

# Convergence Models of Genetic Algorithm Selection Schemes

Dirk Thierens<sup>1</sup> \* and David Goldberg<sup>2</sup>

<sup>1</sup> Department of Electrical Engineering  
ESAT-lab KU.Leuven  
Kardinaal Mercierlaan 94  
B-3001 Leuven, Belgium

<sup>2</sup> Department of General Engineering  
University of Illinois at Urbana-Champaign  
104 South Mathews Avenue  
Urbana, IL 61801, USA

**Abstract.** We discuss the use of normal distribution theory as a tool to model the convergence characteristics of different GA selection schemes. The models predict the proportion of optimal alleles in function of the number of generations when optimizing the bit-counting function. The selection schemes analyzed are proportionate selection, tournament selection, truncation selection and elitist recombination. Simple yet accurate models are derived that have only a slight deviation from the experimental results. It is argued that this small difference is due to the build-up of covariances between the alleles – a phenomenon called linkage disequilibrium in quantitative genetics. We conclude with a brief discussion of this linkage disequilibrium.

## 1 Introduction

Many different genetic algorithm implementations exist and due to the lack of a closed form analytical GA theory it is difficult to compare them objectively. By assuming that the fitness function is normally distributed, we can apply the theory of normal distributions as analysis tool. This kind of analysis is a standard approach in quantitative genetics [2,3] and allows accurate modeling of the GA convergence characteristics.

In [5] a comparative analysis of GA selection schemes was given that modeled the selection behavior in terms of deterministic finite difference equations. The difference equations described the change in proportion of different classes of individuals, assuming fixed and identical fitness function values within each class. Since only selection is used no new fitness values are ever generated. Convergence time is characterized with the takeover time which is the number of

---

\* The first author acknowledges the support provided by the Flemish Community under the Concerted Action Project No. 90/94-4. Correspondence can be sent to [dirk.thierens@esat.kuleuven.ac.be](mailto:dirk.thierens@esat.kuleuven.ac.be).

generations needed for the best individual in the population to take over the entire population.

In this paper we take an alternative view by modeling different selection schemes when optimizing a normal distributed fitness function. There are two main advantages of using this approach. First, restricting ourselves to normal distributed fitness functions allows simple yet accurate modeling of the convergence characteristics of selection and recombination together. In order to maintain the normal distribution assumptions as good as possible, we use uniform crossover which maximally decorrelates the alleles. Since there is no epistatic effect between the genes, the population mean fitness does not change after recombination.

Second, the normal fitness distribution modeling allows us to analyze the build-up of covariances between the alleles caused by selection – the so called linkage disequilibrium [2,3]. In previous work we discussed the mixing problem in genetic algorithms [7,10] and since mixing failure is basically an extreme build-up of linkage disequilibrium it is instructive to model the covariance between the genes.

The convergence models predict the proportion of optimal alleles  $p(t)$  in the total population as a function of the number of generations  $t$ . Experimental results are obtained by optimizing the bit-counting function, which is binomial distributed and can well be approximated with a normal distribution. The population mean fitness at generation  $t$  is given by  $\bar{f}(t) = l p(t)$  and the variance  $\sigma^2(t) = l p(t)(1 - p(t))$ .

The experiments in this paper all use  $l = 100$  and  $n = 200$ . As shown in [6] a population size  $n = 2l$  allows reliable decision making for the bit-counting function.

In the next four sections we analyze different GA selection schemes, namely proportionate selection, tournament selection, truncation selection and finally elitist recombination. We conclude with a brief discussion of the allele covariance build-up or linkage disequilibrium.

## 2 Proportionate Selection

Proportionate selection is the oldest and still best known selection scheme in evolutionary computation and implements the idea that reproduction rates are proportional to the fitness value  $f$  [4,8]. The probability of selecting an individual  $i$  with fitness  $f_i$  and proportion  $P_i(t)$  at generation  $t$  is given by:

$$P_i(t^s) = P_i(t) \frac{f_i}{\bar{f}(t)}$$

with  $\bar{f}(t)$  the mean fitness of the population at generation  $t$ .

The increment in population mean fitness can easily be computed:

$$\begin{aligned}
\overline{f(t^s)} - \overline{f(t)} &= \sum_{i=1}^n P_i(t^s) f_i - \overline{f(t)} \\
&= \sum_{i=1}^n \frac{P_i(t) f_i^2}{\overline{f(t)}} - \overline{f(t)} \\
&= \frac{1}{\overline{f(t)}} (\overline{f^2(t)} - \overline{f(t)}^2) \\
&= \frac{\sigma^2(t)}{\overline{f(t)}}
\end{aligned}$$

Recombination does not change the population mean fitness when optimizing the bit-counting function ( $\overline{f(t+1)} = \overline{f(t^s)}$ ), so the increment of the population average fitness with proportionate selection is proportional to the fitness variance and inverse proportional to the mean fitness.

$$\overline{f(t+1)} - \overline{f(t)} = \frac{\sigma^2(t)}{\overline{f(t)}} \quad (1)$$

Since  $\overline{f(t)} = l p(t)$  and  $\sigma^2(t) = lp(t)(1 - p(t))$  the increase of the proportion  $p(t)$  of optimal alleles is given by

$$l(p(t+1) - p(t)) = \frac{lp(t)(1 - p(t))}{lp(t)} = 1 - p(t)$$

Approximating the difference equation with the corresponding differential equation we obtain a simple convergence model expressing the proportion  $p(t)$  in function of the number of generations  $t$

$$\frac{dp(t)}{dt} = \frac{1}{l}(1 - p(t))$$

which has as solution

$$p(t) = 1 - (1 - p(0)) e^{-t/l}.$$

Starting from a random initial population  $p(0) = 0.5$  the convergence model becomes:

$$\boxed{p(t) = 1 - 0.5e^{-t/l}} \quad (2)$$

To calculate the convergence speed we compute the number of generations  $g_{conv}$  it takes to let the proportion  $p(t)$  come arbitrarily close to 1 or  $p(g_{conv}) = 1 - \epsilon$

$$g_{conv} = -l \ln(2\epsilon)$$

For  $\epsilon = 1/(2l)$  the convergence time is:

$$g_{conv} = l \ln(l) \quad (3)$$

In figure 1 we have plotted the model for a string length  $l = 100$  and compared it with experimental results when implementing proportionate selection with stochastic universal sampling. Obviously the model predicts the experimental results very well. The rate of convergence with proportionate selection is extremely slow and it drastically slows down when approaching the solution.

Instead of calculating the convergence speed by computing how long it takes to let the proportion  $p(t)$  come arbitrarily close to 1, we might also compute the generation at which the global optimum is expected to be found with a given probability. The probability that at least one of the strings in the population consists of all ones is given by:

$$\text{prob}(\text{opt}) = 1 - [1 - p^l(t)]^n$$

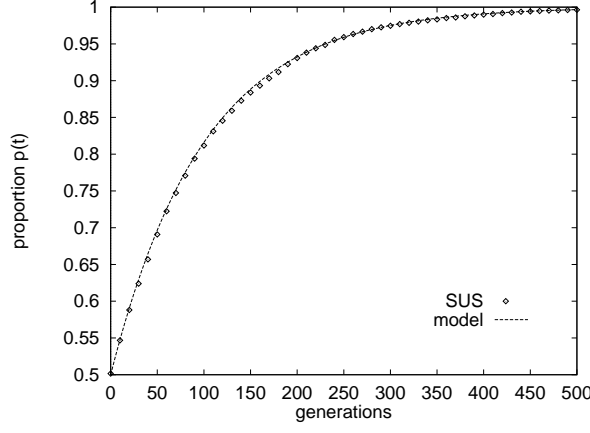
From this equation we can calculate the needed allele proportion  $p(g_{opt})$  and plug this into equation 2 to obtain the number of generations it takes to find the optimum with given probability  $\text{prob}(\text{opt})$ . For instance in our example with string length  $l = 100$ , population size  $n = 200$  and  $\text{prob}(\text{opt}) = 99\%$  we find that  $p(g_{opt}) = 0.96$  and  $g_{opt} = 260$ .

In order to obtain better convergence characteristics for proportionate selection one has to use some form of fitness scaling [4]. Instead of scaling the fitness function explicitly it is also possible to use a selection scheme that does not look at the absolute function value but instead performs selection according to the number of individuals in the population that are better or worse.

Examples of this ranked based selection principle are tournament selection, truncation selection and elitist recombination, which we will discuss in the next sections.

### 3 Tournament Selection

Tournament selection randomly chooses a set of individuals and picks out the best for reproduction. The number of individuals in the set is mostly equal to two but larger tournament sizes can be used in order to increase the selection pressure [10]. Here we consider the case of optimizing the bit-counting function with a tournament size  $s = 2$ . Under the assumption of a normally distributed function the fitness difference between two randomly sampled individuals in each tournament is also normally distributed with mean  $\mu_{\Delta f}(t) = 0$  and variance  $\sigma_{\Delta f}^2(t) = 2\sigma^2(t)$ . Since we are only selecting the best of the two competing individuals we are actually only looking at the absolute value of the fitness difference. The average fitness difference between two randomly sampled individuals is thus given by the mean value of those differences that are greater than  $\mu_{\Delta f} = 0$ , which is equivalent to the mean value of one half of the normal distribution truncated at its mean value 0. The mean value of the right half of a standard normal distribution is given by  $\sqrt{2/\pi} = 0.7979$ .



**Fig. 1.** Convergence model and experimental results of the proportion of optimal alleles when optimizing the bit-counting function with proportionate selection (Stochastic Universal Sampling) and uniform crossover.

Tournament selection selects the best out of every random pair of individuals so the population average fitness increase from one generation to the next is equal to half the mean value of the difference between two randomly sampled individuals:

$$\overline{f(t+1)} - \overline{f(t)} = \frac{1}{2} 0.7979 \sigma_{\Delta f}(t) = \frac{1}{2} 0.7979 \sqrt{2} \sigma(t) = \frac{1}{\sqrt{\pi}} \sigma(t)$$

or

$$\overline{f(t+1)} - \overline{f(t)} = 0.5642 \sigma(t) \quad (4)$$

Since  $f(t) = l p(t)$  and  $\sigma^2(t) = l p(t) (1 - p(t))$  we get

$$p(t+1) - p(t) = \sqrt{\frac{p(t)(1-p(t))}{\pi l}}.$$

Approximating the difference equation with the differential equation

$$\frac{dp(t)}{dt} = \sqrt{\frac{p(t)(1-p(t))}{\pi l}}$$

gives us the solution:

$$p(t) = 0.5(1 + \sin(\frac{t}{\sqrt{\pi l}} + \arcsin(2p(0) - 1))) .$$

For a randomly initialized population  $p(0) = 0.5$  the convergence model becomes:

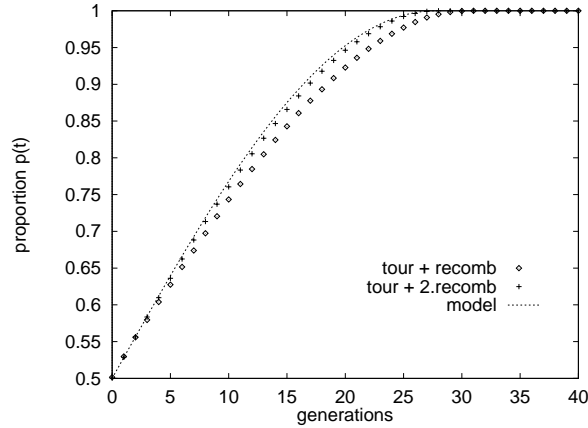
$$\boxed{p(t) = 0.5(1 + \sin(\frac{t}{\sqrt{\pi l}}))} \quad (5)$$

The number of generations  $g_{conv}$  it takes to let the population fully convergence is obtained by putting  $p(g_{conv}) = 1$

$$g_{conv} = \frac{\pi}{2} \sqrt{\pi l} \quad (6)$$

The convergence time complexity for tournament selection is thus  $O(\sqrt{l})$  which compares very favorable to  $O(l \ln(l))$  for unscaled proportionate selection.

The proposed convergence model is compared with experimental results in figure 2. The model slightly overestimates the proportion of optimal alleles for tournament selection and recombination with uniform crossover. In a second experiment we recombine the population twice at each generation, so after the usual procedure of tournament selection followed by recombination we randomly shuffle the population and again recombine the individuals. By doing this the alleles of each individual are more decorrelated and the assumptions we made when building the convergence model are less violated. Figure 2 shows that the model now coincides very well with the experimental results, so the slight overestimation is due to the build-up of covariances between the alleles caused by selection. In the last section we will have a closer look at this phenomenon.



**Fig. 2.** Convergence model and experimental results of the proportion of optimal alleles when optimizing the bit-counting function with tournament selection and uniform crossover.

## 4 Truncation Selection

Truncation selection or block selection ranks all individuals according to their fitness and selects the best ones as parents. In truncation selection a threshold

$T$  is defined such that the  $T\%$  best individuals are selected. Truncation selection has been used extensively in evolution strategies [1]. It is also often used in quantitative genetics where artificial selection performed by breeders is studied. In [9] a specific genetic algorithm - the Breeder Genetic Algorithm - is proposed that incorporates ideas from breeders to perform parameter optimization tasks, and the convergence model for the bit-counting function has also been computed there.

Block selection [10] is equivalent to truncation selection since for a given population size  $n$  one simply gives  $s$  copies to the  $n/s$  best individuals. Both implementations are identical when  $s = 100/T$ .

If the fitness function is normally distributed, quantitative genetics defines the *selection intensity*  $i$  that expresses the selection differential  $S(t) = \overline{f(t^s)} - \overline{f(t)}$  in function of the standard deviation  $\sigma(t)$ :

$$S(t) = i \sigma(t)$$

Using this definition one can easily compute the convergence model:

$$\overline{f(t+1)} - \overline{f(t)} = i \sigma(t) \quad (7)$$

Since  $f(t) = l p(t)$  and  $\sigma^2(t) = l p(t) (1 - p(t))$  we get

$$p(t+1) - p(t) = \frac{i}{\sqrt{l}} \sqrt{p(t)(1 - p(t))}.$$

Approximating the difference equation with the differential equation

$$\frac{dp(t)}{dt} = \frac{i}{\sqrt{l}} \sqrt{p(t)(1 - p(t))}$$

the solution becomes

$$p(t) = 0.5(1 + \sin(\frac{i}{\sqrt{l}}t + \arcsin(2p(0) - 1))).$$

For a randomly initialized population  $p(0) = 0.5$  or:

$$\boxed{p(t) = 0.5(1 + \sin(\frac{i}{\sqrt{l}}t))} \quad (8)$$

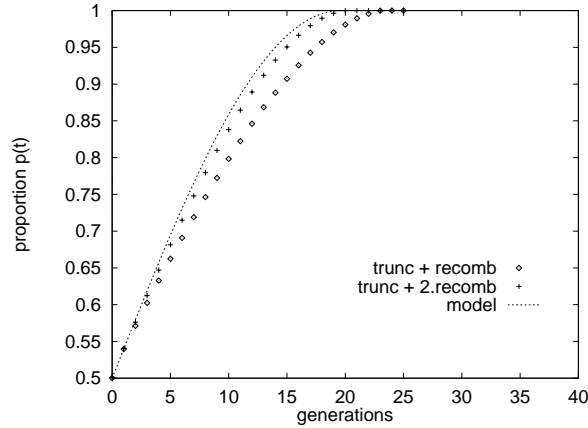
Calculating the number of generations  $g_{conv}$  to convergence ( $p(g_{conv}) = 1$ ) is straightforward:

$$g_{conv} = \frac{\pi}{2} \cdot \frac{\sqrt{l}}{i} \quad (9)$$

It is rather remarkable that the convergence models for tournament selection and truncation selection have the same functional form. The only difference is the magnitude of the selection intensity constant  $i$ , which can be changed by

increasing the truncation threshold or tournament size of the respective algorithms. For the standard tournament selection with tournament size  $s = 2$ , the selection intensity is  $i = 1/\sqrt{\pi} = 0.56$ .

Experimental results are shown in figure 3 for a block size  $s = 2$  or equivalently a truncation threshold  $T = 50\%$ , which gives us a selection intensity  $i = 0.8$  [2,3,9]. Again the model slightly overestimates the proportion  $p(t)$  of optimal alleles and predictions become more accurate when the alleles are decorrelated by repeating the recombination phase.



**Fig. 3.** Convergence model and experimental results of the proportion of optimal alleles when optimizing the bit-counting function with truncation selection and uniform crossover.

## 5 Elitist Recombination

Recently we have introduced an extremely simple GA implementation [11] where the selection and recombination phases are intertwined. Competition for survival takes place at the level of each family — the mating parents and their offspring — which results in a local elitist selection operator, called elitist recombination. For every mating pair two offsprings are created and the best two of these four individuals go to the next generation, so individuals can only be replaced by other individuals with a higher fitness value. Crossover is applied for every mating pair so there is no need to choose a specific value for the crossover probability parameter  $p_c$ .

*Elitist Recombination algorithm:*

1. initialize population



2. for every generation
  - (a) random shuffle population
  - (b) for every mating pair:
    - generate offspring
    - keep best two of each family

It is easy to show that - when optimizing the bit-counting function - the best individual of every mating pair will go to the next generation and the worst parent will be replaced by the best of the two offsprings [11]. As a result the population average fitness increase  $\overline{f(t+1)} - \overline{f(t)}$  is equal to half the average difference between the worst parent and the best child.

Since the children are basically random samples of a binomial distribution with same parameter  $p(t)$  as the parent population, we can compute the population average fitness increase as half the mean fitness difference between two random samples of a normal distribution with mean  $\mu = lp(t)$  and variance  $\sigma^2(t) = lp(t)(1 - p(t))$ .

Obviously this is exactly what we have done when modeling tournament selection so the convergence model for elitist recombination and tournament selection with tournament size  $s = 2$  are the same:

$$\boxed{p(t) = 0.5(1 + \sin(\frac{t}{\sqrt{\pi l}}))} \quad (10)$$

and

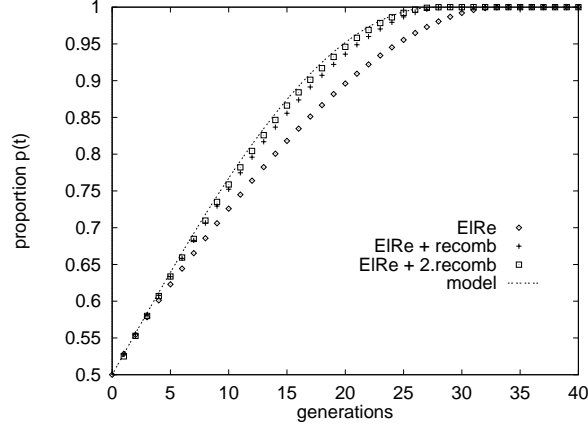
$$g_{conv} = \frac{\pi}{2} \sqrt{\pi l}. \quad (11)$$

Experimental results are shown in figure 4. As with tournament and truncation selection the model slightly overestimates the proportion  $p(t)$ , and predictions become more accurate when the alleles are decorrelated by repeating the recombination phase. Comparing figure 4 with 2 we note that due to the elitist mechanism the build-up of covariance between the alleles is higher for elitist recombination than for tournament selection and standard recombination.

## 6 Linkage disequilibrium

The convergence models discussed in the previous sections all assume that the fitness is normally distributed. In practice however there a number of factors - such as a finite population size - that violate this assumption. From the previous experiments it is clear that the most important factor is the build-up of correlations between the alleles. Repeating the recombination phase during each generation made the model predictions significantly more accurate. In this section we will briefly discuss this problem but due to space limitations a detailed analysis is beyond the scope of this paper and can be found in [12].

Selection picks out individuals that are mainly concentrated at the higher level of the fitness distribution. The variance of the population is reduced but



**Fig. 4.** Convergence model and experimental results of the proportion of optimal alleles when optimizing the bit-counting function with elitist recombination and uniform crossover.

this reduction is not only caused by the change in allele frequency but also by the introduction of covariances between the alleles. If we represent the covariance with  $\sigma_{cov}$  then the variance after selection is given by:

$$\sigma^2(t^s) = \sigma^2(t) + 2\sigma_{cov}(t^s) \quad (12)$$

With truncation selection for instance the variance is reduced with a factor  $k$  such that

$$\sigma^2(t^s) = (1 - k)\sigma^2(t) \quad (13)$$

with  $k$  given by  $k = i(i - x)$  where  $i$  is the selection intensity and  $x$  is the corresponding deviation of the truncation point from the population mean [3]. The covariance after selection can now be computed from equations 12 and 13:

$$\sigma_{cov}(t^s) = -\frac{k}{2}\sigma^2(t)$$

Crossover decorrelates the genes with a decorrelation factor  $\delta = 1 - p_c p_r$  with  $p_r = 2p_x(1 - p_x)$  ( $p_c$  : crossover probability,  $p_r$  : disruption probability and  $p_x$  : allele swap probability for uniform crossover). The covariance after selection and recombination is thus given by:

$$\sigma_{cov}(t + 1) = -\delta \frac{k}{2}\sigma^2(t)$$

Using this expression for the covariance we can more accurately calculate the population fitness variance such that the convergence models predict the experimental results even better (for details see [12]).

## 7 Conclusion

Simple yet accurate convergence models for different selection schemes are derived that allow us to compare the GA selection schemes in terms of their convergence characteristics. The models assume a normally distributed fitness function and predict the proportion of optimal alleles in function of the number of generations. Experimental results for optimizing the bit-counting function show good agreement with the predictions. It is argued that the slight deviations are due to linkage disequilibrium or the build-up of covariances between the alleles which can be included in the models to increase their accuracy.

## References

- [1] Bäck, T., Hoffmeister, F., & Schwefel, H.P. (1991) *A Survey of Evolution Strategies*. Proceedings of the Fourth International Conference on Genetic Algorithms, 2-9. Morgan Kaufmann.
- [2] Bulmer, M.G. (1985). *The Mathematical Theory of Quantitative Genetics*. Oxford University Press.
- [3] Falconer, D.S. (1989). *Introduction to Quantitative Genetics*. Longman Scientific & Technical.
- [4] Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison Wesley Publishing Company.
- [5] Goldberg, D.E., & Deb, K. (1991). *A comparative analysis of selection schemes used in genetic algorithms*. Foundations of Genetic Algorithms I, 69-93.
- [6] Goldberg, D.E., Deb, K., & Clark, J.H. (1991). *Genetic algorithms, noise, and the sizing of populations*. Complex Systems, 6, 333-362.
- [7] Goldberg, D.E., Deb, K., & Thierens, D. (1992). *Toward a better understanding of mixing in genetic algorithms*. Journal of the Society for Instrumentation and Control Engineers, 32(1), 10-16.
- [8] Holland, J.H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan Press.
- [9] Mühlenbein, H., & Schlierkamp-Voosen, D. (1993). *Predictive Models for the Breeder Genetic Algorithm. I. Continuous Parameter Optimization*. Evolutionary Computation 1(1):25-49, MIT Press.
- [10] Thierens, D., & Goldberg, D.E. (1993). *Mixing in Genetic Algorithms*. Proceedings of the Fifth International Conference on Genetic Algorithms, 38-45. Morgan Kaufmann.
- [11] Thierens, D., & Goldberg, D.E. (1993). *Elitist Recombination: an integrated selection recombination GA*. Technical Report ESAT-SISTA/TR 1993-76. To appear in Proceedings of the IEEE World Congress on Computational Intelligence. Orlando, June 1994.
- [12] Thierens, D., & Goldberg, D.E. (1994). *Effects of Linkage Disequilibrium on GA convergence*. Technical Report ESAT-SISTA/TR 1994-41.