ELLIPTICAL AND ARCHIMEDEAN COPULAS IN ESTIMATION OF DISTRIBUTION ALGORITHM

Martin Hyrš, Josef Schwarz

Brno University of Technology
Faculty of Information Technology
Božetěchova 2, Brno, Czech Republic
ihyrs@fit.vutbr.cz, schwarz@fit.vutbr.cz

Abstract: *Estimation of distribution algorithm (EDA) is a variant of evolution algorithms, which is based on construction and sampling of probability model. Nowadays the copula theory is often utilized for the probability model estimation to simplify this process. We made comparison of two classes of copulas – elliptical and Archimedean ones – for the set of standard optimization benchmarks. The experimental results confirm our assumption that the elliptical copulas outperform the Archimedean ones namely in the case of the complex optimization problems.*

Keywords: *Estimation of distribution algorithms, copula theory, multivariate copula sampling, Clayton copula, Gumbel copula, Frank copula, Gaussian copula, Student t-copula.*

1 Introduction

Estimation of distribution algorithms (EDAs) belong to the advanced evolutionary algorithms. Instead of standard genetic algorithms, EDA does not use genetic operators (such as crossover and mutation), but it is based on the estimation and sampling probabilistic model. The important advantage of EDAs is their capability to represent an existing dependency among the variables in individuals using a joint probability distribution.

The EDA algorithms can be ordered according to the complexity of the probability model. The well-known EDAs are UMDA algorithms [11], BMDA [10], MIMIC [4], and BOA [9]. EDAs for the both discrete and continuous domains are described thoroughly in [5]. The main advantage of these algorithms is the capability to discover the variable linkage which results in a successful solution of the complex optimization problem.

Copulas are special probability distribution functions. Due to their properties it is possible to use them to modeling correlations in multivariate problems – the join distribution is separated to the univariate marginal distributions and to the correlation structure expressed by copula function. Different joint distribution functions can include the same marginal functions.

The copula theory has mostly been utilized in the financial and statistical areas [8, 3]. In only recent years the copula theory was imported into the probability model of EDAs [15, 12, 2, 16]. A copula EDA algorithm has the capability to reduce the execution time and the variable dependency can be modeled more precisely.

The paper is organized as follows. In Section 2, the basis of copula theory is given. In section 3, the utilization of copulas in EDA is described, and sampling algorithms for copulas are presented. Our experiments are discussed in Section 4. The conclusions are given in Section 5.

2 Copula theory

The copula concept was introduced in 1959 by Sklar [13, 14] in order to separate the effect of dependency of variables from the effect of marginal distributions in a joint distribution. A copula is a function which joins the univariate distribution function and creates multivariate distribution functions. This approach allows us to transform multivariate statistic problems on the univariate problems with the relation represented by just the copula.

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

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

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. $\forall (u_1,\dots,u_d),(v_1,\dots,v_d) \in [0;1]^d, u_i \leq v_i : \sum_{(w_1,\dots,w_d)\in \times_{i=1}^{d}\{u_i,v_i\}} (-1)^{|\{i|w_i \equiv u_i\}|} C(w_1,\dots,w_d) \geq 0$

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

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

If $F_1,\dots,F_d$ are continuous, then C is unique. Conversely, if C is a d-dimensional copula and $F_1,\dots,F_d$ 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. 1, $W(u,v) = \max(u+v-1,0)$, $\Pi(u,v) = uv$, $M(u,v) = \min(u,v)$. Π copula is independence copula. W, M copulas are called Fréchet-Hoeffding bounds, for every copula $C(u,v)$ holds $W(u,v) \leq C(u,v) \leq M(u,v)$ [6].

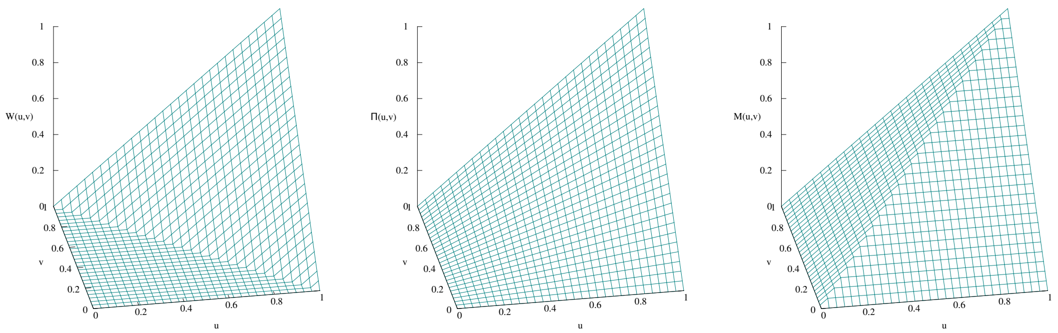

Figure 1: Examples of copula functions: W copula (left), Π copula (middle), M copula (right).

2.1 Archimedean copulas

A particular group of copulas is the Archimedean class. Archimedean copulas are very popular because they are easily derived and they are capable of capturing wide ranges of dependence. The definition of the Archimedean copula is based on the generator function:

Definition. A generator is a function φ: $[0;\infty) \to [0;1]$ with the following properties:

1. $\varphi(0) = 1$ and $\lim_{t\to\infty} \varphi(t) = 0$
2. $\varphi(t)$ is continuous
3. $\varphi(t)$ is decreasing on $[0;\infty)$ and strictly decreasing on $[0; \inf\{t > 0 : \varphi(t) = 0\})$, where $\inf(\emptyset) := 0$

Definition. D-variate Archimedean copulas take the form

$$C_\varphi = \varphi(\varphi^{-1}(u_1) + \varphi^{-1}(u_2) + \ldots + \varphi^{-1}(u_d)) \tag{2}$$

where $\varphi(t)$ is generator function and $\varphi^{-1}(t)$ denotes its inverse.

The reason of Archimedean copulas popularity in empirical applications is that Eq. 2 produces wide ranges of dependence properties for different choices of the generator function. Archimedean copulas are also relatively easy to estimate. There are many existing Archimedean copulas and many more that could be created.

The three Clayton, Gumbel and Frank appear regularly in statistics literature (see Table 1), they are popular because they model different patterns of dependence and have relatively simple functional form. Parameter θ in Table 1 represents the dependency strength. In Fig. 2 there is shown scatterplot of copulas using standard sampling process.

2.2 Elliptical copulas

The elliptical copulas differ from the Archimedean classes of copulas in the approach that only implicit analytical expression is available. These copulas are derived from the related elliptical distribution (e.g. normal distribution, Student t-distribution).

Table 1: Example of bivariate Archimedean copulas and their generators.

copula	$C(u,v)$	$\varphi(t)$	$\varphi^{-1}(t)$
Clayton	$C(u,v) = \left(\max\left\{u^{-\theta} + v^{-\theta} - 1; 0\right\}\right)^{-1/\theta}$	$\varphi(t) = (1+t)^{-1/\theta}$	$\varphi^{-1}(t) = t^{-\theta} - 1$
Gumbel	$C(u,v) = \mathrm{e}^{-\left((-\log(u))^{\theta} + (-\log(v))^{\theta}\right)^{1/\theta}}$	$\varphi(t) = \mathrm{e}^{\left(-t^{1/\theta}\right)}$	$\varphi^{-1}(t) = (-\log(t))^{\theta}$
Frank	$C(u,v) = \frac{-1}{\theta}\log\left(1 + \frac{(\mathrm{e}^{-\theta u}-1)(\mathrm{e}^{-\theta v}-1)}{\mathrm{e}^{-\theta}-1}\right)$	$\varphi(t) = \frac{-1}{\theta}\log(\mathrm{e}^{-t}(\mathrm{e}^{-\theta}-1)+1)$	$\varphi^{-1}(t) = -\log\left(\frac{\mathrm{e}^{-\theta t}-1}{\mathrm{e}^{-\theta}-1}\right)$

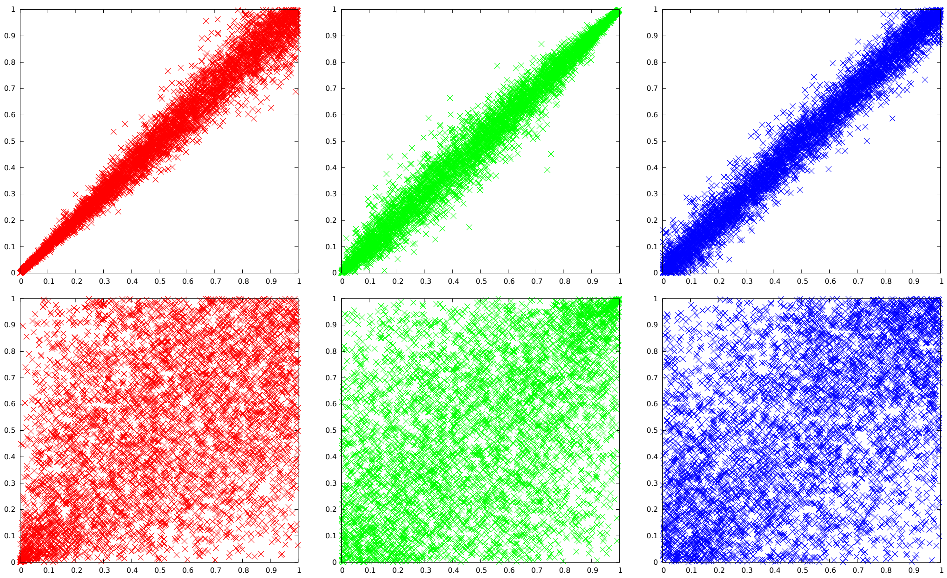

Figure 2: Scatterplots of bivariate Archimedean copulas: Clayton (left), Gumbel (middle) and Frank (right) with dependency strength 0.9 (top) and 0.3 (bottom).

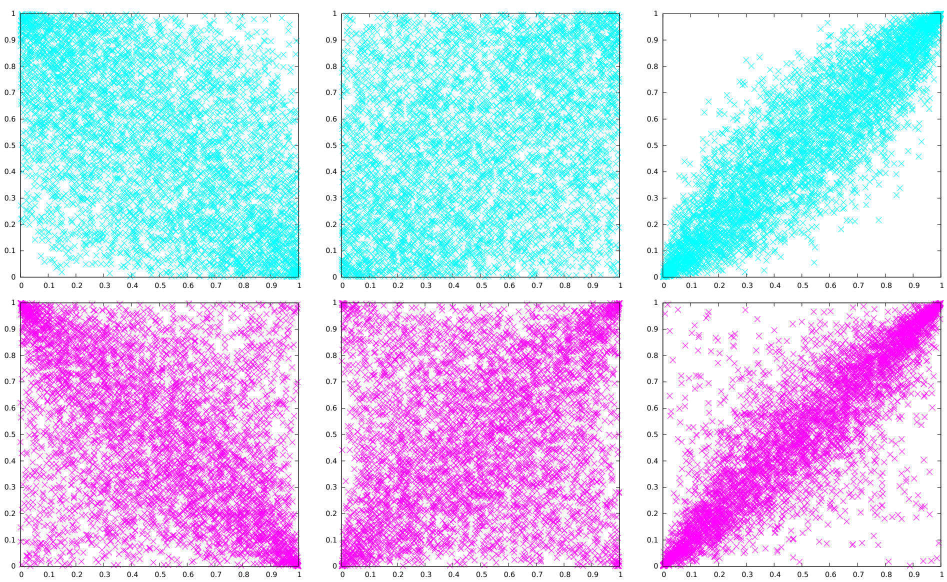

Figure 3: Scatterplots of bivariate eliptical copulas: Gaussian (top) and Student, $\nu = 2$ (bottom), with dependency strength -0.5 (left), 0.3 (middle) and 0.9 (right).

Definition. Elliptical copula has analytical form (which is obtained via Sklar's theorem):

$$C_R(u_1, \dots, u_d) = F_R\left(F^{-1}(u_1), \dots, F^{-1}(u_d)\right) \tag{3}$$

where $F_R(x_1, \dots, x_d)$ is elliptical distribution with correlation matrix R and F^{-1} are the quantile functions.

The matrix R with $\frac{1}{2}(d^2 - d)$ parameters in case of d-variate elliptic copula, expresses the dependency structure of the copula. Considering this fact, elliptic copulas generally have the ability to express more complex dependency structure.

The first example of elliptical copula is **Gaussian copula**, which belongs to normal distribution,

$$C_R^{Gauss}(u_1, \dots, u_d) = F_R\left(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)\right) \tag{4}$$

where F_R is the joint normal distribution and Φ^{-1} is the quantile function of the univariate standard normal distribution.

The second example is **Student copula**, which belongs to t-distribution

$$C_R^{Student}(u_1, \dots, u_d) = t_{\nu,R}\left(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d)\right) \tag{5}$$

where $t_{\nu,R}$ is the joint Student distribution and t_ν^{-1} is the quantile function of the univariate Student distribution with ν degrees of freedom (see Fig. 3).

3 Copula-based Estimation of Distribution Algorithm

Estimation of distribution algorithms belong to the advanced evolutionary algorithms. Solving the numerical optimization problem vector $\mathbf{x} = (x_1, \dots, x_d)$ of the optimal solution is searched out.

Canonical EDA consists of three main steps:

1. Selection subpopulation of promising solutions from the current population.
2. Creating probability model.
3. Sampling the probability model and generation of the new population.

The step 1 is quite straightforward, so we will focus on the step 2 and step 3. In case of copula-based EDA it is necessary to choose the proper type of copula and derive the copula parameters and the marginal distribution parameters.

The principle of sampling scheme for generating the new individuals is described in Algorithm 1:

Algorithm 1 Sampling the copulas and generating the new individuals

1. Obtain copula sample $(u_1, \dots, u_d) \sim C$, where $u_i \in [0; 1]$.
2. Derive the vector $\mathbf{x}$ of the searched solution using inverse marginal distributions, $x_i = F_i^{-1}(u_i)$.

3.1 Identification of copula probability model

The copula-based probability model includes two parts: univariate marginal distributions and copula function. Marginal distributions can be identificated separately for each variable and copula includes the correlation between variables.

For marginal distribution in each dimension $i = 1, \dots, d$ we used normal distribution, which is parameterized by mean value μ_i and standard deviation σ_i.

For assessing parameters of the copulas we used Kendall τ correlation coefficient. It is rank correlation coefficient.

In case of Archimedean copulas the following relations hold for parameter θ (in case of d-variate copulas, $d \geq 3$ we use average $\bar{\tau}$; for $d = 2$ standard pairwise τ is used):

- for Clayton $\theta_{Clayton} = \frac{2\tau}{1-\tau}$.
- For Gumbel $\theta_{Gumbel} = \frac{1}{1-\tau}$.
- for Frank (approximation) $\theta_{Frank} \doteq (10^{\arcsin(\tau)} - 1)\,\mathrm{e}$.

For elliptical copulas, we calculate pairwise coefficients ρ as elements of correlation matrix R using the Kendall τ correlation $\rho = \sin\frac{\pi\tau}{2}$.

3.2 Copula sampling algorithms

Now we specify step 1 from Alg. 1 for each type of copulas in more details.

Algorithm for sampling **Archimedean copula** uses random value J which is obtained from distribution given by inverse of Laplace transform $\mathcal{L}^{-1}$ of generator [6, 1, 7].

Algorithm 2 Archimedean copula sampling

1. generate value $J \sim \mathcal{L}^{-1}[\varphi(t)]$
2. generate uniformly distributed random numbers $r_i \sim U(0,1)$ (for $i = 1, \dots, d$)
3. return $u_i = \varphi\left(\frac{-\log(r_i)}{J}\right)$ (for $i = 1, \dots, d$)

According to [1, 7] the value of J can be derived:

- for Clayton copula by Gamma distribution $J \sim \mathit{Gamma}\left(\frac{1}{\theta}, \theta\right)$.
- for Gumbel copula by Levy skew alpha-Stable distribution $J \sim \mathit{Stable}\left(\frac{1}{\theta}, 1, \left(\cos\frac{\pi}{2\theta}\right)^{\theta}, 0\right)$.
- for Frank copula by logarithmic series distribution $J \sim \mathit{Logarithmic}\left(1 - \mathrm{e}^{-\theta}\right)$.

Sampling scheme for **Gaussian and Student copula** uses Cholesky decomposition of given correlation matrix R to obtain lower triangular matrix L, such that $LL^T = R$ [6]. Student copula is specified furthermore by degrees of freedom, we use $\nu = (N-1)d$ (where N is population size and d number of dimensions).

Algorithm 3 Gaussian copula sampling

1. compute L
2. generate random numbers $z_i \sim \mathit{No}(0,1)$ with standard normal distribution (for $i = 1, \dots, d$)
3. calculate $s_i = \sum_{j=1}^{i} L_{i,j} z_j$ (for $i = 1, \dots, d$)
4. return $u_i = \Phi(s_i)$ (for $i = 1, \dots, d$)

Algorithm 4 Student copula sampling

1. compute L
2. generate $V \sim \chi^2(\nu)$
3. generate random numbers $z_i \sim \mathit{No}(0,1)$ with standard normal distribution (for $i = 1, \dots, d$)
4. calculate $s_i = \sqrt{\frac{\nu}{V}} \sum_{j=1}^{i} L_{i,j} z_j$ (for $i = 1, \dots, d$)
5. return $u_i = t_\nu(s_i)$ (for $i = 1, \dots, d$)

Step 2 from Alg. 1 is straightforward, for each sample u_i of the used copula the components of solution vector $\mathbf{x}$ are calculated using inverse of marginal distributions, $x_i = F_i^{-1}(u_i)$.

4 Experiment and results

4.1 Benchmarks

Several well-known benchmarks from the area of numerical optimization are used. All these benchmarks are adapted to be minimization task with optimal fitness value 0.

- Elliptic Function: $f(\mathbf{x}) = \sum_{i=1}^{d} (10^6)^{\frac{i-1}{d-1}} x_i^2, \quad x_i \in [-100; 100]$
- Rastrigin's Function: $f(\mathbf{x}) = \sum_{i=1}^{d} \left(x_i^2 - 10\cos(2\pi x_i) + 10\right), \quad x_i \in [-5; 5]$
- Ackley's Function: $f(\mathbf{x}) = -20\,\mathrm{e}^{-0.2\sqrt{\frac{1}{d}\sum_{i=1}^{d} x_i^2}} - \mathrm{e}^{\frac{1}{d}\sum_{i=1}^{d}\cos(2\pi x_i)} + 20 + \mathrm{e}, \quad x_i \in [-32; 32]$

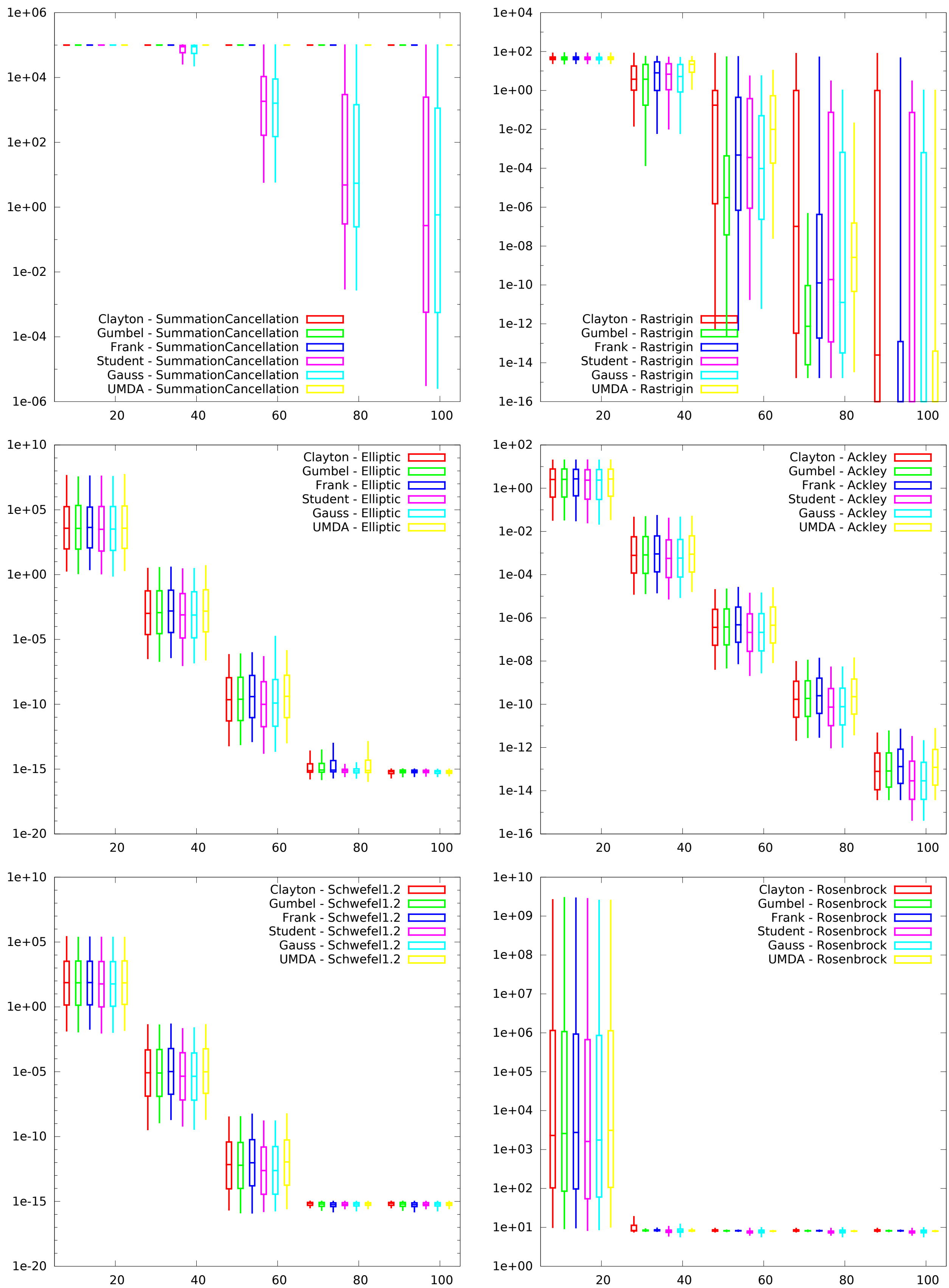

Figure 4: Experimental results – best fitness versus the count of generations for six benchmark problems (Summation Cancellation, Rastrigin, Elliptic, Ackley, Schwefel 1.2, Rosenbrock) in 10 dimensions. Each boxplot is presented after 20 generation epoche, the boxplot is calculated from 50 runs.

- Schwefel's Problem 1.2: $f(\mathbf{x}) = \sum_{i=1}^{d} \left(\sum_{j=1}^{i} x_i \right)^2, \quad x_i \in [-100; 100]$
- Rosenbrock's Function: $f(\mathbf{x}) = \sum_{i=1}^{d-1} \left(100(x_i^2 - x_{i+1})^2 + (x_i - 1)^2 \right), \quad x_i \in [-100; 100]$
- Summation Cancellation: $f(\mathbf{x}) = 10^5 - \frac{1}{10^{-5} + \sum_{i=1}^{d} |\sum_{j=1}^{i} x_j|}, \quad x_i \in [-1; 1]$

We have used shifted optima position, $fitness(\mathbf{x}) = f(\mathbf{x}')$, where $x_i' = x_i + \frac{i}{100N}, \quad i = 1, \ldots, d.$
The following settings is chosen:

- Problem size: 10 variables/dimensions for all problems (different number for additional experiments in case of Summation Cancellation problem).
- Population size: 500.
- Selection: We used truncation selection, with the rate 0.2, i.e. 100 individuals is chosen.
- Maximum number of generation: 100.
  Evolution is stopped if precision of best solution is better than 10^{-15}.
- Number of independent runs for each copula variant: 50.

4.2 Results and discussion

Experimental results can be seen in Fig. 4. We arranged comparison of copula EDA for five kinds of copulas and well known UMDA algorithm for six specified benchmarks. (UMDA is a simplified variant of EDA, it uses only univariate marginal distributions without dependency structure.)

From the Fig. 4 it follows that in case of Elliptic, Ackley, Schwefel 1.2 and Rosenbrock benchmarks the elliptic copulas (Gaussian and Student) are slightly better than Archimedean ones, but the difference is not very significant. It is remarkable that for these functions UMDA performs quite well, it has almost the same efficiency as the copula-based EDA.

In case of Rastrigin function the fitness function proves large variances, but every variant of copula-EDA algorithm was able to find optima in most of runs – see median position in the boxplot.

In Fig. 5 are shown experimental results for the case of Summation Cancellation problem with lower number of dimensions to increase the prospect for Archimedean copula. It can be seen that only elliptical copulas (Gaussian and Student) are able to find best solution. Archimedean copulas are unable to find optimal solution.

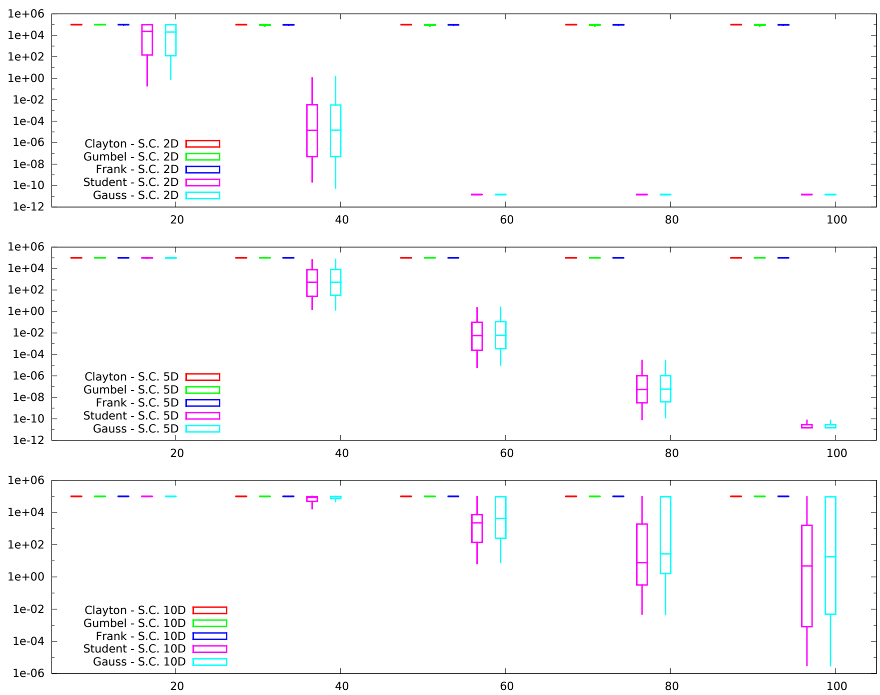

Figure 5: Experimental results – best individual fitness versus the count of generations for Summation Cancellation problem for 2, 5, 10 dimensions. Each boxplot is presented after 20 generation epoche, the boxplot is calculated from 50 runs.

5 Conclusion

In this paper, we have studied the utilization of five types of copulas in copula-based EDA. We have dealt with Archimedean copulas (Clayton, Gumbel and Frank) and with elliptical copulas (Gaussian and Student). We have presented the necessary theoretical bases and namely the proposed algorithms for effective sampling of each considered copula.

In order to compare the performance of copula-based EDAs based on different types of copulas, a few well-known problems of optimization in continuous domain were used. For the most benchmarks used, there is no significant difference in efficiency of copula-based EDA with applied copulas.

However in case of Summation Cancellation problem the only elliptical copulas are efficient, while one-parameter Archimedean copulas were not successful.

We have confirmed our assumption that elliptic copulas with their complex correlation matrix are more appropriate for complex optimization problems with stronger dependencies of variables.

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

References

[1] Aas, K.: Modelling the dependence structure of financial assets: A survey of four copulas (2004)

[2] Aas, K., Czado, C., Frigessi, A., Bakken, H.: Pair-copula constructions of multiple dependence. Insurance: Mathematics and Economics **44**(2), 182 – 198 (2009)

[3] Cherubini, U., Luciano, E., Vecchiato, W.: Copula Methods in Finance. John Wiley & Sons, Hoboken, NJ (2004)

[4] 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)

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

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

[7] Melchiori, M.R.: Tools for sampling multivariate archimedean copulas. YieldCurve, April (2006)

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

[9] 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)

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

[11] 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)

[12] Salinas-Gutiérrez, R., Hernández-Aguirre, A., Villa-Diharce, E.R.: Using copulas in estimation of distribution algorithms. In: MICAI 2009: Advances in Artificial Intelligence, *Lecture Notes in Computer Science*, vol. 5845, pp. 658–668. Springer Berlin Heidelberg (2009)

[13] 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)

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

[15] Soto, M., González-Fernández, Y., Ochoa, A.: Modeling with copulas and vines in estimation of distribution algorithms. CoRR **abs/1210.5500** (2012)

[16] Wang, L.F., Zeng, J.C., Hong, Y.: Estimation of distribution algorithm based on copula theory. In: Evolutionary Computation, 2009. CEC '09. IEEE Congress on, pp. 1057–1063 (2009)