



## Population Sizing for Entropy-based Model Building in Genetic Algorithms

Tian-Li Yu, Kumara Sastry, David E. Goldberg, and Martin Pelikan

MEDAL Report No. 2006006

June 2006

### Abstract

This paper presents a population-sizing model for the entropy-based model building in genetic algorithms. Specifically, the population size required for building an accurate model is investigated. The effect of the selection pressure on population sizing is also incorporated. The proposed model indicates that the population size required for building an accurate model scales as  $\Theta(m \log m)$ , where  $m$  is the number of substructures and proportional to the problem size. Experiments are conducted to verify the derivations, and the results agree with the proposed model.

### Keywords

Genetic algorithms, linkage learning, model building, population sizing, selection pressure, scalability.

### Note

Also published as IlliGAL Report No. 2006020 at the Illinois Genetic Algorithms Laboratory at <http://www-illigal.ge.uiuc.edu/>.

Missouri Estimation of Distribution Algorithms Laboratory (MEDAL)  
Department of Mathematics and Computer Science  
University of Missouri–St. Louis  
One University Blvd., St. Louis, MO 63121  
E-mail: [medal@cs.ums1.edu](mailto:medal@cs.ums1.edu)  
WWW: <http://medal.cs.ums1.edu/>

# Population Sizing for Entropy-based Model Building in Genetic Algorithms

Tian-Li Yu<sup>1</sup>, Kumara Sastry<sup>1</sup>, David E. Goldberg<sup>1</sup>, and Martin Pelikan<sup>2</sup>

<sup>1</sup>Illinois Genetic Algorithms Laboratory

University of Illinois at Urbana-Champaign, IL, USA

<sup>2</sup>Missouri Estimation of Distribution Algorithms Laboratory

University of Missouri at St. Louis, MO, USA

{tianliyu,kumara,deg}@illigal.ge.uiuc.edu, pelikan@cs.umsl.edu

June 22, 2006

## Abstract

This paper presents a population-sizing model for the entropy-based model building in genetic algorithms. Specifically, the population size required for building an accurate model is investigated. The effect of the selection pressure on population sizing is also incorporated. The proposed model indicates that the population size required for building an accurate model scales as  $\Theta(m \log m)$ , where  $m$  is the number of substructures and proportional to the problem size. Experiments are conducted to verify the derivations, and the results agree with the proposed model.

## 1 Introduction

Genetic evolutionary computation (GEC) researchers have long realized the importance of population sizing on the success and efficiency of GEC. While using a smaller population usually yields low-quality solutions, using a population of size larger than required leads to wastage of computational resources. Therefore, facetwise models, such as initial-supply (Goldberg, Sastry, & Latoza, 2001) and decision-making models (Goldberg, Deb, & Clark, 1992a; Harik, Cantú-Paz, Goldberg, & Miller, 1997), have been developed to model different bounds on population sizing required for genetic algorithm (GA) success.

The issue of population sizing is equally critical, if not more, in model-building GAs, such as the estimation of distribution algorithms (EDAs) (Larrañaga & Lozano, 2002), which build linkage models for the given problems and utilize the knowledge gained from linkage models to efficiently recombine new solution candidates. For model-building GAs, the population should be sized properly not only to satisfy the needs of initial-supply and decision-making, but also to ensure the accuracy of the model-building.

Pelikan, Sastry, and Goldberg (2003) showed that the population size required for the Bayesian optimization algorithm (BOA) to build an accurate model scales as a polynomial of the problem size of the order between 1.05 and 2.1.

$$\Theta(m^{1.05}) \leq n \leq \Theta(m^{2.1}). \quad (1)$$

These bounds also apply to many other model-building GAs, and empirical results show that  $n$  roughly scales as  $\Theta(m^{1.4})$  (Sastry & Goldberg, 2004). However, we need a more refined model to

explain the empirical results and better understand population sizing. In addition, empirical results also indicate that the selection pressure affects the population size, and that an optimal selection pressure exists for model building.

The purpose of this paper is to develop a facetwise population-sizing model for the modeling building in GAs. The model is anticipated to better explain the scalability of model-building GAs and capture the effect of the selection pressure on population-sizing requirements.

The paper is organized as follows. First we derive the distribution of the entropy measurement used in the model builder. Next we show how the selection pressure affects the distribution. Finally, we derive the population-sizing model based on the decision-making over the entropy measurement.

## 2 Population-Sizing Models

Facetwise and dimensional models have been very effective not only in the design of genetic algorithms, but also in understanding GA dynamics and mechanisms. Since our methodology depends on the facetwise models of population sizing, we briefly outline the models dictated by building block supply, decision making, and accurate linkage learning in the remainder of this section.

### 2.1 Building-Block Supply Model

The first step towards understanding population sizing is to tackle the issue of building-block (BB) supply, where the minimum population size required to ensure the presence of at least one copy of all raw schemata is modeled. Holland (1975) estimated the number of BBs that receive at least a specified number of trials using Poisson distribution. A later study (Goldberg, 1989) calculated the same quantity more exactly using binomial distribution and studied their effects on population sizing in serial and parallel computation. Reeves (1993) proposed a population sizing model for supply of alphabets with fixed cardinality. Recently, Goldberg, Sastry, and Latoza (2001) developed facetwise models for ensuring BB supply in the initial population for genetic algorithms. They considered a population of fixed-length strings consisting alphabets of arbitrary cardinality  $\chi$ . Goldberg *et al* predicted that the population size required to ensure the presence of all competing building blocks with a tolerance of  $\epsilon = 1/m$  is given by

$$n = \chi^k (k \log \chi + \log m), \quad (2)$$

where  $\chi$  is the alphabet cardinality,  $k$  is BB size, and  $m$  is the number of BBs.

### 2.2 Gambler's Ruin Population-Sizing Model

Goldberg, Deb, and Clark (1992b) proposed population-sizing models for correctly deciding between competing BBs. They incorporated noise arising from other partitions into their model. However, they assumed that if wrong BBs were chosen in the first generation, the GAs would be unable to recover from the error. Harik, Cantú-Paz, Goldberg, and Miller (1999) refined the above model by incorporating cumulative effects of decision making over time rather than in first generation only. Harik et al. (1999) modeled the decision making between the best and second best BBs in a partition as a gambler's ruin problem. Here we use an approximate form of the population-sizing model proposed by Harik et al. (1999):

$$n = \frac{\sqrt{\pi} \sigma_{BB}}{2} \frac{\sigma_{BB}}{d} 2^k \sqrt{m} \log m, \quad (3)$$

where  $k$  is the BB size,  $m$  is the number of BBs,  $d$  is the size signal between the competing BBs, and  $\sigma_{BB}$  is the fitness variance of a building block. building blocks. The above equation assumes a failure probability,  $\alpha = 1/m$ .

### 2.3 Model-Building+Decision Making Population Sizing

Facetwise modes for incorporating the effects of model building and BB-wise decision making on the population size have been analyzed for estimation of distribution algorithms (EDAs) in general, and Bayesian optimization algorithm and extended compact genetic algorithm in particular (Pelikan, Goldberg, & Cantú-Paz, 2000; Pelikan, Sastry, & Goldberg, 2003; Sastry & Goldberg, 2000; Sastry & Goldberg, 2004). The population-sizing model which incorporates the effect of model-building and its accuracy on the population sizing of the GA, and predicts the population size required to solve a problem with  $m$  building blocks of size  $k$  with a failure rate of  $\alpha = 1/m$ , is given by

$$n = c_n \cdot 2^k \left( \frac{\sigma_{BB}}{d} \right)^2 m \log m, \quad (4)$$

where  $n$  is the population size,  $c_n$  is a problem-dependent constant,  $k$  is the BB length,  $\alpha$  is the probability of failure.

## 3 Distribution of the Entropy Measurement

Many different metrics have been proposed to detect the linkage for model building. One of the most commonly-used measurement is Shannon's entropy (Shannon, 1948). Typical examples for such entropy-based model-building GAs include FDA (Mühlenbien, 1999), eCGA (Harik, 1999), EBNA (Etxeberria & Larrañaga, 1999), BOA (Pelikan, Goldberg, & Cantú-Paz, 1999), and DSMGA (Yu & Goldberg, 2006), and the work of Wright *et al.* (Wright, Poli, Stephens, Landgon, & Pulavarty, 2004).

The loss of the entropy by linking two decision variables together is essentially the mutual information:  $I(X;Y) = H(X) + H(Y) - H(X;Y)$  (Cover & Thomas, 1991). In entropy-based model building, the measurement  $I(X;Y)$  needs to be significant enough for accurately detecting the existence of the linkage between  $X$  and  $Y$ .

In this paper, we focus on model building. The scenario is that after unbiased initialization of the population, the GA performs selection and then builds the model. We investigate the population size required for building an accurate model. We also assume that binary tournament selection is adopted.

Consider two random variables  $X_0$  and  $X_1$ :

$X_0$ : the measurement of the linkage between two *independent* decision variables.

$X_1$ : the measurement of the linkage between two *dependent* decision variables.

Let  $\hat{X}_{0,n}$  and  $\hat{X}_{1,n}$  be the means of  $n$  independent samples of  $X_0$  and  $X_1$ , respectively. For an infinite number of samples, it is easily seen that  $E[\hat{X}_{0,n=\infty}] = E[X_0] = 0$  and  $E[\hat{X}_{1,n=\infty}] = E[X_1]$ . For a finite number of samples, the mean and variance of the mutual information of a pair of independent random variables is known as follows (Hutter & Zaffalon, 2004).

$$E[\hat{X}_{0,n}] = \frac{1}{2n} + \mathcal{O}\left(\frac{1}{n^2}\right) \quad \text{Var}[\hat{X}_{0,n}] = \frac{1}{2n^2} + \mathcal{O}\left(\frac{1}{n^3}\right). \quad (5)$$

To calculate  $E[\hat{X}_{1,n}]$ , first we need to estimate  $E[X_1]$ . The derivation is based on Holland's Royal road function, which serves as a worst case scenario for the model building in GAs. It is extremely difficult because given the minimal signal  $d_{min}$ , the signals between to best schema and all other  $(2^k - 1)$  schemata are all  $d_{min}$ . In other words, for a fixed  $d_{min}$ , the growth of the best schema is the slowest in the Royal road function under the same selection pressure.

The following derivations are based on decision making. Similar derivations can be found in Goldberg, Deb, and Clark (1992b) and Pelikan, Sastry, and Goldberg (2003). Define a subfunction of order  $k$  as

$$R_k(\vec{x}) = \begin{cases} 1, & \text{if } \vec{x} = \underbrace{111 \cdots 1}_k \\ 1 - d, & \text{otherwise.} \end{cases} \quad (6)$$

The fitness of an additively decomposable problem with  $m$  building blocks (BBs) is then defined as  $f(\vec{x}) = \sum_{i=0}^{m-1} R_k(x_{ik+1}x_{ik+2} \cdots x_{ik+k})$ . Since the number of BBs is proportional to the problem size, for simplicity, the terms "the problem size" and "the number of BBs" are interchangeable in this paper.

Without loss of generality, we investigate the competition between alleles 0 and 1 at the first gene. Define two schemata

$$H_0 = 0 \underbrace{*** \cdots *}_{k-1} \quad H_1 = 1 \underbrace{*** \cdots *}_{k-1}. \quad (7)$$

Let  $F_0$  and  $F_1$  be the random variables of the fitness values of  $H_0$  and  $H_1$ , respectively. According to the central limit theory (Feller, 1966), the distributions of  $F_0$  and  $F_1$  are close to a normal distribution. The variances of  $F_0$  and  $F_1$  are the same, which is defined as  $\sigma_F^2$ . By treating other  $(m - 1)$  BBs as external noises, we can express the bounds for  $\sigma_F^2$  as

$$\begin{aligned} (m - 1)\sigma_{BB}^2 &\leq \sigma_F^2 \leq m\sigma_{BB}^2 \\ \Rightarrow \sigma_F^2 &= m\sigma_{BB}^2 \cdot (1 - \mathcal{O}(\frac{1}{m})), \end{aligned} \quad (8)$$

where  $\sigma_{BB}^2$  is the variance of the fitness of a BB.

Define  $Z = F_1 - F_0$ .  $Z$  is also of normal distribution and has the mean and variance as follows.

$$E[Z] = f(H_1) - f(H_0) = \frac{d}{2^{k-1}}. \quad (9)$$

$$Var[Z] = \sigma_F^2 + \sigma_F^2 = 2m\sigma_{BB}^2 \cdot (1 - \mathcal{O}(\frac{1}{m})). \quad (10)$$

The probability that  $H_1$  wins over  $H_0$  in the binary tournament selection is given by  $\Phi(\frac{E[Z]}{\sqrt{Var[Z]}})$ , where  $\Phi$  is the cumulative standard normal distribution. Define

$$z = \frac{E[Z]}{\sqrt{Var[Z]}} = \frac{d}{2^{k-1}\sqrt{2m} \cdot \sigma_{BB}} + \mathcal{O}(m^{-1.5}). \quad (11)$$

For large  $m$ ,  $z$  is small, and  $\Phi(z)$  can be approximated by  $\frac{1}{2} + \frac{z}{\sqrt{2\pi}} - \mathcal{O}(z^3)$  (Abramowitz & Stegun, 1970).

$$\Phi(z) = \frac{1}{2} + \frac{d}{2^k \sqrt{\pi m} \cdot \sigma_{BB}} \pm \mathcal{O}(m^{-1.5}). \quad (12)$$

Define  $p_0$  and  $p_1$  as the proportions of  $H_0$  and  $H_1$  before selection ( $p_0 = p_1 = \frac{1}{2}$ ), and  $p'_0$  and  $p'_1$  are that after selection.

$$p'_0 = p_0^2 + 2p_0p_1\Phi(-z) = \frac{1}{2} - \frac{\Delta_m}{2}, \quad (13)$$

$$p'_1 = p_1^2 + 2p_0p_1\Phi(z) = \frac{1}{2} + \frac{\Delta_m}{2}, \quad (14)$$

$$\text{where } \Delta_m = \frac{d}{2^k \sqrt{\pi m} \cdot \sigma_{BB}}. \quad (15)$$

Equations 13 and 14 describe clearly the change of the proportions of 1's and 0's after selection.

The entropy of a binary random variable is defined as  $H(\vec{p}) = H(p, 1-p) = -p \log_2 p - (1-p) \log_2(1-p)$ . Now we calculate the change of the entropy of one decision variable after selection.

$$H(\vec{p}') - H(\vec{p}) = -\frac{\Delta_m^2}{2 \ln 2} + \mathcal{O}(\Delta_m^4). \quad (16)$$

Now consider two decision variables together. Let  $\vec{q}$  denote the joint probability of two dependent genes. By similar derivations, we obtain

$$\begin{aligned} q'_{11} &= \frac{1}{4} + \frac{3}{4}\Delta_m, \\ q'_{00} = q'_{01} = q'_{10} &= \frac{1}{4} - \frac{1}{4}\Delta_m. \end{aligned} \quad (17)$$

The change of the joint entropy can be expressed as

$$H(\vec{q}') - H(\vec{q}) = -\frac{3\Delta_m^2}{2 \ln 2} + \mathcal{O}(\Delta_m^3). \quad (18)$$

Finally, we are able to calculate the mutual information, the gain of the certainty by linking two dependent genes.

$$\begin{aligned} I(X'; Y') &= 2H(\vec{p}') - H(\vec{q}') \\ &= I(X; Y) + \frac{\Delta_m^2}{2 \ln 2} - \mathcal{O}(\Delta_m^3) \end{aligned} \quad (19)$$

For an unbiased initialization of the initial population,  $I(X; Y) = 0$ . Hence

$$E[X_1] = E[\hat{X}_{1, n=\infty}] = \frac{\Delta_m^2}{2 \ln 2} - \mathcal{O}(\Delta_m^3). \quad (20)$$

When the number of samples are finite, the sampling noise makes the mutual information larger (Hutter & Zaffalon, 2004). Kleiter (1999) showed that the variance of the mutual information can be approximated as  $\frac{2E[I]}{n}$ . Therefore, we have the following relations.

$$E[\hat{X}_{1, n}] \geq \frac{\Delta_m^2}{2 \ln 2} \quad \text{Var}[\hat{X}_{1, n}] \geq \frac{\Delta_m^2}{n \ln 2}. \quad (21)$$

Now we empirically verify the derived means and variances of  $\hat{X}_{0, n}$  and  $\hat{X}_{1, n}$ . The test is done by applying a GA on the  $(m, k)$ -trap (Goldberg, 1987) with  $k = 4$  and  $d = 0.25$ . After the binary tournament selection, we compute the mutual information of the pairs of independent genes and the pairs of dependent genes. All results are averaged over 10000 independent runs.

First we fix the problem size  $m$  to 10 and measure the mutual information by varying the population size  $n$ . Figure 1 shows the results. We can see that  $E[\hat{X}_{0, n}]$  and  $\text{Var}[\hat{X}_{0, n}]$  roughly

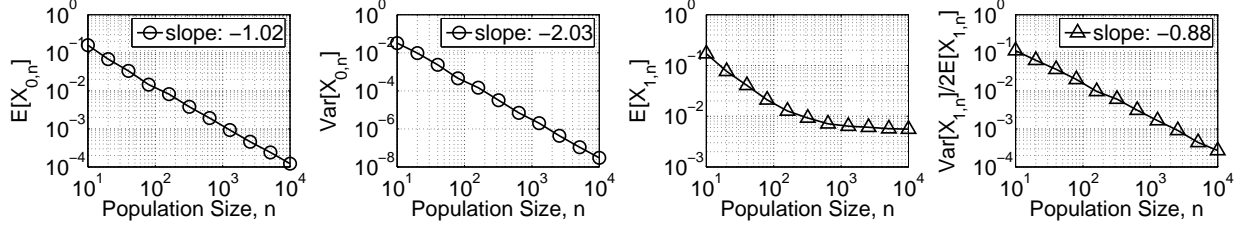


Figure 1: Means and variances versus the population size for a fixed problem size. The mean and variance of  $\hat{X}_{0,n}$  roughly scales as  $\Theta(n^{-1})$  and  $\Theta(n^{-2})$  as Equation 5 predicts. The mean of  $\hat{X}_{1,n}$  is lower bounded by a constant. The variance scales roughly as  $\frac{2E[\hat{X}_{1,n}]}{n^{0.88}}$ .

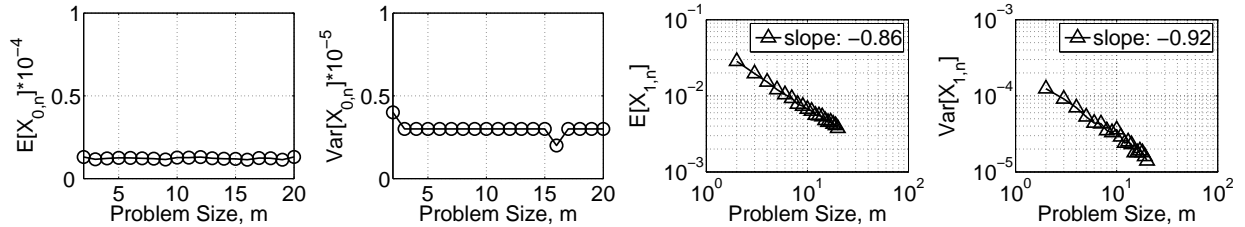


Figure 2: Means and variances versus the problem size for a fixed population size. The mean and variance of  $\hat{X}_{0,n}$  does not depend much on  $m$ . The mean and variance of  $\hat{X}_{1,n}$  roughly scale as  $\Theta(m^{-1})$ , as predicted in Equation 21.

scales as  $\Theta(n^{-1})$  and  $\Theta(n^{-2})$  respectively as Equation 5 predicts.  $E[\hat{X}_{1,n}]$  is lower bounded by some constant. The mean does vary much from  $n = 1000$  to  $n = 10000$ . The variance roughly scales as  $\frac{2E[\hat{X}_{1,n}]}{n^{0.88}}$ , which basically agrees with the model in Kleiter (1999).

Now we fix the population size  $n$  to 1000 and measure the mutual information by varying the problem size  $m$ . Figure 2 shows the results. As indicated in Equation 5,  $\hat{X}_{0,n}$  does not depend on the problem size  $m$ . The mean and variance of  $\hat{X}_{1,n}$  scales as  $\Theta(m^{-0.88})$  and  $\Theta(m^{-0.92})$  respectively. The results indicate that even though that Equation 21 only gives upper bounds, it also serves as a good approximation (Equation 21 predicts that both mean and variance of  $\hat{X}_{1,n}$  scale as  $\Theta(\frac{1}{m})$ ).

## 4 Selection Pressure and Population Size

Here we extend our analysis to an  $s_{to}$ -wise tournament selection. While we acknowledge that using results from order statistics might accurately capture the effect of selection pressure on population sizing, here we approximate by truncation selection to ease the analytical burden.

Consider that we apply the selection for several times. That should give us the similar effect of having an exponentially higher selection pressure. The statement is especially true for the truncation selection. In truncation selection, if we select the best half of the population twice, the effect is exactly the same as if we select the best quarter.

Since all the derivations of the previous selection are based on the tournament selection, now we need to find a transformation from the tournament selection to the truncation selection. Blickle and Thiele (1995) gave the approximation of the selection intensity  $I = \sqrt{2(\ln s_{to} - \ln \sqrt{4.14 \ln s_{to}})}$ , where  $s_{to}$  is the tournament size. On the other hand, Bäck (1995) approximated the selection intensity for the truncation selection with a selection pressure  $s_{tr}$  as  $I = s_{tr} \phi(\Phi^{-1}(1 - \frac{1}{s_{tr}}))$ , where  $\phi$  is the pdf of the standard normal distribution and  $\Phi$  is the cdf.

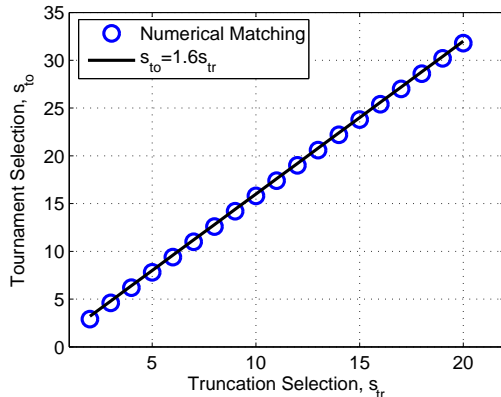


Figure 3: The relationship between the tournament size  $s_{to}$  and the selection pressure of truncation selection  $s_{tr}$  that yield the same selection intensity can be approximated as  $s_{to} \simeq 1.6s_{tr}$ .

By setting the selection intensity to be same, we can solve  $s_{to}$  and  $s_{tr}$  numerically. We get the relationship between  $s_{to}$  and  $s_{tr}$  that yield the same selection pressure as (Figure 3)

$$s_{to} \simeq 1.6s_{tr}. \quad (22)$$

Therefore, applying a binary tournament selection has a similar effect as applying truncation selection with a selection pressure of 1.25. If we apply that truncation selection  $t$  times, the selection pressure becomes  $s_{tr} = 1.25^t$ . When  $t$  is not too large,  $p'_1 = \frac{1}{2} + \frac{t\Delta_m}{2}$  (This can be shown by solving the recursive relation in Equations 13 and 14). The changes in  $\vec{q}$  (Equation 17) is also linearly proportional to  $t$ . As a result, the growth of the entropy measurement is proportional to the square of  $t$ . On the other hand, the number of independent samples reduces from  $n$  to  $\frac{n}{s_{tr}}$  because of the selection procedure. Therefore, we can express the means of the mutual information measurement as

$$E[\hat{X}_{0, \frac{n}{s_{tr}}}] = \frac{s_{tr}}{2n}. \quad E[\hat{X}_{1, \frac{n}{s_{tr}}}] \geq \frac{t^2 \Delta_m^2}{2 \ln 2}. \quad (23)$$

For accurate model building, it is necessary that  $E[\hat{X}_{1, \frac{n}{s_{tr}}}] > E[\hat{X}_{0, \frac{n}{s_{tr}}}]$ . We have shown that Equation 23 is an asymptotic bound (*i.e.*  $E[\hat{X}_{1, n=\infty}] = \frac{t^2 \Delta_m^2}{2 \ln 2}$ ). Therefore, for accurate model building, a necessary condition is that

$$\frac{t^2 \Delta_m^2}{2 \ln 2} > \frac{s_{tr}}{2n} \quad (24)$$

Solving Equation 24 for  $n$  gives us the first bound:

$$n > c_1 \frac{s_{tr}}{\ln^2 s_{tr}} 2^{2k} m \frac{\sigma_{BB}^2}{d^2}, \quad (25)$$

with some constant  $c_1$ . We can also convert Equation 25 back to tournament selection:

$$n > c_2 \frac{s_{to}}{(\ln^2(s_{to}/1.6))} 2^{2k} m \frac{\sigma_{BB}^2}{d^2}. \quad (26)$$

Figure 4 shows the relationship between the tournament size and the population size needed to build an accurate model with  $(m - 1)$  BBs correctly identified. The results agree with the



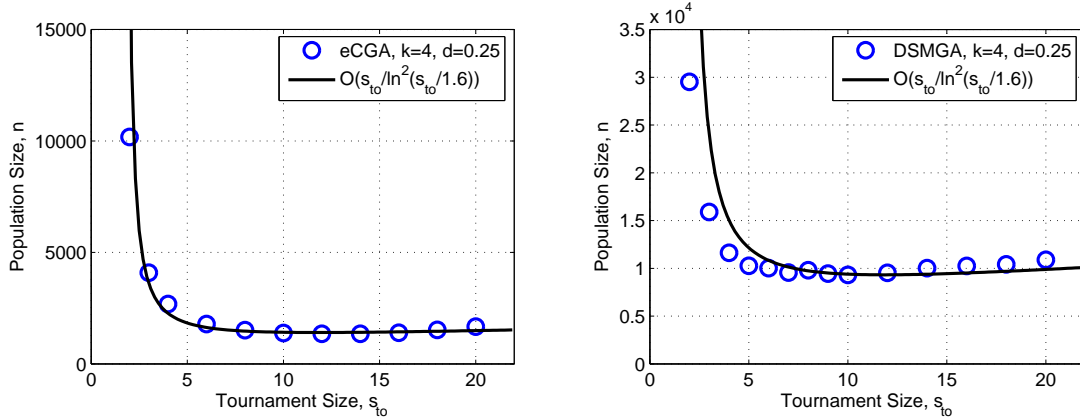


Figure 4: The relationship between the tournament size and the population size. Both the results and model indicates the existence of an optimal  $s_{to}$  around 10.

model qualitatively. Basically, for small and large  $s_{to}$ , a larger population size is needed to build an accurate model. Equation 26 also predicts a fixed optimal  $s_{to}^* \simeq 11.8$ . However, empirically, we observe that the optimal tournament size varies with the problem and model-building procedure. This phenomenon currently is not captured in our model. For example, we can observe some displacement in the  $s_{to}$  dimension of the DSMGA results. The problem might be the approximation of the tournament selection by the truncation selection. Even though the approximation ensures a similar selection intensity, however, in the tournament selection, the number of copies of an individual is proportional to its rank, which is not the case for the truncation selection.

## 5 Population Sizing for Linkage Learning and Model Building

Now that we have means and variances for both  $\hat{X}_{0,n}$  and  $\hat{X}_{1,n}$  (Equations 5 and 21). We are able to derive another bound for population size by the decision-making approach.

According to Hutter and Zaffalon (2004), the distribution of mutual information can be approximated as normal distribution. The decision-making error can be calculated as follows. Let  $Z = \hat{X}_{1,n} - \hat{X}_{0,n}$ . We can then calculate

$$\tau \triangleq \frac{E[Z]}{\sqrt{Var[Z]}} \simeq \frac{\Delta_m \sqrt{n}}{2\sqrt{\ln 2}}. \quad (27)$$

For large  $\tau$ , the decision error  $\epsilon$  can be approximated as

$$\epsilon \simeq \frac{1}{\tau} e^{-\frac{\tau^2}{2}} = \frac{2\sqrt{\ln 2}}{\Delta_m \sqrt{n}} e^{-\frac{\Delta_m^2 n}{4 \ln 2}}. \quad (28)$$

For a problem with  $m$  BBs, there are  $\binom{k}{2} m$  pairs of dependent decision variables and  $\binom{km}{2} - \binom{k}{2} m$  pairs of independent decision variables. If we assume that genes within a BB are maximally dependent, we can then treat one whole BB as one decision variable and there are only  $\binom{m}{2}$  independent decisions to be made correctly. In any case, to build an accurate model, we need

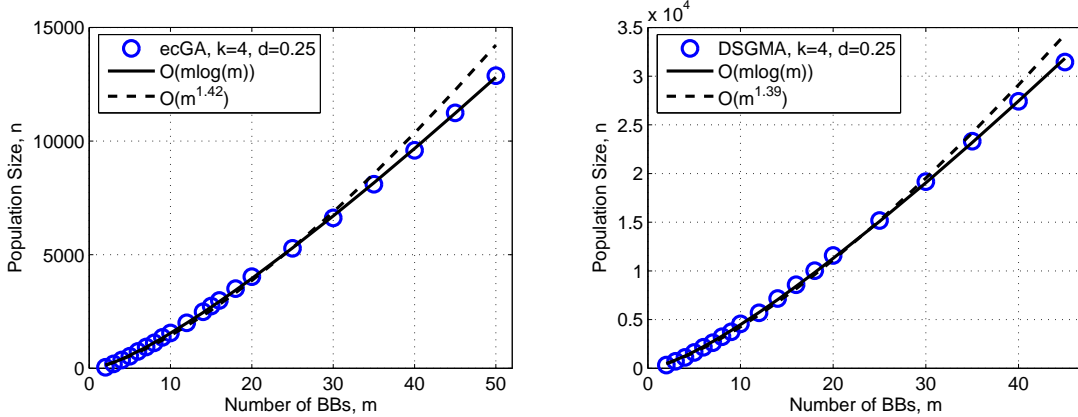


Figure 5: The scalability of the population size.  $n = \Theta(m \log m)$  is a better description of the results of both eCGA and DSMGA than the power-law model.

$\mathcal{O}(m^2)$  independent decisions to be made correctly. Given the model accuracy to be  $(1 - \frac{1}{m})$ , we have

$$(1 - \epsilon)^{\mathcal{O}(m^2)} \geq 1 - \frac{1}{m}. \quad (29)$$

For small  $\epsilon$  and large  $m$ , we have

$$\begin{aligned} \frac{1}{\epsilon} &\geq \mathcal{O}(m^3) \\ \Rightarrow \ln(\Delta_m \sqrt{n}) + \frac{\Delta_m^2 n}{4 \ln 2} &\geq \mathcal{O}(\ln m). \end{aligned} \quad (30)$$

For a large  $n$ , we can omit the first term in Equation 30. By substituting  $\Delta_m$  according to Equation 15, we obtain the following bound.

$$n \geq c_3 2^{2k} m \log(m) \frac{\sigma_{BB}^2}{d^2}. \quad (31)$$

Combining Equations 26 and 31, we can get the following final bound.

$$n \geq c_4 \frac{s_{to}}{\ln^2(s_{to}/1.6)} 2^{2k} m \log(m) \frac{\sigma_{BB}^2}{d^2}. \quad (32)$$

Figure 5 shows the experimental results for eCGA and DSMGA on an  $(m, k)$ -trap function, where  $k = 4$  and  $d = 0.25$ . The power-law curve fitting is done by first-order polynomial fitting on the log-log scale. We can see that  $\Theta(m \log m)$  provides a better description of the data than the power-law model.

## 6 Summary and Conclusions

This paper presents a population-sizing model for entropy-based model-building in GAs. Specifically, the population size required for building an accurate model is investigated. The proposed model refines the required population size for model building from  $\Theta(m^{1.05}) \leq n \leq \Theta(m^{2.1})$  to  $n = \Theta(m \log m)$ . The proposed model also incorporates the effect of selection pressure on the population sizing requirements. Empirical results quantitatively agree well with the proposed model for the scalability on the problem size. The modeling on the selection pressure is qualitatively

verified by the results. To obtain a more accurate modeling on the selection pressure, we may need to utilize results from order statistics on the normal distribution.

Compared with the population-sizing model for EDAs to solve the problem (Pelikan, Sastry, & Goldberg, 2003; Sastry & Goldberg, 2004), the proposed population-sizing model scales the same with the problem size (both scale as  $\Theta(m \log m)$ ). In other words, the population size required to build a near-perfect model is of the same order as that needed for GAs to solve the problem, but with a larger constant (because the GA utilizes a sequence of the population and the model builder utilizes only the current population). Also, it is worth noting that the decision-making model (Goldberg, Deb, & Clark, 1992b) has a similar form. The difference is that in the decision-making model, decisions are made between the two most competing BBs. Here, decisions are made between the dependent and independent variables. The proposed model indicates that for a too small selection pressure, the signal is not strong enough to detect correct linkages; for a too large selection pressure, the sampling noise becomes large enough to cloud the signal that we want to detect. An optimal selection pressure exists somewhere in the middle for the model builder. Finally, although the proposed model is based on the entropy measurement, we believe that a similar procedure should be applied to some other measurements including nonlinearity (Munetomo & Goldberg, 1999) and simultaneity (Aporntewan & Chongstitvatana, 2003).

## Acknowledgements

This work was sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant FA9550-06-1-0096, the National Science Foundation under NSF CAREER grant ECS-0547013 and ITR grant DMR-03-25939 at Materials Computation Center. The work was also supported by the Research Award and the Research Board at the University of Missouri. The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation thereon.

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research, the National Science Foundation, or the U.S. Government.

## References

- Abramowitz, M., & Stegun, L. (1970). *Handbook of mathematical functions*. New York: Dover.
- Aporntewan, C., & Chongstitvatana, P. (2003). Building-block identification by simultaneity matrix. *Proceedings of the Genetic and Evolutionary Computation Conference*, 1566–1567.
- Bäck, T. (1995). Generalized convergence models for tournament and  $(\mu;\lambda)$  selection. *Proceedings of the Sixth International Conference on Genetic Algorithms (ICGA'95)*, 2–8.
- Blickle, T., & Thiele, L. (1995). A mathematical analysis of tournament selection. *Proceedings of the Sixth International Conference on Genetic Algorithms (ICGA'95)*, 9–16.
- Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory* (pp. 18–26). New York: Wiley.
- Etzeberria, R., & Larrañaga, P. (1999). Global optimization using bayesian networks. *Proceedings of the Second Symposium on Artificial Intelligence Adaptive Systems*, 332–339.
- Feller, W. (1966). *An introduction to probability theory*, Volume 2. New York: Wiley.
- Goldberg, D. E. (1987). Simple genetic algorithms and the minimal, deceptive problem. In *Genetic Algorithms and Simulated Annealing* (Chapter 6, pp. 74–88). London: Pitman Publishing.

- Goldberg, D. E. (1989). Sizing populations for serial and parallel genetic algorithms. *Proceedings of the Third International Conference on Genetic Algorithms*, 70–79.
- Goldberg, D. E., Deb, K., & Clark, J. H. (1992a). Genetic algorithms, noise, and the sizing of populations. *Complex Systems*, 6, 333–362.
- Goldberg, D. E., Deb, K., & Clark, J. H. (1992b). Genetic algorithms, noise, and the sizing of populations. *Complex Systems*, 6, 333–362.
- Goldberg, D. E., Sastry, K., & Latoza, T. (2001). On the supply of building blocks. *Proceedings of the Genetic and Evolutionary Computation Conference*, 336–342.
- Harik, G. (1999, February). *Linkage Learning via Probabilistic Modeling in the ECGA* (IlliGAL Report No. 99010). Urbana, IL: University of Illinois at Urbana-Champaign.
- Harik, G., Cantú-Paz, E., Goldberg, D. E., & Miller, B. L. (1997). The gambler’s ruin problem, genetic algorithms, and the sizing of populations. *Proceedings of 1997 IEEE International Conference on Evolutionary Computation*, 7–12.
- Harik, G., Cantú-Paz, E., Goldberg, D. E., & Miller, B. L. (1999). The gambler’s ruin problem, genetic algorithms, and the sizing of populations. *Evolutionary Computation*, 7(3), 231–253.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.
- Hutter, M., & Zaffalon, M. (2004). Distribution of mutual information from complete and incomplete data. *Computational Statistics and Data Analysis*, 48, 633–657.
- Kleiter, G. D. (1999). The posterior probability of bayes nets with strong dependences. *Soft Computing*, 3, 162–173.
- Larrañaga, P., & Lozano, J. (Eds.) (2002). *Estimation of distribution algorithms*. Boston, MA: Kluwer Academic Publishers.
- Mühlenbien (1999). Fda: A scaleable evolutionary algorithm for the optimisation of additively decomposable problems. *Evolutionary Computation*, 7, 353–376.
- Munetomo, M., & Goldberg, D. E. (1999). Identifying linkage groups by nonlinearity/non-monotonicity detection. pp. 433–440.
- Pelikan, M., Goldberg, D. E., & Cantú-Paz, E. (1999). BOA: The Bayesian optimization algorithm. *Proceedings of the Genetic and Evolutionary Computation Conference*, 1, 525–532.
- Pelikan, M., Goldberg, D. E., & Cantú-Paz, E. (2000). Bayesian optimization algorithm, population sizing, and time to convergence. *Proceedings of the Genetic and Evolutionary Computation Conference*, 275–282.
- Pelikan, M., Sastry, K., & Goldberg, D. E. (2003). Scalability of the Bayesian optimization algorithm. *International Journal of Approximate Reasoning*, 31(3), 221–258.
- Reeves, C. (1993). Using genetic algorithms with small populations. *Proceedings of the Fifth International Conference on Genetic Algorithms*, 92–99.
- Sastry, K., & Goldberg, D. E. (2000). On extended compact genetic algorithm. *Late-Breaking Paper at the Genetic and Evolutionary Computation Conference*, 352–359.
- Sastry, K., & Goldberg, D. E. (2004). Designing competent mutation operators via probabilistic model building of neighborhoods. *Proceedings of the 2004 Genetic and Evolutionary Computation Conference*, 2, 114–125.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–423.
- Wright, A., Poli, R., Stephens, C., Landgon, W. B., & Pulavarty, S. (2004). An estimation of distribution algorithm based on maximum entropy. *Proceedings of the Genetic and Evolutionary Computation Conference*, 343–354.
- Yu, T.-L., & Goldberg, D. E. (2006). Conquering hierarchical difficulty by explicit chunking: Substructural chromosome compression. *Proceedings of the Genetic and Evolutionary Computation Conference*. (to appear).