations using higher order moments or cumulants. First steps into this direction have been made by Prügel-Bennet and Shapiro (1994) and Rattray (1995). They have been able to compute the cumulants for quadratic fitness functions using concepts of statistical mechanics. In population genetics Bulmer (1980) already introduced cumulants. His work was extended by Turelli and Barton (1994).

The application of Theorem 9 to a genetic algorithms with two-parent recombination is limited. Therefore we introduce in the next section an algorithm, that keeps the population in Robbins' proportions. This algorithm is completely defined by univariate marginal frequencies.

7 Univariate marginal distribution algorithm

There is a simple recombination scheme that maintains the population in Robbins' proportions; we call it *gene-pool recombination (GPR)* (Mühlenbein & Voigt, 1996). In GPR, the two alleles to be recombined at each locus are chosen *independently* from the gene-pool defined by the selected parent population. The biologically inspired idea of restricting recombination to the alleles of two-parents for each offspring is abandoned.

Definition: *In gene-pool recombination the two "parent" alleles of an offspring are randomly chosen for each locus with replacement from the gene-pool given by the selected parents. The offspring allele is then computed using any of the standard recombination schemes for two-parent recombination.*

For binary functions the bit-based simulated crossover (BSC) of Syswerda (1993) is similar to GPR. However, his implementation merges selection and recombination. An implementation of BSC that separates selection and recombination was empirically investigated by Eshelman and Schaffer (1993). GPR is a generalization of BSC; it can be used for any representation — discrete or continuous. For a discussion of gene-pool recombination and its analysis, see Mühlenbein and Voigt (1996). Gene-pool recombination leads to difference equations for the univariate marginal frequencies $p_i(v)$. Here we generalize this idea and define a conceptual algorithm that does not recombine chromosomes but uses univariate marginal frequencies instead.

The general form of the Univariate Marginal Distribution Algorithm (UMDA) is as follows:

UMDA

- **STEP 0:** Set $t \Leftarrow 1$. Generate $N \gg 0$ points randomly.
- **STEP 1:** Select $M \leq N$ points according to a selection method. Compute the marginal frequencies $r_i(x_i, t)$ of the selected set.
- **STEP 2:** Generate N new points according to the distribution $p(\boldsymbol{x}, t+1) = \prod_{i=1}^{n} r_i(x_i, t)$. Set $t \Leftarrow t+1$.

- **STEP 3:** If termination criteria are not met, go to STEP 1.

From Equation 49 it follows:

Corollary: *For proportionate selection, the UMDA stays in equilibrium iff* $V_A = 0$.

The response to selection is zero if the additive variance is zero. *UMDA only exploits the additive genetic variance.*

The corollary implies UMDA is not a global optimization method for highly non-linear functions characterized by a significant V_2 contribution (see Equation 10) and a small V_A contribution, since evolution stops if $V_A = 0$. This limitation has already been shown by simulation for the case of gene-pool recombination by Mühlenbein and Voigt (1996). *For UMD algorithms the response is zero iff the additive genetic variance* V_A *is zero.*

We have been able to prove a weak form of Fisher's theorem by using an inequality from Baum and Eagon (1967).

Theorem 10 *For UMDA we have* $R(t) > 0$ *unless all univariate marginal frequencies remain the same.*

The Theorem is proven in Appendix 1.

We have not been able to extend Theorem 9 to populations which are in linkage disequilibrium. In the next Section we will in detail compare UMDA with genetic algorithms using two-parent recombination for the case of just two loci.

8 The exact response equation for two loci

In order to investigate Fisher's fundamental theorem rigorously, we will compute in this section exact equations for the response. We assume proportionate selection, linkage equilibrium and for notational convenience binary genes. The computation for two loci is already very tedious, indicating that an exact analysis for three loci by this method would be difficult.

Theorem 11 *Let the fitness be given by* $f(0,0) = f_0, f(1,0) = f(0,1) = f_1, f(1,1) = f_2$. *Let the genotype frequencies at generation* t *be given by* $p(0,0,t) = (1-p)^2$,$p(0,1,t) = p(1,0,t) = p*(1-p)$,$p(1,1,t) = p^2$. *Then for uniform crossover the response is given by*

$$R(t) = \frac{V_A(t) + \frac{1}{2}V_2(t)}{\bar{f}(t)} + p(1-p)(f_2 - 2f_1 + f_0)\frac{V_A}{4\bar{f}(t)^2}. \tag{54}$$

Proof: From the definitions we get

$$\begin{aligned} V &= p^2 f_2^2 + 2p(1-p)f_1^2 + (1-p)^2 f_0^2 - \bar{f}(t)^2 \\ V_A &= 2\left(pF_1^2(1,t) + (1-p)F_1^2(0,t)\right) \\ &= 2p\left((1-p)f_1 + pf_2 - \bar{f}(t)\right)^2 + 2(1-p)\left(pf_1 + (1-p)f_0 - \bar{f}(t))\right)^2. \end{aligned}$$

The computation is straightforward, but tedious. Therefore we show some important intermediate steps. We start with

$$\begin{aligned} V - V_A &= p^2 f_2^2 + 2p(1-p) f_1^2 + (1-p)^2 f_0 - \bar{f}(t)^2 \\ &\quad -2p(1-p)^2 f_1^2 - 2p^3 f_2^2 - 4p^2(1-p) f_1 f_2 + 4p(1-p) f_1 \bar{f}(t) + 4p^2 f_2 \bar{f}(t) \\ &\quad -2(1-p)p^2 f_1^2 - 2(1-p)^3 f_0^2 - 2\bar{f}(t)^2 \\ &\quad -4(1-p)^2 p f_1 f_0 + 4(1-p) p f_1 \bar{f}(t) + 4(1-p)^2 f_0 \bar{f}(t). \end{aligned}$$

Collecting all terms with $\bar{f}(t)$ we obtain

$$\begin{aligned} V - V_A &= p^2 f_2^2 + 2p(1-p) f_1^2 + (1-p)^2 f_0 + \bar{f}(t)^2 \\ &\quad -2(1-p)p^2 f_1^2 - 2(1-p)^3 f_0^2 - 2p(1-p)^2 f_1^2 \\ &\quad -2p^3 f_2^2 - 4p^2(1-p) f_1 f_2. \end{aligned}$$

Combining all coefficients we obtain the expression

$$V_2 = V - V_A = p^2(1-p)^2 (f_2 - 2f_1 + f_0)^2. \tag{55}$$

We now use the expression of $R(t)$ derived in Theorem 4. In the special case considered here, the term $D_s(t)$ is just

$$D_s(t) = p^2(1-p)^2 (f_0 f_2 - f_1^2).$$

Therefore

$$R(t) = \frac{V_A(t) + V_2(t)}{\bar{f}(t)} - \frac{1}{2} \frac{p^2(1-p)^2}{\bar{f}(t)^2} (f_0 - 2f_1 + f_2)(f_0 f_2 - f_1^2).$$

Inserting V_2 into this equation gives

$$R(t) = \frac{V_A(t) + \frac{1}{2} V_2(t)}{\bar{f}(t)} + \frac{1}{2} p^2(1-p)^2 \frac{f_0 - 2f_1 + f_2}{\bar{f}(t)} \left(f_0 - 2f_1 + f_2 - \frac{f_0 f_2 - f_1^2}{\bar{f}(t)} \right).$$

After some computation we obtain

$$(f_0 - 2f_1 + f_2)\bar{f}(t) - f_0 f_2 + f_1^2 = \left(p f_2 + (1-2p) f_1 - (1-p) f_0 \right)^2.$$

We note that V_A can be written as

$$V_A = 2p(1-p) \left(p f_2 + (1-2p) f_1 - (1-p) f_0 \right)^2 \tag{56}$$

Inserting V_A into the above equation completes the proof.
□

Using the decomposition from Theorem 1 we obtain

Corollary: (Robertson's/Price's Theorem) *Under the assumptions of Theorem 11 the response is given for uniform crossover by*

$$R(t) = 2 \frac{cov(F_{mp}(t), F_o(t))}{\bar{f}(t)} + p(1-p)(f_2 - 2f_1 + f_0) \frac{V_A}{4\bar{f}(t)^2}. \tag{57}$$

The approximation

$$R(t) \approx 2\frac{cov(F_{mp}(t), F_o(t))}{\bar{f}(t)}$$

is called Robertson's or Price's version of Fisher's theorem (Lynch & Walsh, 1997;Altenberg, 1995). We have derived the exact equation for two loci. It is not easy to decide whether the second term is small compared to the first term. But it follows that for uniform crossover the regression coefficient is a more accurate estimate for realized heritability than V_A/V. Especially it follows that if $V_A = 0$, the response is exactly $V_2/(2\bar{f}(t))$.

It has to be noted that an equation very similar to Equation 57 has been proven by Nagylaki (1991) for diploid organism and one locus. The reader should consult Nagylaki(1991) and Lynch & Walsh(1997) if he is interested how researchers in quantitative genetics tried to compute an exact response equation.

Next we derive the exact response equation for UMD algorithms.

Theorem 12 *Under the assumptions of Theorem 11 the response for UMD algorithms is given by*

$$R(t) = \frac{V_A(t)}{\bar{f}(t)} + p(1-p))(f_2 - 2f_1 + f_0)\frac{V_A(t)}{2\bar{f}(t)^2} \tag{58}$$

Proof: The difference equation for the univariate marginal frequency is given by

$$p(t+1) = p(t)\frac{(1-p(t))f_1 + pf_2}{\bar{f}(t)}$$

Inserting this expression into $R(t+1) = \bar{f}(t+1) - \bar{f}(t)$ gives the conjecture.
□

We now compare the two Theorems. Let

$$error(t) = p(1-p)(f_2 - 2f_1 + f_0)\frac{V_A(t)}{2\bar{f}(t)^2}.$$

Uniform crossover is an instance of two-parent recombination (TPR). Using this notation we have shown

$$\begin{aligned} R_{TPR}(t) &= \frac{V_A(t) + \frac{1}{2}V_2(t)}{\bar{f}(t)} + \frac{1}{2}error(t), \\ R_{UMDA}(t) &= \frac{V_A(t)}{\bar{f}(t)} + error(t). \end{aligned}$$

The structure of the response equation for TPR and for UMDA is very similar. The error term for TPR is just one half of the error term for UMDA. In particular $error = 0$ if $V_A = 0$ or the function is linear ($f_2 - 2f_1 + f_0 = 0$)

We discuss the result with an example.

Example: $f_0 = 2, f_1 = 1, f_2 = 2$; $p(0) = 0.5$
One computes $\bar{f}(0) = 1.5$,$V_A(0) = 0$, $V_2(0) = 0.25$, $error = 0$. The response is given by $R(0) = V_2(0)/(2\bar{f}(0)) = 1/12$. Equation 54 exactly predicts the response. For TPR

the equilibrium is given by $p(0,0) = p(1,1) = 0.320194$ and $p(0,1) = p(1,0) = 0.179806$. For UMDA $p(0) = 0.5$ is an instable equilibrium.

In order to show the dependency of the genotype frequency dynamics from the fitness function, we disturb the fitness values a tiny fraction ($f_2 = 2.01$).

t	p(0,0)	p(0,1)	p(1,1)	$\bar{f}(t)$	$V(t)$	$V_A(t)$	$V_A(t)/V(t)$	$R(t)/S(t)$
0	0.250	0.250	0.250	1.5025	0.2525	0.000012	0.000050	0.000058
1	0.249	0.250	0.251	1.5025	0.2525	0.000022	0.000088	0.000103
10	0.210	0.248	0.294	1.5064	0.2529	0.003957	0.015647	0.018238
15	0.105	0.219	0.457	1.5664	0.2502	0.056296	0.225002	0.256628
16	0.075	0.199	0.527	1.6074	0.2437	0.083901	0.344156	0.386963
17	0.047	0.170	0.613	1.6661	0.2286	0.111818	0.489138	0.539302
18	0.025	0.134	0.708	1.7401	0.1995	0.127508	0.639117	0.688394
19	0.011	0.095	0.799	1.8190	0.1563	0.120235	0.769219	0.809377
20	0.004	0.060	0.876	1.8886	0.1079	0.093271	0.864725	0.892379
30	0.000	0.000	0.999	2.0098	0.0002	0.000160	0.999845	0.999884

Table 6: $f_0 = 2, f_1 = 1, f_2 = 2.01$, UMDA

In Table 6 the run is shown for UMDA. We observe that for about 10 generations the average of the population and the variance remain almost the same. The additive genetic variance is almost zero, therefore the response is very small. It takes UMDA some time to move away from the equilibrium point. From generation 15 on the genotype frequencies move quickly to the optimum.

In Table 7 data is presented for TPR.

t	p(0,0)	p(0,1)	p(1,0)	p(1,1)	$f(t)$
0	0.250	0.250	0.250	0.250	1.50250
1	0.291	0.208	0.208	0.293	1.58654
2	0.307	0.190	0.190	0.311	1.62189
3	0.313	0.184	0.184	0.320	1.63568
4	0.314	0.181	0.181	0.324	1.64092
5	0.313	0.180	0.180	0.327	1.64292
11	0.287	0.179	0.179	0.355	1.64580
20	0.141	0.151	0.151	0.558	1.70440
30	0.002	0.009	0.009	0.980	1.99225

Table 7: $f_0 = 2, f_1 = 1, f_2 = 2.01$, TPR

We observe that TPR has a quick start. The first response is large. But TPR is heading to the equilibrium defined by the fitness values $f_0 = f_2 = 2$. It spins a long time nearby this equilibrium before it moves to the optimum. In comparison, UMDA is moving faster to the optimum than TPR despite its slow start.

We have many similar results obtained. This leads us to the conclusion: *The dynamics of TPR is more difficult than that of UMDA. But for two loci there is no indication that TPR is more efficient for optimization than GPR.*

In the next section we will derive an exact equation for the response for an arbitrary number of loci n.

9 The exact response equation for proportionate selection

Equation 52 is an exact expression for the response. The equation shows that the response is strongly influenced by the additive genetic variance V_A. But it is difficult to estimate the second term, the error. The error term is a summation over 2^n genotypes. Furthermore $\Delta p(\boldsymbol{x})$ is needed.

In this section we will derive an exact equation for the response by a different method. The equation is a generalization of the equation derived in the previous section. It uses marginal frequencies only. The proof of the equation is based on the multivariate Taylor expansion.

We recall that the average fitness of the population at generation t is given by

$$\bar{f}(t) = \sum_{\boldsymbol{x}} p(\boldsymbol{x}, t) f(x),$$

where

$$p(\boldsymbol{x}, t) = \prod_{i=1}^{n} p_i(x_i, t).$$

For notational convenience we consider binary genes $x_i \in \{0, 1\}$. Then we have $p_i(1, t) = 1 - p_i(0, t)$. We abbreviate $p_i(1, t) = p_i(t)$. If the dependency from t is obvious, we just write p_i. In order to explicitly formulate the dependency of the average from the marginal frequencies p_i, we write

$$W(p_1, \ldots, p_n, t) = \bar{f}(t). \tag{59}$$

For a differentiable function $g(p_1, \ldots, p_n)$ of n variables the multivariate Taylor expansion is given by

$$\begin{aligned} g(p) &= g(a) + \sum_{j=1}^{n} (p_j - a_j) \frac{\partial g}{\partial p_j} \mid_{\mathbf{p}=\mathbf{a}} + \frac{1}{2!} \left(\sum_{j=1}^{n} (p_j - a_j) \frac{\partial}{\partial p_j} \right)^2 g \mid_{\mathbf{p}=\mathbf{a}} \\ &\quad + \frac{1}{3!} \left(\sum_{j=1}^{n} (p_j - a_j) \frac{\partial}{\partial p_j} \right)^3 g \mid_{\mathbf{p}=\mathbf{a}} + \ldots\ldots, \end{aligned} \tag{60}$$

where $\mathbf{p} = (p_1, \ldots, p_n)$, $\mathbf{a} = (a_1, \ldots, a_n)$ and the operators $\partial / \partial p$ are multiplied formally. We are now ready to state the main Theorem.

Theorem 13 *For UMDA with proportionate selection the response to selection is given by*

$$\begin{aligned} R(t) &= \frac{V_A(t)}{W} + \frac{1}{2} \sum_{i \neq j} \frac{p_i(t) F_i(1,t) p_j(t) F_j(1,t)}{W^2} \frac{\partial^2 W}{\partial p_i \partial p_j} \\ &\quad + \frac{1}{3!} \sum_{i \neq j, j \neq k, i \neq k} \frac{p_i(t) F_i(1,t) p_j(t) F_j(1,t) p_k(t) F_k(1,t)}{W^3} \frac{\partial^3 W}{\partial p_i \partial p_j \partial p_k} + \ldots. \end{aligned} \tag{61}$$

Proof: We make a Taylor expansion with $\mathbf{p} = \mathbf{p}(t+1)$ and $\mathbf{a} = \mathbf{p}(t)$. Let $\Delta p_i = p_i(t+1) - p_i(t)$. We recall that

$$\begin{aligned}\Delta p_i &= p_i(t)\frac{F_i(1,t)}{W} \\ \Delta p_i &= -(1-p_i(t))\frac{F_i(0,t)}{W}\end{aligned}$$

Noting that W has a special structure - each p_i occurs only once in $p(\boldsymbol{x},t)$ - the Taylor expansion immediately gives the expressions containing partial derivatives of order two and higher. We are left to prove that the first term contains the additive genetic variance. By simple manipulation we obtain

$$\begin{aligned}V_A(t) &= \sum_{i=1}^{n}(1-p_i(t))F_i(0,t)^2 + p_i(t)F_i(1,t)^2 \\ &= \sum_{i=1}^{n}(1-p_i(t))F_i(0,t)(-W\frac{\Delta p_i}{1-p_i(t)}) + p_i(t)F_i(1,t)W\frac{\Delta p_i}{p_i(1,t)} \\ &= W\sum_{i=1}^{n}\Delta p_i(F_i(1,t)-F_i(0,t)) \\ &= W\sum_{i=1}^{n}\Delta p_i\frac{\partial W}{\partial p_i},\end{aligned}$$

because obviously

$$\frac{\partial W}{\partial p_i} = F_i(1,t) - F_i(0,t).$$

Dividing the equation by W gives the Theorem.
□

Remark: *Because $F_i(1,t) = (1-p_i(t))(F_i(1,t)-F_i(0,t))$ the difference equation for the univariate marginal frequencies can also be written*

$$\Delta p_i = p_i(t)(1-p_i(t))\frac{F_i(1,t)-F_i(0,t)}{W}. \tag{62}$$

Corollary: *For two loci the response of UMDA is given by*

$$R(t) = \frac{V_A(t)}{W} + \frac{p_1(t)F_1(1,t)p_2(t)F_2(1,t)}{W^2}(f(1,1)-f(1,0)-f(0,1)+f(0,0)) \tag{63}$$

For the special case $p_1 = p_2 = p$, $F_1 = F_2$, $f(0,1) = f(1,0)$ we have

$$R(t) = \frac{V_A(t)}{W} + p(1-p)\frac{V_A(t)}{2W^2}(f(1,1)-f(1,0)-f(0,1)+f(00)). \tag{64}$$

Proof: The first equation directly follows from Equation 61. Only equation 64 has to be proven. In this special case we have

$$pF_1(1,t)^2 + (1-p)F_1(0,t)^2 = V_A(t)/2.$$

From

$$F_1(0,t)^2 = \frac{p^2F_1(1,t)^2}{(1-p)^2}$$

we obtain

$$V_A(t)/2 = pF_1(1,t)^2 + p^2\frac{F_1(1,t)^2}{1-p} = \frac{p}{1-p}F_1(1,t)^2.$$

Therefore

$$p^2F_1(1,t)^2 = p(1-p)V_A(t)/2,$$

Noting that for the special case

$$p_1(t)F_1(1,t)p_2(t)F_2(1,t) = p^2F_1(1,t)^2$$

we obtain the conjecture.
□

The corollary shows that for $n = 2$ Theorem 13 correctly gives the equation proven in the previous section. But the proof using the Taylor expansion is much simpler. It is interesting to compute the condition, under which the response is exactly given by V_A/W. Neglecting the trivial case that all $F_i(1,t) = 0$, the necessary and sufficient condition is

$$f(1,1) - f(1,0) - f(0,1) + f(0,0) = 0 \tag{65}$$

It is easy to see that this equation is fulfilled if f is linear. In fact, only linear functions satisfy the equation.

The above technique can be used to explicitly compute the error terms for an arbitrary number of loci. This is left to further research. We just give the error terms for three loci.

10 The response equation for three loci

It is instructive to explicitly compute the response equation for $n = 3$ loci. From Theorem 13 the next corollary can easily be obtained.

Corollary: *Let*

$\alpha_3 = f(1,1,1) - f(1,1,0) - f(1,0,1) - f(0,1,1) + f(1,0,0) + f(0,1,0) + f(0,0,1) - f(0,0,0).$

Then for three loci the response for UMDA is given by

$$\begin{aligned} R(t) \;=\; & \frac{V_A(t)}{W} \\ & + \frac{p_1(t)F_1(1,t)p_2(t)F_2(1,t)}{W^2}(f(0,1,1) - f(0,1,0) - f(0,0,1) + f(0,0,0) + p_3\alpha_3) \\ & + \frac{p_2(t)F_2(1,t)p_3(t)F_3(1,t)}{W^2}(f(1,1,0) - f(1,0,0) - f(0,1,0) + f(0,0,0) + p_1\alpha_3) \\ & + \frac{p_1(t)F_1(1,t)p_3(t)F_3(1,t)}{W^2}(f(1,0,1) - f(1,0,0) - f(0,0,1) + f(0,0,0) + p_2\alpha_3) \\ & + \frac{p_1(t)F_1(1,t)p_2(t)F_2(1,t)p_3(t)F_3(1,t)}{W^3}\alpha_3 \end{aligned} \tag{66}$$

Proof: The proof is straightforward. One just computes the partial derivatives of W.
□

We are now able to compute the conditions under which $R(t) = V_A/W$. Neglecting the trivial case that all $F_i(1,t) = 0$ we obtain the four equations

$$\begin{aligned} 0 &= f(0,1,1) - f(0,1,0) - f(0,0,1) + f(0,0,0) && (67)\\ 0 &= f(1,1,0) - f(1,0,0) - f(0,1,0) + f(0,0,0) && (68)\\ 0 &= f(1,0,1) - f(1,0,0) - f(0,0,1) + f(0,0,0) && (69)\\ 0 &= \alpha_3 && (70) \end{aligned}$$

The first three equations are similar to Equation 65. One has to fix the allele of one of the three loci to 0, then the remaining four fitness values have to fulfill the equation for two loci. If f is a linear function, all four equations are satisfied. It seems that only linear functions fulfill the equations. We motivate this conjecture by counting the number of independent variables. For $n = 3$ loci we have eight fitness values and four equations. This gives four independent variables. These are necessary and sufficient to specify a linear function of three variables.

We have not been able to compute the exact equation for the response of two-parent recombination (TPR). The eight difference equations which describe the evolution of the genotypes of TPR are very long. In order to make a comparison of UMDA and TPR we implemented the difference equations for uniform crossover and made many numerical experiments. The results of the experiments suggest the following conjecture.

Conjecture: *If the fitness function fulfills Equations 67-70 and if the genotypes are in Robbins' proportions, then for a genetic algorithm with uniform crossover the response is given by*

$$R(t) = \frac{V_A(t)}{W}$$

We have proven the conjecture for two loci in Section 8. The conjecture indicates that the structure of the response equation for TPR and UMDA is fairly similar, if genotypes are in Robbins' proportions. We discuss the conjecture and the problem of Robbins' proportions with a numerical example. For notational convenience we sort the genotypes according to their integer value.

Example: Let the fitness function be defined by the values $(5,6,3,4,8,9,6,7)$. It is a linear function. Simulation results for TPR and UMDA are given in Table 8.
For this fitness function the difference between TPR and UMDA is very small. For TPR the linkage disequilibrium term DSQ increases for two generations, at the end it decreases by a factor of 2 each generation. The average of the fitness is very similar. Similar results can be obtained for other fitness functions which fulfill Equations 67-70.

In Table 9 we investigate linkage disequilibrium with selection and without selection. The fitness function for selection is as before. The initial genotype frequencies have been set to $p(0,0,0) = 0.65$, all other frequencies have been set to 0.05.

Without selection linkage disequilibrium DSQ is reduced approximately by a factor of 4 each generation. With selection, the reduction of DSQ is irregular. Nevertheless DSQ is almost the same in both cases up to generation 6. This is the more surprising as the genotype frequencies are already very different at generation 2. Note that TPR without selection has a fixed point at about $p(0,0,0) = 0.512$ as predicted by Geiringer's theorem.

t	p_1	p_2	p_3	$\bar{f}(t)$	$V(t)$	DSQ
0	0.5000	0.5000	0.5000	6.00	3.50	
1	0.5417	0.4167	0.6250	6.58	3.33	
2	0.5794	0.3428	0.7318	7.09	2.91	
3	0.6138	0.2793	0.8149	7.50	2.40	
4	0.6454	0.2256	0.8752	7.82	1.91	
6	0.7019	0.1442	0.9454	8.24	1.17	
0	0.5000	0.5000	0.5000	6.00	3.50	0.000
1	0.5417	0.4167	0.6250	6.58	3.24	7.5E-5
2	0.5701	0.3454	0.7298	7.08	2.82	1.1E-4
3	0.6103	0.2843	0.8113	7.48	2.34	8.1E-5
4	0.6406	0.2320	0.8712	7.79	1.94	4.7E-5
5	0.6690	0.1876	0.9134	8.03	1.44	2.4E-5
6	0.6959	0.1505	0.9423	8.22	1.18	1.2E-5

Table 8: Gene frequencies for UMDA(top) and TPR (bottom)

	Selection		No selection	
t	$p(0,0,0)$	DSQ	$p(0,0,0)$	DSQ
0	0.650	4.00E-02	0.650	4.00E-02
1	0.525	1.31E-02	0.582	1.06E-02
2	0.418	3.87E-03	0.548	2.73E-03
3	0.321	9.30E-04	0.530	6.92E-04
4	0.236	1.60E-04	0.521	1.74E-04
5	0.167	2.39E-05	0.516	4.37E-05
6	0.113	1.34E-05	0.514	1.09E-05
7	0.074	1.20E-05	0.513	2.74E-06
10	0.019	2.26E-06	0.512	4.28E-08

Table 9: Linkage disequilibrium with selection and without selection

In the next section we discuss the differences and similarities of genetic algorithms using TPR with UMD algorithms.

11 Two-parent recombination vs. gene-pool recombination

The relationship between TPR algorithms and UMD algorithms is very intricate. But we see more similarities than differences. We conjecture that the class of fitness function which both algorithms efficiently can solve is very similar. The same is true for the class of fitness functions they are not able to solve.

Let us summarize the results obtained so far. For UMDA algorithms we have proven (Theorem 13)

$$R_{UMDA}(t) = \frac{V_A(t)}{\bar{f}(t)} + error(t).$$

For TPR a corresponding equation could be obtained for $n = 2$ loci only.

$$R_{TPR}(t) = 2\frac{cov(F_{mp}(t), F_o(t))}{\bar{f}(t)} + \frac{1}{2}error(t)$$

The approximate equation

$$R_{TPR}(t) \approx 2\frac{cov(F_{mp}(t), F_o(t))}{\bar{f}(t)}$$

is called Robertson's or Price's Theorem. This approximation can be proven for general n under the assumption that the regression coefficient of the fitness of the selected parents is almost identical to that obtained without selection (Mühlenbein et al. (1994)). This assumption is difficult to verify for a given fitness function.

For genotypes in Robbins' proportions Fisher's variance decomposition can be proven

$$cov(F_{mp}(t), F_o(t)) = \frac{1}{2}V_A(t) + \frac{1}{4}V_2(t) + \ldots + \frac{1}{2^n}V_n(t).$$

Taking these results together we can state that for genetic populations in Robbins' proportions both TPR and UMDA mainly depend on the additive genetic variance $V_A(t)$. The difference between TPR and UMDA is small. The question remains open how important linkage disequilibrium is for TPR. The theorem of Geiringer and our numerical results indicate that for many fitness functions linkage disequilibrium will be small and be not important.

The dependency on the genotype frequencies makes an interpretation of $V_A(t)$ difficult. It is wrong to assume that UMDA can solve only linear fitness functions. It can solve very difficult nonlinear functions if there is always a reasonable $V_A(t)$ contribution in the variance V(t). The same is true for TPR. But both algorithms fail in optimizing fitness functions which have a small $V_A(t)$ contribution. This means that the fitness functions are mainly determined by nonlinear gene interactions.

We give additional empirical evidence for this conjecture. The different sampling strategies of single-point crossover, uniform crossover and gene-pool recombination can informally be characterized as follows. Selection is used to define a population of strings to be used for recombination. If a gene is fixed at a locus, then this gene remains fixed. Recombination/crossover is only sampling the subspace where the alleles of the strings differ. Gene-pool recombination samples this space according to the product of the univariate marginal distributions. Uniform crossover is doing almost the same, but the sampling is biased by the strings contained in the population. Single-point crossover samples a subset of the points sampled by uniform crossover. It is very difficult to describe these sampling strategies formally.

There have been a number of theoretical studies to understand crossover in genetic algorithms. In the analysis of De Jong and Spears (1992) disruption (probability of destroying higher order schemata) as well as recombination potential (probability of creating a higher order schema when the parents contain the necessary lower order schemata) are computed. The authors support evidence that uniform crossover has a higher recombination potential than the other crossover operators. In principle it is possible to use a specific recombination method where the recombination bias matches the nonlinear gene dependencies. But if the bias does not match, the result gets worth than with uniform crossover. So uniform crossover is widely used. But in Section 5 we have shown that

uniform crossover is moving the genotype frequencies very fast to Robbins' proportions. This means that a genetic algorithm with uniform crossover behaves very similar to UMD algorithms.

From the empirical studies of recombination operators, we will only discuss Eshelman and Schaffer (1995) because their empirical findings are backing up our theory. Eshelmann and Schaffer define two sampling biases: *recombinative bias* and *schema bias*. Recombinative bias is the expected proportion of differing bits that a recombination operator copies to a child from its furthest parent (in terms of Hamming distance). Schema bias is defined as follows: a recombination operator has no schema bias if all schemata of the same order are equally likely to be disrupted in a single mating.

A large recombinative bias creates a large standard deviation in the fitness of the offspring. Recombinative bias is related to the standard deviation of offspring's fitness. Schema bias cannot be related to a concept introduced in this paper.

Single-point crossover has weak recombinative bias and strong schema bias. Uniform crossover and HUX, which swaps *exactly half* of the differing bits, have strong recombinative biases and weak schema biases. For a fairly large set of problems, Eshelman and Schaffer show empirically that HUX is the best performer and that single-point and two-point crossover only perform well on the very artificial "needles on a plateau" problem. This result is valid if a single recombination operator is used. Eshelman and Schaffer suggest using a mechanism to switch between operators based on the progress of the search, although no such mechanisms is likely to be ideal.

In summary: *Recombination operators differ in their search biases. All numerical results obtained so far indicate that a high recombinative bias is necessary for a good search and this is achieved by an algorithm that uses only univariate marginal frequencies. No numerical or theoretical evidence has been provided that two-parent recombination detects and explores useful gene interactions in a systematic way.*

12 Analysis of binary tournament selection

The exact response equations (Theorems 9 and 13) have been derived under the assumption of proportionate selection. From these equations the approximate breeders' equation (16) can be derived by using $I = V^{1/2}/\bar{f}(t)$ (Mühlenbein et al. (1994)). Now we investigate whether the breeders' equation 16 is also a good approximation for other selection schemes. In quantitative genetics this is taken for granted.

Numerical experiments with the Breeder Genetic Algorithm BGA suggest that Equation 16 can indeed be used for truncation selection. But the numerical experiments have also revealed that for binary tournament selection this approximation can give a poor estimate. In this section we will explain why this is the case.

In order to keep the analysis simple, we first consider binary fitness functions of class *unitation*, where the fitness values are equal for all chromosomes having the same number of 1's. Let $h(\boldsymbol{x}) = \sum_{i=1}^{n} x_i$. Then $f(\boldsymbol{x}) = g(h(\boldsymbol{x}))$. Furthermore, we assume that the univariate marginal distributions p_i of all loci are equal in the initial population:

$$p_1(1) = p_2(1) = \cdots = p_n(1) = p. \tag{71}$$

Then a genotype $\mathbf{x}$ with k $1's$ is contained in the initial population with probability

$$p(\boldsymbol{x}) = p^k(1-p)^{n-k}.$$

Theorem 14 *Let the fitness values obey the relation*

$$g(0) < g(1) < \cdots < g(n) \tag{72}$$

where $g(i)$ denotes the fitness of a genotype with i 1's. Let p be the univariate marginal frequency at generation t. Then the marginal frequency p' of an UMD algorithm at generation $t+1$ is given by

$$\begin{aligned} p' &= p + p(1-p)\left(2\sum_{k=1}^{n}\sum_{j=0}^{k-1}\binom{n-1}{k-1}\binom{n}{j}p^{k+j-1}(1-p)^{2n-k-j-1}\right. \\ &\left. +\sum_{k=1}^{n-1}\binom{n-1}{k-1}\binom{n}{k}p^{2k-1}(1-p)^{2n-2k-1} - \sum_{j=0}^{2n-2}p^j\right) \end{aligned} \tag{73}$$

Proof: The frequency P_k of a specific genotype with k $1's$ is given by

$$P_k = p^k(1-p)^{n-k}.$$

There are $\binom{n}{k}$ different genotypes with k $1's$. In a binary tournament, P_k wins all tournaments with $P_j, j < k$. There is a draw if P_k meets another genotype with k $1's$. In this case the winner is randomly determined. Therefore on the average P_k wins half of these tournaments. Furthermore, P_k will be the winner of a tournament with itself. The frequency P_k^s after all tournaments have been done is given by:

$$P_k^s = \left(2\sum_{j=0}^{k-1}\binom{n}{j}P_j + \binom{n}{k}P_k\right)P_k. \tag{74}$$

From P_k^s we can compute the univariate marginal frequency p^s by summing up all appropriate genotypes, e.g., all genotypes having allele 1 at locus 1. This gives

$$p^s = \sum_{k=1}^{n}\binom{n-1}{k-1}P_k^s.$$

This equation can be easily understood. We just explain it for $n = 3$ loci. In this case the four genotypes $(1,0,0),(1,0,1),(1,1,0),(1,1,1)$ have to be summed, giving one genotype with a single 1, two genotypes with two 1's and one genotype with three 1's. In order to keep the population in linkage equilibrium we have to set

$$p' = p^s.$$

This gives equation (73).
□

Remark: If tournament selection is used, all fitness functions obeying the relation (72) will lead to the same evolution of the univariate marginal distribution. The absolute fitness values do not have any influence; only the order relation is important for tournament selection. The linear function $ONEMAX(n)$, defined by $g(i) = i$, obeys the order relation in Equation 72. Therefore, the evolution of the gene frequencies is equal to the $ONEMAX(n)$ dynamics for all functions of this class.

For $ONEMAX(n)$ we obtain from equation (73) using $R(t) = np' - np$

$$\begin{aligned} R_n(t) &= np(1-p)\left(2\sum_{k=1}^{n}\sum_{j=0}^{k-1}\binom{n-1}{k-1}\binom{n}{j}p^{k+j-1}(1-p)^{2n-k-j-1}\right. \\ &\left. +\sum_{k=1}^{n-1}\binom{n-1}{k-1}\binom{n}{k}p^{2k-1}(1-p)^{2n-2k-1} - \sum_{j=0}^{2n-2}p^j\right) \end{aligned}$$

Corollary: *For $ONEMAX(2)$ we have*

$$p' = p + p(1-p)(1-p+p^2), \tag{75}$$
$$R_2(t) = 2p(1-p)(1-p+p^2) \tag{76}$$

For $ONEMAX(3)$ we obtain

$$p' = p + p(1-p)(1-2p+4p^2-4p^3+2p^4), \tag{77}$$
$$R_3(t) = 3p(1-p)(1-2p+4p^2-4p^3+2p^4). \tag{78}$$

In order to compute realized heritability, we need an estimate of the selection differential $S(t) = \bar{f}_s(t) - \bar{f}(t)$. This has been done for binary tournament selection and the function $ONEMAX$ already in Section 3. From Equations 22 and 24 we observe that $R(t) = S(t)$. This means that realized heritability is 1 for $n = 2, 3$ as expected. Note that for proportionate selection it follows from Fisher's theorem 9 that realized heritability is one for any linear functions. For tournament selection the proof that the realized heritability is one for $ONEMAX$, is elementary but lengthy. We recall from Section 3 that for binary tournament selection

$$S_n(t) = \sum_{i=0}^{n-1}(1-D^2(i)) - np, \quad where \tag{79}$$
$$D(i) = \sum_{j=0}^{i}\binom{n}{j}p^j(1-p)^{n-j}. \tag{80}$$

Theorem 15 *For $ONEMAX(n)$ the realized heritability of the univariate marginal distribution algorithm with binary tournament is equal to 1, i.e.*

$$R_n(t) = S_n(t)$$

The proof is given in Appendix 2.
It is difficult to solve Equation 73 analytically. But we can solve it approximately by using selection intensity. Recall from Section 3 that $S(t) \approx I_2\sigma(p(t))$

Corollary: *Under the assumptions of Theorem 14 the univariate marginal frequency p is approximately given by the difference equation*

$$p' \approx p + \frac{I_2}{n}\sqrt{np(1-p)} \tag{81}$$

The solution of this equation is given by

$$p(t) = 0.5\left(1 + \sin\left(\frac{I_2}{\sqrt{n}}t + \arcsin\left(2p(0)-1\right).\right)\right) \tag{82}$$

Proof: For $ONEMAX$ we have $\sigma(p) = \sqrt{np(1-p)}$ for UMD algorithms. This gives the RS equation

$$R(t) \approx I_2\sqrt{np(1-p)}$$

From $R(t) = np' - np$ the difference equation is obtained. The difference equation (81) has been approximately solved by Mühlenbein et al. (1993).
□

Numerical simulations have confirmed that Equation 82 is a good prediction for univariate marginal frequency algorithms. It is also a good approximation for a genetic algorithms with uniform crossover. In this case $p(t)$ converges more slowly to 1, because the fitness distribution is slightly different from a binomial distribution. This has already been observed by Mühlenbein et al. (1994) for truncation selection.

Equation 82 has subsequently been used by Thierens and Goldberg (1994), Bäck (1995), and Miller and Goldberg (1996). Their simulations confirm that the approximate solution is in excellent agreement with empirical results. But in these papers, the difficult theoretical derivation of the approximate solution is missing. To summarize the major steps: First, we have shown that the selection intensity of the binomial distribution and of the corresponding normal distribution are almost identical, even for a small number of loci. Second, for UMDA the fitness distribution is binomial, and third we had to show that realized heritability is 1.

Recalling our earlier remark concerning binary tournament selection, we remind the reader of the surprising fact that Equation 82 is valid for *all* fitness function obeying the order relation (72). The function $ONEMAX(n)$ was only needed in order to apply the breeders' equation.

It is fairly straightforward to compute the univariate marginal frequencies for other order relations. For simplicity we just state the results for some instances with two loci.

Theorem 16 *Let the fitness values obey the order relation*

$$(I) \quad g(0) < g(1) < g(2),$$

then the marginal frequency is given by

$$p' = p + p(1-p)(1-p+p^2). \tag{83}$$

Let the order relation be

$$(II) \quad g(0) = g(1) < g(2),$$

then

$$p' = p + p(1-p)p^2. \tag{84}$$

Let the order relation be

$$(III) \quad g(1) < g(0) < g(2),$$

then

$$p' = p + p(1-p)(2p-1). \tag{85}$$

Let the order relation be

$$(IV) \quad g(0) < g(2) < g(1),$$

then

$$p' = p + p(1-p)(1-p-p^2). \tag{86}$$

Proof: The first case was proven before. We sketch only the proof for the last case, the other cases can be proven in the same manner. Counting the tournaments we obtain:

$$\begin{aligned}
p^s(0,0) &= p(0,0)p(0,0) = (1-p)^4 \\
p^s(0,1) &= (2p(0,0) + 2p(1,1) + 2p(0,1))p(0,1) = 2((1-p)^2 + p^2 + p(1-p))p(1-p) \\
p^s(1,0) &= p^s(0,1) \\
p^s(1,1) &= (2p(0,0) + p(1,1))p(1,1) = (2(1-p)^2 + p^2)p^2
\end{aligned}$$

From

$$p' = p^s(0,1) + p^s(1,1),$$

we obtain after some computation the equation

$$p' = p + p(1-p)(1-p-p^2). \qquad \square$$

Note that for order relation (III) $p = 0.5$ is an isolated fixed point. For order relation (IV), there is a stable attractor in the interior at about $p = 0.61803$. This value is the root of $p^2 + p - 1 = 0$. Table 10 gives some numerical results. For comparison, results for $ONEMAX$ and proportionate selection are also given. Here p obeys the difference equation (Mühlenbein et al., 1994)

$$p' = p + \frac{1}{2}(1-p).$$

t	$p:(I)$	$p:(prop.)$	$p:(II)$	$p:(III)$	$p:(IV)$
0	0.100000	0.100000	0.100000	0.100000	0.100000
1	0.181900	0.550000	0.100900	0.028000	0.180100
2	0.308567	0.775000	0.101824	0.002308	0.296380
3	0.476401	0.887500	0.102772	0.000016	0.424794
4	0.663622	0.943750	0.103746	0.000000	0.521250
5	0.837019	0.971875	0.104747	0.000000	0.572919
6	0.954827	0.985937	0.105775	0.000000	0.597105
7	0.996099	0.992969	0.106834	0.000000	0.608258
8	0.999970	0.996484	0.107923	0.000000	0.613444
9	1.000000	0.998242	0.109044	0.000000	0.615873
10	1.000000	0.999121	0.110199	0.000000	0.617015

Table 10: Results for binary tournament selection (order relations (I) till (IV)) and proportionate selection $ONEMAX$

Binary tournament selection does not take the fitness values into account; only the order relations are relevant. This leads to the following behavior. The function $g(0) = 0, g(1) = \epsilon, g(2) = 1$ is contained in class (I). The function $g(0) = 0, g(1) = 0, g(2) = 1$ is contained in class (II). They are mathematically almost identical. Nevertheless, the difference equations for p are very different.

From Theorem 16 another important result can be derived. The fitness function $g(0) = \epsilon, g(1) = 0, g(2) = 1$ is contained in class (III). If the univariate marginal frequency p of the initial population is less than 0.5, then p will converge to $p = 0$. Now the

average fitness of the population is given by $\bar{f} = p^2$. Therefore the response is negative.

Remark: *For binary tournament selection the average fitness of the population may decrease.*

For $ONEMAX$ realized heritability is 1, both for proportionate selection and tournament selection. But it can be very different which we show with a contrived example.

Example: Let the fitness function be given by $g(0) = 0, g(1) = \epsilon, g(2) = 2$ with $\epsilon << 1$. For proportionate selection obviously $R(t) = S(t) = 2 - 2p^2$. Thus realized heritability is 1, independent of p. For binary tournament selection one computes $S(t) = 2p^2(1-p^2)$ and $R(t) = 2p^2(1-p)(1-p+p^2)(3-2p+2p^2-p^3)$. Therefore realized heritability is given by

$$b_2(t) = \frac{(1-p+p^2)(3-2p+2p^2-p^3)}{1+p} \qquad p \neq 0, 1.$$

Thus $b_2(t) \to 3$ for $p \to 0$ and $b_2(t) \to 1$ for $p \to 1$.

Realized heritability depends on the selection method. For binary tournament selection it might even be greater than 1. Therefore estimates for realized heritability for proportionate selection can be very poor for tournament selection.

In the next section we compute an exact response equation for arbitrary fitness functions.

13 The exact response for binary tournament selection

Tournament selection uses only the order relation of the fitness values. Therefore, the evolution of the univariate marginal frequencies depends on the order relation only. In the computation of the *additive variance* the fitness values play a major role. This indicates that the additive genetic variance may be of limited value for predicting the behavior of genetic algorithms using binary tournament selection. We show this by defining a modified fitness function b. With b we can formulate tournament selection as an instance of proportionate selection. Let us first define "payoff" coefficients

$$a_{xy} = \begin{cases} 2 & f(\boldsymbol{x}) > f(\boldsymbol{y}) \\ 1 & f(\boldsymbol{x}) = f(\boldsymbol{y}) \\ 0 & f(\boldsymbol{x}) < f(\boldsymbol{y}) \end{cases}$$

We model tournament selection as a game. Two individuals with genotype $\boldsymbol{x}$ and $\mathbf{y}$ "play" against each other. The one with the larger fitness gets a payoff of 2. If the fitness values are equal, both will win half of the games. This gives a payoff of 1. Because the payoff matrix is derived from a game one can show (x and $\boldsymbol{x}$ denote the same variable)

$$\sum_{\boldsymbol{x}} \sum_{\boldsymbol{y}} p(\boldsymbol{x}, t) a_{xy} p(\boldsymbol{y}, t) = 1.$$

After a round of tournaments the genotype frequencies are given by

$$p^s(\boldsymbol{x}, t+1) = p(\boldsymbol{x}, t) \sum_{\boldsymbol{y}} a_{xy} p(\boldsymbol{y}, t). \tag{87}$$

If we set

$$b(\boldsymbol{x},t) = \sum_{\boldsymbol{y}} a_{xy} p(\boldsymbol{y},t),$$

then the above equation defines proportionate selection for the function b. But b depends on the genotype frequencies. Furthermore the average remains constant, $\bar{b}(t) \equiv 1$.

The difference equations for the univariate marginal frequencies can be written as in Equation 62

$$\Delta p_i = p_i(t)(1-p_i(t))\frac{B_i(1,t) - B_i(0,t)}{W},$$

where B_i is given by (see Equation 48 and following)

$$B_i(1,t) = \sum_{\boldsymbol{x}|x_i=1} b(\boldsymbol{x},t) \prod_{\substack{j=1 \\ j\neq i}}^{n} p_j(x_j,t) - 1$$

B_i can be very different from the terms F_i used for proportionate selection and the additive variance. We now formulate the exact response equation for binary tournament selection.

Theorem 17 *Let $\tilde{B}_i(t) = B_i(1,t) - B_i(0,t)$ and $\tilde{F}_i(t) = F_i(1,t) - F_i(0,t)$. Then for UMDA with binary tournament selection the reponse to selection is given by*

$$\begin{aligned} R(t) \;=\; & \sum_{i=1}^{n} p_i(t)(1-p_i(t))\frac{\tilde{B}_i(t) * \tilde{F}_i(t)}{W} \\ & + \frac{1}{2}\sum_{i\neq j} \frac{p_i(t)(1-p_i(t))\tilde{B}_i(t) p_j(t)(1-p_j(t))\tilde{B}_j(t)}{W^2}\frac{\partial^2 W}{\partial p_i \partial p_j} \\ & + \frac{1}{3!}\sum_{i\neq j, j\neq k, i\neq k} \frac{p_i(t)(1-p_i(t))\tilde{B}_i(t) p_j(t)(1-p_j(t))\tilde{B}_j(t) p_k(t)(1-p_k(t)\tilde{B}_k(t)}{W^3} \\ & \frac{\partial^3 W}{\partial p_i \partial p_j \partial p_k} + ... \end{aligned} \tag{88}$$

Proof: See Theorem 13.
□

For proportionate selection we have $B_i(1,t) = F_i(1,t)$. But for tournament selection the terms B_i are usually different from F_i. In this case F_i cannot be used to estimate the behavior of binary tournament selection. Particularly the response can be different from 0, even if the additive genetic variance $V_A(t)$ is 0.

We summarize the major points concerning selection. The three most popular selection methods - proportionate selection, truncation selection and tournament selection - have their strong and their weak points. The result of tournament selection is independently of the fitness values. Only the order relation is used. Proportionate selection selects too weakly when the population approaches the optimum. This can be observed in Table 10. At the beginning, p increases at a much faster rate than for binary tournament selection, giving a faster convergence to the optimum $p = 1$. But when p approaches 1, the increase of p gets smaller and smaller.

Truncation selection is a compromise between proportionate and tournament selection. It uses the fitness values to determine the truncation point, but all selected points

are treated equally. Therefore heritability for proportionate and truncation selection are much more similar than for proportionate and binary tournament selection. Therefore the breeders' equation can be used for proprotionate as well as for truncation selection. For binary tournament selection the first term of Equation 88 should be used.

14 An incremental UMDA implementation

The investigations so far have indicated that the UMD algorithm is as plausible as any genetic algorithm using some kind of two-parent recombination. In order to implement UMDA, estimates for the univariate marginal distributions are necessary. These estimates are provided by the selected parents. But one can also design a simpler algorithm that takes previous marginal frequencies into account as well. An algorithm, which does this, has in fact already been proposed independently from the theory presented in this paper (Baluja et al., 1995). In this algorithm the univariate marginal frequencies are updated according to the rule

$$p_i(x_i, t+1) = p_i(x_i, t) + \lambda(r_i(x_i, t) - p_i(x_i, t)), \tag{89}$$

where $r_i(x_i, t)$ are the marginal frequencies of the selected points and λ is a control parameter. We call the resulting algorithm the incremental univariate marginal distribution algorithm (IUMDA).

IUMDA

- **STEP0:** Set $t \Leftarrow 1$. Set $p_i(x_i, 1)$.
- **STEP1:** Generate N new points according to the distribution $p(\boldsymbol{x}, t) = \prod_{i=1}^{n} p_i(x_i, t)$.
- **STEP2:** Select $M \leq N$ points according to a selection schedule. Compute the marginal frequencies $r_i(x_i, t)$ of the selected set.
- **STEP3:** Update the marginal frequencies according to equation (89). Set $t \Leftarrow t + 1$.
- **STEP4:** If termination criteria not met, go to STEP1.

Note that λ influences the speed of convergence. The smaller λ is, the slower the convergence speed. Before we show some computational results, we qualitatively analyze the algorithm. In Equation 89 only the univariate marginal frequency of loci i is used for updating. Therefore we simplify the notation by omitting the index i, i.e. $p_i(x_i, t) \equiv p(t)$ and $r_i(x_i, t) \equiv r(t)$. We start with the simplest case.

Theorem 18 *Assume that* $r(t) \equiv c$ *with* $0 \leq c \leq 1$. *Then*

$$p(t+1) = p(1)(1-\lambda)^t - c(1-\lambda)^t + c \qquad t = 0, 1, \ldots \tag{90}$$

The proof is straightforward and will be omitted. We obviously have

$$lim_{t \to \infty} p(t) = c.$$

In real applications $r(t)$ will oscillate. A qualitative analysis of the IUMDA algorithm has been done by Kvasnicka et al. (1995).

Theorem 19 *If the difference equation (89) can be approximated by the differential equation*

$$\frac{dp(t)}{dt} = \lambda \left(r(t) - p(t) \right), \tag{91}$$

the solution is given by

$$p(t) = p(1)e^{-\lambda t} + \lambda e^{-\lambda t} \int_0^t r(\tau) e^{\lambda \tau} d\tau. \tag{92}$$

Equation 92 cannot be used for prediction because $r(t)$ is not known in advance, but it explains the qualitative behavior of the algorithm. One often observes that IUMDA consists of two phases. In the first phase ($0 \leq t \leq t_1$) $r(t)$ more or less randomly oscillates about a mean $< r(t) >_0^{t_1}$. Then it moves to either 0 or 1, forcing $p(t)$ also to move in this direction.

In Table 11 we give numerical results for the linear function $ONEMAX$. Note how λ influences the convergence speed. For $ONEMAX$ $\lambda = 0.25$ leads to a much faster convergence than $\lambda = 0.1$. Because the size of the population N was fixed and very large, the speed of convergence is almost independent of the size of the problem, n. For difficult multi-modal fitness functions, the success of IUMDA critically depends on the parameters λ and N. We omit a detailed discussion here. It is obvious that IUMDA suffers from the problem all algorithms using univariate marginal distributions have: they are not able to handle higher-order gene interactions.

	$\lambda = 0.1$		$\lambda = 0.25$					
	n=30		n=30		n=60		n=90	
t	$\bar{p}$	$std(p)$	$\bar{p}$	$std(p)$	$\bar{p}$	$std(p)$	$\bar{p}$	$std(p)$
10	0.726	0.049	0.952	0.024	0.887	0.086	0.834	0.122
20	0.893	0.025	0.997	0.001	0.993	0.005	0.985	0.014
30	0.963	0.009	1.000	0.000	1.000	0.001	0.999	0.001

Table 11: $IUMDA$ Results for $ONEMAX$: $N = 1024$

15 From recombination to the estimation of distributions

Practical and theoretical investigations have shown the limitations of simple genetic algorithms. Therfore new methods have been tried or are being developed to detect and exploit nonlinear gene interactions. They can be classified as follows:

- Adaptive recombination
- Explicit detection of relations (Kargupta & Goldberg, 1997)
- Dependency trees (Baluja & Davies, 1997)
- Estimation of distributions (Mühlenbein & Paaß, 1996, De Bonet et al., 1997))

Adaptive recombination uses a number of heuristics to modify two-parent recombination. Kargupta's (1996) Gene Expression Messy Genetic Algorithm (GEMGA) tries to detect dependency relations by manipulating individual substrings. GEMGA has only a local view of the data. Kargupta and Goldberg (1997) support our view concerning the limitations of Mendelian recombination: "Unless GAs do a better job in linkage learning, they will continue to search poorly in relation space." The last two methods use all the statistical information contained in the population of selected points to detect dependencies. They have a global view of the data. Conceptually they can be described as follows (estimation of dependency algorithm (EDA)).

EDA

- **STEP 0:** Set $t \Leftarrow 1$. Generate $N \gg 0$ points randomly.
- **STEP 1:** Select $M \leq N$ points according to a selection method. Estimate the distribution $p^s(\boldsymbol{x})$ of the selected set.
- **STEP 2:** Generate N new points according to the distribution $p^s(\boldsymbol{x})$. Set $t \Leftarrow t+1$.
- **STEP 3:** If termination criteria are not met, go to STEP 1.

The estimation of distributions is a notoriously difficult statistical problem. Furthermore, for optimization there is a trade-off. If lots of computing time has to be used to get a good estimate of the distribution, then this effort has to pay off. It has to lead to a substantial reduction of function evaluations in order to beat a simple algorithm like UMDA.

At least for continuous genes, one method that efficiently detects second order interactions does exist; the method is known as principal component analysis PCA. The technique of first doing a principal component analysis and then performing gene-pool recombination in the transformed space, has been successfully used for the optimization of difficult fitness functions by Voigt and Mühlenbein (1995). It is interesting to note that in evolution strategies the need for second order models has been recognized very early (Bäck & Schwefel, 1993).

For discrete genes an obvious extension of univariate marginal distribution algorithms are multivariate ones. But it is difficult to generate $p(\boldsymbol{x})$ from multivariate marginal distributions. We just demonstrate the problem with an example. For $n = 4$ loci for instance, we may use $p(\boldsymbol{x}) = p(x_1, x_2)\, p(x_3, x_4)$. But then four of the six bivariate distribution are left out. There exist methods to solve this problem by solving a system of equations, but it seems easier to start with *conditional distributions* $p(x_i|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ to reconstruct interactions between the variables.

With $\boldsymbol{x}_{-i} := (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ let $p(x_i|\boldsymbol{x}_{-i})$ denote the probability of x_i given $\boldsymbol{x}_{-i}$. Besag (1974) has proven that the n different conditional distributions $p(x_i|\boldsymbol{x}_{-i})$, $i = 1, \ldots, n$, completely determine the joint distribution $p(\boldsymbol{x})$.

An algorithm based on the above conditional distributions is presented in (Mühlenbein & Paaß, 1996). It is computationally so expensive that it is of theoretical value only. A more pragmatic way is to limit the conditional distributions to only pairwise conditional probabilities $p(x_i|x_j)$. Then one should generate samples that match as closely as possible the true joint distribution $p^s(\boldsymbol{x})$. This method has been used by De Bonet et al. (1997).

By introducing memory it is possible to incrementally change the sampling distribution instead of relying only on the distribution $p^s(\boldsymbol{x})$. This method was introduced with the IUMDA algorithm. IUMDA has been extended by Baluja and Davis (1997). Like Bonet et al. (1997) they restrict the estimation to pairwise conditional probabilities. These probabilities define a conditional dependency tree. This representation is more general than the chain used by De Bonet et al. (1997).

Future research will show which of the different methods are of practical relevance for optimization. It has to be investigated if the effort to compute second or even higher order interactions really pays off, either in getting much better solutions or in reducing the number of function evaluations to get the same quality of solutions. For the more theoretical analysis it is an open research question whether exact response equations can be computed for some of the new methods.

16 Conclusion

In this paper we have computed exact equations for the response to selection under the assumption that the genotypes are in Robbins' proportions. For proprotionate selection it follows that UMD algorithms mainly exploit the additive genetic variance. But the exact response equation for tournament selection differs from that of proportionate selection already in the first term. It can even be wrong that the response is zero if the additive genetic variance is zero. Furthermore realized heritability can be different for tournament selection and proprotionate selection. These results weaken the classical concept of heritability at least in the context considered in this paper - discrete genes with arbitrary fitness contributions.

Our results have been derived under the assumption of an infinite population. Furthermore mutation has been neglected. It has been shown that for genetic algorithms with a reasonable population size and small mutation the infinite population size equations are a reasonable approximation (Mühlenbein et al., 1994). But our results cannot be extended to genetic algorithms using a small population and a high mutation rate. Here stochastic effects have to be modelled. Given the mathematical difficulty of the infinite population size model, we doubt that a mathematical analysis of finite populations will be possible.

We have supplied a number of arguments that genetic algorithms with two-parent recombination are not so much different from UMD algorithms. They are also not able to detect nonlinear gene interactions in a systematic way. This result explains why genetic algorithms using two-parent recombination have difficulties in optimizing nonlinear fitness functions with interacting genes. We have outlined new methods to detect interacting genes in nonlinear fitness functions. These methods rely on the estimation of empirical distributions, a difficult problem of statistics. Future research will show if one of the new methods will be of practical relevance for optimization.

Appendix 1

Here we prove that for UMDA the response is always greater or equal to zero. The Theorem follows from an inequality proven by Baum and Eagon (1967). We just state their inequality in our notation. For notational convenience we assume binary genes. Then we have for each loci two marginal frequencies $p_{i1} = p_i(1)$ and $p_{i2} = p_i(0) =$

$1 - p_i(1)$. We furthermore define the average of the population as

$$W(p_1, \ldots, p_n, t) = \bar{f}(t)$$

Theorem 20 (Baum, Eagon) *Let $W(\mathbf{p}) = W(\{p_{ij}\})$ be a polynomial with nonnegative coefficients homogeneous of degree n in its variables p_{ij}, $i = 1, \ldots n$, $j = 1, 2$. Let $\mathbf{p} = \{p_{ij}\}$ be any point of the domain $D : p_{ij} \geq 0, p_{i1} + p_{i2} = 1$. For $\mathbf{p} \in D$ let $\mathbf{p}'$ denote the point of D whose i, j coordinate is given by*

$$p'_{ij} = \frac{p_{ij} \frac{\partial W}{\partial p_{ij}} \mid_{\mathbf{p}}}{p_{i1} \frac{\partial W}{\partial p_{i1}} \mid_{\mathbf{p}} + p_{i2} \frac{\partial W}{\partial p_{i2}} \mid_{\mathbf{p}}} \tag{93}$$

Then $W(\mathbf{p}') > W(\mathbf{p})$ unless $\mathbf{p}' = \mathbf{p}$.

In order to apply the Theorem we just have to show that Equation 93 is identical to our Equation 47 for the univariate marginal frequencies. Obviously

$$\begin{aligned} \frac{\partial W}{\partial p_{i1}} \mid_{\mathbf{p}} &= \bar{f}_i(1, t) \\ \frac{\partial W}{\partial p_{i2}} \mid_{\mathbf{p}} &= \bar{f}_i(0, t). \end{aligned}$$

Furthermore the identity

$$p_{i1} \bar{f}_i(1, t) + p_{i2} \bar{f}_i(0, t) = W(p_1, \ldots, p_n, t)$$

is valid. Combining the above equations shows that the frequencies of the UMDA algorithm obey Equation 93. Therefore Theorem 10 follows from the above Theorem.

Appendix 2

Let $R_n(t)$ be defined by Equation 75 and $S_n(t)$ by Equation 79.

Theorem *For $ONEMAX(n)$ the realized heritability of the univariate marginal distribution algorithm with binary tournament is equal to 1, i.e.*

$$R_n(t) = S_n(t)$$

Proof: We introduce

$$S_n^*(t) = \frac{S_n(t) - np}{n} = 1 - \frac{1}{n} \sum_{i=0}^{n-1} B_i^2(n, p),$$

with

$$B_i(n, p) = \sum_{j=0}^{i} \binom{n}{j} p^j (1-p)^{n-j}, \quad i = 0, 1, \ldots n.$$

Similarly we use

$$R_n^*(t) = \frac{R_n(t) - np}{n}.$$

We compute

$$
\begin{aligned}
R_n^*(t) &= \sum_{i=1}^{n} \binom{n-1}{i-1} p^i (1-p)^{n-i} \left(2B_{i-1}(n,p) + \binom{n}{i} p^i (1-p)^{n-i} \right) \\
&= \frac{1}{n} \sum_{i=1}^{n} i \binom{n}{i} p^i (1-p)^{n-i} \left(2B_i(n,p) - \binom{n}{i} p^i (1-p)^{n-i} \right).
\end{aligned}
$$

Let us introduce

$$
P_{n,j} = \binom{n}{j} p^j (1-p)^{n-j}.
$$

Now $R_n^*(t) = S_n^*(t)$ holds if

$$
n = \sum_{i=0}^{n-1} B_i^2(n,p) + \sum_{i=1}^{n} i P_{n,i} \left(2B_i(n,p) - P_{n,i} \right).
$$

Because of $B_n^2(n,p) = 1$, the equation

$$
n+1 = \sum_{i=0}^{n} \left(B_i^2(n,p) + 2iP_{n,i} B_i(n,p) - iP_{n,i}^2 \right)
$$

has to be proven. Using $B_i(n,p) = \sum_{j=0}^{i} P_{n,j}$ we have to show

$$
\begin{aligned}
n+1 &= \sum_{i=0}^{n} \left(\sum_{j=0}^{i} P_{n,j}^2 + 2 \sum_{0 \le j < k \le i} P_{n,j} P_{n,k} + 2iP_{n,i}^2 \right. \\
&\quad \left. + 2i \sum_{j=0}^{n-1} P_{n,i} P_{n,j} - iP_{n,i}^2 \right).
\end{aligned}
$$

Evaluating the first sum we obtain

$$
n+1 = (n+1) \sum_{i=0}^{n} P_{n,i}^2 + \sum_{i=0}^{n} \left(2 \sum_{0 \le j < k \le i} P_{n,j} P_{n,k} + 2i \sum_{j=0}^{n-1} P_{n,i} P_{n,j} \right).
$$

Using the identity

$$
1 = \left(\sum_{i=0}^{n} P_{n,i} \right)^2 = \sum_{i=0}^{n} P_{n,i}^2 + 2 \sum_{0 \le j < k \le n} P_{n,j} P_{n,k},
$$

we have to prove

$$
\sum_{i=0}^{n} \left(2 \sum_{0 \le j < k \le i} P_{n,j} P_{n,k} + 2i \sum_{j=0}^{n-1} P_{n,i} P_{n,j} \right) = 2(n+1) \sum_{0 \le j < k \le n} P_{n,j} P_{n,k}.
$$

By carefully counting the number of instances of $P_{n,j}$ the above identity follows. This completes the proof.

□

Acknowledgments: The research has been supported by the Real World Computing Partnership. J. Bendisch computed the selection intensities and proved Theorem 15. Uday Chakraborty made the numerical simulations for recombination without selection. Thanks also to Mark Ring, who improved the clarity of the paper substantially. Especially I have to thank Ken DeJong, who made many valuable suggestions for improving the paper.

References

Aarts, E. & Korts, J. (1989). *Simulated Annealing and Boltzmann Machines.* New York:Wiley.

Altenberg, L. (1995) The Schema Theorem and Price's Theorem. In L.D. Whitley and M.D. Vose, editors. *Foundations of Genetic Algorithms 3*,pp. 23–50, San Francisco:Morgan-Kaufman.

Arnold, B.C., Balakrishan, N. & H.N. Nagaraja, H.N. (1992). *A First Course in Order Statistics.* New York:Wiley.

Asoh, H. & Mühlenbein, H. (1994a). Estimating the heritability by decomposing the genetic variance. In Y. Davidor, H.-P. Schwefel, and R. Männer, editors, *Parallel Problem Solving from Nature*, Lecture Notes in Computer Science 866, pp. 98–107. Springer-Verlag:New York.

Asoh, H. & Mühlenbein, H. (1994b). On the Mean Convergence Time of Evolutionary Algorithms without Selection and Mutation In Y. Davidor, H.-P. Schwefel, and R. Männer, editors, *Parallel Problem Solving from Nature*, Lecture Notes in Computer Science 866, pp. 88–97. Springer-Verlag:New York.

Bäck, Th. (1995). Generalized convergence models for tournament and (μ, λ) selection. In L. Eshelman, editor, *Sixth Int. Conf. on Genetic Algorithms*, pp. 2–8, San Francisco:Morgan Kaufmann.

Bäck, Th. & Schwefel H.-P. (1993). An Overview of Evolutionary Algorithms for Parameter Optimization. *Evolutionary Computation*, 1:pp.1–24.

Baluja, S. & Caruana, R. (1995). Removing the genetics from the standard genetic algorithm. In *Proc. of the 12th Intern. Conf. on Machine Learning*, Lake Tahoe.

Baluja, S. & Davies, S. (1997). Using Optimal Dependency-Trees for Combinatorial Optimization: Learning the Structure of the Search Space. *Carnegie Mellon Report CMU-CS-97-107.*

Baum, L.E. & Eagon, J.A. (1967). An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bull. Am. Math. Soc.*, 73:pp360–363

Besag, J.(1974). Spatial interaction and the statistical analysis of lattice systems. *J. Royal Statistical Society, Series B*, pp. 192–236.

Blickle, T. & Thiele, L. (1997) A Comparison of Selection Schemes Used in Evolutionary Algorithms. *Evolutionary Computation*, 4:pp.361–394.

Booker, L.B. (1993). Recombination distributions for genetic algorithms. In L.D. Whitley, editor, *Foundations of Genetic Algorithms 2*, pp. 29–44, San Mateo:Morgan-Kaufman.

Bulmer, M.G. (1980). *The Mathematical Theory of Quantitative Genetics.* Oxford:Clarendon Press.

Crow, J.F., & Kimura,M. (1970). *An Introduction to Population Genetics Theory.* New York:Harper and Row.

De Bonet, J.S.& Isbell, Ch. L. & Viola, P. (1997). MIMIC: Finding Optima by Estimating Probability Densities. to appear in *Advances in Neural Information Processing Systems 1996.*

De Jong, K. A., & Spears, W.M. (1992). A Formal Analysis of the Role of Multi-Point Crossover in Genetic Algorithms. *Annals of Math. and Art. Int*, 5:pp.1–26.

Eshelman, L.J., & Schaffer, J.D. (1993). Crossover's niche. In S. Forrest, editor, *Fifth Int. Conf. on Genetic Algorithms*, pp. 9–14, San Mateo:Morgan Kaufmann.

Eshelman, L.J., & Schaffer, J.D. (1995). Productive Recombination and Propagating and Preserving Schemata. In L.D. Whitley, editor, *Foundations of Genetic Algorithms 3*, pp. 299–314, San Francisco:Morgan Kaufmann.
Ewens, W. J. (1989). An interpretation and proof of the fundamental theorem of natural selection. *Theoretical Population Biology*, 36:pp.167–180.
Ewens, W. J. (1992). On the role of mathematical statistics in population genetics. In K.V. Mardia, editor, *The Art of Statistical Science*, pp. 233–260. New York:Wiley.
Ewens, W. J. (1995). Aspects of optimality behavior in population genetics theory. In W. Banzhaf and F.E. Eeckman, editors, *Evolution and Biocomputation*, pp. 7–17, New York: Springer.
Falconer, D. S. (1981). *Introduction to Quantitative Genetics.* London:Longman.
Fisher, R. A. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Trans. of the Roy. Soc. of Edinburh*, 52:pp. 399–433.
Fisher, R. A. (1958). *The Genetical Theory of Natural Selection.* New York:Dover.
Geiringer, H. (1944). On the probability theory of linkage in Mendelian heredity. *Annals of Math. Stat.*, 15:pp. 25–57.
Haldane, J. B. S. (1932). *The Causes of Evolution.* London:Longmans Green.
Holland, J. (1992). Adaptation in Natural And Artificial Systems.
Cambridge:MIT Press.
Kargupta, H. (1996). Computational Processes in Evolution: The SEARCH Perspective. *Annual SIAM Conference*, Kansas City.
Kargupta, H. & Goldberg, D.E. (1997). SEARCH, Blackbox Optimization, And Sample Complexity. to appear in *Foundations of Genetic Algorithms 4.*
Kvasnicka, V., Pelikan, M., & Pospichal, J. (1995). Hill-climbing with learning: An abstraction of genetic algorithm. Technical report, Bratislava:Slovak Technical University.
Lynch, M. & Walsh, B. (1997). *Fundamentals of Quantitative Genetics* to be published by Sinauer Associates; parts available: *http://nitro.biosci.arizona.edu/zbook/book.html*
Manderick, B., de Wegner, M. & Spiessens,P. (1991). The Genetic Algorithm and the Structure of the Fitness Landscape. In Belew,R.K. & Booker, L. editors, *Procedings of the Fourth International Conference on Genetic Algorithms*, pp. 45–52, San Mateo:Morgan Kaufman.
Miller, B.L & Goldberg, D.E. (1995). Genetic Algorithms, Tournament Selection and the Effects of Noise. *Complex Systems* 9:pp. 193–212.
Mühlenbein, H. & Schlierkamp-Voosen, D. (1993). Predictive Models for the Breeder Genetic Algorithm I. Continuous Parameter Optimization *Evolutionary Computation*, 1:pp. 25–49.
Mühlenbein, H. & Schlierkamp-Voosen, D. (1994). The science of breeding and its application to the breeder genetic algorithm. *Evolutionary Computation*, 1:pp. 335–360.
Mühlenbein, H. & Voigt, H.-M. (1996). Gene pool recombination in genetic algorithms. In J.P. Kelly and I.H Osman, editors, *Metaheuristics: Theory and Applications*, pp. 53–62, Norwell:Kluwer Academic Publisher.
Mühlenbein, H. & Paaß, G. (1996). From Recombination of Genes to the Estimation of Distributions I. Binary Parameters. In Voigt, H.-M et al. (eds.)*Lecture Notes in Computer Science 1141: Parallel Problem Solving from Nature - PPSN IV)*, pp. 178-187, Berlin:Springer.
Nagaraja, H.N. (1982). Selection differentials. In E. Kotz, N.L. Johnson, and C.B. Read, editors, *Encyclopedia of Statistical Science*, pp. 334–336, New York:Wiley.
Nagylaki, T. (1991). Error bounds for the fundamental and secondary theorems of nat-

ural selection. *Proc. Natl. Acad. Sci. USA*, 88:pp. 2401–2406.
Nagylaki, T. (1992). *Introduction to Theoretical Population Genetics.* Berlin:Springer.
Pearson, K. (1904). On a generalized theory of alternative inheritance with special reference to Mendel's laws. *Phil. Trans. Royal Soc*, 203:pp. 55-86.
Prügel-Bennet, A. & J.L. Shapiro, J.L. (1994). An Analysis of Genetic Algorithms Using Statistical Mechanics. *Physical Rev. Letters*, 72:p. 1305.
Rao,C.R. (1973). *Linear Statisticcal Inference and Its Application.* New York:Wiley.
Rattray, L.M. (1995). The Dynamics of a Genetic Algorithm under Stabilizing Selection. *Complex Systems*, 9:pp. 213-234.
Reeves, C. & Wright, Ch. (1995) An Experimental Design Perspective on Genetic Algorithms In L.D. Whitley and M.D. Vose, editors. *Foundations of Genetic Algorithms 3*,pp. 7–22, San Francisco:Morgan-Kaufman.
Robbins, R.A. (1918). Some applications of mathematics to breeding problems III. *Genetics*, 3:pp. 375–389.
Syswerda, G. (1989). Uniform crossover in genetic algorithms. In H. Schaffer, editor, *3rd Int. Conf. on Genetic Algorithms*, pp. 2–9, San Mateo:Morgan Kaufmann.
Syswerda, G. (1993). Simulated crossover in genetic algorithms. In L. D. Whitley, editor, *Foundations of Genetic Algorithms 2*, pp. 239–255, San Mateo:Morgan Kaufmann.
Thierens,D. & Goldberg, D. (1994). Convergence Models of Genetic Algorithm Selection Schemes. In Davidor,Y., Schwefel, H.P. & Männer,R. editors, *Parallel Problem Solving from Nature*, pp 119—129, Lecture Notes in Computer Science 866, New York:Springer-Verlag.
Tietze, H. (1923). Über das Schicksal gemischter Populationen nach den Mendelschen Vererbubgsgesetzen. *Zs. Angew. Math. U. Mech.*, 3:pp. 362–393.
Turelli, M. & Barton, N.H. (1994). Genetic and Statistical Analyses of Strong Selection on Polygenic Traits: What, Me Normal. *Genetics*, 138:pp. 913–941.
Voigt, H.-M. & Mühlenbein, H. (1995). 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*, pp. 172–177, New York:IEEE Press.
Voigt, H.-M., Mühlenbein, H. & Cvetković, D. (1995). Fuzzy recombination for the continuous breeder genetic algorithm. In L. J. Eshelman, editor, *Proc. of the Sixth Int. Conf. on Genetic Algorithms*, pp. 104–112, San Francisco:Morgan Kaufmann.
M. D. Vose, M. D. & Wright, A. D. (1994). Simple Genetic Algorithms with Linear Fitness. *Evolutionary Computation*, 2: pp. 347-368.