

# The optimal population size for uniform crossover and truncation selection

Dragan Cvetković\*  
GMD, Schloß Birlinghoven  
D-53754 St. Augustin

Heinz Mühlenbein  
GMD, Schloß Birlinghoven  
D-53754 St. Augustin

## Abstract

In this paper the optimal population size  $N^*$  is empirically computed for the ONEMAX function and truncation selection.  $N^*$  depends on the size of the problem  $n$ , the probability  $p_0$  of the advantageous allele in the initial population and the selection intensity  $I$ . The dependency of  $N^*$  on  $I$  is very complex. By numerically fitting the data the following formula could be obtained:  $N^* = 1 + f(I) \cdot \sqrt{n} \cdot \ln n \cdot (1/\sqrt{p_0} - 1)$ . Furthermore it is shown that the minimal number of function evaluations  $FE^*$  needed to find the optimum is fairly constant in the range  $1.0 \leq I \leq 1.4$ .

## 1 Introduction

In [MSV93] the breeder genetic algorithm was introduced. For a simplified model a predictive theory could be developed ([MSV94]). This model assumes additive gene effects and uniform crossover. The most simple example of additive gene effects is the fitness function ONEMAX of size  $n$ . ONEMAX gives just the number of 1's in the string. The model needs five parameters to describe the initial state of the population and the selection process. These are the population size  $N$ , the initial frequency of the desirable allele  $p_0$ , the number of loci  $n$ , the mutation rate  $m$  and the intensity of selection  $I$ . In this paper we will assume that the mutation rate is 0 and the crossover rate 1.

For the above model the expected number of generations  $GEN_e$  until convergence has been computed ([MSV94]).  $GEN_e$  is independent of the population size if  $N \geq N^*$ , where  $N^* = N^*(p_0, n, I)$  is the minimal population size needed in order to converge to the optimum with a high probability. We conjectured that  $N^*$  has the following form

$$N^* \approx \sqrt{n} \cdot \ln(n) \cdot f_1(p_0) \cdot f_2(I) \quad (1)$$

---

\*Technical report GMD-AS-TR-94-11

From  $\text{GEN}_e$  the number of function evaluations FE till convergence can be easily computed by  $\text{FE} = N \cdot \text{GEN}_e$ . The selection intensity  $I$  can be chosen by the user, all other parameters are fixed. Therefore the minimum number is given by

$$\text{FE}^* = \min_I \{N^* \cdot \text{GEN}_e\}$$

In this paper we will try to confirm the conjecture concerning  $N^*$ . We will mainly use simulations. All tests are averages over 100 runs. We used elitist truncation selection. The success ratio of converging to the optimum was set to 95%. A different, more qualitative approach for the sizing of the population has been proposed by Goldberg ([Gol92]).

The outline of the paper is as follows. First we investigate how  $N^*$  depends on  $p_0$ , then the dependency on  $n$  and  $I$  is investigated. In section 5 the final expression is obtained. Some comparisons to proportional selection and other fitness functions are made. In section 6 we show that the minimum number of function evaluations is achieved in the area  $1.0 \leq I \leq 1.4$ .

## 2 Dependency of $N^*$ on $p_0$

The first tests were made for  $n \in \{16, 32, 64, 128\}$ , for truncation selection parameter  $T = 0.5$ , and for  $p_0 \in \{\frac{k}{n} \mid 1 \leq k < n\}$ . Some of the results are presented in table 1. The results were obtained in the following way: for each population size the test was run 100 times. If the population did not converge to the optimum at least 95 times, the population size was increased and the test repeated.

$n$	$p_0$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
16	$N^*$	152	80	60	44	28	20	14	9	7
32	$N^*$	224	128	88	64	46	33	24	17	8
64	$N^*$	368	216	144	104	66	53	40	28	15
128	$N^*$	576	448	240	160	113	90	59	41	22

Table 1: Dependency of  $N^*$  on  $p_0$  for different  $n$

The above table suggest a fit of the form  $N^* = a + b/\sqrt{p_0}$ . Using least mean square fit the results of table 2 have been obtained.

The first and the second parameter are (in absolute value) not very distinct. In order to eliminate one parameter, the following function has been chosen as a model:

$$N^* = f_n(p_0) = 1 + a_n \cdot \left(\frac{1}{\sqrt{p_0}} - 1\right) \quad (2)$$

$n$	$a$	$b$
16	-65.96	67.54
32	-100.295	104.12

$n$	$a$	$b$
64	-161.842	167.061
128	-260.823	268.538

Table 2: Table of parameters  $a$  and  $b$  for different  $n$

Figure 1 compares the population size  $N^*$  predicted by the above formula with data obtained by simulations. The fit is fairly good.

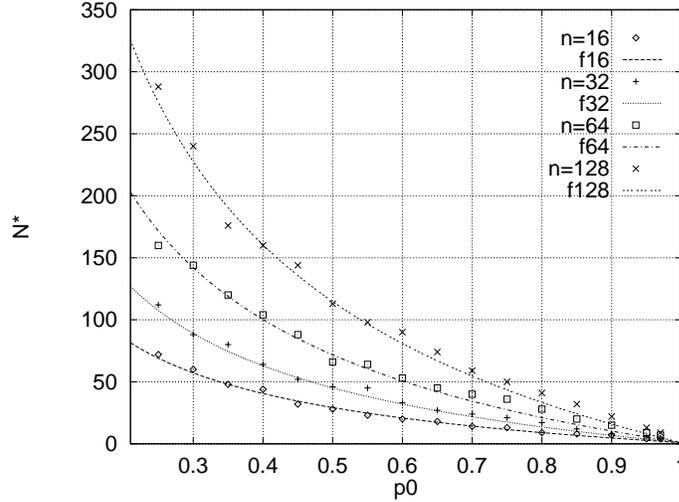


Figure 1: Dependency of  $N^*$  on  $p_0$  for  $n = 16, 32, 64, 128$

### 3 Dependency of $N^*$ on $n$

The next step is to find a functional relation between  $n$  and the constant  $a_n$  in equation (2). Postulating  $a_n = C \cdot \sqrt{n} \cdot \ln n$  we get  $C = 5.11$ , that is

$$a_n = 5.11 \cdot \sqrt{n} \cdot \ln n \quad (3)$$

This leads to the expression

$$N^* = 1 + 5.11 \cdot \sqrt{n} \cdot \ln n \cdot \left( \frac{1}{\sqrt{p_0}} - 1 \right) \quad T = 0.5 \quad (4)$$

Plotting the results together with function (3) we get figure 2. The fit is fairly good.

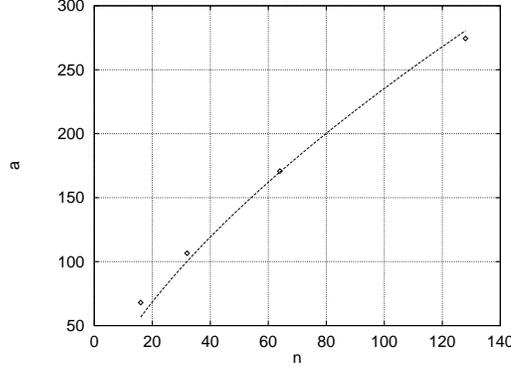


Figure 2: Parameter  $a_n$  for different  $n$  and function that interpolates it

## 4 Dependency of $N^*$ on the selection intensity

This investigation is the most difficult part. We have already mentioned in [MSV94] that  $N^*$  increases for small  $I$  and for large  $I$ . Let us first recall the relation between the truncation threshold  $T$  and the selection intensity  $I$ . If the phenotypes are normally distributed then ([Bul80]):

$$I = \frac{1}{T\sqrt{2\pi}} e^{-t_0^2/2}$$

where  $t_0$  is defined so that:

$$T = \begin{cases} \frac{1}{2} - \frac{1}{\sqrt{\pi}} \int_0^{t_0} e^{-t^2} dt, & \text{if } T \leq \frac{1}{2}; \\ \frac{1}{2} + \frac{1}{\sqrt{\pi}} \int_0^{t_0} e^{-t^2} dt, & \text{if } T > \frac{1}{2}. \end{cases}$$

The selection intensity can be used to estimate the selection differential  $S$ , which is the difference between the average of the selected parents and the average of the population [MSV93]. We computed  $I$  numerically for different  $T$ . Some numerical values are given in table 3. Note that the above formula is valid for an infinite population size, but the values for smaller population sizes differ not very much.

$T$	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
$I$	2.665	2.063	1.755	1.4	1.159	0.966	0.798	0.644	0.496	0.350

Table 3: Table of values for  $I$  for different  $T$

If we substitute  $C_I$  for 5.11 in equation (4) we get

$$C_I = \frac{N^* - 1}{\sqrt{n} \cdot \ln n \cdot \left(\frac{1}{\sqrt{p_0}} - 1\right)} \quad (5)$$

Extensive tests were performed for estimating  $C_I$  for different  $I$  as well as for different  $n$  and  $p_0$ . However, the problem seems to be very hard. The functional dependency can be divided into 4 regions as presented in figure 3(a) on page 5. The first region is defined by small selection intensities  $I \in [0.0, 0.2]$  where it is very difficult to determine  $N^*$ . The second area is given by  $I \in [0.2, 0.8]$  and the third area by  $I \in [0.8, 1.6]$ . In the fourth region  $N^*$  and  $C_I$  increase very fast. For the strongest selection possible,  $N^*$  can be calculated by probability theory. If only the best individual is selected, then we have immediate convergence. Therefore the optimum has to be contained in the initial population. Let  $a$  be the probability that the optimum is contained in the initial population. Then

$$N^* = \frac{\ln(1 - a)}{\ln(1 - p_0^n)}$$

For instance, for  $p_0 = 0.5$ ,  $a = 0.95$  and  $n = 32$  we have that  $C_\infty \approx 1.58 \cdot 10^9$ . Very strong selection reduces the number of generations, but the optimal population size increases dramatically. Therefore strong selection gives no efficient genetic algorithm. We will show in the next section that the third region (i.e.  $I \in [0.8, 1.6]$ ) will contain the selection intensity giving the minimum number of function evaluations for obtaining the optimum.

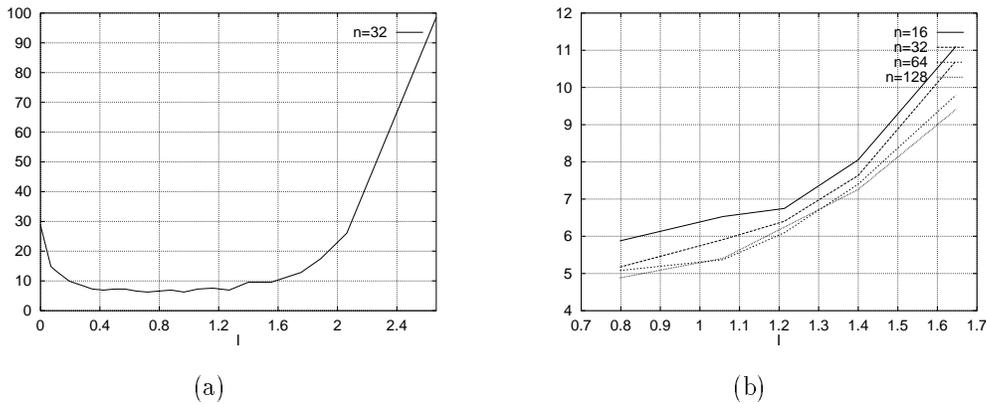


Figure 3: (a) Dependency of ‘constant’  $C_I$  on selection intensity  $I$  for  $n = 32$  and for  $p_0 = 0.75$ ; (b) Graph of results of dependency of  $C_I$  on  $I$  in interval  $[0.8, 1.6]$  for different  $n$  and for  $p_0 = 0.5$

$n$	$I$	0.797	1.058	1.213	1.400	1.647
16	$C_I$	5.878	6.531	6.749	8.054	11.102
32	$C_I$	5.172	5.911	6.403	7.635	10.713
64	$C_I$	5.079	5.370	6.095	7.401	9.796
128	$C_I$	4.882	5.409	6.245	7.257	9.412

Table 4: Dependency of  $C_I$  on  $I$  in interval  $[0.8, 1.6]$  for different  $n$  and for  $p_0 = 0.5$

The region ( $I \in [0.8, 1.6]$ ) is more closely investigated in figure 3(b) and table 4.

We next try to approximate the data of region 3 with a function of the type  $C_I = a - bI + cI^2$  for some constants  $a$ ,  $b$  and  $c$ . For  $n = 64$  and  $n = 128$ , we present the results graphically in figure 4.

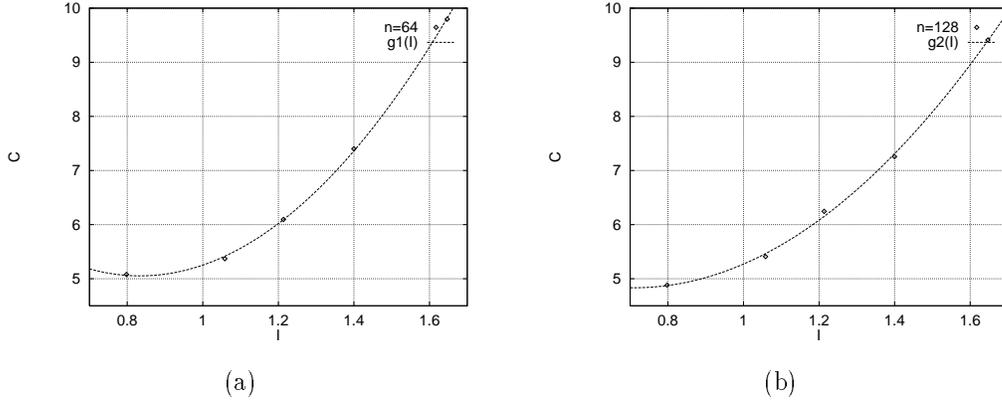


Figure 4: Results of approximation for  $n = 64$  (a) and for  $n = 128$  (b) where  $g_1(I) = 10.061 - 12.012I + 7.203I^2$  and  $g_2(I) = 7.455 - 7.386I + 5.202I^2$

The above constants depend on the size of the problem  $n$ . But this is violating our assumption. Therefore we try to find an expression for  $f(I)$  which is independent of  $n$ . The result is the function  $f(I) = 10.28 - 12.07I + 7.3I^2$ . A comparison between the function values and empirical data is shown in figure 5.

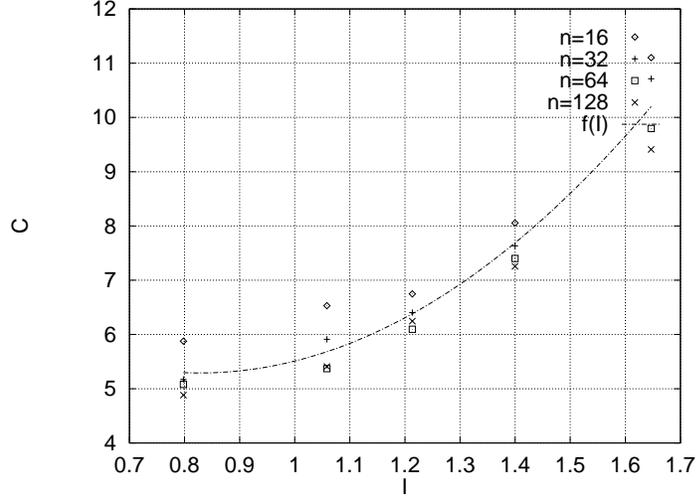


Figure 5: Approximation of all data with function  $f(I) = 10.28 - 12.07I + 7.3I^2$

## 5 The final expression

Combining the results obtained so far we get the following expression for  $N^*$ :

$$N^* = 1 + (10.28 - 12.07I + 7.30I^2)\sqrt{n} \cdot \ln n \cdot \left(\frac{1}{\sqrt{p_0}} - 1\right) \quad 0.8 \leq I \leq 1.4 \quad (6)$$

In order to see how well our function predicts the optimal population size  $N^*$ , we ran some tests for random problem instances. The results of the tests are presented in table 5(a). The predicted  $N^*$  is calculated using formula (6).

Two predictions are very good. The third prediction has an error of 30%. The reason for this is the high value of  $p_0$ . In this case many more runs have to be made, because the population is converging very fast.

All the above results depend on truncation selection and the ONEMAX function. We also made some tests with proportionate selection and other additive fitness functions. The results for proportional selection are presented in table 5(b) together with results for truncation selection with  $I = 1.06$  (i.e. with  $T = 0.35$ ). All test were made with  $p_0 = 0.5$  and with uniform crossover with probability  $p_c = 1$ . Note that the optimal popsize for proportional selection is almost twice that of truncation selection.

Next we investigate an additive fitness function with non-uniform gene contributions. This function plays an important role for floating point representations.

$$F_1(x_1, \dots, x_n) = \sum_{i=1}^n x_i 2^{-i}, \quad x_i \in \{0, 1\}.$$

$n$	$p_0$	$I$	$N_{\text{predicted}}^*$	$N_{\text{obtained}}^*$
128	0.75	1.05	49	53
256	0.9	1.4	38	52
512	0.75	0.8	116	120

(a)

$n$	$N_{\text{prop}}^*$	$N_{\text{trunc}}^*$
16	56	31
32	88	49
64	144	75

(b)

Table 5: (a) Results of tests with different parameters; (b) Comparison of results for  $N_{\text{prop}}^*$  obtained with proportional selection with those for  $N_{\text{trunc}}^*$  obtained with truncation selection with  $I = 1.06$

In the table 6(a) the results are displayed for the BGA with elitist truncation selection ( $T = 0.2$  and  $T = 0.4$ ), no mutation and with uniform crossover with probability 1. The optimal population size is dramatically higher than for the ONEMAX function. The reason for this behavior is the fact that the selection process fixes the genes which make a small contribution to the fitness function.

The next function is more moderate. This function has also been used by Goldberg [Gol92].

$$F_2(x_1, \dots, x_n) = \sum_{i=1}^n c_i x_i$$

where  $x_i \in \{0, 1\}$  and

$$c_i = \begin{cases} \delta, & \text{if } i \equiv 5 \pmod{10}; \\ 1, & \text{otherwise;} \end{cases}$$

In table 6(b) and (c) the results are given for the BGA with elitist truncation selection ( $T = 0.4$  and  $T = 0.2$ ), no mutation and uniform crossover with probability 1. The optimal population size  $N^*$  is much less for this function, but still considerably higher than for ONEMAX.

$n$	20	50
$T = 0.2$	205	682
$T = 0.4$	137	424

(a)

$n$ ( $T = 0.4$ )	20	50
$\delta = 0.2$	38	85
$\delta = 0.4$	39	78

(b)

$n$ ( $T = 0.2$ )	20	50
$\delta = 0.2$	56	122
$\delta = 0.4$	53	105

(c)

Table 6: (a)  $N^*$  for the function  $F_1$ ; (b)  $N^*$  for the function  $F_2$  with  $T = 0.4$ ; (c)  $N^*$  for the function  $F_2$  with  $T = 0.2$

The above results confirm our theoretical analysis in [MSV94] that recombination by crossover should not be used alone, but in connection with mutation.

## 6 Minimum number of function evaluations

The number of function evaluations FE is given by  $FE = GEN_e \cdot N$  where  $GEN_e$  is the number of generations until convergence. FE is minimized for the optimal population size  $N^*$ .  $GEN_e$  has been computed in [MSV94]. It is given by

$$GEN_e \approx \left( \frac{\pi}{2} - \arcsin(2p_0 - 1) \right) \cdot \frac{\sqrt{n}}{I} \quad (7)$$

$GEN_e$  is reduced inversely proportional to the selection intensity  $I$ . The question now arises which selection intensity gives the least number of function evaluations? Using equation 6 we easily obtain that  $I = 1.18$  minimizes  $N^* \cdot GEN_e$ . We will confirm this result by simulations. The results are given in table 7. One observes that a selection intensity of  $I = 1.2$  is the best. But all selection intensities in the range between 1.0 and 1.4 (i.e. truncation between 0.2 and 0.4) give almost the same results.

$n$	$I$	0.798	1.058	1.213	1.400	1.647
16	FE*	257.88	199.95	176.00	183.92	213.72
32	FE*	568.89	502.74	456.86	466.20	557.92
64	FE*	1371.72	1095.75	1089.70	1125.79	1283.84
128	FE*	3109.12	2620.12	2634.06	2677.58	2930.45

Table 7: Optimal number of evaluations FE\* as a function of selection intensity  $I$  for  $n \in \{16, 32, 64, 128\}$

## 7 Conclusion

In this paper the problem of the minimal population size needed for convergence to the optimum has been empirically investigated. It could be shown that truncation thresholds in the area from 0.2 to 0.4 will approximate give the minimal number of function evaluations, if the corresponding minimal population size is used. This result indicates that the efficiency of the BGA does not critically depend on an optimum truncation threshold. We have also shown that proportional selection needs a much larger optimal population size. The above result is valid for the ONEMAX function. Recombination by uniform crossover seems to be inefficient for fitness functions with non-uniform gene contributions. In [MSV94] it is shown how an arbitrary fitness function can be decomposed into an additive and no-additive parts. We have to investigate if this decomposition enables us to transfer some of the results obtained in this paper to more general fitness functions.

## References

- [Bul80] M. G. Bulmer. "*The Mathematical Theory of Quantitative Genetics*". Clarendon Press, Oxford, 1980.
- [Gol92] D.E. Goldberg. Genetic algorithms, noise, and the sizing of populations. *Complex Systems*, 6:333–362, 1992.
- [MSV93] H. Mühlenbein and D. Schlierkamp-Voosen. Predictive Models for the Breeder Genetic Algorithm I. Continuous Parameter Optimization. *Evolutionary Computation*, 1:25–49, 1993.
- [MSV94] H. Mühlenbein and D. Schlierkamp-Voosen. The science of breeding and its application to the breeder genetic algorithm. *Evolutionary Computation*, 1, 1994.