

# **Estimation of Distribution Algorithms: Basic Ideas and Future Directions**

**Ying-ping Chen**

NCLab Report No. NCL-TR-2010003

February 2010

Natural Computing Laboratory (NCLab)  
Department of Computer Science  
National Chiao Tung University  
329 Engineering Building C  
1001 Ta Hsueh Road  
HsinChu City 300, TAIWAN  
<http://nclab.tw/>

## ESTIMATION OF DISTRIBUTION ALGORITHMS: BASIC IDEAS AND FUTURE DIRECTIONS

YING-PING CHEN

*Department of Computer Science  
National Chiao Tung University  
HsinChu City 300, Taiwan  
ypchen@nclab.tw*

### ABSTRACT—

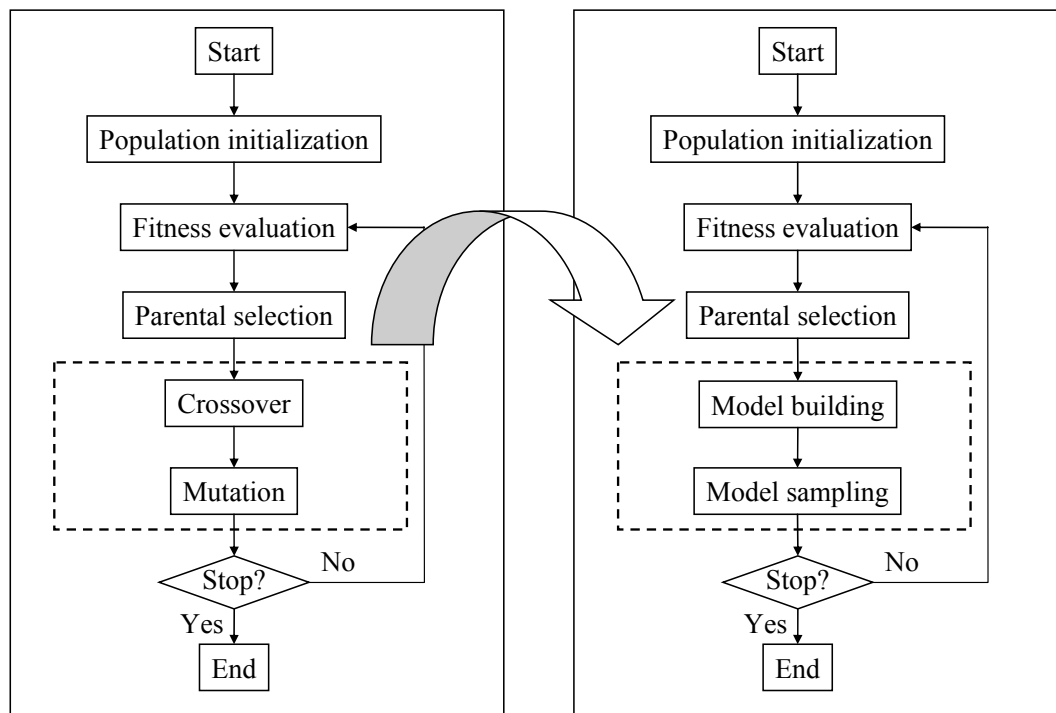
Estimation of distribution algorithms (EDAs) are a class of evolutionary algorithms which can be regarded as abstraction of genetic algorithms (GAs) because in the design of EDAs, the population, one of the GA distinctive features, is replaced by probabilistic models/distributions. Building and sampling from the models substitute for the common genetic operators, such as crossover and mutation. Due to their excellent optimization performance, EDAs have been intensively studied and extensively applied in recent years. In order to interest more people to join the research of EDAs, this paper plays as an entry level introduction to EDAs. It starts with introducing the origination and basic ideas of EDAs, followed by presenting the current EDA frameworks, which are broadly applied in many scientific and engineering disciplines. Finally, this paper also describes some ongoing topics and potential directions in the hope that readers may get further insights into EDAs.

Key Words: Estimation of distribution algorithm, probabilistic model, genetic algorithm, optimization technique, global optimization, evolutionary algorithm, evolutionary computation, computational intelligence.

## 1. INTRODUCTION

Genetic algorithms (GAs) were proposed by Holland [1] with the inspiration of Darwinian view on the evolutionary mechanisms in nature. They were initially designed for generating classifiers in learning classifier systems as well as handling combinatorial optimization problems. Brought to the attention of many researchers by Goldberg's book [2], genetic algorithms have been widely and successfully applied to solving all kinds of search and optimization problems existing in numerous disciplines for the past decades. The proposal of genetic algorithms is remarkably intriguing because it strongly connects several seemingly not-so-related fields, such as biology, mathematical programming (optimization), artificial intelligence, etc., places itself in a unique position among these fields to stir innovations, and makes a major contribution to the creation of evolutionary computation. Similar to the progress of most scientific and engineering development, soon after its birth, the GA taskforce splits and focuses on topics of different origins and requirements. Some researchers explore potential applications of GAs, while others try to improve GA performance by incorporating natural, biological mechanisms or by advancing algorithmic designs with mathematical techniques. Among these attempts to achieve better genetic algorithms, or more broadly, evolutionary algorithms, is the development of estimation of distribution algorithms.

By focusing on the optimization performance and discarding the biological plausibility, estimation of distribution algorithms (EDAs) successfully achieve their design goal can be viewed as abstraction of GAs because in EDAs, the population, which is one of the GA distinctive features, is replaced by certain mathematical construction, and genetic operators for generating offspring solutions are correspondingly changed to work with the adopted mathematical construction. According to the traditional GA performance indicator, function evaluations vs. solution quality, EDAs outperform GAs in most cases because the design of EDAs makes the search explicitly centralized by processing global statistics. Even if the significant computational cost of the mathematical construction is taken into consideration, the performance of EDAs is still usually superior to that of GAs. Thanks to their desirable features and properties, EDAs have been studied, improved, and broadly utilized for around fifteen years. Given the importance of EDAs in evolutionary computation and the usefulness of EDAs in application domains, an entry level introduction to EDAs is needed for those who are interested in getting familiar with and utilizing EDAs in a short time. As a consequence, this paper is written to fulfill this purpose. In particular, basic ideas, existing frameworks,



(a) Genetic Algorithm

(b) Estimation of Distribution Algorithm

**Figure. 1** Diagrams for a simple genetic algorithm and a basic estimation of distribution algorithm

and potential research directions of EDAs are briefly described. Note that this paper is not intended to be a complete survey or to provide details of EDAs. Interested readers may refer to other materials [3, 4].

This paper is organized as follows. Section 2 introduces the origination and basic ideas of EDAs, and section 3 presents existing EDA frameworks according to the adopted mathematical construction. Section 4 describes some of the recent research issues of EDAs as the future directions, followed by section 5 which summarizes and concludes this paper.

## 2. BASIC IDEAS AND ORIGINATION

In this section, we will start with revisiting genetic algorithms and presenting the basic ideas of estimation of distribution algorithms, followed by a brief history of estimation of distribution algorithms.

### 2.1 Genetic Algorithms

Genetic algorithms are a class of evolutionary algorithms developed for conducting search and optimization by mimicking the evolutionary process in biology. GAs, use a population of solutions, called *individuals*, to gather the information regarding the search space and to implicitly process the statistics [5] in order to find the optimal solutions. Figure 1(a) shows a genetic algorithm in its simplest form. In the beginning, a solution population is initialized by random generation. Each of the individuals is evaluated by the fitness function to indicate how well it “fit” the environment, i.e., the optimization problem at hand. The individuals with better fitness have better chances to reproduce their offspring, and the parental selection process implements the idea of natural selection on the procreation side. The process to create new individuals is designed by emulating the recombination (crossover) and alteration (mutation) of genetic materials. After the next generation of individuals is created, each individual is also evaluated by the fitness function, and the GA procedure repeats until certain stop criterion is satisfied. The operation can be considered as explicitly sampling the search space and implicitly exploiting the obtained information.

## 2.2 Probabilistic Models vs. Populations

As we can see in Figure 1(a), the components of which the functionality is implicitly processing and exploiting the information, i.e., the individuals, in a distributed manner are identified by a dashed box. Discarding the biological plausibility, one possible algorithmic way to improve GA performance is to make the implicit mechanism explicit. In order to achieve the explicit processing of the obtained information, probabilistic models are the chosen mathematical construction to “describe” populations. Since the process to find a probabilistic model for a given population is to estimate the probabilistic distributions on decision variables, such algorithms are called *estimation of distribution algorithms*, or sometimes, *probabilistic model building genetic algorithms* (PMBGAs). After gathering and mining the information existing in the form of individuals, the offspring individuals are then created by sampling the built probabilistic model to implement the process of information exploitation. Figure 1(b) shows an EDA in its simplest form. We can see between Figure 1(a) and Figure 1(b) that the key differences between GAs and EDAs are using probabilistic distributions to model populations and replacing genetic operators with the functionally equivalent mechanisms—probabilistic model building and sampling.

## 2.3 Estimation of Distribution Algorithms

There have been numerous variants of estimation of distribution algorithms proposed in the literature. Some of them adopt probabilistic models of different types or complexities, while others employ different techniques to build model. All these studies and developments on estimation of distribution algorithms started after the proposal of *population-based incremental learning* (PBIL) by Baluja [6] in 1994, while the name of “estimation of distribution algorithms” was firstly proposed by H. Mühlenbein and G. Paaß [7] in 1996. PBIL uses a probability vector to replace the population. Slightly different from most existing EDAs in which the probabilistic model is built from scratch at every generation as shown in Figure 1(b), PBIL retains some memory or experience of which the weight can be adjusted by the user. If the weight is set to zero, PBIL becomes a commonly structured EDA which is exactly the *univariate marginal distribution algorithm* (UMDA) proposed by Mühlenbein in [8] 1997. Early studies on EDAs began with simple probabilistic models of which the decision variables of optimization problems were assumed independent of each other. More and more complicated probabilistic models were used in the follow-up work along this line, which will be discussed in the following section.

## 3. EXISTING EDA FRAMEWORKS

This section will introduce some popular EDA frameworks proposed in the literature and widely used in both research and practice. Since probabilistic models are the key component in EDAs, we will introduce the EDAs employing simple models first and then those adopting complex models. Although the EDAs that adopt complex models usually provide excellent performance, one must keep in mind that complex models themselves may induce spurious variable relationships. If such spurious relationships become an obstacle which prevents the EDA from solving problems, EDAs with simpler models, i.e., more suitable for the problem structure, should be used to obtain better performance. Because the frameworks described in this section will be only a fraction of all existing EDAs, interested readers should consult other materials [3, 4].

### 3.1 All Variables Are Considered Independent

The simplest, reasonable probabilistic model to work with EDAs is assuming that no interaction exists between variables. EDAs employing such a model estimate the probabilistic distributions of values in different ways, including PBIL [6], UMDA [8], and the *compact genetic algorithm* (cGA) [9]. These EDAs work very well on problems composed of building blocks of order one and may encounter difficulties when facing problems consisting of longer, misleading building blocks.

### 3.2 Interactions between Two Variables Are Considered

In order to take into account the interactions between variables, probabilistic models considering pairwise interactions are intuitive choices. The *mutual information maximization for input clustering* (MIMIC) [10] algorithm assumes that the pairs of interacting variables are chained by their relationships,

while the *combining optimizers with mutual information trees* (COMIT) [11] algorithm models all the pairwise relationships with a dependency tree. The *bivariate marginal distribution algorithm* (BMDA) [12] further considers that all the pairwise relationships can be modeled with several independent dependency trees, i.e., a forest.

### **3.3 Interactions among More Than Two Variables Are Considered**

Finally, the probabilistic models considering multivariate dependencies are adopted in EDAs. As a rule of thumb, EDAs with more general, complicated probabilistic models are able to handle more difficult problems as long as the adopted models do not induce harmful spurious dependencies. The *extended compact genetic algorithm* (ECGA) clusters variables into separate linkage groups and considers the joint distribution for each group. With the help of human experts, the *factorized distribution algorithm* (FDA) utilizes a fixed model as the problem structure and provides excellent, theoretically proven performance. Adopting Bayesian networks as the probabilistic model, the Bayesian optimization algorithm (BOA) [13] and the estimation of Bayesian networks algorithm (EBNA) [14] uses different criteria to judge the quality of candidate Bayesian networks.

## **4. ISSUES AND FUTURE DIRECTIONS**

In this section, we will describe several important research issues and potential future directions of EDAs. Because the design of EDAs is based on the properties and characteristics of probabilistic models, knowing the intrinsically embedded limitations and reducing the computational cost are no doubt essential. Moreover, obtaining information by examining the built models and hybridizing EDAs with techniques of other origins are promising research directions. Please note that the materials included in this section are far from complete. Many other topics worth pursuing are available in the recent literature.

### **4.1 Can Models Be Misleading Or Always Partially Meaningful?**

Since probabilistic models are used in EDAs as tools for optimization, an obvious question rises: Is it possible that we build an appropriate probabilistic model according to a given population, while the built model leads us away from the optimal solution? This question is about the intrinsic properties of the problems that we want to solve by using EDAs. If some problems upon which the probabilistic model built correctly is actually misleading, EDAs, no matter what kinds of probabilistic models are adopted, will not be able to handle these problems. Coffin and Smith [15, 16] investigated whether the parity functions are such deal breakers. Furthermore, Chen and Yu [17] theorized the difficulty of probabilistic model building with mathematical formalization and obtained certain theoretical results. Another question regarding problem intrinsic properties is: Is it possible that, for certain problems, the built model is always partially meaningful? Chuang and Chen [18, 19] demonstrated that the problems composed of disparate importance weights might render EDAs building partially correct models at any time. In addition to proposing the concepts of *linkage sensibility* and *effective distributions*, they provided a technique to work with ECGA.

### **4.2 Can Models Be Built More Easily?**

The main computational cost of EDAs is apparently caused by building probabilistic models. Research along this line is always active and important. To know EDAs better, Chen et al. [20] analyzed the average time complexity of EDAs. Techniques that can build models more efficiently were proposed by Ding et al. [21], Echegoyen et al. [22], and Iclănzan et al. [23]. For BMDA, probability model migration [24] and aggregation [25] were proposed to be used in a parallel configuration. For BOA, in order to reduce the model building cost, previously built Bayesian networks, were utilized to predict next network structures [26, 27] or were viewed as a prototype for incremental changes [28].

### **4.3 Can Models Provide Useful Information?**

After building and using the probabilistic models, it seems wasteful to put the models aside. As a consequence, looking into the built probabilistic models to collect useful information is worth trying. In addition to getting information for help building the subsequent models as aforementioned [26, 27], Santana et al. [29] tried to conduct data mining on the built probabilistic models, and Echegoyen et al. [30] investigated the interaction as well as relationship between the optimization problem and the probabilistic model via analyzing the probability to the optimal solutions.

#### 4.4 Can EDAs Be Hybridized with Other Techniques?

A common feature of evolutionary algorithms is their flexibility to work or to interface with all kinds of methods from other realms. EDAs are no exception. In order to enhance EDAs for different purposes, a host of mechanisms, methodologies, and frameworks have been integrated, including niching [31], adaptive variance scaling [32], Spearman's rank correlation index [33], particle swarm optimization [34], etc.

### 5. CONCLUSIONS

In this paper, estimation of distribution algorithms (EDAs) as a popular class of evolutionary algorithms have been reviewed. EDAs can be regarded as abstraction of genetic algorithms (GAs) because in EDAs, the population, one of the GA distinctive features, is replaced by probabilistic models, and the common genetic operators, e.g., crossover, mutation, etc., are replaced by building and sampling from the adopted probabilistic model. By pursuing optimization performance instead of insisting on biological plausibility, EDAs successfully accomplish their design goal and become more and more popular in recent years. This paper was written with the intention to provide an entry level introduction to EDAs for researchers and practitioners who are in need and interested in knowing and using EDAs in a short time. Basic ideas, existing frameworks, and potential research directions of EDAs were briefly described in the hope that more and more taskforces will join the research as well as applications of EDAs.

### ACKNOWLEDGMENTS

The work was supported in part by the National Science Council of Taiwan under Grant NSC-98-2221-E-009-072. The author is grateful to the National Center for High-performance Computing for computer time and facilities.

### REFERENCES

1. J.H. Holland, *Adaptation in natural and artificial systems*, University of Michigan Press, 1975.
2. D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley Publishing Co., 1989.
3. P. Larrañaga and J.A. Lozano, eds., *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation*, Kluwer Academic Publishers, 2002.
4. M. Pelikan, et al., "A Survey of Optimization by Building and Using Probabilistic Models," *Computational Optimization and Applications*, vol. 21, 2002, pp. 5-20.
5. D.E. Goldberg, et al., "Genetic algorithms, noise, and the sizing of populations," *Complex Systems*, vol. 6, no. 4, 1992, pp. 333-362.
6. S. Baluja, *Population-based incremental learning: A method for integrating genetic search based function optimization and competitive learning*, Tech. Rep. No. CMU-CS-94-163, Carnegie Mellon University, Pittsburgh, PA, 1994.
7. H. Mühlenbein and G. Paaß, "From recombination of genes to the estimation of distributions I. Binary parameters," *Proceedings of the Fourth Parallel Problem Solving from Nature (PPSN IV)*, 1996, pp. 178-187.
8. H. Mühlenbein, "The Equation for Response to Selection and Its Use for Prediction," *Evolutionary Computation*, vol. 5, no. 3, 1997, pp. 303-346.
9. G.R. Harik, et al., "The compact genetic algorithm," *Proceedings of the International Conference on Evolutionary Computation (ICEC'98)*, 1998, pp. 523-528.
10. J.S. De Bonet, et al., "MIMIC: Finding optima by estimating probability densities," *Advances in Neural Information Processing Systems*, vol. 9, 1997, pp. 424.
11. S. Baluja and S. Davies, "Using optimal dependency-trees for combinatorial optimization: Learning the structure of the search space," *Proceedings of the 14th International Conference on Machine Learning*, 1997, pp. 30-38.
12. M. Pelikan and H. Mühlenbein, "The bivariate marginal distribution algorithm," *Advances in Soft Computing—Engineering Design and Manufacturing*, 1999, pp. 521-535.
13. M. Pelikan, et al., "BOA: The Bayesian optimization algorithm," *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-99)*, 1999, pp. 525-532.
14. R. Etxeberria and P. Larrañaga, "Global optimization with Bayesian networks," *Proceedings of the Second Symposium on Artificial Intelligence (CIMAF-99)*, 1999, pp. 332-339.

15. D.J. Coffin and R.E. Smith, "The Limitations of Distribution Sampling for Linkage Learning," *Proceedings of 2007 IEEE Congress on Evolutionary Computation (CEC 2007)*, 2007, pp. 364-369.
16. D.J. Coffin and R.E. Smith, "Why Is Parity Hard for Estimation of Distribution Algorithms," *Proceedings of ACM SIGEVO Genetic and Evolutionary Computation Conference 2007 (GECCO-2007)*, 2007, pp. 624.
17. S.-C. Chen and T.-L. Yu, "Difficulty of Linkage Learning in Estimation of Distribution Algorithms," *Proceedings of ACM SIGEVO Genetic and Evolutionary Computation Conference 2009 (GECCO-2009)*, 2009, pp. 397-404.
18. C.-Y. Chuang and Y.-p. Chen, "On the Effectiveness of Distributions Estimated by Probabilistic Model Building," *Proceedings of ACM SIGEVO Genetic and Evolutionary Computation Conference 2008 (GECCO-2008)*, 2008, pp. 391-398.
19. C.-Y. Chuang and Y.-p. Chen, "Sensibility of Linkage Information and Effectiveness of Estimated Distributions," *Evolutionary Computation*, 2010. (Accepted).
20. T. Chen, et al., "On the Analysis of Average Time Complexity of Estimation of Distribution Algorithms," *Proceedings of 2007 IEEE Congress on Evolutionary Computation (CEC 2007)*, 2007, pp. 453-460.
21. N. Ding, et al., "Reducing Computational Complexity of Estimating Multivariate Histogram-Based Probabilistic Model," *Proceedings of 2007 IEEE Congress on Evolutionary Computation (CEC 2007)*, 2007, pp. 111-118.
22. C. Echegoyen, et al., "Exact Bayesian Network Learning in Estimation of Distribution Algorithms," *Proceedings of 2007 IEEE Congress on Evolutionary Computation (CEC 2007)*, 2007, pp. 1051-1058.
23. D. Iclănzan, et al., "Correlation Guided Model Building," *Proceedings of ACM SIGEVO Genetic and Evolutionary Computation Conference 2009 (GECCO-2009)*, 2009, pp. 421-428.
24. J. Jaroš and J. Schwarz, "Parallel BMDA with Probability Model Migration," *Proceedings of 2007 IEEE Congress on Evolutionary Computation (CEC 2007)*, 2007, pp. 1059-1066.
25. J. Jaroš and J. Schwarz, "Parallel BMDA with an Aggregation of Probability Models," *Proceedings of 2009 IEEE Congress on Evolutionary Computation (CEC 2009)*, 2009, pp. 1683-1690.
26. M. Hauschild, et al., "Using Previous Models to Bias Structural Learning in the Hierarchical BOA," *Proceedings of ACM SIGEVO Genetic and Evolutionary Computation Conference 2008 (GECCO-2008)*, 2008, pp. 415-422.
27. M.W. Hauschild and M. Pelikan, "Intelligent Bias of Network Structures in the Hierarchical BOA," *Proceedings of ACM SIGEVO Genetic and Evolutionary Computation Conference 2009 (GECCO-2009)*, 2009, pp. 413-420.
28. M. Pelikan, et al., "iBOA: The Incremental Bayesian Optimization Algorithm," *Proceedings of ACM SIGEVO Genetic and Evolutionary Computation Conference 2008 (GECCO-2008)*, 2008, pp. 455-462.
29. R. Santana, et al., "Mining Probabilistic Models Learned by EDAs in the Optimization of Multi-objective Problems," *Proceedings of ACM SIGEVO Genetic and Evolutionary Computation Conference 2009 (GECCO-2009)*, 2009, pp. 445-452.
30. C. Echegoyen, et al., "Analyzing the Probability of the Optimum in EDAs Based on Bayesian Networks," *Proceedings of 2009 IEEE Congress on Evolutionary Computation (CEC 2009)*, 2009, pp. 1652-1659.
31. H. Handa, "Estimation of Distribution Algorithms with Niche Separation Mechanism," *Proceedings of 2007 IEEE Congress on Evolutionary Computation (CEC 2007)*, 2007, pp. 110-126.
32. P.A.N. Bosman and D. Thierens, "Adaptive Variance Scaling in Continuous Multi-Objective Estimation-of-Distribution Algorithms," *Proceedings of ACM SIGEVO Genetic and Evolutionary Computation Conference 2007 (GECCO-2007)*, 2007, pp. 500-507.
33. A.H. Aguirre, et al., "An Estimation Distribution Algorithm with the Spearman's Rank Correlation Index," *Proceedings of ACM SIGEVO Genetic and Evolutionary Computation Conference 2008 (GECCO-2008)*, 2008, pp. 469-470.
34. C.W. Ahn and H.-T. Kim, "Estimation of Particle Swarm Distribution Algorithms: Bringing Together the Strengths of PSO and EDAs," *Proceedings of ACM SIGEVO Genetic and Evolutionary Computation Conference 2009 (GECCO-2009)*, 2009, pp. 1817-1818.