

**Convergence-Time Models for the Simple Genetic Algorithm with Finite  
Population**

**Alessio Ceroni, Martin Pelikan and David E. Goldberg**

IlliGAL Report No. 2001028  
November 2001

Illinois Genetic Algorithms Laboratory  
University of Illinois at Urbana-Champaign  
117 Transportation Building  
104 S. Mathews Avenue Urbana, IL 61801  
Office: (217) 333-2346  
Fax: (217) 244-5705

# Convergence-Time Models for the Simple Genetic Algorithm with Finite Population

**Alessio Ceroni, Martin Pelikan and David E. Goldberg**

Illinois Genetic Algorithms Laboratory

Department of General Engineering

University of Illinois at Urbana-Champaign

Urbana, IL - 61801

{aceroni, pelikan, deg}@illigal.ge.uiuc.edu Phone: 1-217-333-2346

November, 2001

## Abstract

This paper presents various convergence models for the simple genetic algorithm (SGA) in the case of finite population. A piecewise convergence-time model is derived using ideas from two existing convergence models. The factors affecting the convergence with small population size are explained and used to construct a correct model of the variance in fitness for the OneMax problem. This knowledge is included in the existing asymptotic model to derive the embedded convergence-time model. The model is extended to a different environment and modified to include an unexpected behavior that makes the SGA converge solely by genetic drift.

## 1 Introduction

The development of convergence time models for the simple genetic algorithm (SGA) is important for understanding the efficiency and scaling capabilities of the algorithm. Some of the first efforts in this direction were made by Goldberg (1989b) and Goldberg and Deb (1991). Both studies developed estimates of takeover times in genetic algorithms (GAs) focusing on the dynamics of the best individual. A convergence-time model was developed by Mühlenbein and Schlierkamp-Voosen (1993) using the concept of selection intensity. The validity of this model is limited to the case of infinite population. This selection-intensity-based model provides a description of the dynamics of the fitness mean. To extend the model to finite populations, separate works by Blickle and Thiele (1995) and Shapiro, Prügel-Bennett, and Rattray (1994) addressed the problem of modeling higher-order moments. Blickle and Thiele (1995) introduced a model of the evolution of the fitness variance under the effect of the selection strategy. Shapiro, Prügel-Bennett, and Rattray (1994) used a statistical mechanics approach to model the evolution of cumulants of the fitness distribution. In contrast to studies that use selection intensity, the last approach does not yield closed-form solutions. Also, it does not reduce to convenient forms. On the other side, the "gambler's ruin" model (Goldberg, Deb, & Clark, 1992) addresses the effect on the quality of having a finite population.

The purpose of this paper is to construct a convergence-time model for the SGA with finite populations. We use an approach based on selection intensity, going in a different direction from the studies based on statistical mechanics. This approach lets us

1. form an algebraically simple model that can be easily applied in certain situations

2. gain physical insight from the process of model construction that illuminates complex phenomena, and
3. account for the population effect on quality.

To obtain these results we decompose the dynamics of the genetic algorithm and address the various facets of the convergence. We construct a model of the evolution of the variance that uses results from population sizing. In this way we obtain an analytical model to predict the convergence time given a desired population size.

We start with a general discussion of the work, giving a description of parameters and genetic operators. In section 3 we introduce the piecewise convergence-time model, constructed using ideas from two existing convergence models. In section 4 the factors affecting the convergence with small populations are explained and used to construct a correct model of the fitness' variance for the OneMax problem. In section 5 this knowledge is included in the existing asymptotic model to derive the embedded convergence-time model. In section 6 the model is extended to a different environment and modified to include an unexpected behavior that makes the SGA converge solely by genetic drift. In section 7 we discuss a possible way to further improve this model, and we present some possible applications.

## 2 Description of the Work

The GA considered in this paper manages a population of strings of  $l$  bits. The strings are randomly initialized with a certain probability  $P_0$  of having a bit equal to 1. Every string is divided into  $l/k$  groups of  $k$  bits. Therefore, we can see the strings as a composition of  $l/k$  elements, where each element can assume  $2^k$  configurations. The fitness is given by the number of bits equal to 1 (the OneMax problem). Individuals are selected at each generation using binary tournament selection without replacement (Goldberg, 1989a). The selected individuals are recombined using uniform-building-block population crossover (Mühlenbein, 1997, gene pool recombination). The recombination operator acts on the whole population of selected individuals (mating pool) instead of the standard two-parent operators, working on pairs of individuals. The elements of the string are randomly exchanged among the members of the mating pool so that the elements are uniformly spread over the population. The crossover point is always between elements. This implies that is not possible to destroy an existing element or create a new one. Therefore, we can assert that this GA combines building blocks of order  $k$  (Goldberg, 1989a), where each element of a string represents a single building block. We have chosen this operator because it has a higher mixing capability than the standard two-parent recombination (Mühlenbein, 1997). Moreover the correlation among the strings introduced by the selection strategy, which duplicates the fittest individuals, is eliminated, preventing its effect on the convergence (Thierens & Goldberg, 1994). No mutation operator is used. The term *time to convergence* is defined as the time the SGA needs to obtain a population of equal individuals.

We study the convergence of the SGA applied to the OneMax problem for various values of  $k$  and  $P_0$ . We can thus extend the results of our work to a different class of problems (Miller, 1997). In the Royal-Road domain, one building block of  $k$  bits has fitness 1, while all other configurations have fitness 0. The optimal configuration has a probability of  $1/2^k$  of appearing in the randomly initialized population in a certain position. This scenario is equivalent to the OneMax problem if we use  $P_0 = 1/2^k$  and the same number of building blocks in the string. On the other side, using various values of  $k$  lets us extend the work to other separable problems with fixed contributions to fitness from different blocks.

### 3 The Piecewise Convergence-Time Model

In the literature we find two interesting convergence-time models, useful for two different environments. The asymptotic convergence-time model (Mühlenbein & Schlierkamp-Voosen, 1993; Thierens & Goldberg, 1994) successfully describes the behavior of an SGA with an infinite population. Conversely the drift model (Asoh & Mühlenbein, 1994) describes the behavior of an SGA with no selection strategy for finite populations.

Our goal is to combine these two models to derive a logical description of the dynamics of the SGA with finite populations, when the selection strategy is applied. We introduce this piecewise convergence-time model in the last part of this section. First we further explain these two models, which constitute the base of our work.

#### 3.1 An Overview of the Asymptotic Convergence-Time Model

The assumption made by Mühlenbein and Schlierkamp-Voosen (1993) and Thierens and Goldberg (1994) in constructing this model is that all the variables are normally distributed. This is correct for the OneMax problem, given that the fitness is binomially distributed and that, for large samples, the binomial distribution is well approximated by a normal distribution. For this reason the model can be considered valid also for large but finite populations. We can use  $2l$  as an upper bound to the critical dimension that ensures the validity of the model (Goldberg, Deb, & Clark, 1991). We now explore all the steps followed by Thierens and Goldberg (1994) to derive an analytical expression of the convergence time, because we use the same steps in the next sections to obtain our extension to this model.

We are using tournament selection with 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  $\sigma_{\Delta f}^2(t) = 2\sigma^2(t)$ , where  $\sigma^2(t)$  is the variance of the distribution of the fitness values. Because we are selecting the best of only two competing individuals, we are actually 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 0, which is equivalent to the mean value of one half of the normal distribution. For the case of a standard normal distribution, this value is given by  $\sqrt{2/\pi}$ . Tournament selection selects the best out of every random pair of individuals, so the increase in population average fitness from one generation to the next is equal to half the mean value (because half of the population is substituted) of the difference between two randomly sampled individuals:

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

We define  $p(t)$  as the proportion of optimal building blocks in the population. Because the fitness is binomially distributed for the OneMax problem,  $\bar{f}(t) = lp(t)$  and  $\sigma^2(t) = lp(t)(1-p(t))$ . If we substitute these expressions into the previous equation, we obtain

$$p(t+1) - p(t) = \sqrt{\frac{p(t)(1-p(t))}{\pi l}} \quad (2)$$

Approximating the difference equation with the differential equation

$$\frac{dp(t)}{dt} = \sqrt{\frac{p(t)(1-p(t))}{\pi l}} \quad (3)$$

gives us the solution

$$p(t) = \frac{1}{2} \left[ 1 + \sin \left( \frac{t}{\sqrt{\pi l}} + \arcsin(2P_0 - 1) \right) \right] \quad (4)$$

Equation (4) can be used to calculate the time  $\tau$  until the GA converges to a population formed by copies of the optimal individual. Solving for  $p(\tau) = 1$ ,

$$\tau = \left( \frac{\pi}{2} - \arcsin(2P_0 - 1) \right) \sqrt{\pi l} \quad (5)$$

### 3.2 Overview of the Drift Model of the Convergence Time

The other important result about the convergence of an SGA is derived from the theory of genetic drift. Genetic drift is the random fluctuation of gene frequencies from generation to generation in finite populations. Given that there is no selection pressure, the convergence is obtained only by random accumulation of copies of a particular element in each position of the strings. Eventually, in every position, all the strings will contain the same element, and the population will have converged. This effect has been studied by Asoh and Mühlenbein (1994). In their work they demonstrate that the mean convergence time is almost proportional to the population size  $N$ . In case of uniform crossover, its value can be calculated as

$$t_{drift} = C_0(a \log l + 1)^b N \quad (6)$$

where the values of the constants  $C_0$ ,  $a$ , and  $b$  depend on the initial proportion of bit 1. If this proportion is equal to 1/2, these are  $C_0 = 1.4$ ,  $a = 0.5$ , and  $b = 1.1$ . The genetic drift is always acting with finite populations, but usually the time needed by the selection strategy is much lower. This is not true with very small populations, where the drift makes the population converge prematurely. This effect is called *drift stall* (Thierens, Goldberg, & Pereira, 1998), and it imposes a limit on the convergence time:

$$t_{convergence} \leq t_{drift} \quad (7)$$

### 3.3 Derivation of the Piecewise Convergence-Time Model

We now combine these two results to derive a piecewise convergence-time model (Fig 1). When the population size is small, the drift stall limits the convergence time to be less than or equal to the drift time. If we increase the population size, the convergence time increases, always limited by the drift time. Eventually, the selection will have enough time to act, and the convergence will not be forced by the drift stall. The result is a convergence time lower than the drift time. After the population size has reached its critical value, the convergence time is constant and is equivalent to the asymptotic value. Between 0 and the critical size of the population, we have an undiscovered region where the exact value of the convergence time is unknown.

The next step is to run experiments to compare the real behavior of the SGA with our model. The average value of the convergence time of 100 independent runs is plotted. We can see from the graph (Fig 2) that the piecewise model correctly bounds the experimental value of the convergence time. However, when the population size is small, the convergence time is significantly lower than the drift time. We therefore must argue that the convergence is not simply forced by the drift stall.

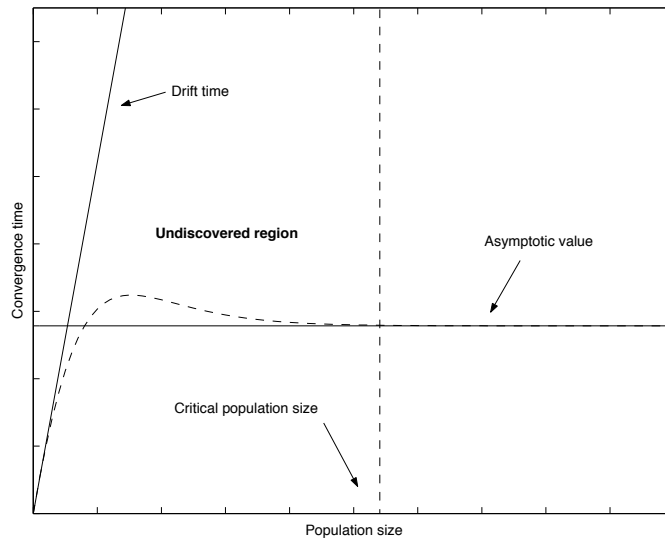


Figure 1: The piecewise convergence-time model. When the population size is small, the convergence time is given by the drift model. With increasing population size, the convergence time increases, limited by the drift time. After the population size has reached the critical value, the convergence time is constant and is equivalent to the asymptotic value. The interval ranging from very small population sizes to the critical size is labeled the *undiscovered region* because the exact value of the convergence time is unknown.

## 4 Finite Populations and Fixation

While the experiments have confirmed the validity of the piecewise model, they show us that random drift is not an exhaustive explanation of the dynamics of the convergence for small populations. The convergence time for small populations is much lower than the drift time. Thus, there must be other factors affecting the convergence with finite populations.

In this section we identify these factors and use this knowledge to build a correct model of the variance of the fitness for finite populations. In the next section, this model is used to extend the asymptotic convergence-time model to finite populations.

### 4.1 Sample Distribution

Before studying in detail the convergence of the SGA, we need to assert the statistical assumptions we use to work with finite populations. All the equations derived in the next sections must be considered valid only as descriptions of the average behavior. The population of randomly initialized individuals represents a sample of size  $N$  extracted from the set of all possible strings of size  $l$  (the *universe*). Thus, the distribution of the fitness of the population must be estimated using the sample mean  $\bar{f}$  and the sample variance  $s^2$ :

$$\bar{f} = \frac{\sum f_i}{N}, \quad s^2 = \frac{\sum (f_i - \bar{f})^2}{N - 1} \quad (8)$$

where  $f_i$  is the value of the fitness of the  $i$ th individual in the population. The sample variance is an unbiased estimator for the variance of the universe:

$$\overline{s^2} = \sigma^2 \quad (9)$$

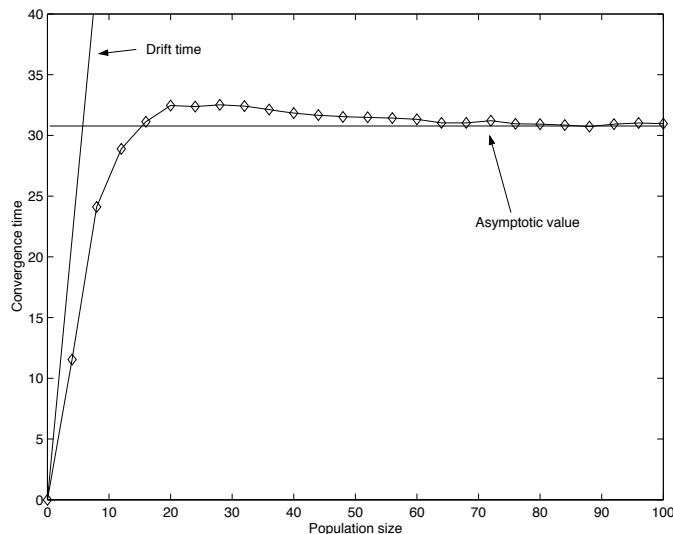


Figure 2: Comparison of the piecewise model with experimental results. The piecewise model correctly bounds the value of the convergence time obtained from the experiments; however, for small population sizes, the convergence time is significantly lower than the value expected from the drift theory.

Now we can calculate the average value of the fitness variance of the randomly initialized population:

$$\overline{\sigma_N^2} = \frac{N-1}{N} s^2 = \frac{N-1}{N} \sigma^2 \quad (10)$$

Therefore, we need the correction factor  $(N-1)/N$  for all the equations involving the variance. Unfortunately, this equivalence cannot be directly used to derive the convergence model. We cannot simply substitute  $\overline{\sigma_N^2}$  in the expression used to derive the asymptotic model, since the assumption of normality is broken by the bias introduced by the genetic operators after the first generation. However, we can use this formula to calculate the initial value of the variance, which will be useful in the following sections:

$$\overline{\sigma_N^2}(0) = \frac{N-1}{N} l P_0 (1 - P_0) \quad (11)$$

## 4.2 Fixation

When the population is finite, it can converge after a certain time determined by the effect of the genetic drift. The convergence is given by the total loss of diversity in the population, which is now composed by copies of the same individual. During the run the diversity is partially lost in the case that all the individuals have the same configuration in a certain position. This event, called *fixation*, is common in biological genetics. The configuration (allele) that appears in all the individuals in a specific position (locus) is said to be *fixed* (Falconer & Mackay, 1996). This is a direct consequence of the finiteness of the population and leads to the possibility of the complete disappearance of a certain configuration.

If we impose a selection pressure, this still happens but tends to increase the fitness. Thus, it is more probable to have a position fixed with the fittest building block. The "gambler's ruin" model (Goldberg, Deb, & Clark, 1992; Feller, 1967) gives us a perfect picture of this behavior. The convergence can be seen as the process of accumulating the optimal building blocks through

the selection strategy. The selection is a competition between two individuals composed of many building blocks. Choosing the best configuration from the individuals' fitnesses is difficult because of the noise introduced by all the building blocks. In the decision model the selection is reduced to a competition between the two better building blocks. The fitness of the two competitors is assumed to be normally distributed. The distance between the means is called the *signal* and is given by the difference between the contributions of the two configurations to the fitness. The variance is given by the building blocks in the other positions. The probability of choosing the optimal configuration is then

$$p = \phi \left( \frac{d}{\sigma \sqrt{2(\frac{l}{k} - 1)}} \right) \quad (12)$$

where  $\phi(x)$  is the cumulative distribution function for the standard normal distribution. In this expression  $d$  is the signal, while the denominator represents the noise. If the noise is too high or the signal is too low, we obtain  $p = 1/2$ . That means the selection method cannot decide between the two building blocks, and the result is random drift. When the population is finite, the accumulation of a configuration can result in the saturation of a position and the complete disappearance of the other competitors. However, the noise introduced by the other configurations implies that the optima do not always win the competition, but can possibly lose it. In case of random drift, the time needed for a configuration to saturate a position is  $t_{drift}$ . Otherwise the decision is biased in the direction of the best building block, and it is easier and faster to accumulate it.

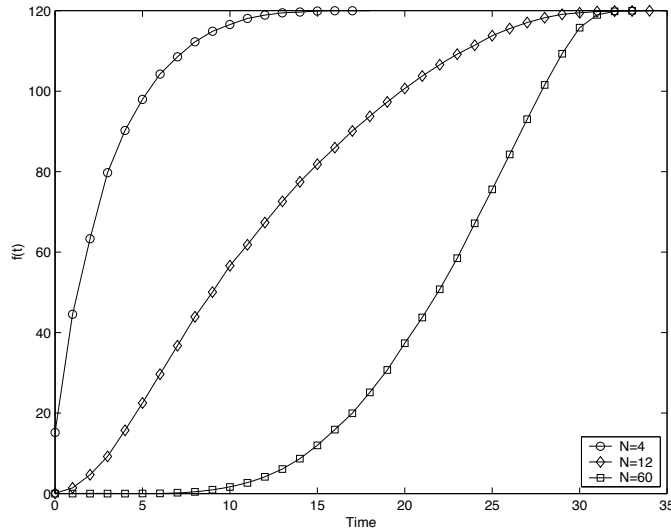


Figure 3: Value of the fixation function  $f(t)$  for different population sizes. The function counts the number of positions fixed at a certain generation. The initial value is given by the probability that the randomly initialized population had fixed positions. The rate of growth is a function of the population size and varies during the run.

To measure this effect, we can define a fixation function  $f(t)$  that counts the number of positions fixed at a certain generation. We can calculate its value from the experiments (Fig 3). At time  $t = 0$ , the value of  $f(t)$  is  $l/2^{N-1}$  if  $P_0 = 1/2$ , where  $1/2^{N-1}$  is the probability that all the individuals had the same value in a certain position. For  $t > 0$ , the number of fixed positions increases until the population converges. The bigger the population the fewer the positions fixed at a certain time, given that is more difficult for all the configurations but



one to be lost in a given position.

### 4.3 Correct Model of the Fitness Variance

We can now use the knowledge about fixation to improve the modeling of the convergence. In the asymptotic convergence-time model, the key step to formulate the dynamics of  $p(t)$  is the construction of the differential equation using the fitness variances. The variance  $\sigma^2(t)$  is supposed to be equal to  $lp(t)[1 - p(t)]$ . This expression is consequence of the assumption that each of the  $l$  positions contributes to the variance by an equal amount  $p(t)[1 - p(t)]$ . But when the population is finite, some positions become fixed during the run and do not contribute anymore to the variance. To derive the correct estimation of the variance, we must then account for these fixed positions and use the modified formula:

$$\overline{\sigma_N^2}(t) = \frac{N-1}{N}[l - f(t)]p(t)[1 - p(t)] \quad (13)$$

where  $f(t)$  is the fixation function. Comparing the model with the experimental results (Fig 4), we can see that where the asymptotic model fails completely in predicting the value of  $\overline{\sigma_N^2}(t)$  for a small population (in this case  $N = 12$  for  $l = 120$ ), the corrected model, which accounts for the effect of fixation, leads to a correct explanation of the dynamics of the fitness variance.

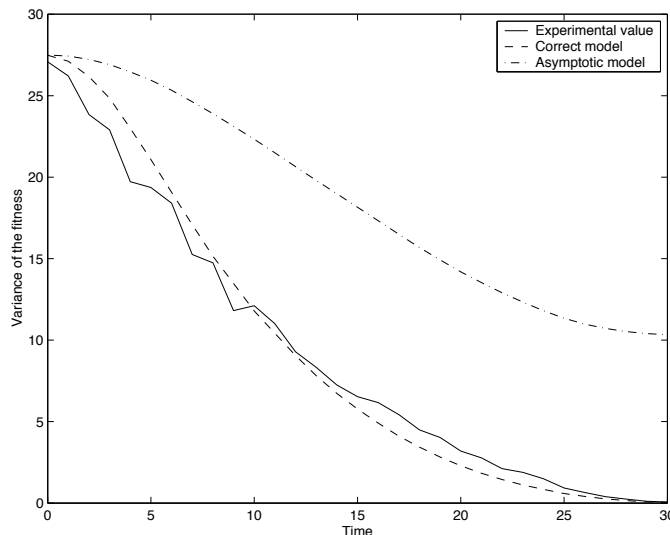


Figure 4: Correct model of the fitness variance. The asymptotic model fails in predicting the value of  $\overline{\sigma_N^2}(t)$  for small population sizes (in this case  $N = 12$  and  $l = 120$ ); correcting the model to account for the effect of fixation leads to a correct prediction of the dynamics of the fitness variance.

## 5 Embedded Convergence-Time Model

In the previous section we identified fixation as the factor affecting the convergence with finite populations. We verified that knowledge of the fixation function is sufficient to correctly model the variance of the fitness. This is a key point, given that we could simply plug the correct model of the variance into the asymptotic convergence model. The problem is that we cannot formulate an analytical expression for the fixation. In the following subsections, we solve this problem using an approximated model of the fitness variance. We introduce this approximated

expression in the asymptotic model and derive the embedded convergence-time model. We prove that this model is an accurate prediction of the behavior of the SGA with finite populations.

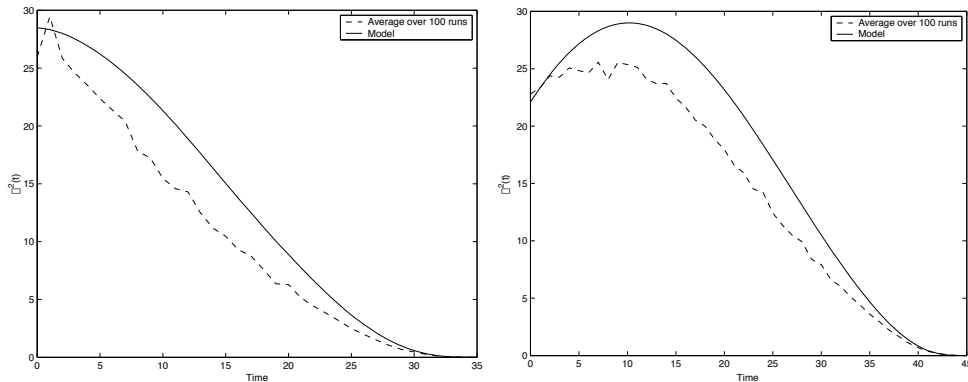


Figure 5: Approximated model of the fitness variance. Comparing the approximated model with the experimental values shows that the error introduced by the approximation is small. The two figures display the dynamics of  $\overline{\sigma_N^2}(t)$  for two different parameter settings of the OneMax problem with  $l = 120$ : (a)  $N = 20$ ,  $P_0 = 0.5$ ; (b)  $N = 50$ ,  $P_0 = 0.25$ .

### 5.1 The Approximated Model of the Fitness Variance

To derive an analytical expression of the convergence time following the same steps used in the derivation of the asymptotic model, we need an analytical formulation of  $\overline{\sigma_N^2}(t)$ . Unfortunately, we do not know how to formulate  $f(t)$ , which is a complex function of the population size and the problem's parameters. To solve this problem, we can use an approximated model of the fitness variance. To construct this model, we can start from the three conditions that this expression must satisfy:

1. Match the value of the variance for the randomly initialized population.
2. Match the value of the variance at the convergence.
3. Account for the effect of population size.

If we do not account for the proportion of fixed positions at the initialization (which is significant only when  $l \gg 2^{N-1}$ ) the value of the fitness variance at  $t = 0$  is given by equation (11). We know that at convergence all the individuals in the population are the same; thus, they have the same fitness and the value of the fitness variance at convergence will be zero. The problem is then to identify when the population has converged. We know from the "gambler's ruin" model that, for a given population size  $N$ , the proportion of optimal building blocks at the end of the run can be calculated with the equation

$$P_n = \frac{1 - \left(\frac{1-p}{p}\right)^{NP_0^k}}{1 - \left(\frac{1-p}{p}\right)^N} \quad (14)$$

(The meaning of  $p$  was explained in the section about fixation.) If we use the proportion of optimal building blocks as an indicator of the state of the run, the convergence is reached when  $p(\tau) = P_n$ . Given these three conditions, we can introduce the following approximated expression for the fitness' variance:

$$\overline{\sigma_N^2}(t) = \frac{N-1}{N} l p(t) [P_n - p(t)] \frac{1 - P_0}{P_n - P_0} \quad (15)$$

It is easy to demonstrate that this expression satisfies all three conditions. Moreover, it converges to the asymptotic model if  $N$  goes to infinity, since  $P_n$  converges to 1. We can now compare this approximated model with the value obtained from experiments. The plot shows that the error introduced by the approximation is small, confirming the validity of this model (Fig 5).

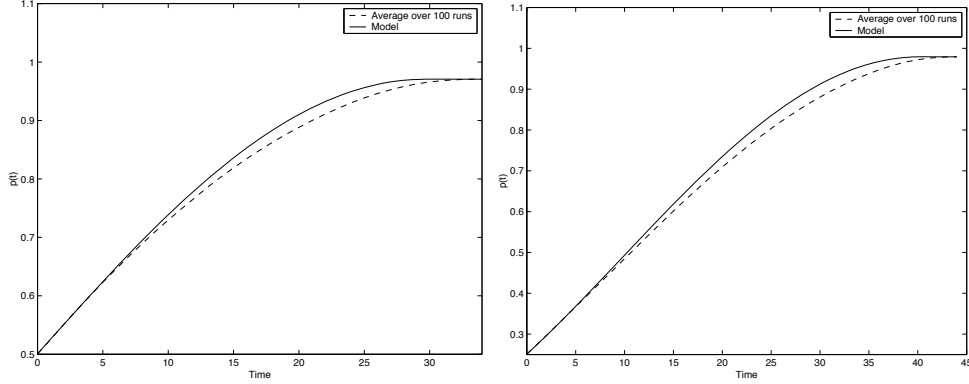


Figure 6: Using the approximated model of the fitness variance, it is possible to predict the dynamics of the proportion of correct building blocks during the run. The two figures display the dynamics of  $p(t)$  for two different parameter settings of the OneMax problem with  $l = 120$ : (a)  $N = 20$ ,  $P_0 = 0.5$ ; (b)  $N = 50$ ,  $P_0 = 0.25$ .

## 5.2 Embedding the Approximated Model of the Fitness Variance in the Asymptotic Convergence Model

With the approximated model of the fitness variance, we can derive the analytical expression of  $p(t)$ . If we follow the same steps given for the asymptotic convergence model, using  $\overline{\sigma_N^2}(t)$  instead of  $\sigma^2(t)$ , we obtain

$$\frac{dp(t)}{dt} = \sqrt{\frac{p(t)[P_n - p(t)]}{\pi l \frac{N}{N-1} \frac{P_n - P_0}{1 - P_0}}} \quad (16)$$

The differential equation can be solved, yielding

$$p(t) = \frac{P_n}{2} \left[ 1 + \sin \left( \frac{t}{\sqrt{\pi l \frac{N}{N-1} \frac{P_n - P_0}{1 - P_0}}} + \arcsin \left( \frac{2P_0 - P_n}{P_n} \right) \right) \right]. \quad (17)$$

By comparing this expression with the value obtained from experiments, we can verify that the approximated model of the fitness variance makes it possible to predict the dynamics of the proportion of correct building blocks during the run (Fig 6). We can use the derived expression to calculate the convergence time. Because at convergence the proportion of optimal building blocks is  $P_n$ , we solve the equality for  $p(\tau) = P_n$  and obtain

$$\tau = \left[ \frac{\pi}{2} - \arcsin \left( \frac{2P_0 - P_n}{P_n} \right) \right] \sqrt{\pi l \frac{N}{N-1} \frac{P_n - P_0}{1 - P_0}} \quad (18)$$

The comparison of the embedded model with the experimental results shows that the model predicts the value of the convergence time with a small error (Fig 7). The approximation

introduced in the model of the fitness variance accounts for the residual error, leading to a failure in predicting the peak in the convergence time.

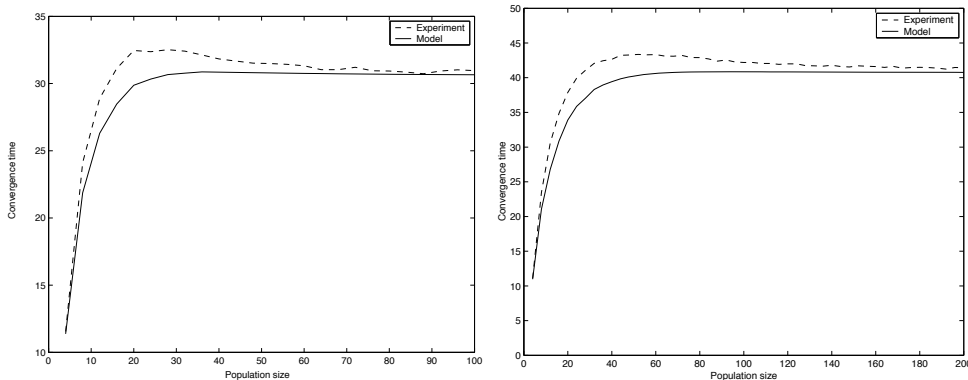


Figure 7: Embedded convergence-time model. The comparison of the embedded model with the experimental results shows that the model predicts the value of the convergence time with a small error; the approximation introduced in the model of the variance accounts for the failure in predicting the peak. The two figures display the value of the convergence time for two different parameter settings of the OneMax problem with  $l = 120$ : (a)  $N = 20$ ,  $P_0 = 0.5$ ; (b)  $N = 50$ ,  $P_0 = 0.25$ .

### 5.3 An Insight from Introducing Approximations into the Embedded Model

We have seen from the plots that the model has still inaccuracy: failing to predict the peak in the convergence time. The nature of this residual error can be explained by reformulating the approximated model of the fitness variance to obtain an equation of the same form as the correct model. We can think of the approximated model as utilizing an approximated fixation function  $f^*(t)$ :

$$\overline{\sigma_N^2(t)} = \frac{N-1}{N} l p(t) [P_n - p(t)] \frac{1 - P_0}{P_n - P_0} = \frac{N-1}{N} [l - f^*(t)] p(t) [1 - p(t)] \quad (19)$$

The expression of  $f^*(t)$  can be obtained from the previous equation:

$$f^*(t) = l \left[ 1 - \frac{[P_n - p(t)](1 - P_0)}{(P_n - P_0)[1 - p(t)]} \right] \quad (20)$$

The value of the approximated fixation function  $f^*(t)$  is compared with the value of the real fixation function  $f(t)$  (Fig 8). The plot shows that the approximated function underestimates the experimental value. Underestimating the value of the fixation leads to an overestimation of the fitness variance (the model calculates a higher number of varying positions). The comparison of the approximated model with the experimental results (Fig 5) confirms this finding. The value of the variance is used in the differential equation to calculate the value of  $p(t)$ . Then the embedded model predicts a faster growth of  $p(t)$ , which results in a lower convergence time.

The next step in the direction of improving the convergence model should be the formulation of a more accurate model of the fixation function. However, to be useful, this model should be easy to embed in the convergence-time model, so that it is still possible to calculate the analytical solution of the differential equation.

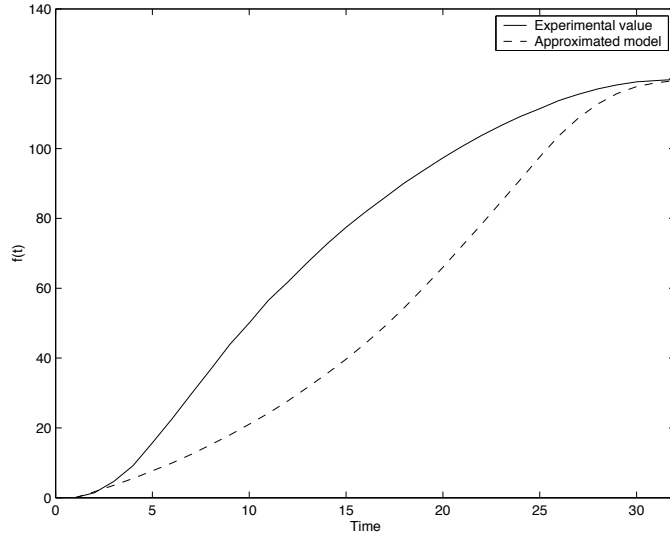


Figure 8: The approximated fixation function. Comparing the approximated value of the fixation function with the value obtained from the experiments leads to the conclusion that the approximated model of the variance of the fitness underestimates the number of fixed positions during the run.

## 6 Extension of the Embedded Model to Different Building-Block Sizes

What we have discussed to this point is valid only for the OneMax problem with  $k = 1$ , irrespective of the proportion of 1 bits in the initial population. Other modifications are needed to extend this model to the case of  $k > 1$ . The key idea for this generalization is that the probability of having the optimal building block is the probability of having  $k$  bits equal to 1. Since the values of the bits can be considered independent, this probability is the  $k$ th power of the probability of having a single bit equal to 1 (Miller, 1997). Therefore we can state that

$$p_k(t) = p_1(t)^k \iff p_1(t) = p_k(t)^{\frac{1}{k}} \quad (21)$$

where the subscripts indicate the dimension of the building block. We can now substitute this value in the previous expressions to obtain the following formula for the variance (where we have eliminated the subscripts since  $p(t)$  now refers only to building blocks of size  $k$ ):

$$\overline{\sigma_n^2(t)} = \frac{N-1}{N} l p^{1/k}(t) (P_n^{1/k} - p^{1/k}(t)) \frac{1 - P_0}{P_n^{1/k} - P_0} \quad (22)$$

The proportion of correct building blocks at time  $t$  become

$$p(t) = \frac{P_n}{2^k} \left( 1 + \sin \left( \frac{t}{\sqrt{\pi l \frac{N}{N-1} \frac{P_n^{1/k} - P_0}{1 - P_0}}} + \arcsin \left( \frac{2P_0 - P_n^{1/k}}{P_n^{1/k}} \right) \right) \right)^k. \quad (23)$$

From this expression we can calculate the value of the convergence time:

$$\tau = \left[ \frac{\pi}{2} - \arcsin \left( \frac{2P_0 - P_n^{1/k}}{P_n^{1/k}} \right) \right] \sqrt{\pi l \frac{N}{N-1} \frac{P_n^{1/k} - P_0}{1 - P_0}} \quad (24)$$

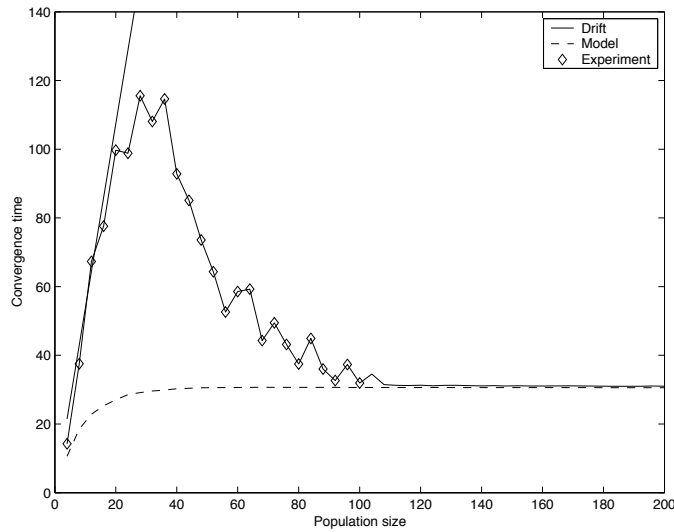


Figure 9: The extension of the embedded model is not sufficient to correctly predict the value of the convergence time. The convergence time initially grows like the drift time, resulting in an unexpected peak.

### 6.1 Post-convergence Drift

If we compare the extended embedded model with the experimental results (Fig 9), we find a large unexpected peak in the convergence time. The slope of the initial portion of the peak is equal to the value predicted by the genetic drift theory. The easiest explanation for this peak is that, when  $k > 1$ , at small population sizes the convergence is dominated by the genetic drift.

To explain this behavior, we need to study the dynamics of the SGA during the run. If we plot the variance of the fitness and the proportion of correct building blocks during a single run (Fig 10), we can see that it takes a considerable amount of time to converge after the fitness variance has gone to zero without discovering better solutions. This does not happen when the population is large enough to find the optimum.

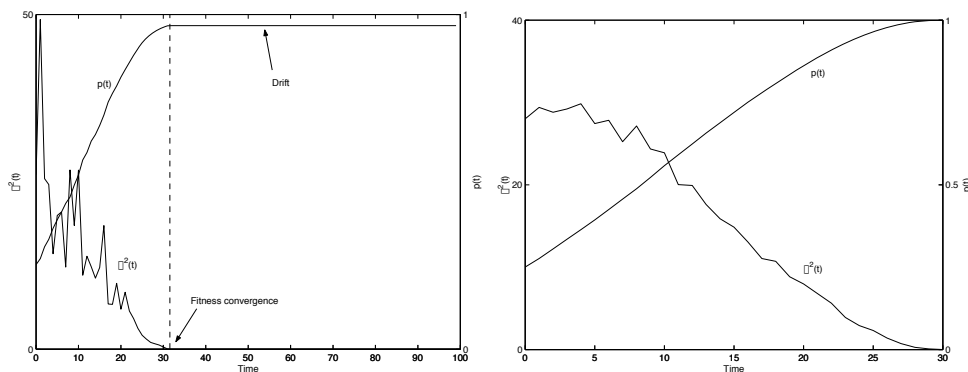


Figure 10: These plots show the difference between the dynamics of the SGA with a small population size ( $N = 24$ ) and with a large population ( $N = 1000$ ). When the population is small, the variance of the fitness goes to zero long before the population converges. The time spent after the convergence of the fitness can be attributed to drift.

With small populations the proportion of correct building blocks at the convergence is  $P_n < 1$ . Because all the individuals at the convergence are equal, this implies that the optimal

configuration is not present in  $l/k(1 - P_n)$  positions. It is easy to verify that, when the variance is zero, in these positions many suboptimal configurations survive (Fig 11), while the other positions are already fixed to the optimal configuration. If we compare the contribution to the fitness of all the possible configurations for  $k = 2$  (Table 1), we can see that the surviving configurations have the same value. At this point the population is then composed of individuals with different genetic codes but with the same fitness. Thus, the population has not converged even if the fitness is already converged (its variance is zero).

Schema	***11***	***10***	***01***	***00***
Fitness	2	1	1	0

Table 1: Contribution to the Fitness of All the Possible Configurations for  $k = 2$

The selection strategy cannot choose between individuals with the same fitness. Every time two individuals are picked to participate in the tournament, one of them is selected completely at random. Thus, the tournament selection degenerates to a random selection strategy. This implies that the convergence is now subject only to the genetic drift, and the residual time needed to make the population converge after the fitness converges, is the drift time.

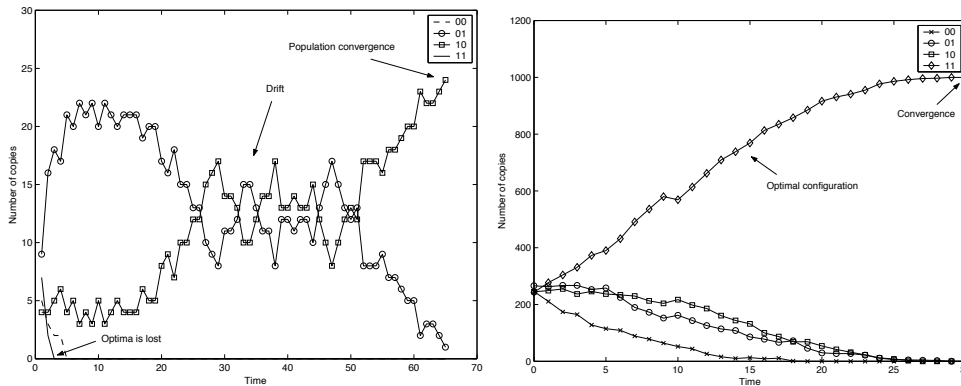


Figure 11: These plots compare the evolution of the distribution of building blocks with a small population size ( $N = 24$ ) and with a large population ( $N = 1000$ ). When the population is small, the optima can be lost in some positions. After this occurs, those positions are occupied by configurations with the same fitness. Thus, convergence can be reached only by drift.

## 6.2 Introducing Post-convergence Drift into the Embedded Model

We can experimentally verify that the embedded model correctly predicts the dynamics of the SGA from  $t = 0$  to the time of the convergence of the fitness (Fig 12). We have explained that the time from the convergence of the fitness to the convergence of the population is the drift time. Thus, the convergence time is given by the sum of the two contributions

$$t_{convergence} = \tau + t_{drift} \quad (25)$$

We know that when the fitness converges a proportion  $P_n$  of the positions are fixed to the optimal configuration. Thus, the other  $l/k(1 - P_n)$  positions contain suboptimal configurations and are responsible for the drift time. We assume that once the optimum disappears from a position, only the configurations with  $k - 1$  bits equal to 1 survive. Thus, the number of configurations (alleles) that are drifting is  $\binom{k}{k-1} = k$ . If  $k = 2$ , we can directly use the equation proposed in Asoh and Mühlenbein (1994):

$$t_{drift} = 1.4 \left[ 0.5 \log \left( \frac{l}{k} (1 - P_n) + e^{-2} \right) + 1 \right]^{1.1} N \quad (26)$$

We have added the constant  $e^{-2}$  to satisfy the condition  $t_{drift} = 0$  when  $P_n = 1$ . If  $k > 2$ , the preceding equation is not valid, and we must use an alternative formula for multiple alleles. For the case of a single position drifting, we find in Asoh and Mühlenbein (1994) that  $t_{drift} = C_0 N = 2N$ . We assume that we can simply substitute the value of  $C_0$  in the formula used for  $l > 1$ ; then

$$t_{drift} = 2 \left[ 0.5 \log \left( \frac{l}{k} (1 - P_n) + e^{-2} \right) + 1 \right]^{1.1} N \quad (27)$$

Comparing the values predicted by the corrected model with the experimental results shows us that the embedded model is now able to predict with great accuracy the peak in the convergence time (Fig 13).

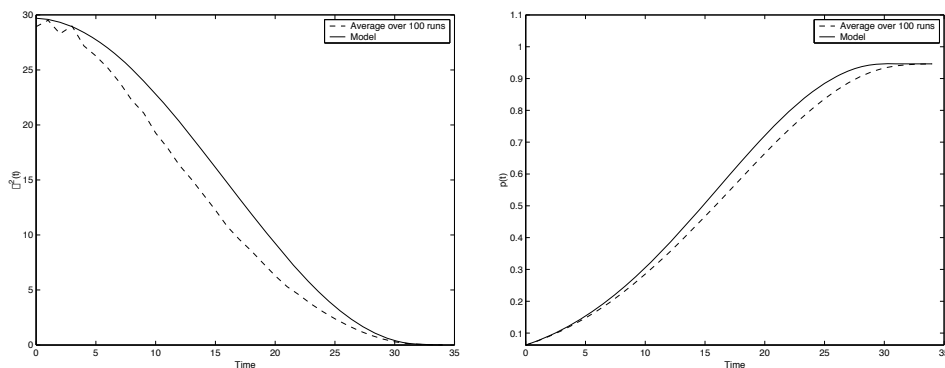


Figure 12: These plots show that the extension to the embedded model is useful for predicting the dynamics of (a)  $\sigma_N^2(t)$  and (b)  $p(t)$  before the convergence of the fitness. The parameters of the problem are  $l = 120$ ,  $k = 4$ ,  $N = 100$ .

## 7 Applications of the Embedded Convergence-Time Model

We propose some possible applications of the embedded convergence-time model, in which knowledge of the convergence time can be useful for an optimal choice of the algorithm's parameters:

1. Continuation techniques
2. Coarse-grain parallel GAs

Continuation techniques are used to improve the quality of the SGA when, due to an insufficient population size, the optimum is not reached. The basic idea of these techniques is to provide a continuation operator that reinitializes the population, restarting the search for the optimum after an unsuccessful convergence. Every run of the GA between two restarts is called an epoch. The way the population is reinitialized is a unique characteristic of each technique. The population size is an important parameter to consider while applying these techniques: the bigger the population, the better the quality reached in every epoch and fewer the times the search must be restarted. Goldberg (1999) proposed the existence of a *continuation problem*.



This problem arises from the trade-off given by the number of times the continuation operator must be applied to reach the optimum and the overhead introduced by this operator. The target is the optimal choice of population size to minimize the time needed to discover the optimum. When the continuation problem was proposed, the author assumed the convergence time was a constant value. In this paper we have demonstrated that this is not true. In the equation for the optimization of the population size, the convergence time is another variable that must be accounted for, moving the trade-off in the direction of smaller populations (capable of faster convergence). The embedded model can thus be used to derive an analytical formulation of the optimal population size given the parameters of the problem.

Another possible application of the embedded model is the optimization of the parameters of parallel GAs (Cantú-Paz, 1999). In particular the coarse-grain GA is the architecture that should most benefit from the modeling of the convergence. The coarse-grain GA is a parallel architecture in which the population is divided in small chunks (demes) and distributed over different communicating processors. As in the continuation techniques, the search is performed during many epochs, each of which can be seen as a single run of an SGA. At the end of each epoch, the population is reinitialized, with the individuals migrating from the communicating demes. Again, the size of every deme is a key factor affecting the number of epochs needed to converge to the optimum. Using the embedded model introduces the trade-off between the number of epochs needed and the duration of every epoch. An optimal population size would minimize the total time used by the parallel GA to discover the optimum.

## 8 Summary and Conclusions

This paper introduced different convergence models for the SGA in the case of finite populations. First, we introduced the piecewise model as a crossover of existing ideas about the convergence; we demonstrated that this model is capable of correctly bounding the convergence time and can be useful for a first approximation of the behavior of the SGA. Then we discussed the key role of the fixation in the convergence of the SGA with small population sizes. We used this knowledge to build an approximated model of the variance of the fitness. We included this expression in the existing asymptotic convergence model to obtain the embedded convergence-time model. Experimental results show the validity of this model. We extended the model to the case of building-block sizes greater than 1. We saw that in this environment an unexpected behavior arises, which we explained as a post-fitness-convergence drift. The explanation of this behavior was introduced into the embedded model to correctly predict the convergence time in this environment.

These models have a variety of applications where knowledge of the convergence time is a key factor for an optimal choice of parameters. There are still some small differences between the embedded model and the experimental values. We discussed the origin of these differences and the necessary steps to improve the embedded model. However, the model we have proposed is useful for predicting the convergence time with only a small error of an SGA with finite population size, using only the problem's parameters.

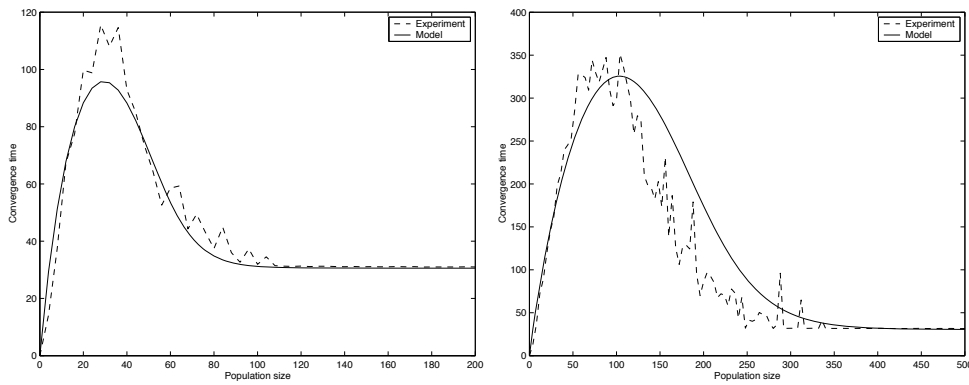


Figure 13: Embedded model with post-convergence drift. Correcting the embedded model to account for the post-convergence drift leads the model to accurately predict the convergence time for building block sizes greater than 1; in this case, the parameters of the problem are  $l = 120$  and (a)  $k = 2$ , (b)  $k = 4$ .

## References

- Asoh, H., & Mühlenbein, H. (1994). On the mean convergence time of evolutionary algorithms without selection and mutation. *Parallel Problem Solving from Nature, PPSN III*, 88–97.
- Blickle, T., & Thiele, L. (1995). A mathematical analysis of tournament selection. *Proceedings of the Sixth International Conference on Genetic Algorithms*, 9–16.
- Cantú-Paz, E. (1999). *Designing efficient and accurate parallel genetic algorithms* (IlliGAL Report No. 99017). Urbana, IL: University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory. Phd Thesis.
- Falconer, D. S., & Mackay, F. C. T. (1996). *Introduction to quantitative genetic*. Edinburgh Gate, Harlow, UK: Addison Wesley Longman Limited.
- Feller, W. (1967). *An introduction to probability theory and its applications*. John Wiley and Sons.
- Goldberg, D. E. (1989a). *Genetic algorithms in search, optimization, and machine learning*. Reading, MA: Addison-Wesley.
- Goldberg, D. E. (1989b). Sizing populations for serial and parallel genetic algorithms. *Proceedings of the Third International Conference on Genetic Algorithms*, 70–79. (Also TCGA Report 88004).
- Goldberg, D. E. (1999). *Using time effectively: Genetic-evolutionary algorithms and the continuation problem* (IlliGAL Report No. 99002). Urbana, IL: University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory.
- Goldberg, D. E., & Deb, K. (1991). A comparative analysis of selection schemes used in genetic algorithms. *Foundations of Genetic Algorithms, 1*, 69–93. (Also TCGA Report 90007).
- Goldberg, D. E., Deb, K., & Clark, J. H. (1991). *Genetic algorithms, noise, and the sizing of populations* (IlliGAL Report No. 91010). Urbana, IL: University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory.
- Goldberg, D. E., Deb, K., & Clark, J. H. (1992). Genetic algorithms, noise, and the sizing of populations. *Complex Systems, 6*, 333–362.

- Miller, B. L. (1997). *Noise, sampling, and efficient genetic algorithms*. doctoral dissertation, University of Illinois at Urbana-Champaign, Urbana. Also IlliGAL Report No. 97001.
- Mühlenbein, H. (1997). The equation for response to selection and its use for prediction. *Evolutionary Computation*, 5(3), 303–346.
- Mühlenbein, H., & Schlierkamp-Voosen, D. (1993). Predictive models for the breeder genetic algorithm: I. Continuous parameter optimization. *Evolutionary Computation*, 1(1), 25–49.
- Shapiro, J., Prügel-Bennett, & Rattray, M. (1994). A statistical mechanical formulation of the dynamics of genetic algorithms. *Evolutionary Computing: AISB Workshop 1994*, 17–27.
- Thierens, D., & Goldberg, D. E. (1994). Convergence models of genetic algorithm selection schemes. *Parallel Problem Solving from Nature, PPSN III*, 119–129.
- Thierens, D., Goldberg, D. E., & Pereira, Â. G. (1998). Domino convergence, drift, and the temporal-salience structure of problems. *Proceedings of 1998 IEEE International Conference on Evolutionary Computation*, 535–540.