

Evolutionary Algorithm using Marginal Histogram Models in Continuous Domain

Shigeyoshi Tsutsui*, Martin Pelikan**, and David E. Goldberg**

* Department of Management and Information Science, Hannan University, 5-4-33 Amamihigashi, Matsubara, Osaka 580-5802 Japan, phone: +81-723-32-1224, e-mail: tsutsui@hannan-u.ac.jp

**Illinois Genetic Algorithms Laboratory, University of Illinois at Urbana-Champaign, 117 Transportation Building, 104 S. Mathews, Urbana, IL 61801, U.S.A, phone: +1-217-333-0897, e-mail: {pelikan, deg}@illigal.ge.uiuc.edu

Abstract: In this paper, we propose an evolutionary algorithm using marginal histograms to model the parent population in a continuous domain. We propose two types of marginal histogram models: the fixed-width histogram (FWH) and the fixed-height histogram (FHH). The results showed that both models worked fairly well on test functions with no or weak interactions among variables. Especially, FHH could find the global optimum with very high accuracy effectively and showed good scale-up with the problem size.

1. Introduction

Recently, there has been a growing interest in developing evolutionary algorithms based on probabilistic models. In this scheme, the offspring population is generated according to the estimated probabilistic model of the parent population instead of using traditional recombination and mutation operators. These algorithms are called the probabilistic model-building genetic algorithms (PMBGAs) [Pelikan 00]. In PMBGAs, better individuals are selected from an initial randomly generated population similar to standard GAs. Then, the probability distribution of the selected set of individuals is estimated and new individuals are generated according to this estimate to create candidate solutions for the next generation. The process is repeated until the termination conditions are satisfied. It is beyond the scope of this paper to give an overview of recent work on PMBGAs. For a survey, please see [Pelikan 00].

Several attempts to apply PMBGAs in continuous domain have been made [Sebag 98], [Servet 97], [Gallagher 99]. All these do not cover any interactions among the variables. In the estimation of Gaussian networks algorithm (EGNA) [Larranaga,99], a Gaussian network is learned to estimate a multivariate Gaussian distribution of the parent population. In [Bosman 99], two normal density estimation models and the histogram distribution are discussed. These models are intended to cover multivariate interaction among variables. In [Bosman 00a], it is reported that the normal distribution models have shown good performance.

In this paper, we propose an evolutionary algorithm using marginal histograms to model promising solutions in a continuous domain. We propose two types of marginal histogram models: the fixed-width histogram (FWH) and the fixed-height histogram (FHH). The results showed that both models worked fairly well on test functions which have no

or weak interactions among variables. FHH could find the global optimum with high accuracy effectively and showed good scale-up behavior.

2. Evolutionary Algorithms using Marginal Histogram Models

This section starts by describing the general algorithm. It continues by discussing how marginal histograms can be used to (1) model promising solutions and (2) generate new solutions by simulating the learned model. We consider two types of histograms.

2.1 General description of the algorithm

The algorithm starts by generating an individual population of candidate solutions at random. Promising solutions are then selected using any selection scheme. A marginal histogram model for the selected solutions is constructed and new solutions are generated according to the built model. New solutions replace some of the old ones and the process is repeated until the termination criteria are met. Therefore, the algorithm differs from traditional GAs only by the method used to process promising solutions in order to create the new ones.

2.2 Marginal fixed-width histogram (FWH)

In the FWH, we divide the search space $[\min_j, \max_j]$ of each variable x_j ($j=1, \dots, n$) into H ($h_j=0, 1, \dots, H-1$) bins. Then the estimated density of each variable is calculated according to the current vectors in the population.

Since we consider a marginal histogram, the total size of all the bins should be $H \times n$. Figure 1 shows an example of marginal FWH on variable x_j . The bin width ϵ can be determined by choosing appropriate number of bins depending on required precision for a given problem.

For each new individual, each value is generated as follows: first a bin is selected according to the probability of each bin. Then the value of variable is generated by generating a number from the selected bin with uniform distribution. This is repeated until all individuals are obtained. The simplest method to sample each bin is to use a roulette wheel (RW). However, RW has a certain amount of stochastic sampling error.

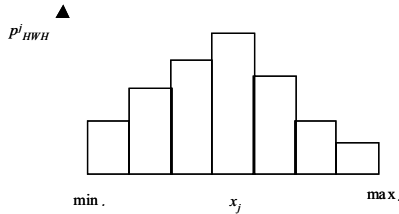


Figure 1: Marginal fixed width histogram (FWH)

One well known sampling method to reduce this sampling error is Baker’s stochastic universal sampling (SUS), which was proposed for the proportional selection operator [Baker 87]. We extend this method to sample from marginal histogram models. We call this method *E-SUS* (extended SUS) (see Figure 2). In the experiments in Section 3, we use both the RW and the E-SUS.

```

1 // S: sampling individual number, n: number of variables, H: number of bins
2 // p[j][h]: probability density of bin h of variable x_j
3 // f[j][h]: left edge position of bin h of variable x_j
4 // Rand(): generate random number from [0, 1.0]
5 V[S][n]; // array for sampled vectors
6 xh[S]={0,1,2,...,S-1}; // array for random permutation of bin position
7 for(int j=0; j<n; j++){
8     Shuffle(xh); // get a permutation xh[] by shuffling for each parameter
9     double ptr = Rand();
10    double sum = 0.0;
11    int k = 0;
12    for(int h=0; h<H; h++){
13        double expected = p[j][h]*S;
14        for(sum += expected; sum>ptr; ptr++){
15            V[xh[k++]][j] = f[j][h+(H[j][h+1]-H[j][h])]*Rand();
16        }
17    }

```

Figure 2: Pseudo C code of the E-SUS for FWH

2.3 Marginal fixed-height histogram (FHH)

In the FWH each bin has the same width. On the other hand, in the fixed-height histogram (FHH) each bin has the same height. That means that each bin contains the same number of points. The important feature of the FHH is that the bins in dense regions are narrower and thus the accuracy of the model in important regions increases. In context of evolutionary algorithms, the width of bins around high peaks decreases as more sample points are located in these areas. Since the probability of generating a point from each bin in the FHH is the same, we expect both the density as well as the accuracy improve in promising region of the search space. This is the distinguished characteristics of the FHH. Figure 3 (a) is an example of the FHH for a population which

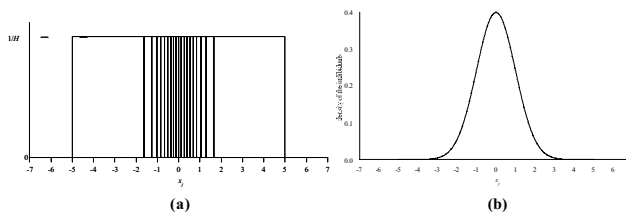


Figure 3: An example of the FHH of normal density

has a normal distribution $N(0, 1)$ for variable x_j in the range $[-5, 5]$ as shown in Figure 3 (b). Twenty (20) bins are used. In this example, more points are sampled around $x_j = 0.0$ because the bin width around $x_j = 0.0$ is narrower and therefore there are more bins for the same area.

Sampling method for the FHH is basically the same as that of FWH except the probability density for FHH has the same value $1/H$ for all bins. The pseudo C code of the E-SUS for FHH is the same with that for FWH except line 13 in Figure 2. The line 13 for the FHH should be:

```
“double expected = (double)S/H;”.
```

3. Empirical Study

To evaluate the marginal histogram models proposed in Section 2, we ran two histogram models using four test functions. The experimental methodology, test functions used, and experimental analysis are described in this section.

3.1 Experimental methodology

The basic evolutionary model we used in these experiments is similar to that of $(\mu+\lambda)$ -ES [Schwefel 95]. Let the population size be N , and let it, at time t , be represented by $P(t)$. The population $P(t+1)$ is produced as follows:

1. Marginal histogram model is developed from $P(t)$
2. KN new individuals are sampled according to the built model
3. The new individuals are evaluated
4. Individuals in $P(t)$ and KN new individuals are ranked and the best N individuals are selected forming $P(t+1)$.

We used $K = 1$ in our experiments in this section. The test functions we used are the 20-variable two peaks function $F_{TwoPeaks}$, the 20-variable Rastrigin function ($F_{Rastrigin}$), the 10-variable Griewank function $F_{Griewank}$, and the 5-variable Schwefel function $F_{Schwefel}$.

Function $F_{TwoPeaks}$ is the sum of sub-functions similar to the commonly used deceptive function in binary domain and has a broader and lower peak, and a more narrow and higher peak for each variable x_j as shown in Figure 4. This function is designed to see the mixing capability and the scale-up behavior of the algorithm. This function has no interactions among variables. Other three functions are commonly used in the evolutionary computation literature. Rastrigin function has also no interactions among variables. Griewank function has weak interactions among variables. Schwefel function has medium-scale interactions among variables.

As to the number of bins (H) for each variable, we set H so that the initial bin width ϵ is 0.1.

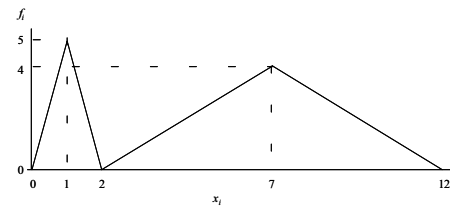


Figure 4: A sub-function of function $F_{TwoPeaks}$

3.2 Analysis of results

The experiments in this section focus on (1) convergence properties of the FWH and the FHH with resolution of bin width 0.1, (2) convergence properties of the FHH with high precision, and the (3) scale-up behavior of the FHH on function $F_{TwoPeaks}$ for problem size ranging from 10 to 100 variables.

3.2.1 Convergence properties of the FWH and the FHH with resolution of 0.1

The accuracy of the FWH is at most ϵ (i.e., 0.1). In our experiments, we evaluated the algorithms by measuring the number #OPT of runs in which the algorithm succeeded in finding the global optimum with resolution of 0.1 and the mean number of function evaluations (MNE) to find the global optimum in those runs where it did find the optimum. Twenty (20) runs were performed. Each run continued until the global optimum was found or a maximum of 200,000 function evaluations was reached.

First, on almost all experiments with some exceptions, the performance with sampling method E-SUS, which has a smaller stochastic sampling error, is much better than the performance with sampling method RW, which has a larger stochastic sampling error. For example, on function $F_{TwoPeaks}$, the FWH with RW (shown as FWH/RW) found optimal solution 20 times with the population size of 600. The MNE is 14,620.9. On the other hand, the FWH with E-SUS (shown as FWH/E-SUS) found optimal solution 20 times with the population size of 300 and the MNE is 7,178.3, almost the half of the FWH/RW. Next, all of the results of the FHH both with RW and E-SUS showed better performance than FWH. For more detail, please refer to [Tsutsui 01].

3.2.2 Convergence properties of the FHH

As described in Section 2.2, the bin width of FHH around the global optimum becomes narrower as the evolution proceeds. As a result, the FHH can find much more accurate solution. Figure 5 shows a typical change of bin positions of variable x_1 in a run with FHH/RW, which success to converge to the global optimum, on function $F_{Rastrigin}$. We

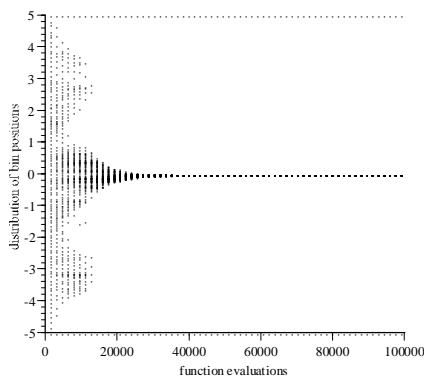


Figure 5: Typical changes of bin positions in parameter x_1 with FHH/RW on function $F_{Griewank}$

can observe that the FHH gathers bin positions to the converged point as evolution proceeds. The population size in the above experiment was set to 800.

To see the convergence properties of the FHH, we ran FHH model on the test functions. Population size of 800 was used for all functions in this experiments. Figure 10 shows the changes of mean functional value of the populations until function evaluations reached 100,000. Ten (10) runs has been performed.

With RW, the FHH converged to local optima several times on functions $F_{TwoPeaks}$, $F_{Rastrigin}$, and $F_{Griewank}$. However, with E-SUS the FHH converged to global optima 9 out of 10 times on function $F_{Rastrigin}$. On function $F_{Schwefel}$, the FHH failed to converge to the global optimum.

Functions $F_{TwoPeaks}$ and $F_{Rastrigin}$ have no interactions among variables and function $F_{Griewank}$ has weak interactions among variables. Function $F_{Schwefel}$ has medium-scale interactions among variables. With a marginal histogram, it is clearly difficult to solve problems which have medium-scale or stronger interactions among variables.

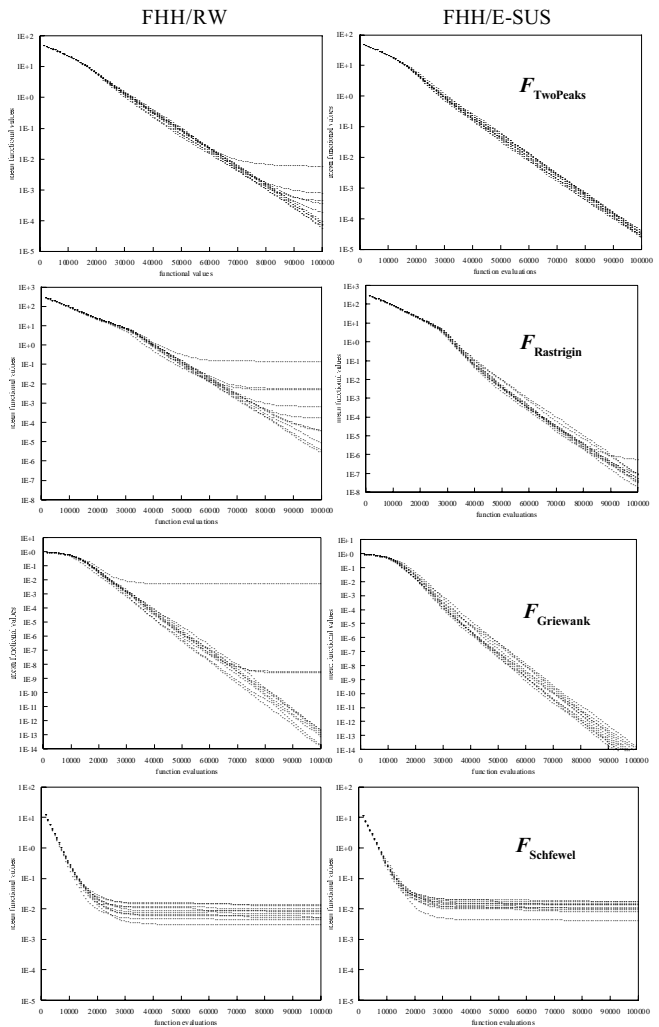


Figure 6: Convergence property of the FHH

4.2.3 Scale-up behavior of the FHH

This section focus on the scale-up behavior of the FHH with E-SUS using function F_{TwoPeaks} with the number of variables from 10 to 100 with step 10. We used a different evolutionary model, so called MGG [Satoh 96]. The optimum of function F_{TwoPeaks} is located at $(1.0, 1.0, \dots, 1.0)$. We assumed the optimum solution was found if the best individual in the population has a vector value of

$$\forall_{i=1, \dots, n} 1.0 - 0.005 \leq x_i < 1.0 + 0.005$$

We did experiment for each number of parameters increasing population size from 200 with step 200 until optimal solution was obtained 10 times in 10 runs.

Figure 7 shows the mean number of function evaluations (MNE). The MNE increased with almost $O(n^{1.56})$.

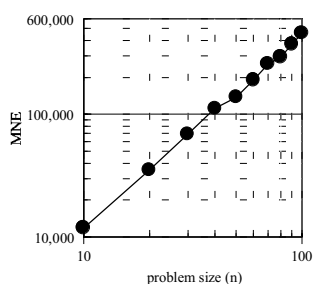


Figure 7: The MNE to find optimal 10 times in 10 runs

5. Conclusions

In this paper, we have proposed an evolutionary algorithm using marginal histograms in continuous domain. Two types of marginal histogram models, the FWH and the FHH, were proposed. As the sampling methods, we used both the RW (roulette wheel) and the E-SUS (extended stochastic universal sampling).

Both the FWH and the FHH showed good performance on functions which have no or weak interactions among parameter. Comparing two sampling methods, the performance with the E-SUS, which has a smaller stochastic sampling error, was much better than the performance with the RW, which has a larger stochastic sampling error. The FHH/E-SUS showed very good performance. It could find solutions with a very high precision which is being evolved along with individual solutions. The FHH/E-SUS also scale-up well and mixes solutions very effectively.

However, the marginal histograms did not work well on function which have a medium (or larger) level of interactions among variables. Introducing models considering interactions among variables remains a topic for future work. In this paper, although we focused on an evolutionary algorithm using probabilistic model with marginal histograms models, it is reported that the probabilistic model using the normal mixture model shows good performance [Bosman 00b]. To combine the histogram models with such a model might enhance the performance of those models because the histogram model has a high mixing capability of sub-solutions.

Acknowledgments

David Goldberg's and Martin Pelikan's contribution was sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant F49620-00-0163. Research funding for this work was also provided by a grant from the National Science Foundation under grant DMI-9908252. Support was also provided by a grant from the U. S. Army Research Laboratory under the Federated Laboratory Program, Cooperative Agreement DAAL01-96-2-0003. The US 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, the U. S. Army, or the U.S. Government.

References

- [Bosman 99] Bosman, P. and Thierens, D.: An algorithmic framework for density estimation based evolutionary algorithms, *Tech. Rep. No. UU-CS-1999-46*, Utrecht Univ. (1999).
- [Bosman 00a] Bosman, P. and Thierens, D.: Continuous iterated density estimation evolutionary algorithms within the IDEA framework, *Proc. of OBUPM Workshop at the GECCO-2000*, pp.197-200 (2000).
- [Bosman 00b] Bosman, P. and Thierens, D.: Mixed IDEAs, *Tech. Rep. No. UU-CS-2000-45*, Utrecht Univ. (2000).
- [Larranaga 99] Larranaga, P., Etxeberria, R., Lozano, J.A., and Pena, J.M.: Optimization by learning and simulation of bayesian and gaussian networks, *University of the Basque Country Tech. Rep. EHU-KZAAIK -4/99* (1999).
- [Pelikan 00] Pelikan, M., Goldberg, D. E., and Lobo, F. G. : A survey of optimization by building and using probabilistic models, *Computational Optimization and Applications*, Kluwer Academic Publishers (in press).
- [Satoh 96] Satoh, H., Yamamura, M., and Kobayashi, S.: Minimal generation gap model for considering both exploration and exploitation, *Proc. of the IIZUKA'96*, pp. 494-497 (1996).
- [Sebag 98] Sebag, M. and Ducoulombier, A.: Extending population-based incremental learning to continuous search spaces, *Proc. of the PPSN V*, pp. 418-427 (1998).
- [Servet 97] Servet, I. L., Trave-Massuyes, L., and Stern, D.: Telephone network traffic overloading diagnosis and evolutionary computation techniques, *Proc. of the 3rd European Conf. on Artificial Evolution (AE 97)*, pp. 137-144 (1997).
- [Schewefel 95] Schewefel, H.-P.: *Evolution and optimum seeking*, Sixth-generation Computer Technology Series, Wiley (1995).
- [Tsutsui 01] Tsutsui, S, Pelikan, M., and Goldberg, D. E.: Evolutionary algorithm using marginal histogram models in continuous domain, *IlligAL Tech. Rep. 2001019*, Univ. of Illinois (2001).