An Estimation of Distribution Algorithm Based on Clayton Copula and Empirical Margins*

L.F. Wang, Y.C. Wang, J.C. Zeng, and Y. Hong

Collage of Electrical and Information Engineering, Lanzhou University of Technology,
Lanzhou, 730050, China
Complex System and Computational Intelligence Laboratory,
Taiyuan University of Science and Technology, Taiyuan, 030024, China
wlf1001@163.com

Abstract. Estimation of Distribution Algorithms (EDAs) are new evolutionary algorithms which based on the estimation and sampling the distribution model of the selected population in each generation. The way of copula used in EDAs is introduced in this paper. The joint distribution of the selected population is separated into the univariate marginal distribution and a function called copula to represent the dependence structure. And the new individuals are obtained by sampling from copula and then calculating the inverse of the univariate marginal distribution function. The empirical distribution and Clayton copula are used to implement the proposed copula Estimation of Distribution Algorithm (copula EDA). The experimental results show that the proposed algorithm is equivalent to some conventional continuous EDAs in performance.

Keywords: Estimation of distribution algorithms (EDAs), copula theory, the joint distribution, the marginal distribution, Clayton copula.

1 Introduction

Estimation of Distribution Algorithms (EDAs) [1] are a class of evolutionary algorithms(EAs) in which the main operators are the estimation of a probability distribution from the selected solutions and the subsequent sampling from the estimated distribution. EDAs can be understood and further developed without the background of Genetic Algorithms (GAs) though they were introduced as an extension of GAs. It has been shown in previous work that EDAs can outperform traditional evolutionary algorithms on a number of difficult benchmark problems [2].

Many kinds of EDAs [3,4] were studied by scholars which uses different approaches to estimating the probability distributions of the selected population. But the computational costs to estimating the distribution are large in many algorithms [5]. This paper introduces how the copula theory can be used in an EDA. The benefit of the copula EDA is that only marginal distributions need to be estimated from the population of solutions, thus promising efficiency savings.

Copula theory [6] is a recently developed mathematical theory for multivariate probability analysis. It separates joint probability distribution function into the product of univariate margins and a copula which represents the dependency structure of random variables. Copulas are functions that join the multivariate distribution functions to their one-dimensional marginal distribution functions. On the grounds of it, given a joint distribution with margins, there exists a copula uniquely determined; conversely, the joint distribution function is determined for a given copula and margins. The study about EDAs utilizing two dimensional copula can be found in [7]and [8].

This paper is organized as follows. It is discussed in section 2 how copula theory can be used in EDA. Section 3 talks about a sampling algorithm for Clayton copula proposed by Marshall and its application in the copula EDA. The experimental results are shown in section 4. The paper ends with concluding remarks in section 5.

2 The Framework of Copula EDA

Consider a continuous random vector $(x_1, x_2, \ldots, x_n)$ with the joint cumulative distribution function H and margins $F_1, F_2, \ldots, F_n$. The copula representation of H is given by $H(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n))$ [6], where C is a unique cumulative distribution function having uniform margins on $(0,1)$. The copula function is the joint distribution of the probability integral transforms. Therefore, the copula function is also known as the "dependence function." The copula function contains all the information regarding the cross-dependence structures of $(x_1, x_2, \ldots, x_n)$.

By virtue of copula theory, if the joint distribution of random vector $(x_1, x_2, \ldots, x_n)$ is denoted by a copula C and n marginal distribution functions $F_1, \ldots, F_n$, i.e., the joint is $C(F_1(x_1), \ldots, F_n(x_n))$, we need only generate a vector $(u_1, u_2, \ldots, u_n)$ of observation of uniform $(0,1)$ random variables $(U_1, U_2, \ldots, U_n)$ whose joint distribution function is C, and then transform the variables $u_1, u_2, \ldots, u_n$ to $x_1, x_2, \ldots, x_n$ via $x_i = F_i^{-1}(u_i)$. There are many algorithms about sampling $(u_1, u_2, \ldots, u_n)$ from copula[9][10]. In this paper, we focus on the following generating method proposed by Marshall and Olkin [9].

Algorithm 1: sampling algorithm from copula

1) Sample $v \sim F = \mathcal{LS}^{-1}\left[\phi^{-1}\right]$, where $\mathcal{LS}^{-1}\left[\phi^{-1}\right]$ is the inverse Laplace-Stieltjes transform of the generator ϕ^{-1}.

2) Sample *i.i.d.* $x_i \sim U[0,1]$, $i=1,\ldots,n$.

3) Return $(u_1, \ldots, u_n)$, where $u_i = \phi^{-1}\left((-\log x_i)/v\right), i = 1, \ldots, n$.

EDAs are algorithm implemented by the iterative run containing three main steps. The first step is to select a subpopulation, which contains the distribution information of the better-performed population, according to a certain selection strategy. Many selection strategy used in GAs can be used in EDAs too. The second step is to estimate the distribution model of the selected population. The distribution model is denoted in different ways, for example, the Baysian network or multivariate Gaussian distribution. The joint distribution function is a good way to exactly reflect the relationship of random variables and the distribution characters of each random variable. But the work to directly estimate the joint is very difficult and time consuming. Copula theory provides the theory basis to simplifying the work as to estimate the marginal distribution functions and a

copula. Obviously, estimating one-dimensional distribution is easier than estimating the multivariate joint distribution. The Gaussian distribution, empirical distribution and other distributions can be used to estimate the margins. There are many known copulas and many studies about them. We can use one of the known copulas to denote the dependence structure of the selected solution. Then the joint distribution can be obtained according to copula theory. Thus the second step is finished. The third step is to sample from the estimated distribution. On the grounds of copula theory, the sampling work can be done by doing the following two works. First, sample a vector in $[0,1]^n$ from the copula. Second, get the value of each random variable according to the inverse function of its marginal distribution function. In other words, if the optimization problem is

$$\min f(X) = f(x_1, x_2, \ldots, x_n), \; x_i \in [a_i, b_i] \;\; (i=1,2,\ldots,n) \tag{1}$$

Denote the corresponding random vector as $(X_1, X_2, \ldots, X_n)$ and the estimated margins as $F_1, F_2, \ldots, F_n$. C is the copula. Sample a vector $(u_1, u_2, \ldots, u_n)$ from C firstly. Obviously $u_i \in [0,1]$, $i=1,2,\cdots n$. Calculate $x_i = F_i^{-1}(u_i)$ $(i=1,2,\cdots n)$ secondly. And then the $(x_1, x_2, \ldots, x_n)$ is the sampling result of $(X_1, X_2, \ldots, X_n)$ whose joint distribution is $C(F_1, F_2, \ldots, F_n)$.

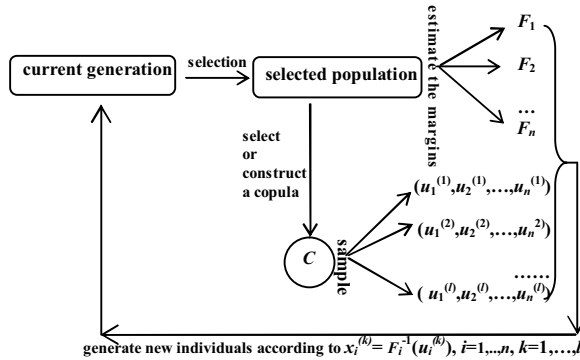Concluding the above discussion, the framework of the copula EDA is expressed as in Fig.1.

Fig. 1. The flow chart of copula EDA

As shown in Fig.1, the copula EDA is an iterative run starting with a randomly generated population of solutions. The better solutions are selected from the current population of solutions. Let $\mathbf{x}=\{x^j=(x_1^j, x_2^j, \ldots x_n^j), \; j=1,2,\ldots,s\}$ denote the selected population. Therefore, the set $\{x_i^j, j=1,2,\ldots,s\}$ can be considered as the sample of the random variable X_i, $i=1,2,\ldots,n$. The following two operators can be implemented in parallel. One is to estimate the margins F_i for each random variable $X_i(i=1,2,\ldots,n)$ according to the samples $\{x_i^j, j=1,2,\ldots,s\}$. The other is to select or to construct a copula C according to $\mathbf{x}$, and then sample from C. supposing the sampled vectors are $\{\mathbf{u}^{(k)}=(u_1^{(k)}, u_2^{(k)}, \ldots, u_n^{(k)}), \; k=1,2,\ldots,l\}$, the new individuals $\{\mathbf{x}^{(k)}=(x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)}), \; k=1,2,\ldots,l\}$ are obtained by calculating $x_i^{(k)} = F_i^{-1}(u_i^{(k)})$, $i=1,\ldots,n$, $k=1,\ldots,l$. copula EDA

replaces some old individuals of the original population with the new generated individuals and progresses the new evolution.

3   Copula-EDA Based on Clayton Copula

According to the frame work of copula EDA, we use Clayton copula to estimate the dependence structure of variables and use the empirical distribution to estimate the marginal distribution of each variable. Because the generator of Clayton copula is $\phi(t) = (t^{-\theta} - 1)/\theta$, the inverse function is $\phi^{-1}(t) = (1 + \theta t)^{-1/\theta}$ and the inverse Laplace-Stieltjes transform of the generator ϕ^{-1} is $\mathcal{LS}(\phi^{-1}) = F(v) = \dfrac{(1/\theta)^{1/\theta}}{\Gamma(1/\theta)} e^{-v/\theta} \cdot v^{1/\theta - 1}$.

For the selected population $\mathbf{x}$, the samples of the ith variable X_i are $\{ x_i^1, x_i^2,\ldots, x_i^s\}$. the sorted list of these s samples is $x_i^{<1>} \leq x_i^{<2>} \leq \ldots \leq x_i^{<s>}$. Denote the search space of X_i is $[x_i^{<0>}, x_i^{<s+1>})$, then the empirical distribution function of X_i is

$$F_i(x) = j/s, (x_i^{<j>} \leq x < x_i^{<j+1>}), \tag{2}$$

and the inverse function of $F_i(x)$ used in this paper is

$$x = F_i^{-1}(u) = \begin{cases} rand[x_i^{<j>}, x_i^{<j+1>}) & if \quad \lfloor u \times s \rfloor = j \quad and\ x_i^{<j>} \neq x_i^{<j+1>} \\ rand(x_i^{<j>}; \delta) & if \quad \lfloor u \times s \rfloor = j \quad and\ x_i^{<j>} = x_i^{<j+1>} \end{cases}, \tag{3}$$

where, $rand[x_i^{<j>}, x_i^{<j+1>})$ denote the random number in the interval $[x_i^{<j>}, x_i^{<j+1>})$, and $rand(x_i^{<j>}; \delta)$ denote the random number in the δ-neighbor of $x_i^{<j>}$.

Concluding the analysis presented above, the pseudo code for copula EDA is shown in Algorithm 3.

Algorithm 2: Pseudo code for copula EDA using Clayton copula and empirical margins
1) Initialize randomly the population P_0 with m individuals and set $g \leftarrow 0$.
2) Select a subpopulation S_t of size s from P_g according to certain select-strategy.
3) Estimate the unvariate marginal distribution function F_i for each dimension from S_t.
4) For $k=1$ to l do

4.1) Sample $v \sim F(v) = \dfrac{(1/\theta)^{1/\theta}}{\Gamma(1/\theta)} e^{-v/\theta} \cdot v^{1/\theta - 1}$.

4.2) Sample *i.i.d.* $v_i \sim U[0,1]$, $i=1,\ldots,n$, get u_i from

$u_i = \phi^{-1}((-\log v_i)/v) = (1 - \dfrac{\theta \log v_i}{v})^{-1/\theta}$.

4.3) Get new individual $(x_1^{(k)}, x_2^{(k)},\ldots, x_n^{(k)})$ by calculating

$$x_i^{(k)} = F_i^{-1}(u_i) = \begin{cases} rand[x_i^{<j>}, x_i^{<j+1>}) & if \quad \lfloor u_i \times s \rfloor = j \quad and\ x_i^{<j>} \neq x_i^{<j+1>} \\ rand(x_i^{<j>}; \delta) & if \quad \lfloor u_i \times s \rfloor = j \quad and\ x_i^{<j>} = x_i^{<j+1>} \end{cases}. \tag{4}$$

5) Replace the old individuals of P_g with the new individuals, set $g \leftarrow g+1$.
6) If stopping criterion is not reached go to step (2).

4 Experiments

The three benchmark functions in Table 1 are used to test the performance of the proposed algorithm in this paper. They are also used in reference [11] and the parameters are set as the same as in [11], *i.e.* the dimensions are 10, the population size are 2000, 2000 and 750 respectively for each function, and the maximal evaluation times are 300000.

Table 1. The description of the test functions

Function ：	$F_1(x) = -\{10^{-5} + \sum\limits_{i=1}^{n} \mid y_i \mid\}^{-1},$
	where, $\quad y_1 = x_1, y_i = y_{i-1} + x_i$
Search space ：	$-0.16 \leq x_i \leq 0.16, i = 1,...,10$
Minimum value ：	$F(0) = 100000$
Function ：	$F_2(x) = \sum\limits_{i=1}^{n} [(x_1 - x_i^2)^2 + (x_i - 1)^2]$
Search space ：	$-10 \leq x_i \leq 10, i = 1,...,10$
Minimum value ：	$F(0) = 0$
Function ：	$F_3(x) = 1 + \sum\limits_{i=1}^{n} \dfrac{x_i^2}{4000} - \Pi_{i=1}^{n} \cos(\dfrac{x_i}{\sqrt{i}})$
Search space ：	$-600 \leq x_i \leq 600, i = 1,...,10$
Minimum value ：	$F(0) = 0$

The mutation operator is also used in the experiments. The selection rate is 0.2 and the mutation rate is 0.05. To increase the diversity of the population, 1/3 individuals of the selected population are selected by truncation selector and the other 2/3 are selected by roulette selector. The mutation operator is performed by randomly producing individuals in the neighbor areas of the selected population. The parameter θ of Clayton copula is set to 1. The experimental results are shown in Table 2. The total run time is 50. The StdVar, min and max in Table 2 represent respectively the standard variance, the minimal value and the maximal value of the 50 search results.

The copula EDA performs well in function F_2 but it fails in function F_3, and its performance in function F_1 is only a little better than $UMDA_c^G$, $MIMIC_c^G$ and ES. From the standard variance of the search results, we know that the performance of the proposed algorithm is not robust. The copula EDA encounter the same problem as other EDAs that is the search result is affected by the initial population. The parameter θ of Clayton copula influence the shape of the copula, it is need to be studied in details. Clayton copula is one of the Archimedean copulas, it may not reflect correctly the relationship of random variables for all problems. The diversity of population is limited because of the property of the empirical distribution which is used to estimate the marginal distribution.

Table 2. The performances of copula EDA and some conventional evolutionary algorithms

Algorithm	mean	StdVar	min	max
F1				
$UMDA_c^G$	-53460			
$MIMIC_c^G$	-58775			
$EGNA_{ee}$	-100000			
$EGNA_{BGe}$	-100000			
ES	-5910			
copula EDA	-81944	19609	-93394	-31875
F2				
$UMDA_c^G$	0.13754			
$MIMIC_c^G$	0.13397			
$EGNA_{ee}$	0.09914			
$EGNA_{BGe}$	0.0250			
ES	0			
copula EDA	8.7786e-005	1.4235e-004	3.7552e-007	3.7416e-004
F3				
$UMDA_c^G$	0.011076			
$MIMIC_c^G$	0.007794			
$EGNA_{cc}$	0.008175			
$EGNA_{BGe}$	0.012605			
ES	0.034477			
copula EDA	0.038551	0.0239	0	0.074655

5 Conclusions

EDAs are evolutionary algorithms by iteratively estimating the distribution model of the selected population and sample new individuals from the model. The precision of the model and the fitness between the distribution of the ancestors and those of the offspring affects the convergence of the algorithm. The joint distribution is an entire and precise way to describe the distribution of random vector. By using of copula theory, the joint distribution is simplified to a copula and the univariate marginal distribution. Therefore, the complexity of EDA will decrease by using of copula. We propose the frame work of copula EDA and implement it by the exchangeable Clayton copula and the empirical marginal distributions. The sampling algorithm from copula is used the algorithm proposed by Marshall and Olkin. The experimental results show that the performance of the proposed algorithm is equivalent to the conventional evolutionary algorithms in performance. The following topics need to be studied in the future.

1) The choose criteria for copulas. For the given selected population, the criteria decide which copula is the most proper function to reflect the dependence of the random variables.

2) The choice of the parameter of copulas. The copula with different parameter values correspond to different dependence structure which influence the precision of the joint distribution reflecting the selected population.

3) The marginal distribution in the copula EDA.

References

1. Mühlenbein, H., Paaß, G.: From combination of genes to the estimation of distributions: Binary parameters. In: Ebeling, W., Rechenberg, I., Voigt, H.-M., Schwefel, H.-P. (eds.) PPSN 1996. LNCS, vol. 1141, pp. 178–187. Springer, Heidelberg (1996)
2. Larranaga, P., Lozano, J. (eds.): Estimation of distribution algorithms. A new tool for evolutionary computation. Kluwer Academic Publishers, Dordrecht (2002)
3. Pelikan, M.: Hierarchical Bayesian Optimization algorithm: Towards a new generation of evolutionary algorithms. Springer, New York (2005)
4. Shakya, S.: DEUM: A Framework for an Estimation of Distribution Algorithm based on Markov Random Fields. PhD thesis, The Robert Gordon University, Aberdeen, UK (April 2006)
5. Duque, T.S., Goldberg, D.E., Sastry, K.: Enhancing the efficiency of the ECGA. In: Rudolph, G., et al. (eds.) PPSN 2008. LNCS, vol. 5199, pp. 165–174. Springer, Heidelberg (2008)
6. Nelsen, R.B. (ed.): An introduction to copulas, 2nd edn. Springer, New York (2006)
7. Wang, L.F., Zeng, J.C., Hong, Y.: Estimation of Distribution Algorithm Based on Copula Theory. In: Proceedings of the IEEE Congress on Evolutionary Computation (CEC 2009), Trondheim, Norway, May 18-21, pp. 1057–1063 (2009)
8. Salinas-Gutierrez, R., Hernandez-Aguirre, A., Villa-Diharce, E.R.: Using Copulas in Estimation of Distribution Algorithms. In: Hernandez Aguirre, A., et al. (eds.) MICAI 2009. LNCS (LNAI), vol. 5845, pp. 658–668. Springer, Heidelberg (2009)
9. Marshall, A.W., Olkin, I.: Families of Multivariate Distributions. Journal of the American statistical association 83, 834–841 (1988)
10. Whelen, N.: Sampling from Archimedean copulas. Quantitative Finance 4(3), 339–352 (2004)
11. Larranaga, P., Etxeberria, R., Lozano, J.A., Pena, J.M.: Optimization in continuous domains by learning and simulation of Gaussian networks. In: Proceedings of the GECCO 2000 Workshop in Optimization by Building and Using Probabilistic Models, pp. 201–204. Morgan Kaufmann, San Francisco (2000)