Explicit Exploration in Estimation of Distribution Algorithms

Rogelio Salinas-Gutiérrez[1], Ángel Eduardo Muñoz-Zavala[1], Arturo Hernández-Aguirre[2], and Manuel Alonso Castillo-Galván[1]

[1] Universidad Autónoma de Aguascalientes, Aguascalientes, México
{rsalinas,aemz}@correo.uaa.mx, gerber.heelflip@gmail.com
[2] Center for Research in Mathematics (CIMAT), Guanajuato, México
artha@cimat.mx

Abstract. This work proposes an Estimation of Distribution Algorithm (EDA) that incorporates an explicit separation between the exploration stage and the exploitation stage. For each stage a probabilistic model is required. The proposed EDA uses a mixture of distributions in the exploration stage whereas a multivariate Gaussian distribution is used in the exploitation stage. The benefits of using an explicit exploration stage are shown through numerical experiments.

Keywords: Estimation of Distribution Algorithm, Exploration stage, Exploitation stage.

1 Introduction

Estimation of Distribution Algorithms (EDAs) [10] are metaheuristics designed for searching good solutions in optimization problems. Similar to other metaheuristics of Evolutionary Computation (EC), EDAs are iterative algorithms based on the use of populations. However, an important characteristic of EDAs is the incorporation of probabilistic models in order to represent the dependencies among the decision variables of selected individuals. Once a probabilistic model is learnt by an EDA, it is possible to replicate dependencies in the new population by sampling from the model.

Algorithm 1 shows a pseudocode for EDAs. According to step 4, the dependencies among decision variables are taken into account by means of the probabilistic distribution $\mathcal{M}_t$. Step 5 shows how the dependence structure of the selected individuals is transferred to the new population, which greatly modifies the performance of an EDA.

As shown in Algorithm 1, step 4 involves an important and critical procedure in EDAs. For this reason, much of the research in EDAs has been focused precisely on proposing and enhancing new probabilistic models with many contributions in discrete and continuous domains [9,12,3]. Some of these probabilistic models are based on Bayesian and Markov networks [14,5,11]. Other EDAs have

F. Castro, A. Gelbukh, and M. González (Eds.): MICAI 2013, Part II, LNAI 8266, pp. 81–92, 2013.

Algorithm 1. Pseudocode for EDAs

1: Initialize the generation counter $t \longleftarrow 0$
Generate the initial population $\mathcal{P}_0$ with N individuals at random.
2: Evaluate population $\mathcal{P}_t$ using the cost function.
3: Select a subset $\mathcal{S}_t$ from $\mathcal{P}_t$ according to the selection method.
4: Estimate a probabilistic model $\mathcal{M}_t$ from $\mathcal{S}_t$.
5: Generate the new population $\mathcal{P}_{t+1}$ by sampling from the model $\mathcal{M}_t$
Assign $t \longleftarrow t + 1$.
6: If stopping criteria are not reached go to step 2.

used Gaussian assumptions [6,7,1,2], such as Gaussian kernels, Gaussian mixture models and the multivariate Gaussian distribution. The interested reader is referred to [8,4] for knowing more about the probabilistic models used in EDAs.

Although the active research in EDAs has been oriented to model adequately dependencies among decision variables [13], the generation of individuals in the exploration stage has not been investigated. This observation gives an opportunity for proposing a new exploration procedure and for studying its effects in EDAs.

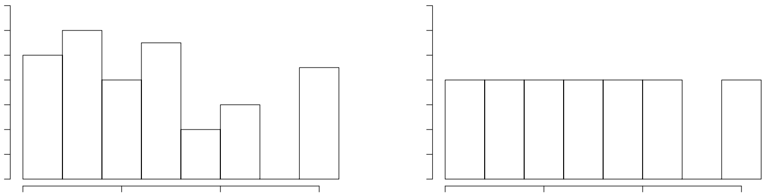
The structure of the paper is the following: Section 2 describes the proposal of this work, Section 3 shows some preliminary results of the implementation of the exploration stage, Section 4 presents the experimental setting to solve five test global optimization problems, and Section 5 resumes the conclusions.

2 The Exploration Stage

According to Algorithm 1, the initial population is generated at random. This means that the first population is generated by sampling from the uniform distribution. However, once the first population is generated, the following populations are generated by sampling from a probabilistic model $\mathcal{M}_t$ which is in general different than the uniform distribution. A common practice in EDAs is that the probabilistic model $\mathcal{M}_t$ is selected beforehand from a family of probabilistic distributions. Therefore, the immediate transition between the uniform distribution and the probabilistic model $\mathcal{M}_t$ could affect the performance of the exploration stage. This work investigates the effects of having an explicit separation between the exploration stage and the exploitation stage.

The proposal of incorporating an explicit exploration stage in EDAs requires the support of an adequate estrategy. Firstly, the number of generations for the exploration stage must be defined in advance. For example, the number of generations can be given by a fixed number. Secondly, a probabilistic model is needed in order to generate populations in the exploration stage. The natural choice for exploration purposes is a probabilistic distribution with high variance. However, the progress of the exploration stage must be reflected in the variance of the probabilistic model.

This work proposes the incorporation of a mixture of distributions for the exploration stage. The mixture is formed with the uniform distribution and with a distribution based on a modified histogram. The uniform distribution allows to generate individuals with high variance. The histogram is a statistics tool for density estimation and its implementation is well known. The histogram is used as a model for the selected individuals in each generation within the exploration stage. However, in order to favor the generation of individuals with high variance, we propose the use of a histogram with similar height for all bars. Figure 1 illustrates this idea. The total area of each histogram, (a) and (b), is normalized to 1.



(a) Histogram (b) Modified histogram

Fig. 1. The modified histogram (b) is based on the initial histogram (a) and its rectangles have the same height

The expression for the proposed mixture of distributions is given by:

$$\mathcal{E}_t = w_t \cdot \mathcal{U} + (1 - w_t) \cdot \mathcal{H} \ \text{, with } w_t \in [0, 1]. \tag{1}$$

The mixture (1) offers the following characteristics:

1. The initial weight of the uniform distribution $\mathcal{U}$ is the highest possible whereas the weight of the modified histogram $\mathcal{H}$ is the lowest. This allows to start the exploration stage with individuals sampled from a distribution of high variance.
2. According to the advance of the exploration stage, the weight of the uniform distribution $\mathcal{U}$ is decreased whereas the weight of the modified histogram $\mathcal{H}$ is increased.

Algorithm 2 shows the inclusion of a procedure for the exploration stage in EDAs. It can be noted that the number of 100 generations (step 2) and the rule for decreasing the weight w_t (step 7) are defined in this way to indicate the extension of the exploration stage. Both the number of generations and the expression for the weight can be changed by other values. On the other hand, it can be also note that the exploitation stage has elitism whereas the exploration stage has not elitism. However, the best individuals found during the process of the exploration stage are used as the initial population for the exploitation stage.

Algorithm 2. Pseudocode for EDA with explicit exploration

1: Exploration stage
Initialize the weight $w_t \longleftarrow 1$
2: **for** $t = 1 \rightarrow 100$ **do**
3: Generate the population $\mathcal{P}_t$ with N individuals by sampling from the model $\mathcal{E}_t$ (see Eq. (1))
4: Evaluate population $\mathcal{P}_t$ using the cost function.
5: Select a subset $\mathcal{S}_t$ from $\mathcal{P}_t$ according to the selection method.
6: Estimate a modified histogram $\mathcal{H}$ from $\mathcal{S}_t$.
7: Assign $w_t \longleftarrow 1 - (t/100)$.
8: Select the best N individuals from all the previous generations and record them in $\mathcal{B}$.
9: **end for**
10: Exploitation stage
Assign $\mathcal{P}_t \longleftarrow \mathcal{B}$.
11: Evaluate population $\mathcal{P}_t$ using the cost function.
12: Select a subset $\mathcal{S}_t$ from $\mathcal{P}_t$ according to the selection method.
13: Estimate a probabilistic model $\mathcal{M}_t$ from $\mathcal{S}_t$.
14: Generate the new population $\mathcal{P}_{t+1}$ by sampling from the model $\mathcal{M}_t$
15: Set $\mathcal{P}_{t+1}$ with the best N individuals from $\mathcal{P}_{t+1} \cup \mathcal{P}_t$
Assign $t \longleftarrow t + 1$.
16: If stopping criteria are not reached go to step 2.

3 Preliminary Results

In order to gain some insight about how the inclusion of the exploration stage modifies the performance of an EDA, we compare two EDAs in two test problems. The comparison is done through the Estimation of Multivariate Normal Algorithm (EMNA) and the EMNA with the exploration stage (EMNA+E). The test problems are the Rosenbrock and Sphere functions. These test functions are described in Fig. 2.

The benchmark test suite includes separable functions and non-separable functions, from which there are unimodal and multimodal functions. In addition, the search domain is asymmetric. All test functions are scalable. We use test problems in 10 dimensions. Each algorithm is run 30 times for each problem. The population size is 100 and the maximum number of generations is 150.

A graphical comparison between EMNA and EMNA+E is shown in Figure 3. According to these graphical results, the EMNA has a better performance than the EMNA+E in the first 100 generations. However, after the exploration stage is done, the performance of the EMNA+E outperforms the performance of the EMNA.

4 Experiments

Five test problems are used to compare an EDA with exploration against a typical EDA without explicit exploration. These algorithms are, respectively, the

Description	
Ackley	
$-20 \cdot \exp\left(-0.2\sqrt{\frac{1}{d} \cdot \sum_{i=1}^{d} x_i^2}\right) - \exp\left(\frac{1}{d} \cdot \sum_{i=1}^{d} \cos(2\pi x_i)\right) + 20 + \exp(1)$; $\boldsymbol{x} \in [-10, 30]^d$	
Properties: Multimodal, Non-separable	Global Minimum: $f(\mathbf{0}) = 0$
Griewangk	
$1 + \sum_{i=1}^{d} \frac{x_i^2}{4000} - \prod_{i=1}^{d} \cos\left(\frac{x_i}{\sqrt{i}}\right)$; $\boldsymbol{x} \in [-200, 1000]^d$	
Properties: Multimodal, Non-separable	Global Minimum: $f(\mathbf{0}) = 0$
Rastrigin	
$\sum_{i=1}^{d}(x_i^2 - 10\cos(2\pi x_i) + 10)$; $\boldsymbol{x} \in [-10, 30]^d$	
Properties: Multimodal, Separable	Global Minimum: $f(\mathbf{0}) = 0$
Rosenbrock	
$\sum_{i=1}^{d-1}[100 \cdot (x_{i+1} - x_i^2)^2 + (1 - x_i)^2]$; $\boldsymbol{x} \in [-10, 30]^d$	
Properties: Unimodal, Non-separable	Global Minimum: $f(\mathbf{1}) = 0$
Sphere Model	
$\sum_{i=1}^{d} x_i^2$; $\boldsymbol{x} \in [-200, 1000]^d$	
Properties: Unimodal, Separable	Global Minimum: $f(\mathbf{0}) = 0$

Fig. 2. Names, mathematical definition, search domains, global minimum and properties of the test functions

EMNA+E and the EMNA. The multivariate Gaussian distribution is incorporated as probabilistic model to the EMNA and the same distribution is used for the exploitation stage in the EMNA+E. Algorithm 1 is the basis for the EMNA whereas Algorithm 2 is the corresponding basis for the EMNA+E. In order to make a fair comparison, the elitism in the exploitation stage of EMNA+E is also included in the EMNA.

The test problems used in the experiments are the Ackley, Griewangk, Rastrigin, Rosenbrock, and Sphere functions. Fig. 2 describe the test functions. The algorithms are tested in different dimensions and asymmetric search domain. Each algorithm is run 30 times for each problem. The population size is ten times the dimension ($10 * d$). The maximum number of evaluations is 100,000. However, when convergence to a local minimum is detected the run is stopped. Any improvement less than 1×10^{-6} in 25 iterations is considered as convergence. The goal is to reach the optimum with an error less than 1×10^{-4}.

The results in dimensions 4, 6, 8, 10, 15 and 20 for non-separable functions are reported in Table 1, whereas the results for separable functions are reported

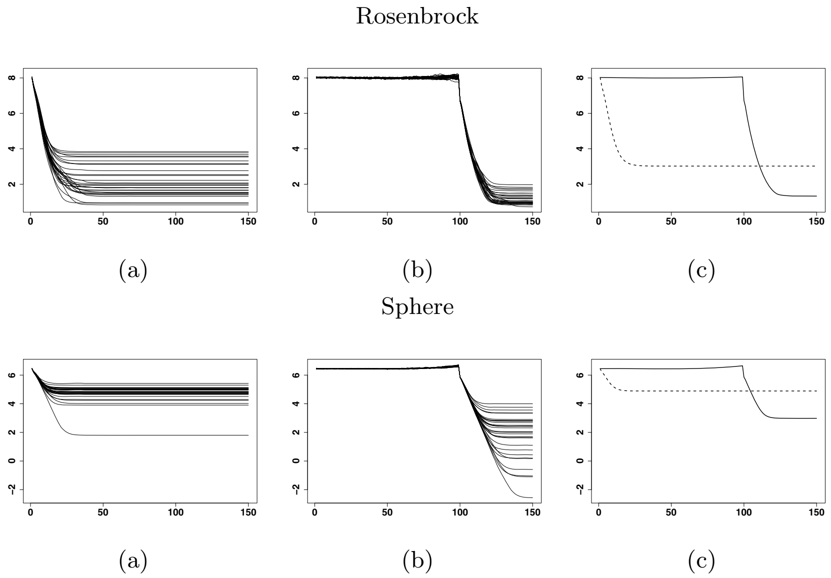


Fig. 3. The horizontal axis represents the generation and the vertical axis represents the fitness in logarithmic scale (base 10). (a) The fitness performance of EMNA. (b) The fitness performance of EMNA+E. (c) The dashed line is used for the average performance of EMNA and the solid line is used for the average performance of EMNA+E.

in Table 2. Both tables report descriptive statistics for the fitness values reached in the all runs. The fitness value corresponds to the value of a test problem. For each algorithm and dimension, the minimum, median, mean, maximum, standard deviation and success rate are shown. The minimum (maximum) value reached is labelled best (worst). The success rate is the proportion of runs in which an algorithm found the global optimum.

Besides the descriptive results shown in Tables 1 and 2, a hypothesis test is conducted to properly compare the performance of EMNA+E against EMNA. The statistical comparisons are for the algorithms with the same test problem and the same dimension. The t-test is employed to compare the fitness average between EMNA+E and EMNA. When a hypothesis test indicates that EMNA+E is significantly better than the EMNA, the corresponding average in in Tables 1 and 2 is marked with an asterix (*).

Another measure that can help in the comparisons of the algorithms is the success rate. Tables 1 and 2 show respectively the success rate in each dimension for non-separable and separable functions. If the success rate of EMNA+E is greater than the success rate of EMNA, it is marked with a dagger (†).

Table 1. Descriptive results of the fitness for non-separable functions

Algorithm	d	Best	Mean	Median	Worst	Std. Dev.	Success Rate
Ackley							
EMNA	4	1.86E-7	5.28E-1	1.42E-3	4.30E+0	1.11E+0	0.40
	6	5.01E-5	1.57E+0	8.50E-1	5.60E+0	1.72E+0	0.00
	8	1.65E-2	3.20E+0	3.22E+0	9.14E+0	2.41E+0	0.00
	10	6.86E-1	3.85E+0	3.12E+0	8.53E+0	1.97E+0	0.00
	15	1.34E+0	5.21E+0	5.54E+0	8.36E+0	1.60E+0	0.00
	20	4.81E+0	7.09E+0	7.34E+0	1.00E+1	1.33E+0	0.00
EMNA + E	4	2.71E-7	6.29E-3**	9.24E-7	1.49E-1	2.77E-2	0.60 †
	6	6.46E-7	4.57E-2**	2.22E-3	5.76E-1	1.14E-1	0.23 †
	8	7.26E-7	1.57E-1**	2.69E-2	1.32E+0	3.11E-1	0.03 †
	10	7.12E-6	3.59E-1**	1.60E-1	1.59E+0	4.42E-1	0.00
	15	2.15E-2	1.96E+0**	1.81E+0	3.25E+0	8.35E-1	0.00
	20	2.10E+0	3.84E+0**	3.66E+0	6.44E+0	1.23E+0	0.00
Griewangk							
EMNA	4	5.26E-2	1.64E+0	1.68E-1	1.28E+1	3.31E+0	0.00
	6	9.41E-2	3.58E+0	4.67E-1	3.39E+1	7.25E+0	0.00
	8	2.49E-1	9.65E+0	5.24E+0	9.45E+1	1.73E+1	0.00
	10	9.86E-1	1.33E+1	8.47E+0	5.01E+1	1.36E+1	0.00
	15	7.46E+0	5.97E+1	5.46E+1	1.21E+2	2.44E+1	0.00
	20	6.36E+1	1.08E+2	1.06E+2	2.03E+2	3.05E+1	0.00
EMNA + E	4	4.95E-2	1.27E-1**	1.23E-1	2.43E-1	5.09E-2	0.00
	6	1.49E-1	2.86E-1**	3.01E-1	4.10E-1	7.07E-2	0.00
	8	3.36E-5	3.59E-1**	3.53E-1	6.70E-1	1.96E-1	0.00
	10	1.37E-3	5.08E-1**	1.89E-1	2.90E+0	7.50E-1	0.00
	15	1.54E+0	9.27E+0**	7.03E+0	2.73E+1	7.28E+0	0.00
	20	4.32E+0	2.97E+1**	3.04E+1	6.46E+1	1.49E+1	0.00
Rosenbrock							
EMNA	4	3.32E-7	3.01E+0	1.23E+0	4.32E+1	7.76E+0	0.03
	6	3.26E-5	1.24E+1	4.28E+0	1.19E+2	2.34E+1	0.00
	8	1.26E-1	7.82E+2	4.61E+1	1.22E+4	2.31E+3	0.00
	10	7.24E+0	3.24E+3	1.28E+2	3.71E+4	8.25E+3	0.00
	15	4.36E+2	3.14E+4	1.47E+4	1.64E+5	4.50E+4	0.00
	20	7.12E+2	1.05E+5	6.40E+4	5.45E+5	1.30E+5	0.00
EMNA + E	4	4.13E-5	1.77E+0	1.37E+0	9.71E+0	2.01E+0	0.00
	6	1.01E+0	4.53E+0*	4.01E+0	1.32E+1	2.12E+0	0.00
	8	2.05E+0	1.56E+1*	8.20E+0	4.44E+1	1.38E+1	0.00
	10	7.03E+0	3.51E+1*	1.29E+1	2.72E+2	5.09E+1	0.00
	15	1.46E+1	1.35E+2**	7.45E+1	1.31E+3	2.37E+2	0.00
	20	7.73E+1	8.23E+2**	4.00E+2	4.26E+3	8.83E+2	0.00

* denotes EMNA+E is significantly better than the EMNA, at $\alpha = 0.05$
** denotes EMNA+E is significantly better than the EMNA, at $\alpha = 0.01$
† denotes that the EMNA+E has greater success rate than the EMNA

Table 2. Descriptive results of the fitness for separable functions

Algorithm	d	Best	Mean	Median	Worst	Std. Dev.	Success Rate
Rastrigin							
EMNA	4	3.57E-7	4.74E+0	3.69E+0	2.09E+1	4.86E+0	0.07
	6	1.99E+0	1.24E+1	1.04E+1	3.32E+1	6.76E+0	0.00
	8	9.95E+0	3.10E+1	2.72E+1	8.77E+1	1.63E+1	0.00
	10	2.17E+1	5.01E+1	4.81E+1	1.01E+2	1.94E+1	0.00
	15	6.45E+1	1.15E+2	1.07E+2	2.16E+2	3.83E+1	0.00
	20	1.35E+2	2.37E+2	2.27E+2	4.58E+2	7.26E+1	0.00
EMNA + E	4	1.30E+0	3.91E+0	3.92E+0	7.91E+0	1.36E+0	0.00
	6	4.79E-7	1.08E+1	1.04E+1	1.57E+1	3.30E+0	0.03 †
	8	1.30E+1	1.99E+1**	1.92E+1	2.75E+1	4.15E+0	0.00
	10	8.26E-2	3.27E+1**	3.39E+1	4.58E+1	8.23E+0	0.00
	15	5.16E+1	7.22E+1**	7.38E+1	8.48E+1	9.23E+0	0.00
	20	8.67E+1	1.20E+2**	1.22E+2	1.54E+2	1.45E+1	0.00
Sphere							
EMNA	4	4.46E-7	3.70E+3	3.13E+1	4.83E+4	9.38E+3	0.17
	6	9.20E-7	1.27E+4	4.76E+3	1.02E+5	2.08E+4	0.03
	8	1.16E-2	3.66E+4	2.90E+4	2.02E+5	4.52E+4	0.00
	10	1.14E+4	9.86E+4	7.30E+4	3.09E+5	8.14E+4	0.00
	15	9.27E+4	2.04E+5	1.94E+5	3.65E+5	7.13E+4	0.00
	20	1.43E+5	3.94E+5	3.80E+5	6.51E+5	1.42E+5	0.00
EMNA + E	4	7.18E-8	6.21E-3*	8.64E-7	1.61E-1	2.94E-2	0.53 †
	6	2.08E-7	1.25E+0**	3.66E-4	2.58E+1	4.70E+0	0.33 †
	8	4.99E-7	9.94E+1**	6.34E+0	2.11E+3	3.83E+2	0.03 †
	10	3.29E-2	9.01E+2**	1.24E+2	7.68E+3	1.82E+3	0.00
	15	1.16E+3	3.25E+4**	3.12E+4	8.48E+4	2.22E+4	0.00
	20	7.24E+3	1.26E+5**	1.20E+5	2.28E+5	6.04E+4	0.00

* denotes EMNA+E is significantly better than the EMNA, at $\alpha = 0.05$
** denotes EMNA+E is significantly better than the EMNA, at $\alpha = 0.01$
† denotes that the EMNA+E has greater success rate than the EMNA

Tables 1 and 2 show a total of 30 comparisons. Out of the 18 comparisons for the non-separable functions, the EMNA+E excels in 17 cases. Similarly, out of the 12 comparisons for the separable functions, the EMNA+E excels in 10 cases. These results give an evidence of the benefits achieved by the incorporation of the exploration stage in EDAs. However, regarding the number of evaluations, Tables 3 and 4 show that EMNA+E requires more function evaluations than the EMNA.

Table 3. Descriptive results of the number of evaluations for non-separable functions

Algorithm	d	Best	Mean	Median	Worst	Std. Dev.
Ackley						
EMNA	4	1.52E+3	2.07E+3	2.20E+3	2.56E+3	3.78E+2
	6	3.42E+3	4.11E+3	4.02E+3	5.58E+3	4.19E+2
	8	5.52E+3	6.34E+3	6.12E+3	8.48E+3	6.95E+2
	10	7.60E+3	8.80E+3	8.35E+3	1.54E+4	1.56E+3
	15	1.17E+4	1.52E+4	1.46E+4	2.16E+4	1.93E+3
	20	2.02E+4	2.37E+4	2.33E+4	2.94E+4	2.22E+3
EMNA + E	4	5.04E+3	5.46E+3	5.20E+3	6.00E+3	4.00E+2
	6	8.64E+3	9.45E+3	9.60E+3	9.90E+3	3.94E+2
	8	1.24E+4	1.34E+4	1.34E+4	1.38E+4	2.52E+2
	10	1.69E+4	1.73E+4	1.72E+4	1.80E+4	2.31E+2
	15	2.73E+4	2.81E+4	2.78E+4	2.99E+4	6.78E+2
	20	3.86E+4	4.01E+4	4.00E+4	4.22E+4	9.20E+2
Griewangk						
EMNA	4	1.40E+3	2.64E+3	2.52E+3	5.32E+3	9.65E+2
	6	2.34E+3	4.64E+3	3.90E+3	1.01E+4	1.85E+3
	8	4.00E+3	8.52E+3	8.92E+3	1.17E+4	2.22E+3
	10	5.20E+3	1.01E+4	1.03E+4	1.38E+4	1.93E+3
	15	8.40E+3	1.33E+4	1.43E+4	1.56E+4	2.00E+3
	20	1.28E+4	1.81E+4	1.78E+4	2.38E+4	2.15E+3
EMNA + E	4	5.00E+3	5.95E+3	5.64E+3	8.92E+3	9.07E+2
	6	7.98E+3	9.30E+3	9.15E+3	1.20E+4	1.00E+3
	8	1.10E+4	1.39E+4	1.29E+4	2.04E+4	2.96E+3
	10	1.42E+4	2.11E+4	2.11E+4	2.58E+4	2.54E+3
	15	2.25E+4	2.75E+4	2.81E+4	3.08E+4	2.32E+3
	20	3.64E+4	3.75E+4	3.70E+4	4.04E+4	1.22E+3
Rosenbrock						
EMNA	4	1.64E+3	2.70E+3	2.60E+3	3.52E+3	3.73E+2
	6	3.84E+3	4.71E+3	4.47E+3	6.30E+3	6.22E+2
	8	3.52E+3	6.91E+3	6.68E+3	9.92E+3	1.11E+3
	10	7.80E+3	1.05E+4	1.01E+4	1.71E+4	1.84E+3
	15	1.25E+4	1.69E+4	1.66E+4	2.42E+4	2.31E+3
	20	1.60E+4	2.45E+4	2.46E+4	2.88E+4	2.49E+3
EMNA + E	4	5.48E+3	6.19E+3	6.16E+3	7.04E+3	2.64E+2
	6	9.78E+3	1.02E+4	1.00E+4	1.14E+4	4.09E+2
	8	1.37E+4	1.44E+4	1.44E+4	1.57E+4	5.07E+2
	10	1.79E+4	1.90E+4	1.88E+4	2.11E+4	9.03E+2
	15	2.49E+4	3.13E+4	3.14E+4	3.47E+4	1.85E+3
	20	3.64E+4	4.39E+4	4.29E+4	5.38E+4	4.02E+3

Table 4. Descriptive results of the number of evaluations for separable functions

Algorithm	d	Best	Mean	Median	Worst	Std. Dev.
Rastrigin						
EMNA	4	1.44E+3	2.49E+3	2.26E+3	5.80E+3	9.40E+2
	6	2.46E+3	4.27E+3	3.93E+3	8.64E+3	1.42E+3
	8	2.96E+3	5.50E+3	5.40E+3	1.16E+4	1.96E+3
	10	4.30E+3	7.18E+3	6.90E+3	1.37E+4	2.44E+3
	15	7.05E+3	1.11E+4	1.03E+4	2.58E+4	3.91E+3
	20	8.60E+3	1.65E+4	1.49E+4	2.64E+4	5.04E+3
EMNA + E	4	5.00E+3	5.73E+3	5.60E+3	8.80E+3	8.19E+2
	6	7.62E+3	9.21E+3	8.82E+3	1.51E+4	1.53E+3
	8	1.08E+4	1.28E+4	1.28E+4	1.45E+4	1.07E+3
	10	1.33E+4	1.58E+4	1.51E+4	2.65E+4	2.54E+3
	15	2.01E+4	2.37E+4	2.34E+4	2.96E+4	2.43E+3
	20	2.84E+4	3.36E+4	3.38E+4	4.04E+4	3.16E+3
Sphere						
EMNA	4	1.24E+3	2.51E+3	2.82E+3	3.32E+3	6.04E+2
	6	2.46E+3	4.68E+3	4.80E+3	5.16E+3	5.30E+2
	8	6.32E+3	7.07E+3	7.12E+3	7.44E+3	2.51E+2
	10	9.00E+3	9.48E+3	9.50E+3	9.90E+3	2.05E+2
	15	1.49E+4	1.60E+4	1.61E+4	1.64E+4	2.78E+2
	20	2.28E+4	2.34E+4	2.34E+4	2.40E+4	2.73E+2
EMNA + E	4	4.60E+3	5.21E+3	4.76E+3	6.08E+3	5.54E+2
	6	7.86E+3	9.19E+3	9.60E+3	1.02E+4	9.05E+2
	8	1.14E+4	1.40E+4	1.42E+4	1.46E+4	5.57E+2
	10	1.77E+4	1.85E+4	1.86E+4	1.90E+4	3.72E+2
	15	3.00E+4	3.04E+4	3.05E+4	3.11E+4	2.39E+2
	20	4.22E+4	4.27E+4	4.26E+4	4.32E+4	2.98E+2

5 Conclusions

This work has introduced an explicit exploration stage for EDAs. In particular, the numerical implementation of the exploration stage has been done with continuous decision variables in a well known EDA (EMNA). According to the numerical experiments, the explicit separation between the exploration stage and the exploitation stage (EMNA+E) can help achieving better fitness values. Nonetheless, the benefit of including an exploration stage requires an increase of function evaluations.

An important contribution of this paper is the design of a probabilistic model for the exploration stage. The goal of the proposed model in the exploration stage is to provide a new tool for finding an set of individuals that can be used as initial population in the exploitation stage.

Although the statistical comparisons clearly indicate that the EDA with the exploration stage has better performance than the typical EDA, the success rate shows that more experiments are necessary in order to identify where the exploration stage have a positive impact in EDAs.

Acknowledgments. The first author acknowledges financial support from the Programa de Mejoramiento del Profesorado (PROMEP) of México.

References

1. Bosman, P.: Design and Application of Iterated Density-Estimation Evolutionary Algorithms. Ph.D. thesis, University of Utrecht, Utrecht, The Netherlands (2003)
2. Bosman, P., Grahl, J., Thierens, D.: AMaLGaM IDEAs in Noiseless Black-Box Optimization Benchmarking. In: Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers, GECCO 2009, pp. 2247–2254. ACM (2009)
3. De Bonet, J., Isbell, C., Viola, P.: MIMIC: Finding Optima by Estimating Probability Densities. In: Advances in Neural Information Processing Systems, vol. 9, pp. 424–430. The MIT Press (1997)
4. Hauschild, M., Pelikan, M.: An introduction and survey of estimation of distribution algorithms. Swarm and Evolutionary Computation 1(3), 111–128 (2011)
5. Larrañaga, P., Etxeberria, R., Lozano, J., Peña, J.: Combinatorial optimization by learning and simulation of Bayesian networks. In: Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence, pp. 343–352 (2000)
6. Larrañaga, P., Etxeberria, R., Lozano, J., Peña, J.: Optimization in continuous domains by learning and simulation of Gaussian networks. In: Proceedings of the Optimization by Building and Using Probabilistic Models OBUPM Workshop at the Genetic and Evolutionary Computation Conference, GECCO 2000, pp. 201–204 (2000)
7. Larrañaga, P., Lozano, J., Bengoetxea, E.: Estimation of Distribution Algorithm based on multivariate normal and Gaussian networks. Tech. Rep. EHU-KZAA-IK-1/01, Department of Computer Science and Artificial Intelligence, University of the Basque Country (2001)
8. Larrañaga, P., Lozano, J. (eds.): Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation. Genetic Algorithms and Evolutionary Computation. Kluwer Academic Publishers (2002)
9. Mühlenbein, H.: The Equation for Response to Selection and its Use for Prediction. Evolutionary Computation 5(3), 303–346 (1998)
10. Mühlenbein, H., Paaß, G.: From recombination of genes to the estimation of distributions I. Binary parameters. In: Voigt, H.-M., Ebeling, W., Rechenberg, I., Schwefel, H. (eds.) PPSN 1996. LNCS, vol. 1141, pp. 178–187. Springer, Heidelberg (1996)
11. Pelikan, M., Goldberg, D., Cantú-Paz, E.: BOA: The Bayesian Optimization Algorithm. In: Banzhaf, W., Daida, J., Eiben, A., Garzon, M., Honavar, V., Jakiela, M., Smith, R. (eds.) Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 1999, vol. 1, pp. 525–532. Morgan Kaufmann Publishers (1999)
12. Pelikan, M., Mühlenbein, H.: The Bivariate Marginal Distribution Algorithm. In: Roy, R., Furuhashi, T., Chawdhry, P. (eds.) Advances in Soft Computing - Engineering Design and Manufacturing, pp. 521–535. Springer, London (1999)

13. Salinas-Gutiérrez, R., Hernández-Aguirre, A., Villa-Diharce, E.: Estimation of Distribution Algorithms based on Copula Functions. In: GECCO 2011: Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation, Graduate Students Workshop, pp. 795–798. ACM (2011)
14. Soto, M., Ochoa, A., Acid, S., de Campos, L.: Introducing the polytree approximation of distribution algorithm. In: Ochoa, A., Soto, M., Santana, R. (eds.) Second International Symposium on Artificial Intelligence. Adaptive Systems, CIMAF 1999, pp. 360–367, Academia, La Habana (1999)