ESTIMATION OF DISTRIBUTION ALGORITHM WITH COPULA PROBABILISTIC MODEL: A SHORT INTRODUCTION

Martin Hyrš, Josef Schwarz

Brno University of Technology
Department of Computer Systems
Brno, Božetěchova 2
Czech Republic
ihyrs@fit.vutbr.cz, schwarz@fit.vutbr.cz

Abstract: *A new approach of probabilistic modeling used in Estimation of Distribution Algorithms (EDAs) based on Copula theory is described. By means of copulas it is possible to separate the structure of dependence from the marginal distributions in a joint distribution. Two dimensional Gaussian copula is depicted in more details including the sampling of the copula based on the conditional probabillity density function. The use of copulas for modeling joint distributions in EDAs is illustrated on several benchmarks.*

Keywords: *Estimation of distribution algorithms, Copula Theory, Sklar's theorem, Gaussian copula, optimization problems.*

1 Introduction

EDAs belong to the advanced evolutionary algorithms based on the estimation and sampling of graphical probabilistic models. They do not suffer from the disruption of building blocks known from the theory of standard genetic algorithms. The canonical sequential EDA is described in Fig. 1.

EDAs often surpass classical EAs in the number of required fitness function evaluations. However, the absolute execution time is still limiting factor which determines the size of practically tractable problems. Referring to Fig. 1 the most time consuming task is the estimation of probability model for many problems.

Set $t \leftarrow 0$;
Generate initial population $D(0)$ with N individuals;
While termination criteria is false **do**
begin
    Select a set $D_s(t)$ of $K < N$ promising individuals;
    Construct the probability model M from $D_s(t)$;
    Sample offspring $O(t)$ from M;
    Evaluate $O(t)$;
    Create $D(t+1)$ as a subset of $O(t) \cup D(t)$ with cardinality N;
    $t \leftarrow t + 1$;
end

Figure 1: The pseudo code of canonical EDA

The EDAs algorithms can be assorted according the complexity of the probability model. The simplest EDAs are UMDA algorithms [8], BMDA [7], MIMIC [2], and BOA [6], in the discrete domain. EDAs for the both discrete and continuous domains are described thoroughly in [3]. The main advantage of these algorithms is the capability to discover the variable linkage which results in successful solution of complex optimization problem. But there are two problems that which must be taken into consideration. The first problem is the need of the model complexity option and the relation to probability model overspecification. The second one is the time complexity of the probability model design.

The copula theory was utilizated so far in the financial and statistical areas [5, 1]. Only in few recent years the copula theory has been imported into the probabilty model of EDAs. Simply expressed copulas join the multivariate distribution functions to their univariate distribution functions that are uniform on the interval $[0; 1]$. Copula EDA algorithm therefore has the capability to reduce the execution time and the variable dependency can be modeled more exact.

The paper is organized as follows. In Section 2 basics of the copula theory is presented. The structure of Gaussian copula EDA is described in section 3. The eperimental results are shown in Section 4. The conclusions are provided in Section 5.

2 Copula theory

The copula concept was introduced 55 years ago by Sklar [9] to separate the effect of dependence from the effect of marginal distributions in a joint distribution. Copula is a function which joins the univariate distribution function and creates multivariate distribution functions. This approach allows to transform multivariate statistic problems on the univariate problems with the relation represented just by copula.

Definition. A copula C is a multivariate probability distribution function for which the marginal probability distribution of each variable is uniform in $[0;1]$.

A copula is a function $C : [0;1]^d \to [0;1]$ with the following properties [10]:

1. $C(u_1, u_2, \ldots, u_d) = 0$ for at least one $u_i = 0$
2. $C(1, 1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for all $i = 1, 2, \ldots, d$
3. $\forall (u_1, \ldots, u_d), (v_1, \ldots, v_d) \in \langle 0; 1\rangle^d, u_i \leq v_i : \sum_{(w_1,\ldots,w_d) \in \times_{i=1}^d \{u_i, v_i\}} (-1)^{|\{i: w_i = u_i\}|} C(w_1, \ldots, w_d) \geq 0$

Theorem. *Sklar's theorem [9]: Let F be a d-dimensional distribution function with margins $F_1, \ldots, F_d$. Then there exists a d-dimensional copula C such that for all $(x_1, \ldots, x_d) \in \mathbb{R}^d$ it holds that*

$$F(x_1, \ldots, x_d) = C\left(F_1(x_1), \ldots, F_d(x_d)\right) \tag{1}$$

If $F_1, \ldots, F_d$ are continuous, then C is unique. Conversely, if C is a d-dimensional copula and $F_1, \ldots, F_d$ are univariate distribution functions, then the function F defined via (1) is a d-dimensional distribution function.

Examples of bivariate copula functions can be seen in Fig. 2. M copula has form $M(u, v) = \min(u, v)$, W copula has form $W(u, v) = \max(u + v - 1, 0)$. These copulas are called Fréchet-Hoeffding bounds, for every copula $C(u, v)$ holds $W(u, v) \leq C(u, v) \leq M(u, v)$ [4].

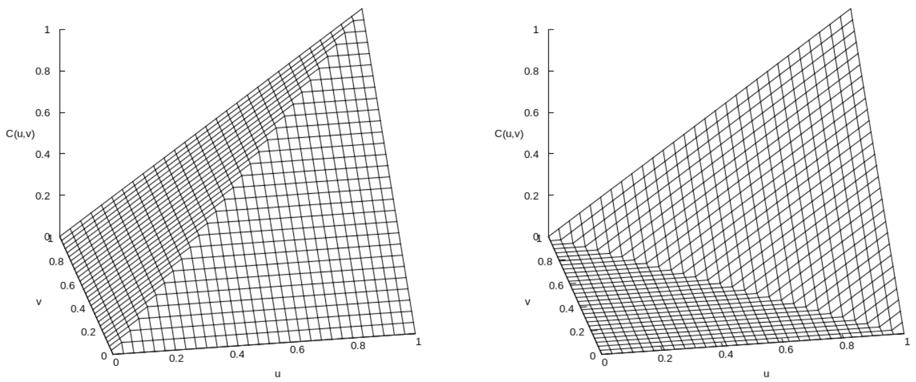

Figure 2: Examples of copula functions: M copula (left), W copula (right)

2.1 Gaussian copula

One type of copula is the Gaussian copula function, it is the member of elliptical copula family. Gaussian copula is the copula associated to the joint standard Gaussian distribution. We will deal with two-dimensional copula more in detail.

Definition. Bivariate Gaussian copula has the following functional form:

$$C_\rho(u, v) = \Phi_\rho\left(\Phi^{-1}(u), \Phi^{-1}(v)\right) \tag{2}$$

where $\Phi_\rho(x, y) = P(X < x, Y < y) = \int_{-\infty}^{x} \int_{-\infty}^{y} \frac{1}{2\pi\sqrt{1-\rho^2}} \mathrm{e}^{-\frac{s^2+t^2-2\rho st}{2(1-\rho^2)}} \,\mathrm{d}t\,\mathrm{d}s$ is the joint bivariate normal cumulative distribution with zero mean and with correlation coeficient ρ and $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}} \,\mathrm{d}t$ is cumulative distribution function of the standard normal distribution, $\Phi^{-1}(x)$ is it's quantile function.

3 Gaussian copula EDA

We used two-dimensional Gaussian copula for modeling dependencies and normal distribution as marginal distribution functions.

Similarly as EDA algorithm in Fig. 1 the core of Gaussian EDA includes the selection of promising individuals, identification of the probability model on the level of parameters and a mechanism of the model sampling.

3.1 Selection of promising solutions

For a set of promising solutions, we select K best solutions according to their fitness value. The cardinality K of $D_s(t)$ influences the level of selection pressure and the capability of the model. We used also elitism phenomenon in the phase of the new population creation.

3.2 Identification of copula probability model

Probability model is specified by two parts, parameters of marginal distribution functions and parameters of copula. These parameters are derived from the set $D_s(t)$ of selected promising solutions.

For parametrization of marginal distributions in each dimension i, we used mean value μ_i and standard deviation σ_i. The marginal univariate distribution function is expressed by

$$F_i(x_i) = \int_{-\infty}^{x_i} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu_i)^2}{2\sigma_i^2}} \,\mathrm{d}t = \Phi\left(\frac{x_i - \mu_i}{\sigma_i}\right) \tag{3}$$

The Gaussian copula function is parametrized by correlation matrix R, for two-dimensional copula the matrix has form $R = \begin{pmatrix} 1 & R_{1,2} \\ R_{2,1} & 1 \end{pmatrix}$. Because the matrix is symmetric, holds $R_{1,2} = R_{2,1} = \rho_S$, so the copula has one parameter ρ_S, which express correlation between two dimensions. For this correlation coeficient we used Spearman's correlation coefficient ρ_S.

The Spearman ρ_S is defined as the Pearson correlation coefficient between the ranked variables. For a sample of size n, the n original values x_i, y_i are converted to ranks p_i, q_i, and ρ_S is computed from these:

$$\rho_S = 1 - \frac{6\sum_i (p_i - q_i)^2}{n(n^2-1)} \tag{4}$$

3.3 Sampling offspring from copula

The main task is to obtain the copula sample $(u, v) \sim C$, then due to virtue of Sklar's theorem the new values of variables x_1, x_2 can be determined using inverse of marginal distribution

$$x_1 = F_1^{-1}(u) \qquad\qquad x_2 = F_2^{-1}(v) \tag{5}$$

In [4] is published general copula sampling methodology in low dimension based on conditional distribution. In this metodology, one variable is sampled independently, $u \sim U[0;1]$. It is computed conditional distribution $F_{V|U=u}(v) = \frac{\partial}{\partial u} C(u, v)$, the inverse of this distribution is $F^{-1}_{V|U=u}(\omega)$. Then is simulated independent value $\omega \sim U[0;1]$ and variable v is calculated using $v = F^{-1}_{V|U=u}(\omega)$.

Let's start with bivariate Gaussian copula distribution

$$C_\rho(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}}\, \mathrm{e}^{-\frac{s^2+t^2-2\rho st}{2(1-\rho^2)}} \,\mathrm{d}t\,\mathrm{d}s \tag{6}$$

Utilizing the known formula of fundamental theorem of calculus $\frac{\mathrm{d}}{\mathrm{d}x}\left(\int_c^{g(x)} f(t)\,\mathrm{d}t\right) = f(g(x))g'(x)$ leads to

$$\omega = \frac{\partial}{\partial u} C_\rho(u, v) = \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}}\, \mathrm{e}^{-\frac{\Phi^{-1}(u)^2+t^2-2\rho\Phi^{-1}(u)t}{2(1-\rho^2)}} \left(\Phi^{-1}\right)'(u)\,\mathrm{d}t \tag{7}$$

After some proper transformations:

$$\omega = \left(\Phi^{-1}\right)'(u) \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}}\, \mathrm{e}^{-\frac{\left(t-\rho\Phi^{-1}(u)\right)^2+\Phi^{-1}(u)^2-\rho^2\Phi^{-1}(u)^2}{2(1-\rho^2)}} \,\mathrm{d}t = \tag{8}$$

$$= \left(\Phi^{-1}\right)'(u)\,\mathrm{e}^{-\frac{\Phi^{-1}(u)^2}{2}} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}}\, \mathrm{e}^{-\frac{\left(t-\rho\Phi^{-1}(u)\right)^2}{2(1-\rho^2)}} \,\mathrm{d}t \tag{9}$$

The next step is calculating derivative $\left(\Phi^{-1}\right)'(u)$. For derivative of inverse function stands $\left(\Phi^{-1}\right)'(u) = \frac{1}{\Phi'(\Phi^{-1}(u))}$. Derivative Φ' can be done using fundamental theorem of calculus

$$\Phi'(x) = \frac{\mathrm{d}}{\mathrm{d}x} \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \, \mathrm{e}^{-\frac{t^2}{2}} \, \mathrm{d}t = \frac{1}{\sqrt{2\pi}} \, \mathrm{e}^{-\frac{x^2}{2}} \tag{10}$$

after substitution into equation for derivative of inverse

$$\left(\Phi^{-1}\right)'(u) = \frac{1}{\frac{1}{\sqrt{2\pi}} \, \mathrm{e}^{-\frac{\Phi^{-1}(u)^2}{2}}} = \frac{\sqrt{2\pi}}{\mathrm{e}^{-\frac{1}{2}\Phi^{-1}(u)^2}} \tag{11}$$

The last step is substitution $\left(\Phi^{-1}\right)'(u)$ into equation (9)

$$\omega = \sqrt{2\pi} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \, \mathrm{e}^{-\frac{\left(t-\rho\Phi^{-1}(u)\right)^2}{2(1-\rho^2)}} \, \mathrm{d}t = \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \, \mathrm{e}^{-\frac{\left(t-\rho\Phi^{-1}(u)\right)^2}{2(1-\rho^2)}} \, \mathrm{d}t \tag{12}$$

Now the equation (12) is similar to cumulative distribution function of the normal distribution $\int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \, \mathrm{d}t = \Phi\left(\frac{x-\mu}{\sigma}\right)$ with coeficients $\sigma = \sqrt{1-\rho^2}$ and $\mu = \rho\Phi^{-1}(u)$, let's rewrite the equation into the form

$$\omega = \Phi\left(\frac{\Phi^{-1}(v) - \rho\Phi^{-1}(u)}{\sqrt{1-\rho^2}}\right) \tag{13}$$

Using inverse operation we get v as

$$v = \Phi\left(\sqrt{1-\rho^2}\Phi^{-1}(\omega) + \rho\Phi^{-1}(u)\right) \tag{14}$$

Couple (u, v) obtained by this computation is the required copula sample. So generating the new offspring from copula probability model has these three steps:

1. Randomly generate variables $u \sim U[0;1]$ and $\omega \sim U[0;1]$.
2. Calculate variable $v = \Phi\left(\sqrt{1-\rho^2}\Phi^{-1}(\omega) + \rho\Phi^{-1}(u)\right)$.
3. Determine $x_1 = F_1^{-1}(u)$ and $x_2 = F_2^{-1}(v)$. In our case, as we use univariate marginal normal distribution, the previous equations have form $x_1 = \sigma_1\Phi^{-1}(u) + \mu_1, \quad x_2 = \sigma_2\Phi^{-1}(v) + \mu_2$.

4 Experimental results

Five well known benchmarks for optimization (finding of minimum extreme) according to [11] are used:

1. Shifted Elliptic Function:

$$f(\mathbf{z}) = \sum_{i=1}^{D} (10^6)^{\frac{i-1}{D-1}} z_i^2, \quad x_i \in [-100; 100] \tag{15}$$

2. Shifted Rastrigin's Function:

$$f(\mathbf{z}) = \sum_{i=1}^{D} \left(z_i^2 - 10\cos(2\pi z_i) + 10\right), \quad x_i \in [-5; 5] \tag{16}$$

3. Shifted Ackley's Function:

$$f(\mathbf{z}) = -20\,\mathrm{e}^{-0.2\sqrt{\frac{1}{D}\sum_{i=1}^{D} z_i^2}} - \mathrm{e}^{\frac{1}{D}\sum_{i=1}^{D}\cos(2\pi z_i)} + 20 + \mathrm{e}, \quad x_i \in [-32; 32] \tag{17}$$

4. Shifted Schwefel's Problem 1.2:

$$f(\mathbf{z}) = \sum_{i=1}^{D} \left(\sum_{j=1}^{i} z_i\right)^2, \quad x_i \in [-100; 100] \tag{18}$$

5. Shifted Rosenbrock's Function:

$$f(\mathbf{z}) = \sum_{i=1}^{D-1} \left(100(z_i^2 - z_{i+1})^2 + (z_i - 1)^2\right), \quad x_i \in [-100; 100] \tag{19}$$

where D is number of dimensions and $\mathbf{z}$ is shifted candidate solution $\mathbf{z} = \mathbf{x} - \mathbf{o}_{shift}$.

The minimum for all problems in global optimum $\mathbf{x}^*$ is $f(\mathbf{x}^*) = 0$. The reported results were obtained from 20 independent runs, with the following parameters:

- Population size: 500.
- Maximum number of fitness evaluations: 50,000.
- Selection: We used $K = 0.2N$, i.e. 100 individuals.
- Problem size: 2 variables for all problems.

After evaluation of whole population, when the total number of fitness evaluations exceeded 1000, 2000, etc., the fitness of best individuum was recorded. Average results are presented in Tables 1–5. The evolution was stopped after the maximum number of fitness evaluations was reached or when the fitness value of best individual reached the known optimum (with possible precision of used floating point number representation).

Table 1: Experiment results for Shifted Elliptic Function

fitness eval.	1000	2000	5000	10000	20000	after stopping
mean	3.14084E-04	6.20478E-06	3.70212E-13	—	—	4.68629E-17
std. dev.	2.14157E-04	9.69165E-06	5.64220E-13	—	—	3.44622E-17
min	9.67161E-06	1.97070E-07	7.42473E-15	—	—	1.12245E-18
max	8.57832E-04	4.25024E-05	2.57162E-12	—	—	1.25340E-16

Table 2: Experiment results for Shifted Rastrigin's Function

fitness eval.	1000	2000	5000	10000	20000	after stopping
mean	9.65771E-01	3.71404E-01	5.45110E-02	1.32083E-04	—	0.00000E+00
std. dev.	5.51695E-01	4.04482E-01	4.99525E-02	3.40353E-04	—	0.00000E+00
min	4.31849E-02	2.43028E-03	2.43028E-03	4.49827E-10	—	0.00000E+00
max	1.73662E+00	1.29294E+00	2.10099E-01	1.49188E-03	—	0.00000E+00

Table 3: Experiment results for Shifted Ackley's Function

fitness eval.	1000	2000	5000	10000	20000	after stopping
mean	1.17129E+00	8.89131E-02	1.24053E-05	4.83663E-10	9.99310E-11	7.25176E-11
std. dev.	6.80348E-01	4.49678E-02	8.68315E-06	5.08244E-10	2.75852E-10	2.79930E-10
min	2.12168E-01	5.92512E-03	6.02814E-07	9.26410E-11	8.49543E-13	7.18092E-13
max	2.54969E+00	1.82530E-01	3.40700E-05	1.91008E-09	1.29229E-09	1.29229E-09

Table 4: Experiment results for Shifted Schwefel's Problem 1.2

fitness eval.	1000	2000	5000	10000	20000	after stopping
mean	2.29068E-05	3.24285E-08	—	—	—	7.00368E-17
std. dev.	4.00942E-05	1.08713E-07	—	—	—	6.16074E-17
min	7.29431E-08	5.96988E-11	—	—	—	4.84636E-22
max	1.44324E-04	5.04142E-07	—	—	—	1.72475E-16

Table 5: Experiment results for Shifted Rosenbrock's Function

fitness eval.	1000	2000	5000	10000	20000	after stopping
mean	9.80988E+00	2.71917E+00	1.89456E-01	4.82955E-03	—	8.76696E-17
std. dev.	8.22324E+00	2.57893E+00	2.46643E-01	9.78643E-03	—	7.34051E-17
min	1.28253E-01	5.57618E-03	4.99794E-03	6.98564E-07	—	1.86577E-18
max	3.50877E+01	8.77529E+00	1.01694E+00	4.24769E-02	—	2.04121E-16

5 Conclusion

In this paper we have introduced the utilization of bivariate Gaussian copula as a variant of probability model in Estimation of Distribution Algorithm. We have presented the main theoretical grounds and the approach of constructing and sampling of the copula. Relatively much of effort was spent to derive the right approach of application of conditional distribution function to sample Gaussian copula. To illustrate the performance of Copula EDA algorithm a few known benchmarks of optimization were used. From the experimental results it is clear that the proposed allgorithm is effective. The next target of our research is the parallelization of algorithm and implementation of different univariate marginal distributions.

Acknowledgement: This work was supported by the Brno University of Technology project FIT-S-14-2297.

References

[1] Cherubini, U., Luciano, E., Vecchiato, W.: Copula Methods in Finance. The Wiley Finance Series. John Wiley & Sons Ltd (2004)

[2] De Bonet, J.S., Isbell, C.L., Viola, P.A.: Mimic: Finding optima by estimating probability densities. In: Advances in Neural Information Processing Systems, vol. 9, pp. 424–430. The MIT Press, Cambridge (1997)

[3] Larrañaga, P., Lozano, J.A.: Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation. Kluwer Academic Publishers, Norwell, MA, USA (2001)

[4] Mai, J., Scherer, M.: Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications, *Series in quantitative finance*, vol. 4. Imperial College Press (2012)

[5] Nelsen, R.B.: An Introduction to Copulas. Springer Series in Statistics. Springer New York (2006)

[6] Pelikan, M., Goldberg, D., Cantú-Paz, E.: Boa: The bayesian optimization algorithm. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-99), vol. I, pp. 525–532 also IlliGAL Report no. 99003 (1999)

[7] Pelikan, M., Mühlenbein, H.: The bivariate marginal distribution algorithm. In: Advances in Soft Computing, pp. 521–535. Springer London (1999)

[8] Pelikan, M., Mühlenbein, H.: Marginal distributions in evolutionary algorithms. In: In Proceedings of the International Conference on Genetic Algorithms Mendel 98, pp. 90–95 (1999)

[9] Sklar, A.: Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris **8**, 229–231 (1959)

[10] Sklar, A.: Random variables, joint distribution functions, and copulas. Kybernetika **9**(6), 449–460 (1973)

[11] Tang, K., Li, X., Suganthan, P.N., Yang, Z., Weise, T.: Benchmark Functions for the CEC'2010 Special Session and Competition on Large-Scale Global Optimization. Tech. rep., Nature Inspired Computation and Applications Laboratory, USTC, China (2009). URL http://nical.ustc.edu.cn/cec10ss.php