An EDA Based on Bayesian Networks Constructed with Archimedean Copulas

Mario Rogelio Flores Méndez and Ricardo Landa
Information Technology Laboratory
CINVESTAV Tamaulipas
Cd. Victoria, Tamaulipas, MEXICO
Email: mflores@tamps.cinvestav.mx, rlanda@tamps.cinvestav.mx

Abstract—In this paper, an estimation of distribution algorithm that adopts a copula Bayesian network as probabilistic graphic model is presented. Multivariate Archimedean copula functions with one parameter are used to model the dependences between variables and the beta distribution is used to describe the univariate marginals. The learning process of the Bayesian network is assisted through a simple technique that relies on the associative property of Archimedean copulas, the use of Kendall's tau coefficient for measuring relations between variables and the relation between tau coefficients and bivariate Archimedean copulas. This paper presents the proposal, together with some initial experiments, which show encouraging results.

Keywords-numerical optimization; estimation of distribution algorithm; Archimedean copulas

I. Introduction

Optimization is a task present in many fields of human activity. Optimization problems range from very easy to solve to quite challenging problems, especially for classical methods. This is the case when metaheuristics are valuable.

A possible source for hardness of problems is when the number of decision variables increases, particularly if the interdependences between variables are strong.

Estimation of distribution algorithms (EDAs) are a special type of metaheuristics, which aim is to extract information about the probability distribution an evolutionary algorithm would follow when generating new solutions (sampling), building a probabilistic model. Additionally, the model can be constructed trying to capture the relations between variables, and thus taking advantage of the characteristics of the problem.

However, in the literature of state of the art, the reader can notice the predominant use of probabilistic graphical models in EDAs due to the good results they exhibit, but at the same time, exists a lack of them in EDAs that use copulas. Similarly, most of the work up to the date in EDAs that use copulas, only adopts one copula throughout the entire algorithm, which restricts the kind of dependences that can be described and exploited by the algorithm.

Such situations are the main motivations of this work. Here, we propose the use of a copula bayesian network, which is a probabilistic graphical model recently proposed and not yet implemented in numerical EDAs. However, results of their use on combinatorial EDAs are very promising.

Other type of networks are also good options for numerical optimization, such as Bayesian networks or Gaussian networks, so it is worth trying a new probabilistic model. On the other hand we employ three copulas from the Archimedean family to describe different types of dependences, with the aim of the algorithm to capture a wide range of relations between variables.

The rest of the paper is organized as follows: in Section II, the state of the art is summarized; in Section III, some concepts needed for the current proposal are enumerated; in Section IV, the proposed approach is described; in Section V are shown some preliminary results obtained by this proposal; finally, in Section VI, some conclusions are metioned, together with some work that remains to be done and future research ideas.

II. Previous Work

In this section a brief review of the state of the art is presented. Such techniques are grouped according to the approach for representing the search space and the complexity for representing the relations between variables.

A. Discretization of the Search Space

This type of approaches make divisions of each decision variable, basically by the use of histograms. In [1], [2], [3] the authors propose the use of fixed length or fixed height histograms. In [4] a technique for discretizing on demand is proposed, on models for combinatorial optimization.

B. Direct Representations

1) Without Dependences: Here, variables are supposed to be independent from one another, so the joint density function of vector $\vec{x} = [x_1, x_2, \ldots, x_n]$ of decision variables can be factorized as $p(\vec{x}) = \prod_{i=1}^{n} p(x_i)$. In [5] the authors kept a vector for the mean of each variable, while a single overall standard deviation is stored. Some extensions of this work include the use of standard deviations for each variable [6], or the use of intervals [7].

Moreover, in [8], [9] is proposed to perform statistical test for each variable at each generation, with the aim to find the density function that best fits each variable. After that, the parameters are estimated.

2) Bivariate Dependences: Relations are represented in these approaches, but only for pairs of variables. A continuous variable version of [10] is proposed in [8], [9], with two-variable Gaussian functions as model.

In [11], [12], the authors propose and approach with Gaussian copulas with a single parameter, obtained by maximum likelihood, and then the sampling of new individuals is performed by the Monte Carlo method. Clayton and Ali-Mikhail-Haq Archimedean copulas and conditional sampling are implemented in [12], [13]. Empirical marginal functions with fixed parameter copulas are proposed in [14] (with Gumbel copulas) and [15] (with Clayton copulas). In [16] the authors propose a dependency tree as graphical model (represented as a string) with bivariate copulas selected and estimated by maximum likelihood.

3) Multivariate Dependences: Here, relations among multiple variables are represented. For example, the algorithms [8] use Gaussian networks which parameters are obtained by maximum likelihood.

Another approach is that in [17], which can be adapted for continuous or combinatorial problems. The distribution is truncated with the value of the worst individual in previous generation, and just a fraction of the population is replaced. Mixed integer problems are also tackled in [18], where each probability distribution is modeled by a dependency tree, and each leaf represents a normal distribution.

In [12] the authors adopted empirical copulas, together with a method [19] for first building a copula, and then use it for sampling new individuals. Archimedean copulas are used in [20], with parameters fixed *a priori*, and a conditional sampling method. In [21] the authors propose the use of a regular vine as a graphical model, Gaussian copulas obtained by maximum likelihood and beta marginal distributions. A more recent approach is [22], with empirical marginals and Archimedean copulas (Gumbel or Clayton).

4) Mixed Distributions: Models with mixed density functions are also possible, aiming for more flexibility in the representation. For example, in [23] each variable is represented as a sum of several Gaussian distribution components. Another proposal [24] is to group solutions and then construct a Gaussian network for each group, also using mixed Gaussian distributions.

5) Adaptive Algorithms: Another promising option is to implement special mechanisms in order to select the different parts of the algorithm, from the type of graphical model or the type of copulas used, to even switch between different algorithms.

The algorithm proposed in [25] selects at each iteration the type of graphical model and the inference method associated with it, in order to approximate the distribution on variables and their dependences.

Selecting from different algorithms at running time fall in the area of hyper-heuristics. So, a hyper-heuristic is proposed in [26], where the selection is from a collection

Clayton	
Generator	$\varphi(u) = \frac{u^{-\theta}-1}{\theta}$
Inverse	$\varphi^{-1}(t) = (1+\theta t)^{-1/\theta}$
Function	$C(u_1, u_2, \ldots, u_n) = \left(\sum_i u_i^{-\theta} - n + 1 \right)^{-1/\theta}$
Parameter	$\theta > 0$
Kendall's tau	$\tau = \frac{\theta}{\theta+2}$
Frank	
Generator	$\varphi(u) = -\ln\left(\frac{e^{-\theta u}-1}{e^{-\theta}-1} \right)$
Inverse	$\varphi^{-1}(t) = -\frac{1}{\theta}\ln(1 + e^{-t}(e^{-\theta}-1))$
Function	$C(u_1, u_2, \ldots, u_n) = -\frac{1}{\theta}\ln\left(1 + \frac{\prod_{i=1}^{n}(e^{-\theta u_i}-1)}{(e^{-\theta}-1)^{n-1}} \right)$
Parameter	$\theta > 0$
Kendall's tau	$\tau = 1 - \frac{1}{\theta}$
Gumbel	
Generator	$\varphi(u) = (-\ln(u))^{\theta}$
Inverse	$\varphi^{-1}(t) = \exp\left(-t^{1/\theta}\right)$
Function	$C(u_1, u_2, \ldots, u_n) = \exp\left(-\left(\sum_i (-\ln u_i)^{\theta} \right)^{1/\theta} \right)$
Parameter	$\theta > 1$
Kendall's tau	$\tau = 1 - \frac{4}{\theta}(1 - D_1(\theta))$

Table I
THREE TYPES OF COPULAS ADOPTED

of different algorithms without dependences, with bivariate dependences, and with multivariate dependences.

III. RELEVANT CONCEPTS

In this section, some concepts related with the proposed work are presented.

A. Archimedean Copulas

Definition 1. *An* Archimedean copula *of n variables is the function* $C(u_1, u_2, \ldots, u_n) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2) + \ldots + \varphi(u_n))$, *where the so-called strict generator* $\varphi(\cdot)$ *is a function* $\varphi : I \to R_{>0}$, *continuous, decreasing and convex, such that* $\varphi(1) = 0$ *and* $\varphi(0) = +\infty$

Archimedean copulas are symmetric: $C(v, z) = C(z, v)$, as well as associative: $C(C(v, z), u) = C(v, C(z, u))$.

Moreover, they are related with Kendall's tau coefficient, as stated in [27]:

$$\tau = 4 \int_I \frac{\varphi(v)}{\varphi'(v)} dv + 1$$

Our proposal is to make use of such property, and use several, adaptively selected, copulas. The most representative copulas were selected: Clayton, Frank and Gumbel. The functions and other properties of such copulas are showed in Table I. The density function of the copulas has the form: $c = \varphi^{-(n)}(C(u_1, u_2, \ldots, u_n)) \prod_{i=1}^{n} \varphi'(u_i)$. The expressions for the i-th derivative of the inverse of the generator are taken from [28].

B. Copula Bayesian Network

As probabilistic graphic model, we adopted a copula Bayesian network. A probabilistic graphic model consists of a directed acyclic graph, which represents the conditional dependences of the variables; and a set of local generalized probability distributions, given the obtained relations between variables.

Definition 2. *A copula Bayesian network $C = (G, \Theta_C, \Theta_f)$ represents the joint density $f_X(x)$. G is a directed acyclic graph, Θ_C is a set of local density functions $c_i(F(x_i), \{F(\mathbf{pa}_{ik})\})$ associated with the nodes of G having at least one parent. Θ_f is the set of parameters representing the marginal densities $f(x_i)$. This way, $f_X(x)$ is parameterized as $f_X(x) = \prod_i R_{c_i}(F(x_i), \{F(\mathbf{pa}_{ik})\}) f(x_i)$.*

Further details can be found in [29].

IV. PROPOSED APPROACH

As mentioned above, we propose the use of a copula Bayesian network as probabilistic model, which allow us the adoption of multiple types of copulas. The method *score+search* is used for learning the structure. Such method is based on a greedy search and the BIC metric for estimating the fitness of the structure [30].

As proposed in [24], the construction of the structure starts with an empty graph, and then test edge by edge, selecting that which contributes with more information according to the metric adopted. Once the edge has been added, it is necessary to remove such edges that form cycles.

Some modifications are proposed here in order to speed up the construction of the structure.

- First, the use of the Kendall's tau coefficients to identify correlation between variables and, consequently, creating edges for new copulas. So, we need to keep a table with the different values of tau for any pair of variables. Since the structure only allows positive relations, only positive values of tau are considered.
  Addition of variables to an existing copula is also possible with this method, computing the mean of the tau values for each pair composed of a variable already present in such copula and a new candidate variable. Again, valid mean values are only positive.
  The BIC metric is used as in the original proposal, for measuring the increment in fitness of the structure.
- The parameter of the copula is commonly estimated by maximum likelihood. But an alternative, less expensive approach can be applied. Certainly, the tau coefficients cannot be used for obtaining the parameter of multivariate copulas in the exact expressions, but, thanks to the associative property of Archimedean copulas, it is possible to average the tau values of all the pairs of variables present in the copula. This is considerably less expensive in CPU time.

- Finally, the adaptive selection of one of the three types of copulas adopted. This can reflect with higher precision, the relations between the affected variables. The type of copula is selected by their likelihood at the time of creation, an then remains fixed by the rest of the process.

For estimating the parameters we adopted the approach proposed in [31], with some minor modifications. The estimation consists on two steps:

1) The estimation of the parameters of the univariate marginal distributions. Such distributions are only beta distributions, and the parameters are the mean μ and the variance σ^2.
   For a softer behavior throughout the process, the estimation is increasingly weighted by the values obtained from previous generations. A method for doing that was proposed in [6]:

$$\mu_{t+1} = (1 - \alpha)\mu_t + \alpha\overline{x}$$

$$\sigma_{t+1}^2 = (1 - \alpha)\sigma_t^2 + \alpha\left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})\right)$$

2) The estimation of the parameters of the copulas. This process is commonly performed by maximum likelihood, but we propose the use of the Kendall's tau, as described above.

The overall process of learning of the structure is depicted in Figure 1.

The sampling process is a standard conditional approach, for the multivariate case.

Given that each variable can belong to several copulas, is needed a heuristic in order to decide which copula will be used for sampling. The rules proposed here are:

- If the variable is a child in a copula, such copula is used for sampling.
- If the variable is not a child in any copula, but belongs to some copulas, the copula with already more sampled variables is used.
- If the variable does not belong to any copula, sampling is made at random.

V. PRELIMINARY RESULTS

A. Experimental Setup

Five of the problems found in [32] were selected. Such functions present diverse characteristics, and are frequently used in the specialized literature. Such problems are and their characteristics are:

1) Shifted elliptic problem: unimodal, separable.
2) Shifted Rastrigin problem: multimodal, separable.
3) Shifted Ackley problem: multimodal, separable.
4) Shifted 1.2 Schwefel problem: unimodal, non-separable.

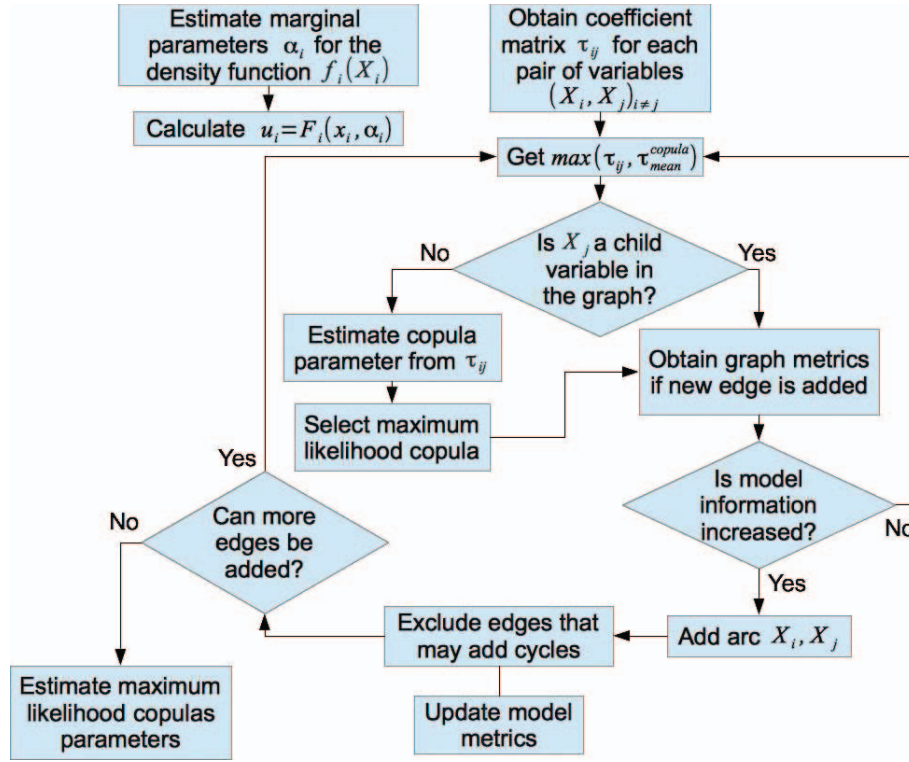Figure 1. Construction of graph structure

5) Shifted Rosenbrock problem: multimodal, non-separable.

The minimum for all problems is $f(X) = 0$. The reported results were obtained from 31 independent runs, with the following parameters:

- Population size: 100.
- Maximum number of generations: 1,000.
- Maximum number of function evaluations: 100,000.
- Maximum number of parents in a copula: 5.
- Selection: 30 individuals.
- Problem size: 10 variables for all problems.
- α for estimation of univariate distributions: from 0.01 to 0.1, with increments of 0.01.

B. Results

The results are shown in Table II. Experiments with higher dimensionalities are currently under way.

The obtained results show that the approach based on multivariate Archimedean copulas is promising, even competitive with other evolutionary algorithms not based on statistical modeling.

VI. CONCLUSIONS AND FUTURE WORK

This work proposes the use of several type of copulas, adaptively selected during the process. This can increase the precision of reflecting the nature of relations between variables in the problem at hand.

Additionally, the use of Archimedean copulas can provide means for decreasing the computational cost associated with the estimation of the parameters of multivariate copulas.

The use of an incremental rule for estimating the parameters of marginal distributions is very important in this approach, because it promotes exploration and maintains diversity, otherwise quickly lost during the process.

Some preliminary results are presented in this paper, with only 10 variables. Experiments with higher dimensionalities (up to 100) are currently under way. They require a considerable amount of time and space in the respective report, so they are planned for a future publication.

As a future work related with the estimation of marginals, is the self-adaptation of the α value, eliminating the necessity of fine tuning or dynamic rules.

Another future work is related with negative correlations, currently discarded because of the probabilistic model. However, the contribute important information which can be incorporated to the model.

REFERENCES

[1] S. Tsutsui, M. Pelikan, and D. E. Goldberg, "Evolutionary algorithm using marginal histogram models in continuous domain," in *in Continuous Domain, Proc. of the 2001 Genetic and Evolutionary Computation Conference Workshop Program*, 2001, pp. 230–233.

Problem	1	2	3	4	5
Mean	0.0060614539	2.387246119	0.0370533613	22.3287542871	10.4979051613
Std. dev.	0.0021301646	1.0597248887	0.006409029	22.8224320997	0.6866521235
Variance	0.0000045376	1.1230168397	0.000041076	520.8634069434	0.4714911387
Best	0.00264332	0.00743969	0.0154605	0.0865839	8.7676
Worst	0.00976279	4.00952	0.0494799	113.528	11.6509

Table II
OBTAINED RESULTS

[2] N. Ding, S. Zhou, and Z. Sun, "Optimizing continuous problems using estimation of distribution algorithm based on histogram model," in *Simulated Evolution and Learning*, ser. Lecture Notes in Computer Science, T.-D. Wang, X. Li, S.-H. Chen, X. Wang, H. Abbass, H. Iba, G.-L. Chen, and X. Yao, Eds. Springer Berlin / Heidelberg, 2006, vol. 4247, pp. 545–552.

[3] N. Ding, S.-D. Zhou, and Z.-Q. Sun, "Histogram-based estimation of distribution algorithm: A competent method for continuous optimization," *Journal of Computer Science and Technology*, vol. 23, no. 1, pp. 35–43, JAN 2008.

[4] C.-H. Chen, W.-N. Liu, and Y.-P. Chen, "Adaptive discretization for probabilistic model building genetic algorithms," in *Proceedings of the 8th annual conference on Genetic and evolutionary computation*, ser. GECCO '06. New York, NY, USA: ACM, 2006, pp. 1103–1110.

[5] S. Rudlof and M. Köppen, "Stochastic hill climbing with learning by vectors of normal distributions," in *Proceedings of the First Online Workshop on Soft Computinf (WSC1)*, 1997, pp. 60–70.

[6] M. Sebag and A. Ducoulombier, "Extending population-based incremental learning to continuous search spaces," in *Proceedings of the 5th International Conference on Parallel Problem Solving from Nature*, ser. PPSN V. London, UK: Springer-Verlag, 1998, pp. 418–427.

[7] I. Servet, L. Travé-Massuyès, and D. Stern, "Telephone network traffic overloading diagnosis and evolutionary computation techniques," in *Artificial Evolution*, ser. Lecture Notes in Computer Science, J.-K. Hao, E. Lutton, E. Ronald, M. Schoenauer, and D. Snyers, Eds. Springer Berlin / Heidelberg, 1998, vol. 1363, pp. 137–144, 10.1007/BFb0026596.

[8] P. Larrañaga, R. Etxeberria, J. A. Lozano, J. Peña, and J. M. Peña, "Optimization by learning and simulation of bayesian and gaussian networks," University of the Basque Country, Tech. Rep. EHU-KZAA-IK-4/99, December 1999.

[9] P. Larrañaga, R. Etxeberria, J. A. Lozano, and J. M. Peña, "Optimization in continuous domains by learning and simulation of gaussian networks," University of the Basque Country, Tech. Rep., 2000.

[10] J. S. D. Bonet, C. L. Isbell, and P. Viola, "Mimic: Finding optima by estimating probability densities," in *Advances in Neural Information Processing Systems*. The MIT Press, 1997, p. 424.

[11] L.-F. Wang, J.-C. Zeng, and Y. Hong, "Estimation of distribution algorithm based on copula theory," in *Evolutionary Computation, 2009. CEC '09. IEEE Congress on*, may 2009, pp. 1057 –1063.

[12] L.-F. Wang and J.-C. Zeng, "Estimation of distribution algorithm based on copula theory," in *Exploitation of Linkage Learning in Evolutionary Algorithms*, ser. Adaptation, Learning, and Optimization, Y.-p. Chen, L. M. Hiot, and Y. S. Ong, Eds. Springer Berlin Heidelberg, 2010, vol. 3, pp. 139–162.

[13] L.-F. Wang, J.-C. Zeng, and Y. Hong, "Estimation of distribution algorithm based on archimedean copulas," in *Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation*, ser. GEC '09. New York, NY, USA: ACM, 2009, pp. 993–996.

[14] L. Wang, X. Guo, J. Zeng, and Y. Hong, "Using gumbel copula and empirical marginal distribution in estimation of distribution algorithm," in *Advanced Computational Intelligence (IWACI), 2010 Third International Workshop on*, aug. 2010, pp. 583 –587.

[15] L. F. Wang, Y. C. Wang, J. C. Zeng, and Y. Hong, "An estimation of distribution algorithm based on clayton copula and empirical margins," in *Life System Modeling and Intelligent Computing*, ser. Communications in Computer and Information Science, K. Li, X. Li, S. Ma, and G. W. Irwin, Eds. Springer Berlin Heidelberg, 2010, vol. 98, pp. 82–88.

[16] R. Salinas-Gutiérrez, A. Hernández-Aguirre, and E. R. Villa-Diharce, "Dependence trees with copula selection for continuous estimation of distribution algorithms," in *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, ser. GECCO '11. New York, NY, USA: ACM, 2011, pp. 585–592.

[17] P. A. N. Bosman and D. Thierens, "Continuous iterated density estimation evolutionary algorithms within the IDEA framework," in *Workshop Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2000)*, 2000, pp. 197–200.

[18] J. Ocenasek, S. Kern, N. Hansen, and P. Koumoutsakos, "A mixed bayesian optimization algorithm with variance adaptation," in *Parallel Problem Solving from Nature - PPSN VIII*, ser. Lecture Notes in Computer Science, X. Yao, E. Burke, J. Lozano, J. Smith, J. Merelo-Guervós, J. Bullinaria, J. Rowe, P. Tino, A. Kabán, and H.-P. Schwefel, Eds. Springer Berlin / Heidelberg, 2004, vol. 3242, pp. 352–361.

[19] J. C. Strelen and F. Nassaj, "Analysis and generation of random vectors with copulas," in *Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come*, ser. WSC '07. Piscataway, NJ, USA: IEEE Press, 2007, pp. 488–496.

[20] A. Cuesta-Infante, R. Santana, J. Hidalgo, C. Bielza, and P. Larrañandaga, "Bivariate empirical and n-variate archimedean copulas in estimation of distribution algorithms," in *Evolutionary Computation (CEC), 2010 IEEE Congress on*, july 2010, pp. 1 –8.

[21] R. Salinas-Gutiérrez, A. Hernández-Aguirre, and E. R. Villa-Diharce, "D-vine eda: a new estimation of distribution algorithm based on regular vines," in *Proceedings of the 12th annual conference on Genetic and evolutionary computation*, ser. GECCO '10.   New York, NY, USA: ACM, 2010, pp. 359–366.

[22] X. Guo, L. Wang, J. Zeng, and X. Zhang, "Copula estimation of distribution algorithm with pmle," in *Natural Computation (ICNC), 2011 Seventh International Conference on*, vol. 2, july 2011, pp. 1077 –1081.

[23] M. R. Gallagher, "Multi-layer perceptron error surfaces: Visualization, structure and modelling models for iterative global optimization," Ph.D. dissertation, University of Queensland, Queensland, Australia, 2000.

[24] P. Bosman and D. Thierens, "Mixed ideas," Department of Information and Computing Sciences, Utrecht University, Tech. Rep. UU-CS-2000-45, 2000.

[25] R. Santana, P. Larrañaga, and J. Lozano, "Adaptive estimation of distribution algorithms," in *Adaptive and Multilevel Metaheuristics*, ser. Studies in Computational Intelligence, C. Cotta, M. Sevaux, and K. Sörensen, Eds.   Springer Berlin / Heidelberg, 2008, vol. 136, pp. 177–197.

[26] F. G. Lobo and C. F. Lima, "Towards automated selection of estimation of distribution algorithms," in *Proceedings of the 12th annual conference companion on Genetic and evolutionary computation*, ser. GECCO '10.   New York, NY, USA: ACM, 2010, pp. 1945–1952.

[27] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1/2, pp. 81–93, June 1938.

[28] F. Wu, E. Valdez, and M. Sherris, "Simulating from exchangeable archimedean copulas," *Communications in Statistics - Simulation and Computation*, vol. 36, no. 5, pp. 1019–1034, 2007.

[29] G. Elidan, "Copula bayesian networks," in *Advances in Neural Information Processing Systems 23*, J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, Eds. Curran Associates Inc, June 2010, pp. 559–567.

[30] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[31] H. Joe and J. Xu, "The estimation method of inference functions for margins for multivariate models." Department of Statistics, University of British Columbia, Tech. Rep. 166, 1996.

[32] K. Tang, X. Li, P. N. Suganthan, Z. Yang, , and T. Weise, "Benchmark functions for the cec 2010 special session and competition on large-scale global optimization," Tech. Rep., January 2010.