Optimal Sampling and Speed-up for Genetic Algorithms on the Sampled OneMax Problem

Tian-Li Yu
David E. Goldberg
Kumara Sastry

IlliGAL Report No. 2003008
February, 2003

Illinois Genetic Algorithms Laboratory (IlliGAL)
Department of General Engineering
University of Illinois at Urbana-Champaign
117 Transportation Building
104 S. Mathews Avenue, Urbana, IL 61801
http://www-illigal.ge.uiuc.edu

Optimal Sampling and Speed-up for Genetic Algorithms on the Sampled OneMax Problem

Tian-Li Yu, David E. Goldberg, and Kumara Sastry
Illinois Genetic Algorithms Laboratory (IlliGAL)
Department of General Engineering
University of Illinois at Urbana-Champaign
104 S. Mathews Ave, Urbana, IL 61801
{tianliyu,deg,kumara}@illigal.ge.uiuc.edu

February 1, 2003

Abstract

This paper investigates the optimal sampling and the speed-up obtained through sampling for the sampled OneMax problem. Theoretical and experimental analyses are given for three different population-sizing models: the decision-making model, the gambler's ruin model, and the fixed population-sizing model. The results suggest that, when the desired solution quality is fixed to a high value, the decision-making model prefers a large sampling size, the fixed population-sizing model prefers a small sampling size, and the sampling makes no difference for the gambler's ruin model. The speed-up obtained by sampling is then empirically verified.

1 Introduction

Over the last few decades, significant progress has been made in the theory, design and application of genetic and evolutionary algorithms. A decomposition design theory has been proposed and several *competent* genetic algorithms (GAs) have been developed (Goldberg, 1999). By competent GAs, we mean GAs that can solve hard problems quickly, reliably, and accurately. Competent GAs not only render problems that were intractable by first generation GAs tractable, but do so by requiring only a polynomial (usually subquadratic) number of function evaluations.

However, in real-world problems, even the subquadratic number of function evaluations can be very high, especially if the function evaluation is a complex model, simulation, or computation. Therefore, GA practitioners have used a variety of efficiency enhancement techniques to alleviate the computational burden. One such technique is evaluation relaxation (Sastry & Goldberg, 2002), in which the accurate, but costly fitness evaluation is replaced by a cheap, but less accurate evaluation. Partial evaluation through sampling is an example of evaluation relaxation scheme, which has been emprically shown to yield significant speed-up (Grefenstette & Fitzpatrick, 1985). Evaluation relaxation through sampling has also been analyzed by developing facetwise and dimensional models (Miller & Goldberg, 1996a; Giguère & Goldberg, 1998). This study extends the theoretical analyses and investigates the utility of sampling as an evaluation-relaxation technique.

The objective of this paper is to extend the work of Giguère and Goldberg (1998) and incorporate the effect of sampling on both convergence time and population sizing of GAs. Specifically, we concentrate on problems with substructures of uniform salience. This paper is composed of four primary parts: (1) background knowledge including previous work and an introduction to the sampled OneMax problem, (2) the derivation of the optimal sampling sizes for the decision-making model, the gambler's ruin model, and the fixed population-sizing model, (3) empirical results, and (4) extensions and conclusions of this work.

2 Past Work

Grefenstette and Fitzpatrick (1985) achieved great success in applying sampling techniques to the image registration problem. Their success motivated Miller and Goldberg (1996b), which gave a theoretical analysis of a related problem. They derived and empirically verified the theoretical optimal sampling size when the fitness function is clouded by an additive Gaussian noise. The sampling methods in the above two papers have a subtle but significant difference. In Grefenstette and Fitzpatrick (1985), the variance becomes zero if full sampling is used. However, in the additive Gaussian noisy OneMax problem in Miller and Goldberg (1996b), the variance of fitness function inversely corresponds to the sampling size. It can be very small when the sampling size is large, but will never be zero. More recently, Giguère and Goldberg (1998) investigated a sampled OneMax (SOM) problem where the fitness value is calculated by sampling. Their results showed that sampling is not really useful when the Gambler's ruin population-sizing model (Harik, Cantú-Paz, Goldberg, & Miller, 1997; Miller, 1997) is adopted. Even though Giguère and Goldberg (1998) have investigated all population-sizing models discussed in this paper, they did not consider convergence time. In addition, detailed analytical models are needed for a better understanding of the sampling schemes as a technique of evaluation relaxation.

3 Sampled OneMax Problem (SOM)

Before starting the derivation of the number of function evaluation model, let us first define SOM. The SOM is basically the OneMax or the counting ones problem, except the fitness value is computed by sampling without replacement.

$$F_n(\bar{x}) = \frac{l}{n}(\Sigma_{i\in S}\ x_i), \tag{1}$$

where n is the sampling size ($0 < n \leq l$), S is a subset of $\{1, 2, \cdots, l\}$ with a restriction that $|S| = n$, and l is the chromosome length. The term $\frac{l}{n}$ is just for normalization so that the expectation of the fully sampled fitness $F(\bar{x}) = \Sigma_{i=1}^{l} x_i$ is the same as the expectation of F_n for all n. The variance of the noise introduced by sampling can be expressed as (Giguère & Goldberg, 1998):

$$\sigma_n^2 = \frac{l^2 p(t)(1-p(t))}{n}\frac{l-n}{l-1}, \tag{2}$$

where $p(t)$ is the performance model defined in Thierens and Goldberg (1994). Initially, $p(0) = 0.5$.

The number of function evaluations is defined as how many bits the GA has visited before convergence. We choose the SOM because (1) it is linear and easy to analyze, and (2) it bounds the speed-up. SOM is considered as a GA-easy problem, the speed-up obtained for this problem should be higher than many other problems.

4 Optimal Sampling for the Sampled OneMax Problem (SOM)

This section gives theoretical and empirical analyses of the optimal sampling sizes for three different population sizing models: the decision-making model, the gambler's ruin model, and the fixed population-sizing model on the SOM.

This section starts with deriving the number of function evaluation model for SOM. Then with the number of function evaluation model derived, the optimal sampling sizes are derived for the three population sizing models. Finally, experiments results are shown to verify the optimal sampling sizes derived.

In this section, we fix the solution quality to $(l-1)$, where l is the chromosome length, and try the minimize the number of function evaluations through sampling. The desired solution quality is set so high that the convergence time model is valid to use.

4.1 Model for Number of Function Evaluation for SOM

Now let us derive the model for number of function evaluations for SOM. Sastry and Goldberg (2002) indicated the time to convergence for those problems with uniformly-scaled BBs corresponds to the squared root of variance of the fitness function.

$$t_{conv} = \frac{\ln(l)}{I}\sqrt{\sigma_f^2 + \sigma_n^2}, \tag{3}$$

where I is selection intensity (Mühlenbein & Schlierkamp-Voosen, 1993). For binary tournament selection, $I = 1/\sqrt{\pi} \sim 0.5642$ (Bäck, 1994). By equation 2 and approximate σ_f^2 by the initial variance $lp(0)(1-p(0))$, t_{conv} can be rewritten as

$$t_{conv} = C_1\sqrt{\frac{l^2}{n} - 1}, \tag{4}$$

where $C_1 = \frac{\ln(l)}{I} \cdot \sqrt{\frac{l}{l-1}p(0)(1-p(0))}$.

The number of function evaluations is given by $n_{fe} = nGN$, where n is the sampling size, G is the number of generations, and N is the population size. By substituting G with t_{conv} obtained in equation 4, the number of function evaluations can be expressed as

$$n_{fe} = C_1 \cdot nN\sqrt{\frac{l^2}{n} - 1} \tag{5}$$

Now that we have the model for the number of function evaluations for SOM, we are ready to derive the optimal sampling for different population sizing models.

4.2 Optimal Sampling for the Decision-Making Model

The decision-making model (Goldberg, Deb, & Clark, 1992) is given by

$$N = \Gamma(\sigma_f^2 + \sigma_n^2), \tag{6}$$

where N is the population size, Γ is the population coefficient as defined in Miller and Goldberg (1996b), σ_f^2 is the initial population fitness variance, and σ_n^2 is the variance of fitness noise.

By equation 2 and substituting N into equation 5, n_{fe} for the decision-making model can be written as

$$n_{fe} = C_2 \cdot n\left(\frac{l^2}{n} - 1\right)^{\frac{3}{2}}, \tag{7}$$

where $C_2 = \frac{\ln(l)}{I}\Gamma\left(\frac{l}{l-1}p(0)(1-p(0))\right)^{\frac{3}{2}}$.

Equation (7) has a minimum at $n = l$, which means no sampling at all (figure 1). This is not surprising because the population sizing model is known as an overestimation than needed for the desired solution quality. When sampling is adopted, the noise introduced makes the population size even larger than needed according to the decision-making model. As a result, the larger population size results a larger number of function evaluations.

4.3 Optimal Sampling for the Gambler's Ruin Model

The gambler's ruin model (Harik, Cantú-Paz, Goldberg, & Miller, 1997; Miller, 1997) is given by

$$N = \Gamma'(\sqrt{\sigma_f^2 + \sigma_n^2}), \tag{8}$$

where N, σ_f^2, and σ_n^2 are defined the same as those in equation (6), and Γ' is another coefficient defined as

$$\Gamma' = -\frac{2^{k_{max}-1}\sqrt{\pi}\ln(\psi)}{d_{min}}. \tag{9}$$

k_{max} is an estimate of the maximal length of building blocks. ψ is the failure rate, defined as the probability that a particular partition in the chromosome that fails to converge to the correct Building Blocks (BBs). In other words, $(1-\psi)$ is the expected proportion of the correct BBs in an individual. d_{min} is an estimate of the minimal signal difference between the best and the second best building block. In other words, d_{min} is the smallest building block signal that GAs can detect. In OneMax domain, both k_{max} and d_{min} are 1, which yields a simpler form for Γ':

$$\Gamma' = -\sqrt{\pi}\ln(\psi) \tag{10}$$

By a similar algebraic process as in the previous subsection, the number of function evaluations is expressed as

$$n_{fe} = C_3 \cdot (l^2 - n), \tag{11}$$

where $C_3 = \frac{\ln(l)}{I}\Gamma'\left(\frac{l}{l-1}p(0)(1-p(0))\right)$.

The minimum of equation 11 occurs at $n = l$, which means, again, no sampling at all (figure 2). The gambler's ruin model still prefers a larger sampling size, but only with a slight preference. This can be shown by comparing the case of $n = 1$ and $n = l$.

$$\frac{n_{fe}(n=1)}{n_{fe}(n=l)} = \frac{l^2-1}{l^2-l} = \frac{l+1}{l} \tag{12}$$

For a 100-bit SOM problem, the difference of n_{fe} is only 1%, which is so small that can be neglected compared with those approximations in derivations. As a conclusion, for the gambler's ruin model, sampling does not make much difference. The conclusion agrees with Giguère and Goldberg (1998).

4.4 Fixed Population-Sizing Model

The fixed population-sizing model is not an accurate model of GAs since it does not account for the effects of problem size and noise. However, it is widely adopted in real-world applications (for example, Grefenstette and Fitzpatrick (1985)). Therefore, it is worthwhile to investigate this model as well. In this section, the assumption is that the fixed population size is larger than needed so that the GA can converge to a $(l-1)$solution quality. This assumption is needed for applying the convergence time model. Since the population size is larger than needed, sampling should obtain speed-up. In addition, we still fix the desired solution quality to be $(l-1)$ because we stop the GA when a $(l-1)$ solution quality is reached.

The number of function evaluations for a fixed population size is given by

$$n_{fe} = C_4 \cdot n \left(\frac{l^2}{n} - 1 \right)^{\frac{1}{2}}, \tag{13}$$

where $C_4 = \frac{\ln(l)}{I} N \left(\frac{l}{l-1} p(0)(1-p(0)) \right)^{\frac{1}{2}}$.

Equation (13) has a minimum at $n = 1$. If the overhead (α) is taken into account, and the cost of sampling one bit is β, the total run duration is expressed as

$$T = C_4 \cdot (\alpha + n\beta) \left(\frac{l^2}{n} - 1 \right)^{\frac{1}{2}} \tag{14}$$

For large l, $\frac{l^2}{n} \gg 1$, equation 14 can be approximated as

$$T = C_4 \cdot l(\alpha n^{-1/2} + \beta n^{1/2}) \tag{15}$$

By differentiating equation 15 by n, and then setting it to be zero, the minimum is found at

$$n_{op} = \frac{\alpha}{\beta} \tag{16}$$

It is interesting to compare equation 16 with what Miller and Goldberg (1996b) got ($\sqrt{\frac{\alpha}{\beta}}\sqrt{\frac{\sigma_n^2}{\sigma_f^2}}$). The term σ_n in Miller and Goldberg's result is vanished because now σ_n is controlled by the sampling size n. If the constant term is ignored, the result here is the square of Miller and Goldberg's optimal sampling size.

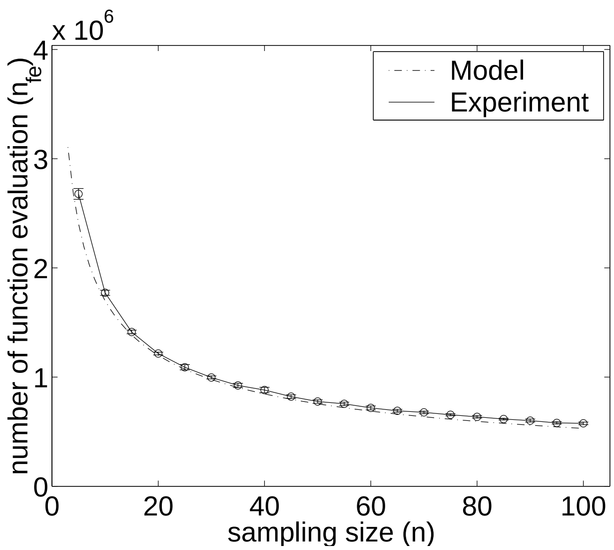

Figure 1: The relationship between n_{fe} and n for the decision-making model.

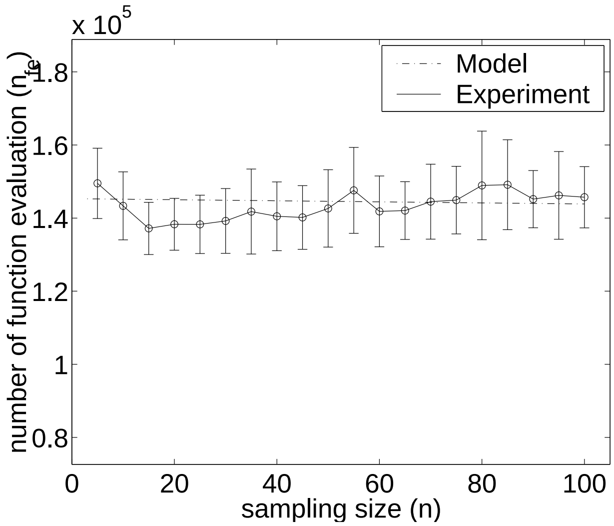

Figure 2: The relationship between n_{fe} and n for the gambler's ruin model.

4.5 Experimental Results

The experimental results are obtained for the SOM problem with chromosome length $l = 100$. Binary tournament selection and uniform crossover are adopted. For all experiments, the crossover probability (p_c) is one and the mutation probability (p_m) is zero, which means, crossover is always taken place and there is no mutation. All experimental results are averaged over 50 independent runs. In Sastry and Goldberg (2002), they used the variance of fitness of the initial population to estimate σ_f^2. This is, of course, an overestimation. In fact, the variance of fitness becomes smaller and smaller during the runs of a GA. For OneMax problem, it almost becomes zero when the average fitness is converged to be $l - 1$ ($p(t) \simeq 1$). Therefore, for a tighter estimation, $p(t) = 0.75$ (half convergence) is used in the calculation of σ_f^2 and σ_n^2.

Figure 1, 2, and 3 shows the relationship of n_{fe} versus n for the decision-making model, the gambler's ruin model, and the fixed population-sizing model, respectively. In figure 1, the population size is obtained from $N = 8(\sigma_f^2 + \sigma_n^2)$ (Goldberg, Deb, & Clark, 1992). Figure 2 uses $N = 9.4\sqrt{\sigma_f^2 + \sigma_n^2}$ (Miller & Goldberg, 1996b). The experimental results show slight minima in-between $N = 10$ and $N = 30$. It agrees with the observation in Giguere and Goldberg (1998), which unfortunately has not be explained by mathematical models so far. The fixed population size is set to be 500 in figure 3. The fixed population size is set so large to prevent the failure of convergence. Finally, the total run duration for $\alpha/\beta = 20$ is shown in figure 4.

The results agree with the models derived in this section. The model of number of function evaluations for the decision-making model especially matches experimental results. The model for fixed population size overestimates a little bit. Nevertheless, the overestimation is of some constant, and figure 4 empirically verifies the argument. Therefore, the overestimation does not impair the accuracy of the optimal sampling size derived in equation 16.

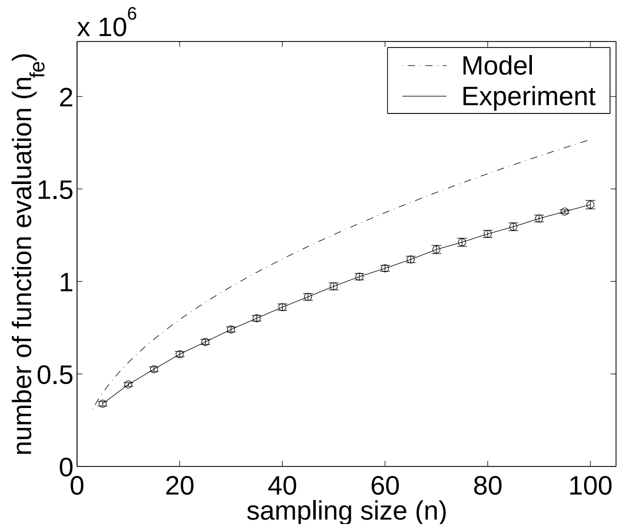

Figure 3: The relationship between n_{fe} and n for fixed population $N = 500$. $\frac{\alpha}{\beta} = 0$.

5 Future work

In this paper, we analyzed the utility of sampling as an evaluation-relaxation technique on problems with substructures of uniform-salience. There are many avenues for future work which are both interesting as well as practically significant. Some of them are listed below.

- For a better understanding of applying sampling techniques, we hope to investigate the limit of sampling schemes: where and when it will work (or fail).
- Further investigation needs to be done to bound the effectiveness of sampling schemes by considering an array of adversarial problems that contain one or more facets of problem difficulty such as deception, noise, and scaling.
- In this paper, the assumption that the GA converges to a high solution quality is needed for applying the convergence time model. We hope to be able to understand the time-critical condition where the number of function evaluations is very limited.

We believe that the analysis and methodology presented this paper should carry over or can be extended in a straightforward manner on most of the above issues.

6 Conclusions

This paper studied the optimal sampling and the speed-up obtained through sampling for the sampled OneMax problem (SOM). Based on Sastry and Goldberg (2002), facetwise models for solution quality as a function of sampling size were derived for three population-sizing models, namely, the decision-making model (Goldberg, Deb, & Clark, 1992), the gambler's ruin model (Harik, Cantú-Paz, Goldberg, & Miller, 1997; Miller, 1997), and the fixed population-sizing model. The optimal sampling and the potential speed-up obtained by sampling are analyzed under the scenario: we fix the solution quality to a very high value and our goal is to obtain it with minimum number of function evaluations. Each of the models were verified with empirical results.

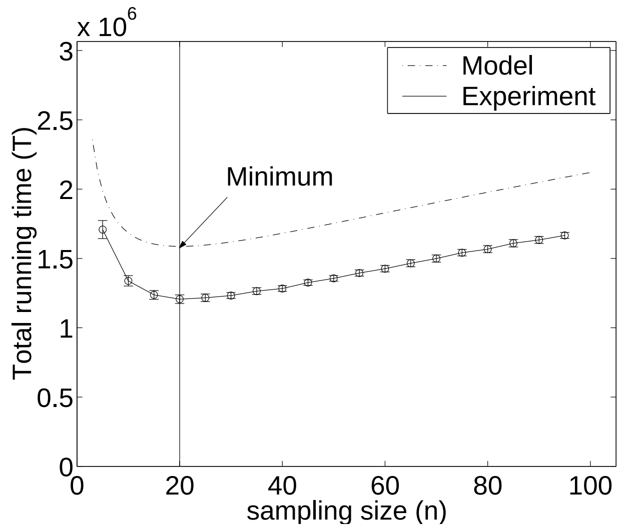

Figure 4: The relationship between T and n for fixed population $N = 500$. $\frac{\alpha}{\beta} = 20$.

When the desired solution quality is fixed, the results suggest that the decision-making model prefers a larger sampling size, and that the fixed population sizing model prefers a smaller sampling size. Sampling does not make much difference for the gambler's ruin model.

The results show that sampling does not give speedup for problems with subsolutions of uniform salience, if the population is sized appropriately to handle both the stochastic decision making and the noise introduced by sampling. On the other hand, if the population size is fixed without accounting for either decision-making or sampling, then the results presented in this paper show that sampling does indeed yield speed-up and an optimal sampling size exists that yields maximum speed-up.

Acknowledgement

This work was sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant F49620-00-0163, the National Science Foundation under grant DMI-9908252. 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

Bäck, T. (1994). Selective pressure in evolutionary algorithms: A characterization of selection mechanisms. *Proceedings of the First IEEE Conference on Evolutionary Computation*, 57–62.

Giguère, P., & Goldberg, D. E. (1998). Population sizing for optimum sampling with genetic algorithms: A case study of the Onemax problem. In *Genetic Programming 98* (pp. 496–503).

Goldberg, D. E. (1999). The race, the hurdle, and the sweet spot: Lessons from genetic algorithms for the automation of design innovation and creativity. *Evolutionary Design by Computers*, 105–118.

Goldberg, D. E., Deb, K., & Clark, J. H. (1992). Genetic algorithms, noise, and the sizing of populations. *Complex Systems*, *6*, 333–362.

Grefenstette, J. J., & Fitzpatrick, J. M. (1985). Genetic search with approximate function evaluations. In *Proceedings of an International Conference on Genetic Algorithms and Their Applications* (pp. 112–120).

Harik, G., Cantú-Paz, E., Goldberg, D. E., & Miller, B. L. (1997). The gambler's ruin problem, genetic algorithms, and the sizing of populations. In *Proceedings of 1997 IEEE International Conference on Evolutionary Computation* (pp. 7–12). Piscataway, NJ: IEEE.

Miller, B. L. (1997). *Noise, sampling, and efficient genetic algorithms.* doctoral dissertation, University of Illinois at Urbana-Champaign, Urbana. Also IlliGAL Report No. 97001.

Miller, B. L., & Goldberg, D. E. (1996a). Genetic algorithms, selection schemes, and the varying effects of noise. *Evolutionary Computation*, *4*(2), 113–131.

Miller, B. L., & Goldberg, D. E. (1996b). Optimal sampling for genetic algorithms. *Proceedings of the Artificial Neural Networks in Engineering (ANNIE '96) conference*, *6*, 291–297.

Mühlenbein, H., & Schlierkamp-Voosen, D. (1993). Predictive models for the breeder genetic algorithm: I. Continuous parameter optimization. *Evolutionary Computation*, *1*(1), 25–49.

Sastry, K., & Goldberg, D. E. (2002). Genetic algorithms, efficiency enhancement, and deciding well with differing fitness variances. *Proceedings of the Genetic and Evolutionary Computation Conference*, 528–535. (Also IlliGAL Report No. 2002002).

Thierens, D., & Goldberg, D. E. (1994). Convergence models of genetic algorithm selection schemes. In *Parallel Problem Solving fron Nature, PPSN III* (pp. 119–129).