

# Estimation-of-Distribution Algorithms: Theory and Applications

Josu Ceberio Uribe<sup>\*1</sup>, Benjamin Doerr<sup>\*2</sup>, Carsten Witt<sup>\*3</sup>, and Vicente P. Soloviev<sup>†4</sup>

1 University of the Basque Country – Donostia, ES. josu.ceberio@ehu.eus

2 Ecole Polytechnique – Palaiseau, FR. doerr@lix.polytechnique.fr

3 Technical University of Denmark – Lyngby, DK. cawi@dtu.dk

4 Polytechnic University of Madrid, ES. vicente.perez.soloviev@alumnos.upm.es

---

## Abstract

The Dagstuhl seminar 22182 *Estimation-of-Distribution Algorithms: Theory and Practice* on May 2–6, 2022 brought together 19 international experts in estimation-of-distribution algorithms (EDAs). Their research ranged from a theoretical perspective, e.g., runtime analysis on synthetic problems, to an applied perspective, e.g., solutions of industrial optimization problems with EDAs. This report documents the program and the outcomes of the seminar.

**Seminar** May 1–6, 2022 – <http://www.dagstuhl.de/22182>

**2012 ACM Subject Classification** Computing methodologies → Search methodologies; Theory of computation → Design and analysis of algorithms

**Keywords and phrases** estimation-of-distribution algorithms, heuristic search and optimization, machine learning, probabilistic model building

**Digital Object Identifier** 10.4230/DagRep.12.5.17

## 1 Executive Summary

*Josu Ceberio Uribe (University of the Basque Country – Donostia, ES)*

*Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR)*

*Carsten Witt (Technical University of Denmark – Lyngby, DK)*

**License**  Creative Commons BY 4.0 International license  
© Josu Ceberio Uribe, Benjamin Doerr, and Carsten Witt

The seminar “Estimation-of-Distribution Algorithms: Theory and Practice” on May 2–6, 2022 brought together 19 international experts in estimation-of-distribution algorithms (EDAs). Their research ranged from a theoretical perspective, e.g., runtime analysis on synthetic problems, to an applied perspective, e.g., solutions of industrial optimization problems with EDAs.

The main aim of the seminar was to narrow the gap between theory and practice in EDAs by bringing together researchers from both sides and stimulating interaction. We facilitated this interaction through longer introductory talks, e.g., on theoretical analyses of EDAs, regular conference-style talks, short flash talks presenting open problems and stimulating discussions and, last but not least, through various breakout sessions and group work. After each talk, we scheduled ample time for discussion, and even adapted the schedule when

---

\* Editor / Organizer

† Editorial Assistant / Collector



discussions had gained momentum and took longer than expected. On the last day of the seminar, all participants joined a 1-hour plenum discussion summarizing the findings of the seminar, discussing open problems and identifying further research topics.

We believe that the seminar has achieved its main aims by making a step towards narrowing the gap between theory and practice in EDAs. This is witnessed by the high number of spontaneous talks given by the participants, allowing almost everyone to present his/her perspective on EDAs, high and stable attendance of participants at talks and group sessions, and lively discussions after basically every talk as well as in the dedicated discussion fora. Several participants shared the feedback with us that they learned new aspects of EDAs during the seminar and had increased their understanding of what the other side (theory/practice) was interested in.

The seminar also identified several open problems related to the design, analysis, and application of EDAs. Details can be found in the summary of the concluding plenum discussion (see further below in the report).

Finally, several participants reported to the organizers that the seminar had helped them to understand current challenges in the theory and practice of EDAs, that they had extended their professional network, and drawn inspiration from the seminar for new research ideas. They also expressed interest in attending a similar seminar in the future.

The only downside of this seminar was that several colleagues, most notably from China, could not attend because of the still ongoing sanitary crisis.

**2 Table of Contents**

**Executive Summary**

*Josu Ceberio Uribe, Benjamin Doerr, and Carsten Witt* . . . . . 17

**Overview of Talks**

An application of EDAs on a random difference equation of tumour growth  
*Carlos Andreu Vilarroig and Josu Ceberio Uribe* . . . . . 21

Comparing EDAs with quantum or quantum-inspired solvers  
*Mayowa Ayodele* . . . . . 21

The question of structure in Markov Network EDAs  
*Alexander Brownlee* . . . . . 22

General univariate EDAs  
*Benjamin Doerr* . . . . . 23

Genetic drift and a smart restart scheme  
*Benjamin Doerr* . . . . . 23

A Gentle Introduction to Information-Geometric Optimization (IGO)  
*Nikolaus Hansen* . . . . . 23

Generating permutations  
*Ekhine Irurozki* . . . . . 24

Universal permutation aggregation  
*Ekhine Irurozki* . . . . . 24

Theoretical Analyses of EDAs – Introduction and Overview  
*Martin S. Krejca* . . . . . 25

Do univariate EDAs cope well with deception/epistasis?  
*Per Kristian Lehre* . . . . . 25

EDAs and more noise  
*Johannes Lengler* . . . . . 25

Where and How Does the Brain Use EDAs?  
*Johannes Lengler* . . . . . 25

Expensive black-box optimization problem in permutation spaces  
*Manuel López-Ibáñez* . . . . . 26

EDAs, epistasis, and structure  
*John McCall* . . . . . 26

EDA applications in real industrial problems  
*Vicente P. Soloviev* . . . . . 28

Semi-parametric Bayesian networks in EDAs  
*Vicente P. Soloviev* . . . . . 28

A bit about algorithm configuration  
*Thomas Stützle* . . . . . 29

Different parameter regimes in the cGA and their effect on performance  
*Dirk Sudholt* . . . . . 29

EDAs, noise and graceful scaling <i>Andrew M. Sutton</i> . . . . .	30
<b>Working groups</b>	
Breakout session: Dealing with constraint search spaces <i>Josu Ceberio Uribe and Ekhine Irurozki</i> . . . . .	30
Breakout session: Univariate EDAs, deception and epistasis <i>Per Kristian Lehre and Johannes Lengler</i> . . . . .	31
Breakout session: Explainability <i>John McCall and Josu Ceberio Uribe</i> . . . . .	31
Breakout session: Do we need to model interactions in EDAs <i>Jonathan L. Shapiro</i> . . . . .	32
Breakout session: Genetic drift <i>Dirk Sudholt and Johannes Lengler</i> . . . . .	32
<b>Panel discussions</b>	
Concluding group discussion of the seminar <i>Josu Ceberio Uribe, Benjamin Doerr, Per Kristian Lehre, and Carsten Witt</i> . . . .	33
<b>Participants</b> . . . . .	36

### 3 Overview of Talks

#### 3.1 An application of EDAs on a random difference equation of tumour growth

Carlos Andreu Vilarroig (Technical University of Valencia, ES) and Josu Ceberio Uribe (University of the Basque Country – Donostia, ES)

License © Creative Commons BY 4.0 International license  
© Carlos Andreu Vilarroig and Josu Ceberio Uribe

In this contribution, a discussion of two possible implementations of an EDA on a tumour growth model has been presented. This model describes the size of a tumour  $X_t$  in each time instant  $t$  through a simple difference equation  $X_{t+1} = (1 + k)X_t$  – with two parameters: the growth rate  $k$  and the initial condition (size)  $X_0$  [1]. Since the parameters have been considered as Gaussian random variables, the aim of the problem is to find the best probability distributions of the parameters that capture the uncertainty of the real tumour size data [2] using an EDA. A first, more classical, approach is to estimate by EDA the means and variances parameters of the probability distributions. In contrast, in a second, more alternative approach, the final distributions of the EDA are the parameter distributions themselves. On this second case, a possible implementation has been developed, in which the cost function is the RMSE between the expected value of the difference equation and the data and, as a stopping criterion to avoid convergence of the variance of the distributions to 0, the algorithm stops at the iteration in which the data starts to fall outside the 95% confidence interval of the model solution.

#### References

- 1 Kathleen MC Tjørve and Even Tjørve. The use of gompertz models in growth analyses, and new gompertz model approach: An addition to the unified-richards family. *PLoS one*, 12(6):e0178691, 2017.
- 2 Andrea Worschech, Nanhai Chen, Yong A Yu, Qian Zhang, Zoltan Pos, Stephanie Weibel, Viktoria Raab, Marianna Sabatino, Alessandro Monaco, Hui Liu, Vladia Monsurro, R Mark Buller, David F Stroncek, Ena Wang, Aladar A Szalay, and Francesco M Marincola. Systemic treatment of xenografts with vaccinia virus GLV-1h68 reveals the immunologic facet of oncolytic therapy. *BMC Genomics*, 10(1):301, 2009.

#### 3.2 Comparing EDAs with quantum or quantum-inspired solvers

Mayowa Ayodele (Fujitsu Research of Europe – Slough, GB)

License © Creative Commons BY 4.0 International license  
© Mayowa Ayodele

According to the observation made by Gordon Moore that the number of transistors in a dense integrated circuit doubles about every two years [1], it has been projected that general purpose computers may struggle to cope with computational demand in the future. There has therefore been research interests in the use of other specialised hardware such as quantum computers. Within the context of solving optimisation problems, quantum computers and quantum-inspired hardware have been developed. Quadratic Unconstrained Binary Optimisation (QUBO) is a common formulation used by quantum and quantum-inspired solvers. Examples of quantum or quantum-inspired QUBO solvers are D-wave's

Advantage released in 2020, which uses quantum hardware and has the capability of 5,000+ quantum bits (qubits) [2], IBM’s Eagle which is being extended from a 27-qubit solver to 127-qubit [3] and Fujitsu’s Digital Annealer (DA) which uses application specific CMOS to find optimum or sub-optimum solution to problems formulated with up to 100,000 bits [4]. The DA has advantages over other quantum-based algorithms such as scale (up to 100,000 bits), precision (64bit gradations), stability (stable operation at room temperature with digital circuits) and connectivity (easy to use with total bit coupling) [4].

Estimation of Distribution Algorithms (EDAs) have been applied to many problem classes including QUBOs. Quantum-inspired EDA (QiEDA) [5] have also been designed for and applied to problems formulated as QUBO e.g., TSP [6] and feature selection [7]. It is however unclear how these algorithms compare to other QUBO solvers such as the DA. The need to design EDAs that are not problem specific and that take advantage of specialised hardware is therefore motivated.

### References

- 1 Schaller, R.R., 1997. *Moore’s law: past, present and future*. IEEE spectrum, 34(6), pp.52-59.
- 2 McGeoch, C. and Farré, P., 2020. *The D-wave advantage system: An overview*. D-Wave Systems Inc., Burnaby, BC, Canada, Tech. Rep.
- 3 Chow, J., Dial, O. and Gambetta, J., 2021. *IBM Quantum breaks the 100-qubit processor barrier*. IBM Research Blog, available in <https://research.ibm.com/blog/127-qubit-quantum-process-or-eagle>.
- 4 Nakayama, H., Koyama, J., Yoneoka, N. and Miyazawa, T., 2021 *Third Generation Digital Annealer Technology*. Fujitsu, available in [https://www.fujitsu.com/jp/documents/digitalannealer/researcharticles\DA\\_WP\\_EN\\_20210922.pdf](https://www.fujitsu.com/jp/documents/digitalannealer/researcharticles\DA_WP_EN_20210922.pdf)
- 5 Chen, M. and Quan, H., 2007, September. *Quantum-inspired evolutionary algorithm based on estimation of distribution*. In 2007 Second International Conference on Bio-Inspired Computing: Theories and Applications (pp. 17-19). IEEE.
- 6 Soloviev, V.P., Bielza, C. and Larranaga, P., 2021, June. *Quantum-Inspired Estimation Of Distribution Algorithm To Solve The Travelling Salesman Problem*. In 2021 IEEE Congress on Evolutionary Computation (CEC) (pp. 416-425). IEEE.
- 7 Soliman, O.S. and Rassem, A., 2012, December. *Correlation based feature selection using quantum bio inspired estimation of distribution algorithm*. In International Workshop on Multi-disciplinary Trends in Artificial Intelligence (pp. 318-329). Springer, Berlin, Heidelberg.

### 3.3 The question of structure in Markov Network EDAs

Alexander Brownlee (*University of Stirling, GB*)

License © Creative Commons BY 4.0 International license  
© Alexander Brownlee

Joint work of Alexander E. I. Brownlee, John A. W. McCall, Qingfu Zhang, Lee A. Christie, Martin Pelikan, D. Brown

Main reference Alexander E. I. Brownlee, John A. W. McCall, Qingfu Zhang: “Fitness Modeling With Markov Networks”, IEEE Trans. Evol. Comput., Vol. 17(6), pp. 862–879, 2013.

URL <https://doi.org/10.1109/TEVC.2013.2281538>

This talk is on the long-running theme of “structure” in EDAs. I will introduce Markov network EDAs, which use an undirected probabilistic graphical model relating binary decision variables to the fitness distribution. This model’s accuracy and, in turn, algorithm efficiency, depend on the neighbourhood structures that are captured within it. That is, does the model treat variables independently, or are bivariate or higher order interactions considered. I will

briefly recap some early results looking at the impact of problem and algorithm structure in this context, covering well-known benchmarks including Onemax, Trap, and MAX3SAT, before looking at a systematic study covering 2 and 3 bit problems which shows that not all structure is essential to problem solution.

### 3.4 General univariate EDAs

*Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR)*

**License** © Creative Commons BY 4.0 International license  
© Benjamin Doerr

**Joint work of** Benjamin Doerr, Marc Dufay

We propose a general model for univariate EDAs that encompasses the four classic EDAs. We prove a genetic drift result for the general model that implies the corresponding known results for the particular algorithms. Finally, we show that the general class of algorithm contains EDAs which are more powerful than the existing ones (with optimal parameters).

### 3.5 Genetic drift and a smart restart scheme

*Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR)*

**License** © Creative Commons BY 4.0 International license  
© Benjamin Doerr

**Joint work of** Benjamin Doerr, Weijie Zheng

**Main reference** Benjamin Doerr, Weijie Zheng: “Sharp Bounds for Genetic Drift in Estimation of Distribution Algorithms”, *IEEE Trans. Evol. Comput.*, Vol. 24(6), pp. 1140–1149, 2020.

**URL** <https://doi.org/10.1109/TEVC.2020.2987361>

Genetic drift, that is, the random movement of sampling frequencies not justified by a fitness signal, usually leads to an inferior performance of EDAs. In this talk, we quantify how the parameters of the four common univariate EDAs lead to genetic drift. This allows to set the parameters in a way that during the desired runtime, the genetic drift is low enough to avoid the undesired effects. We also describe how to use this quantification to define a smart-restart mechanism that, provably, leads to a performance asymptotically equal to the one obtainable from optimal parameter values.

### 3.6 A Gentle Introduction to Information-Geometric Optimization (IGO)

*Nikolaus Hansen (INRIA Saclay – Palaiseau, FR)*

**License** © Creative Commons BY 4.0 International license  
© Nikolaus Hansen

**Main reference** Yann Ollivier, Ludovic Arnold, Anne Auger, Nikolaus Hansen: “Information-Geometric Optimization Algorithms: A Unifying Picture via Invariance Principles”, *J. Mach. Learn. Res.*, Vol. 18, pp. 18:1–18:65, 2017.

**URL** <https://dl.acm.org/doi/10.5555/3122009.3122027>

With invariance as major design principle, we present a canonical way to turn any smooth parametric family of probability distributions (on an arbitrary search space) into a continuous-time black-box optimization method and into an “IGO algorithm” through time discretization. The construction is based upon (i) the natural gradient over the family of distributions to

express (intrinsic) update directions and (ii) the cumulative distribution function of the  $f$ -values under the current sample distribution to express (invariant) preference weights. The construction recovers well-known algorithms like cGA, PBIL or Natural Evolution Strategies and major aspects of CMA-ES.

### 3.7 Generating permutations

*Ekhine Irurozki (Telecom Paris, FR)*

License  Creative Commons BY 4.0 International license  
© Ekhine Irurozki

In this tutorial we show some known and new results in statistical properties and methods for permutations

- Uniform sampling : in this section we show how to generate permutations uniformly (where all the possible permutations have the same probability of being generated) among all those at a given distance for the identity.
- Distributions on permutations: We show the different and most common distribution for permutation data, motivation, main statistics and inference algorithms.
- Depth function: we show a typical function for real-valued data (its novel adaptation for permutation spaces)
- Experiments: in order to contribute we present an easy to use code with the above mentioned methods and operations.

### 3.8 Universal permutation aggregation

*Ekhine Irurozki (Telecom Paris, FR)*

License  Creative Commons BY 4.0 International license  
© Ekhine Irurozki

The problem of permutation-aggregation (PA) is the main optimization problem in statistical community. It is equivalent to the median computation in the real-valued vector spaces. It can be posed under different distances and the complexity varies for these cases: it is known to be NP-hard for some distances, polynomial for others and unknown (although conjectured to be NP-hard for other)

This talk introduces a current work on the problem of aggregating permutations. We introduce an aggregation algorithm that is valid for several distance. This is an approximated algorithm for PA for which we can characterize the conditions for convergence. Moreover, we bound the expected error and the expected run-time. We also introduce simulations that confirm the theoretical results.

### 3.9 Theoretical Analyses of EDAs – Introduction and Overview

*Martin S. Krejca (Sorbonne University – Paris, FR)*

License  Creative Commons BY 4.0 International license  
© Martin S. Krejca

The theoretical analysis of estimation-of-distribution algorithms (EDAs) has gained a lot of momentum during the last decade. In this talk, I introduce the basic concepts and terminology used in the analysis of EDAs, and I summarize most of the latest results. The aim is to provide the audience with a solid foundation for understanding theoretical analyses of EDAs and to give it a good overview about the state of the art.

### 3.10 Do univariate EDAs cope well with deception/epistasis?

*Per Kristian Lehre (University of Birmingham, GB)*

License  Creative Commons BY 4.0 International license  
© Per Kristian Lehre

Recently, there have been mixed theoretical and empirical evidence as to whether univariate estimation of distribution algorithms (EDAs) such as the UMDA perform worse than classical EAs on multimodal/deceptive problems. We discuss empirical investigations of a range of EAs and EDAs on NK-landscapes, as well as theoretical analyses of UMDA end EAs on the Deceptive Leading Ones problem. These results point to the limitation of univariate EDAs under certain parameter regimes.

### 3.11 EDAs and more noise

*Johannes Lengler (ETH Zürich, CH)*

License  Creative Commons BY 4.0 International license  
© Johannes Lengler

This talk picks up the topic of Andrew Sutton's previous discussion of EDAs on noise. I discussed several instances in which Evolutionary Algorithms (EAs) struggle in noisy or dynamic environments. All examples have the fact at their heart that the mutation generator of EAs is intrinsically biased. In contrast, EDAs are unbiased (except for the boundaries, which are exceptional for EDAs). This is the reason why EDAs are often much more robust to noise than EAs.

### 3.12 Where and How Does the Brain Use EDAs?

*Johannes Lengler (ETH Zürich, CH)*

License  Creative Commons BY 4.0 International license  
© Johannes Lengler

In this flash talk, I will present some recent interpretations of low-level conscious processing that resemble the way EDAs operate. Unconscious perceptions have local inconsistencies that can be interpreted as a probability distribution. A conscious perception is then simply a sample from this distribution. Episodic memory seems to be based (or to consist) purely on these samples, while the probabilistic model is also updated by unconscious perceptions.

### 3.13 Expensive black-box optimization problem in permutation spaces

Manuel López-Ibáñez (*University of Málaga, ES*)

License  Creative Commons BY 4.0 International license  
© Manuel López-Ibáñez

Although there is a large amount of research in black-box optimization problems in the continuous domain, there is far fewer algorithms designed for permutation spaces, where the input points in the decision space represent a ranking or an ordering, in particular, when the number of evaluations is extremely limited, e.g., to a number of function evaluations between 100 and 1000. We describe two alternative approaches: CEGO, which is a Bayesian (i.e., surrogate-based) optimiser using Kriging (i.e., Gaussian Processes) regression and a distance metric between permutations, and UMM, an Estimation-of-Distribution algorithm based on the Mallows model. We discuss results on classical combinatorial problems and a real-world-like benchmark inspired by an application to Space exploration and on the effect of starting from a very good initial solution versus completely random ones. We conclude by discussing that there is still room for improvement in existing Bayesian and EDA optimisers in this particular context.

#### References

- 1 Manuel López-Ibáñez, Francisco Chicano, and Rodrigo Gil-Merino. *The Asteroid Routing Problem: A Benchmark for Expensive Black-Box Permutation Optimization*. In J. L. Jiménez Laredo et al., editors, *Applications of Evolutionary Computation*, volume 13224 of *Lecture Notes in Computer Science*. Springer Nature, Switzerland, 2022.
- 2 Ekhine Irurozki and Manuel López-Ibáñez. *Unbalanced Mallows Models for Optimizing Expensive Black-Box Permutation Problems*. In *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2021*. ACM Press, New York, NY, 2021.

### 3.14 EDAs, epistasis, and structure

John McCall (*The Robert Gordon University – Aberdeen, GB*)

License  Creative Commons BY 4.0 International license  
© John McCall

An EDA constructs a probabilistic model that estimates the distribution of good solutions / fitness. Typically the model is a probabilistic graphical model (PGM) composed of structure components with weights. Bayesian and Markov networks are two widely used PGMs [1, 2]. EDAs select good quality solutions from the current population, use them to estimate the topology and parameters of a PGM and then sample the PGM to create a successor population.

That part of the modelling process that relates to learning the topology of the PGM is known as structure learning. Structure components relate directly to subsets of variables (graph components of the PGM) and can be thought of as indicating patterns that make a contribution to fitness depending on the values of the variables. This is strongly related to the concept of epistasis which means strong interactions in determining individual fitness between separated genes. Repeated runs of an EDA may learn different structure and different structure elements may be present in models generated at different stages of a single run. A mainstream outcome of EDA research was to develop a series of EDAs for a wide range of problems that used PGMs with increasingly complex structure to tackle

challenging problems with high degrees of epistasis and with a range of representations e.g. MAXSAT, Ising problems, NK landscapes, continuous benchmark sets, QAP, PFSP, TSP, etc. EDAs that use PGMs are classified as univariate, bivariate or multivariate according to the structural complexity of their models.

A key observation is that fitness may be thought of as a mass distribution and so by normalising, or other means, we can create “true” probabilistic graphical models of the fitness distribution, examine their structure and explore the relationship between this structure and the structure estimated by EDAs. While it is not always possible to determine all of the true structure of a fitness function, structure discovery algorithms exist – e.g. Heckendorn and Wright [3]. It is possible at reasonable cost in evaluations to “probe” for structure of a certain complexity by evaluating carefully selected solutions. One can ask for a given problem, what structure does an EDA need to discover to optimise efficiently?

A folklore intuition exists that, on optimisation problems with a complex true structure, algorithms that can detect and model complex structure will exhibit superior performance to algorithms that cannot do so. Many papers exist showing multivariate algorithms outperforming univariate or bivariate algorithms on benchmarks. In the talk, we present a real world cancer chemotherapy optimisation problem with a high degree of bivariate structure on which univariate EDAs, despite being incapable of detecting this structure, outperform hBOA, one of the most powerful and widely applied multivariate EDAs [4].

The result is important because a drawback of multivariate EDAs is that the computational cost of model building can quickly become significant compared to overall function evaluation costs as structural complexity increases. An understanding of when structure learning is unnecessary to, or even impedes, optimisation may lead to important insights for the design of EDAs that use PGMs. A key observation is that most PGM EDAs operate invariantly to order-preserving function transformations, that is to say they are invariant across welldefined function classes. In this talk we share some results from [5] where we algebraically characterise inessential structure for such EDAs on pseudo-boolean problems.

## References

- 1 Larrañaga, P. and Lozano, J.A. 2002. Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation. Kluwer Academic Publishers, Boston.
- 2 Chen, Y. 2010. Exploitation of Linkage Learning in Evolutionary Algorithms, Springer
- 3 Heckendorn, R.B. and Wright, A.H. 2004. Efficient Linkage Discovery by Limited Probing. *Evolutionary Computation* 12, 517-545.
- 4 Brownlee, Alexander EI; Pelikan, Martin; McCall, John AW; Petrovski, Andrei, 2008, An application of a multivariate estimation of distribution algorithm to cancer chemotherapy, Proceedings of the 10th annual conference on Genetic and evolutionary computation (GECCO), pp463-464. Full technical paper at [arxiv.2205.08438](https://arxiv.org/abs/2205.08438) .
- 5 Christie, L. A., McCall, J. A. W. and Lonie, D. P., 2014. Minimal walsh structure and ordinal linkage of monotonicity-invariant function classes on bit strings. In: Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation (GECCO '14). 12-16 July 2014. New York: ACM. pp. 333-340.

### 3.15 EDA applications in real industrial problems

Vicente P. Soloviev (*Polytechnic University of Madrid, ES*)

License  Creative Commons BY 4.0 International license  
© Vicente P. Soloviev

In this talk four different applications of the Estimation of Distribution Algorithms (EDAs) have been shown. The first application [1] is a complex multivariate optimization problem where a Gaussian Bayesian network is used a probabilistic model where some prior knowledge is involved and fixed as evidence in the model to restrict the search space. The second one is a Feature Subset Selection problem with time series, where not only the optimum subset of variables are chosen, but also de ideal time series transformation that improves the model. An EDA as a wrapper approach is implemented. The third one is a Real Estate portfolio optimization problem where it is aimed to minimize the risk of the investment and to maximize the profit of the portfolio. It is presented how EDA deal with large scale data in this problem. The last optimization problem [2] consists of optimizing some hyperparameters of a concrete quantum circuit. This talk leads to discussion about the probabilistic model complexity needed to approach the real-world optimization problems.

#### References

- 1 Soloviev, V.P., Larranaga, P. and Bielza, C. , 2022, June. *Estimation of distribution algorithms using Gaussian Bayesian networks to solve industrial optimization problems constrained by environment variables*. In Journal of Combinatorial Optimization.
- 2 Soloviev, V.P., Larranaga, P. and Bielza, C. , 2022, June. *Quantum Parametric Circuit Optimization with Estimation of Distribution Algorithms*. In 2022 The Genetic and Evolutionary Computation Conference (GECCO). DOI: <https://doi.org/10.1145/3520304.3533963>

### 3.16 Semi-parametric Bayesian networks in EDAs

Vicente P. Soloviev (*Polytechnic University of Madrid, ES*)

License  Creative Commons BY 4.0 International license  
© Vicente P. Soloviev

Some real applications involve leading with variables that do not fit a normal distribution in the continuous space. Thus, some more complex algorithms should be used rather than the well-known Gaussian Bayesian network. A novel method is presented where EDA decides whether the variables will fit or not a Gaussian distribution during runtime. Some results are shown comparing the approach with some other state-of-the-art methods such as EGNA, JADE or CMA-ES for some specific benchmarks, and also explaining the performance of the approach during runtime.

### 3.17 A bit about algorithm configuration

Thomas Stützle (Free University of Brussels, BE)

License © Creative Commons BY 4.0 International license  
© Thomas Stützle

This is a bit about automated algorithm configuration. It is divided in two aspects. The first highlights the main advantages of addressing algorithm design and configuration by algorithmic techniques and describes the main existing techniques. We show as one of the various available ones the irace algorithm for automated configuration. The second highlights the advantages of automated configuration. In particular, we show how flexible algorithm frameworks can support the automatic design of high-performing hybrid stochastic local search algorithms. We show this by the example of nine permutation flow shop problems for which we obtained in each one of these a new state-of-the-art algorithm even though they have on average around 500 parameters.

### 3.18 Different parameter regimes in the cGA and their effect on performance

Dirk Sudholt (Universität Passau, DE)

License © Creative Commons BY 4.0 International license  
© Dirk Sudholt

Understanding the dynamics of the underlying probabilistic model is a fundamental problem in estimation-of-distribution algorithms. We consider the compact Genetic Algorithm (cGA) on the function ONEMAX and show how the choice of the update strength  $1/K$  affects performance. For small  $K$  updates have a large impact and genetic drift leads to chaotic behaviour and exponential times. For large  $K$ , genetic drift is negligible since updates are small and the algorithm slowly learns good bit values. For medium  $K$ , frequencies may hit their lower borders, but these wrong decisions can be reverted efficiently. The parameter landscape has a bimodal shape with two asymptotically optimal parameter values. A similar, but more extreme effect is observed for the multimodal CLIFF function where the cGA is inefficient for all values of  $K$  and the best (exponential) runtime is obtained for exponential values of  $K$  that prevent genetic drift.

#### References

- 1 Johannes Lengler, Dirk Sudholt, and Carsten Witt. The complex parameter landscape of the compact genetic algorithm. *Algorithmica*, 83(4):1096–1137, 2021.
- 2 Frank Neumann, Dirk Sudholt, and Carsten Witt. A few ants are enough: ACO with iteration-best update. In *Proc. of GECCO '10*, pages 63–70. ACM Press, 2010.
- 3 Frank Neumann, Dirk Sudholt, and Carsten Witt. The compact genetic algorithm struggles on cliff functions. In *Proc. of GECCO '22*. ACM Press, 2022.
- 4 Dirk Sudholt and Carsten Witt. On the choice of the update strength in estimation-of-distribution algorithms and ant colony optimization. *Algorithmica*, 81(4):1450–1489, 2019.

### 3.19 EDAs, noise and graceful scaling

Andrew M. Sutton (University of Minnesota – Duluth, US)

License  Creative Commons BY 4.0 International license  
© Andrew M. Sutton

A search algorithm scales gracefully with noise when a polynomial increase in noise intensity can be overcome by at most a polynomial increase in resource usage. In this talk, I outline a few results related to graceful scaling on additive posterior noise and estimation of distribution algorithms.

## 4 Working groups

### 4.1 Breakout session: Dealing with constraint search spaces

Josu Ceberio Uribe (University of the Basque Country – Donostia, ES) and Ekhine Irurozki (Telecom Paris, FR)

License  Creative Commons BY 4.0 International license  
© Josu Ceberio Uribe and Ekhine Irurozki

There are many problems whose search space is the space that results on applying constraints to  $\{0, 1\}^n$ . For example, the graph partitioning problem on a graph of  $n$  edges. The set of solutions to this problem is the set of binary vectors of length  $n$  in which the number of 1s equals the number of 0s,  $S = \{v : v \in \{0, 1\}^n \wedge \sum_i v(i) = n/2\}$ . Another example is the MaxCut problem.

We consider this set  $S$  as a running example but extensions to different settings with other constraints are straight forward.

In order to use EDAs on  $S$  the naive approach is to consider the whole set  $\{0, 1\}^n$  and then discard the infeasible solutions (aka those that do not belong to  $S$ ). However, this approach is not efficient as the number of infeasible solutions can be a large proportion of the total space and the proportion of discarded solutions increases with  $n$ . This motivates the need for distributions with support on  $S$ .

Based on techniques described in the talks in this seminar we have described a probability model for  $S$ . We describe here the basic methods for EDAs, which are sampling and learning.

Given a sample of vectors  $V \in S$  of cardinality  $m$  the learning process consists on a the estimation of a matrix  $M \in R^{n \times n}$  for which  $M_{ij}$  is the ratio of the vectors for which  $v(i) = 0$  and  $v(j) = 1$  among those in which they have different values, i.e., set  $w_{ij} = |\{v \in V \wedge v(i) = 0 \wedge v(j) = 1\}|$ ,  $M_{ij} = \{w_{ij}/w_{ji}\}$

The sampling process of a vector  $v$  is as follows: (1) let  $i, j$  be random pair in  $\{1, n\}$  and (2)  $v[i] = 0$  and  $v[j] = 1$ .

For the development of an EDA the learning could be skipped (substitute by a lazy-learning) in which the learning is done on-line.

## 4.2 Breakout session: Univariate EDAs, deception and epistasis

*Per Kristian Lehre (University of Birmingham, GB) and Johannes Lengler (ETH Zürich, CH)*

License  Creative Commons BY 4.0 International license  
© Per Kristian Lehre and Johannes Lengler

We discussed the surprising similarity of the cGA and UMDA on OneMax, where the step size  $K$  of UMDA is equated with the population size  $\mu$  of the UMDA, and the sample size  $\lambda$  is in  $\Theta(\mu)$ . The right scaling is to compare  $K$  steps of the cGA with one step of the UMDA. In this scaling, both algorithms use  $\Theta(K)$  function evaluations per round. It turns out that for any distribution, the drift (expected change) in each frequency is essentially the same for UMDA and cGA. The most extremal possible change is also the same: it is just so possible to go from one boundary to the other in one round. However, we found out that the tail distribution for such extremal events is not the same.

The discussion took place with the whole seminar, after a talk on deception and epistasis by Per Kristian Lehre. Martin Krejca speculated that one might see a difference on such functions, because the tail distribution of large steps is important in this case. In the end, we identified several research questions in that direction:

1. For which other fitness functions than OneMax is the drift identical? Can we identify a class of such fitness functions?
2. While cGA and UMDA have the same expected change on OneMax, do they also have the same variance? If not, this should result in a stronger/weaker genetic drift that might be important in some settings
3. Can we formally or experimentally prove that the behaviour of the two algorithms is qualitatively different in the deceptive environments invented by Lehre?

## 4.3 Breakout session: Explainability

*John McCall (The Robert Gordon University – Aberdeen, GB) and Josu Ceberio Uribe (University of the Basque Country – Donostia, ES)*

License  Creative Commons BY 4.0 International license  
© John McCall and Josu Ceberio Uribe

In this breakout session, the explainability of evolutionary computation has been discussed. The chair explained that when it comes to approach real world problems, many times, the solutions that are obtained from the EC algorithm are difficult to accept by the decision makers of the projects (usually people who do not have a scientific background). Some participants of the session have discussed that it is not the responsibility of the scientist or engineer to make the algorithm understandable for the decision maker. On the other side, some participants have explained that when it comes to apply the proposed solution in critical scenarios (medical therapies, healthcare systems, or any other critical infrastructure configuration), then decision makers usually does not trust in solutions that they do not trust, as they are risky in the opinion. The goal of the explainability is to try to understand why the algorithm output a particular solution with rules that are understandable for the human.

In the board, a scheme of types of explainability was sketched. The first branching distinguishes between (a) risk assessment and (b) problem understanding, and from the second two other branches were obtained: (1) sensitivity analysis and (2) problem information from algorithm traces.

Thinking on an analogous problem on machine learning and explainability, it was mentioned that good explainability in general requires looking power in the model/algorithm. No direct link was done with EC.

The audience agreed that there could be scenarios like this, but there was no clue about how to do it. It will take place a workshop in GECCO 2022 on the topic under the acronym ECXAI, that could be a nice discussion point.

#### 4.4 Breakout session: Do we need to model interactions in EDAs

*Jonathan L. Shapiro (University of Manchester, GB)*

License  Creative Commons BY 4.0 International license  
© Jonathan L. Shapiro

This session motivated to some extent by a comment in a talk by John McCall, which said (paraphrasing) sometimes it is not worth the effort to model the interactions, but rather to just search fast using a univariate EDA.

Jon Shapiro started by pointing out that many combinatorial optimization problems exist on undirected connected graphs. These are computationally expensive to sample from, requiring hundreds or thousands of Markov chain Monte Carlo steps. These are also very difficult to model from data, because in complex situations, variables can be “frozen” into correlated or anti-correlated states even without there being direct interactions between them. So would dedicating computational resources to model building be fruitful, or would this computational effort be better used in search?

#### 4.5 Breakout session: Genetic drift

*Dirk Sudholt (Universität Passau, DE) and Johannes Lengler (ETH Zürich, CH)*

License  Creative Commons BY 4.0 International license  
© Dirk Sudholt and Johannes Lengler

We first tried to clarify our understanding of the term of genetic drift, and found a satisfactory definition: it is the difference between the mean from perfect distribution and empirical distribution, in other words it is the effect of finite samples. We also discussed what this implies about genetic drift in case of several optima, or when the fitness signal is zero, for example in the checkerboard benchmark. Finally, in Mallows models, it is less clear what the genetic drift is. Empirically, when the solution stop improving, then the variance steeply goes down. It is unclear which is the cause and which the effect.

We discussed several methods on how to observe it empirically. A conservative way would be to measure or compute for a fitness-neutral frequency whether there is a risk of hitting boundaries. Another option is to either compare a run with normal sample size with a run with a very high sample size, or several runs with normal sample sizes. If the runs are diverging, this is a sign of genetic drift.

We then turned to the effects of genetic drift. One question is whether it leads to a loss of variance, and that depends on the situation. We have also discussed several cases in which genetic drift might help, including tie-breaking and escaping from local optima.

Finally, we discussed how to control and prevent it, though it was debated whether we even need or want to do that, so we also discussed ways to increase it. One way would be to control parameters like learning rate or steps size, which may be interpreted as controlling the amount of genetic drift. It was suggested to design a cGA where we adapt  $K$  in order to control the amount of genetic drift. The sig-cGA achieves a related goal in a similar way, by using statistical test. The sig-cGA is rather artificial and extreme, but it might be promising to look at less extreme version of this algorithm. We discussed that this might be hard in continuous domains with a large or moderate number of dimensions.

In the end, we digressed into discussion what other methods one can use to deal with local optima or premature convergence.

## 5 Panel discussions

### 5.1 Concluding group discussion of the seminar

*Josu Ceberio Uribe (University of the Basque Country – Donostia, ES), Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR), Per Kristian Lehre (University of Birmingham, GB), and Carsten Witt (Technical University of Denmark – Lyngby, DK)*

License © Creative Commons BY 4.0 International license  
© Josu Ceberio Uribe, Benjamin Doerr, Per Kristian Lehre, and Carsten Witt

**Chair:** Per Kristian Lehre

The group discussion on the last day of the seminar involved all participants of the seminar and aimed at summarizing the findings of the seminar and discussing central questions related to the main aim of the seminar, bridging the gap between theory and practice in EDAs. Concretely, the following two main questions along with their subtopics were proposed as items for discussion:

#### 1. What are practice-relevant EDAs?

- When is a univariate model enough?
- When do we have to model dependencies between variables?
- When is it too expensive to employ an EDA?

#### 2. What are the barriers for theory of EDAs

- What is the right theory?
- What problems should theory study?
- What are the major challenges?

Regarding the first subject, practice-relevant EDAs, the seminar discussed the use of restart strategies. Also, the suggestion came up to start with very simple EDAs (e. g. UMDA) in the first place (an idea supported by talks of practitioners during the seminar). One would then proceed to more complex interaction models only if the results of simple EDAs, along with restart strategies, are unsatisfactory.

A challenge in using complex EDAs is the complex parameter space. On the one hand, empirical evidence indicates that parameter tuning for EDAs is more benign than expected [9]; on the other hand, theoretical research has identified examples where the parameter

landscape is surprisingly complex [7]. As an interesting note, state-of-the-art automated parameter configuration tools (e. g. *irace*, [8]) conduct their search in a manner resembling an EDA. As a note of caution, participants pointed out that optimal parameter settings in EDAs and other search heuristics may also vary during the run.

As another domain where EDAs may be the preferred algorithm, the participants discussed the evolution of surrogate models of fitness functions. They pointed out that building these models can be computationally expensive so standard complexity measures like the number of fitness calls are no longer representative.

The seminar discussed briefly that common EDAs focus on building a model of the search space and do not rely on directed search operators like mutation vectors in continuous search spaces. However, no conclusion was reached on how to best model the notion of *direction* in discrete search spaces.

During the second half of the group discussion, the seminar turned to a complementary subject, the barriers for theory mentioned above. As theoretical benchmark problems for multivariate EDAs, the so-called concatenated trap functions [3], checkerboard functions [1], deceiving LeadingOnesBlocks [6], Ising models [4] and generalized models of multimodal problems (e. g., the *SparseLocalOpt* problem class, [2]) were suggested. Several of these may serve as examples where multivariate models may be required for efficient optimization times.

The participants discussed the purpose of simple benchmark functions like ONEMAX, which were defended as building blocks of more complex functions and a basic sanity check for new algorithms. If an algorithm does not optimize ONEMAX efficiently, it should be discarded.

Criticism with respect to the commonly analyzed search space of bit strings  $\{0, 1\}^n$  led the discussion to permutation spaces, which were the subject of several talks during the seminar. Runtime analyses of EDAs in permutation spaces, using simple benchmark problems like sorting were suggested.

More generally, participants also demanded runtime analyses of EDAs on classical combinatorial optimization problems (e. g., minimum spanning trees) and the study of other performance measures than the first hitting time of an optimum. In particular, fixed-budget analyses as common with evolutionary algorithms [5] seem to be missing so far even though existing tools for the analyses should make results in this direction feasible.

Regarding the “right” theory, the suggestion came up to study general properties/characterizations of problems and algorithms that guarantee efficient runtimes. No final conclusion was reached how such a theory could look like.

Towards the end of the session, the use of surrogate models in EDAs was picked up again. Runtime analyses seem very complex; however, a starting point may be to discriminate the structure of surrogate models at two points in time  $t_1, t_2$  using experimental studies of EDAs at these times or even using theoretical means. Such studies of the quality of surrogate models could possibly lead to estimations of the development of the model over time and even the final optimization time.

As a last point in the discussion, the participants talked about runtime analyses of EDAs in noisy optimization where the noise is not Gaussian; for example, dynamic noise that decreases over time or when getting closer to the optimum. This was linked to the observation that the variance of the sampling distribution in EDAs tends to decrease over time.

## References

- 1 Shumeet Baluja and Scott Davies. Using optimal dependency-trees for combinational optimization. In Douglas H. Fisher, editor, *Proceedings of the Fourteenth International Conference on Machine Learning (ICML 1997)*, Nashville, Tennessee, USA, July 8-12, 1997, pages 30–38. Morgan Kaufmann, 1997.
- 2 Duc-Cuong Dang, Anton V. Eremeev, and Per Kristian Lehre. Escaping local optima with non-elitist evolutionary algorithms. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 12275–12283. AAAI Press, 2021.
- 3 Kalyanmoy Deb and David E. Goldberg. Analyzing deception in trap functions. In L. Darrell Whitley, editor, *Proceedings of the Second Workshop on Foundations of Genetic Algorithms. Vail, Colorado, USA, July 26-29 1992*, pages 93–108. Morgan Kaufmann, 1992.
- 4 Simon Fischer and Ingo Wegener. The ising model on the ring: Mutation versus recombination. In Kalyanmoy Deb, Riccardo Poli, Wolfgang Banzhaf, Hans-Georg Beyer, Edmund K. Burke, Paul J. Darwen, Dipankar Dasgupta, Dario Floreano, James A. Foster, Mark Harman, Owen Holland, Pier Luca Lanzi, Lee Spector, Andrea Tettamanzi, Dirk Thierens, and Andrew M. Tyrrell, editors, *Genetic and Evolutionary Computation – GECCO 2004, Genetic and Evolutionary Computation Conference, Seattle, WA, USA, June 26-30, 2004, Proceedings, Part I*, volume 3102 of *Lecture Notes in Computer Science*, pages 1113–1124. Springer, 2004.
- 5 Thomas Jansen and Christine Zarges. Fixed budget computations: a different perspective on run time analysis. In Terence Soule and Jason H. Moore, editors, *Genetic and Evolutionary Computation Conference, GECCO '12, Philadelphia, PA, USA, July 7-11, 2012*, pages 1325–1332. ACM, 2012.
- 6 Per Kristian Lehre and Phan Trung Hai Nguyen. On the limitations of the univariate marginal distribution algorithm to deception and where bivariate edas might help. In Tobias Friedrich, Carola Doerr, and Dirk V. Arnold, editors, *Proceedings of the 15th ACM/SIGEVO Conference on Foundations of Genetic Algorithms, FOGA 2019, Potsdam, Germany, August 27-29, 2019*, pages 154–168. ACM, 2019.
- 7 Johannes Lengler, Dirk Sudholt, and Carsten Witt. The complex parameter landscape of the compact genetic algorithm. *Algorithmica*, 83(4):1096–1137, 2021.
- 8 Manuel López-Ibáñez, Jérémie Dubois-Lacoste, Leslie Pérez Cáceres, Thomas Stützle, and Mauro Birattari. The *irace* package: Iterated racing for automatic algorithm configuration. *Operations Research Perspectives*, 3:43–58, 2016.
- 9 Yasha Pushak and Holger H. Hoos. Algorithm configuration landscapes: – more benign than expected? In Anne Auger, Carlos M. Fonseca, Nuno Lourenço, Penousal Machado, Luís Paquete, and L. Darrell Whitley, editors, *Parallel Problem Solving from Nature – PPSN XV – 15th International Conference, Coimbra, Portugal, September 8-12, 2018, Proceedings, Part II*, volume 11102 of *Lecture Notes in Computer Science*, pages 271–283. Springer, 2018.

## Participants

- Carlos Andreu Vilarroig  
Technical University of  
Valencia, ES
- Mayowa Ayodele  
Fujitsu Research of Europe –  
Slough, GB
- Alexander Brownlee  
University of Stirling, GB
- Josu Ceberio Uribe  
University of the Basque Country  
– Donostia, ES
- Benjamin Doerr  
Ecole Polytechnique –  
Palaiseau, FR
- Nikolaus Hansen  
INRIA Saclay – Palaiseau, FR
- Ekhine Irurozki  
Telecom Paris, FR
- Ata Kaban  
University of Birmingham, GB
- Martin S. Krejca  
Sorbonne University – Paris, FR
- Per Kristian Lehre  
University of Birmingham, GB
- Johannes Lengler  
ETH Zürich, CH
- Manuel López-Ibáñez  
University of Málaga, ES
- John McCall  
The Robert Gordon University –  
Aberdeen, GB
- Vicente Perez Soloviev  
Polytechnic University of  
Madrid, ES
- Jonathan L. Shapiro  
University of Manchester, GB
- Thomas Stützle  
Free University of Brussels, BE
- Dirk Sudholt  
Universität Passau, DE
- Andrew M. Sutton  
University of Minnesota –  
Duluth, US
- Carsten Witt  
Technical University of Denmark  
– Lyngby, DK

