

Estimation-of-Distribution Algorithms: Theory and Applications

Josu Ceberio Uribe^{*1}, Benjamin Doerr^{*2}, John McCall^{*3},
Carsten Witt^{*4}, and Marcus Schmidbauer^{†5}

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

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

3 The Robert Gordon University – Aberdeen, GB. j.mccall@rgu.ac.uk

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

5 Universität Passau, DE. marcus.schmidbauer@uni-passau.de

Abstract

The Dagstuhl Seminar 25092 “Estimation-of-Distribution Algorithms: Theory and Practice” took place on February 23–28, 2025. It brought together 25 international experts in estimation-of-distribution algorithms (EDAs) and related research areas. Their research focus ranged from theoretical aspects like mathematical runtime analyses to efficient solutions of real-world problems in industrial contexts. This report documents the program and the main outcomes of this fruitful seminar.

Seminar February 23–28, 2025 – <https://www.dagstuhl.de/25092>

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.15.2.158


1 Executive Summary

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

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

© John McCall, Josu Ceberio Uribe, Benjamin Doerr, and Carsten Witt

The Dagstuhl Seminar “Estimation-of-Distribution Algorithms: Theory and Practice” from February 23 to 28, 2025, brought together 25 international experts in the field of estimation-of-distribution algorithms (EDAs). EDAs iteratively evolve a distribution on the search space (called probabilistic model of the search space) in such a way that better solutions accumulate probability mass. Consequently, after running an EDA for sufficient time, one can obtain near-optimal solutions simply by sampling from this probabilistic model. The expertise of the seminar participants covered a broad variety of aspects ranging, in particular, from theoretical topics such as mathematical runtime analyses to an applied topics such as the solution of industrial optimization problems through EDAs.

* Editor / Organizer

† Editorial Assistant / Collector



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Estimation-of-Distribution Algorithms: Theory and Applications, *Dagstuhl Reports*, Vol. 15, Issue 2, pp. 158–183
Editors: Josu Ceberio Uribe, Benjamin Doerr, John McCall, Carsten Witt, and Marcus Schmidbauer



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

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 two long introductory talks on theory of EDAs and on typical applications of EDAs, through some regular conference-style talks discussing recent substantial results, short flash talks presenting new ideas and open problems, and breakout sessions for focused discussions. At all times, we left ample time for discussions, and flexibly adjusted the schedule to not cut short interesting discussions.

We believe that the seminar has succeeded its main aims of narrowing the gap between theory and practice in EDAs. This is visible from the high number of talks and the intensive discussions both in plenary and breakout sessions. Moreover, all participants attended essentially all sessions and all stayed for the full five days of the seminar. Given the large number of truly leading experts from academia and industry usually having overloaded schedules, this shows the success of the seminar, and the Dagstuhl Seminar center in general.

The seminar has identified several promising research directions, which can be explored in the remainder of this report. One that deserves particular mention is the relation of model-based genetic algorithms, that use additional mechanisms to understand the search space structure, and multi-variate EDAs, that try to understand the search space via the core working principles of EDAs. Here clearly a more intensive exchange of ideas is very promising, and will surely result from this seminar.

Overall, we feel that this was a very successful seminar, and, also in the name of the participants, we thank Dagstuhl for enabling us to conduct this seminar at Schloss Dagstuhl, which was the ideal place for this type of seminar.

2 Table of Contents

Executive Summary

John McCall, Josu Ceberio Uribe, Benjamin Doerr, and Carsten Witt 158

Overview of Talks

Multi-Armed Bandits and Algorithm Configuration and their connections to EDAs (Self-presentation) <i>Jasmin Brandt</i>	162
Inversion vectors: an alternative representation for permutations <i>Josu Ceberio Uribe</i>	162
Searching Permutations for Constructing Uniformly Distributed Point Sets <i>Carola Doerr</i>	163
An Introduction to Search Trajectory Mining <i>Martin Fyvie</i>	163
It's Always the Step-Size <i>Nikolaus Hansen</i>	164
Can EDAs be useful for Data augmentation and Imputation? <i>Jose Ignacio Hidalgo</i>	164
Is Deep Optimization an EDA? <i>Joshua D. Knowles</i>	165
Estimation-of-Distribution Algorithms for Multi-Valued Decision Variables <i>Martin S. Krejca</i>	166
Introduction to Theoretical Analyses of EDAs <i>Timo Kötzing</i>	166
Faster Optimization Through Genetic Drift <i>Johannes Lengler</i>	167
Introduction to EDAs <i>John McCall</i>	167
Estimation of Distribution Algorithms using generative Machine Learning models <i>Franz Rothlauf</i>	167
From Fitness Landscapes to Problem Factorizations: A Machine Learning Perspective <i>Roberto Santana</i>	168
The use of the Doubly Stochastic Matrix models for the Quadratic Assignment Problem <i>Valentino Santucci</i>	169
Linkage-Learning Evolutionary Algorithms: Recent Theoretical Advances <i>Marcus Schmidbauer</i>	169
Estimation of Distribution Algorithms and the Ising Model <i>Jonathan L. Shapiro</i>	169

Vine Copula-Based Estimation of Distribution Algorithms for Multivariate Continuous Optimization	
<i>Marta Rosa Soto Ortiz</i>	170
Optimal Mixing in Model-based Evolutionary Algorithms	
<i>Dirk Thierens</i>	170
Estimated Distributions and Warm Starts	
<i>Vanessa Volz</i>	170
Runtime Analysis of the Multi-valued cGA	
<i>Carsten Witt</i>	171

Working groups

Breakout Session: Learning Distributions with Deep Learning	
<i>Josu Ceberio Uribe</i>	172
Breakout Session: Exploration versus Exploitation	
<i>Duc-Cuong Dang and Jonathan L. Shapiro</i>	173
Breakout Session: Univariate versus Multivariate 2	
<i>Duc-Cuong Dang and Jonathan L. Shapiro</i>	174
Breakout Session: Genetic Drift 1	
<i>Martin S. Krejca</i>	175
Breakout Session: Univariate versus Multivariate 1	
<i>Timo Kötzing</i>	177
Breakout Session: Genetic Drift 2	
<i>Johannes Lengler and Carsten Witt</i>	177
Breakout Session: Explainability of EDAs	
<i>John McCall and Josu Ceberio Uribe</i>	178
Breakout Session: Linkage Learning	
<i>John McCall and Duc-Cuong Dang</i>	180

Panel discussions

Discussion Forum: Early Insights on Design and Analysis of EDAs	
<i>Duc-Cuong Dang, Timo Kötzing, Marcus Schmidbauer, and Andrew M. Sutton</i> . .	181

Open problems

Sampling of Rare Fitness Values	
<i>Duc-Cuong Dang</i>	182

Participants	183
-------------------------------	-----

3 Overview of Talks

3.1 Multi-Armed Bandits and Algorithm Configuration and their connections to EDAs (Self-presentation)

Jasmin Brandt (Universität Paderborn, DE)

License © Creative Commons BY 4.0 International license
© Jasmin Brandt

Main reference Jasmin Brandt, Viktor Bengs, Björn Haddendorst, Eyke Hüllermeier: “Finding Optimal Arms in Non-stochastic Combinatorial Bandits with Semi-bandit Feedback and Finite Budget”, in Proc. of the Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022, 2022.

URL http://papers.nips.cc/paper_files/paper/2022/hash/820e95997d050178323230e316897c38-Abstract-Conference.html

Main reference Jasmin Brandt, Elias Schede, Björn Haddendorst, Viktor Bengs, Eyke Hüllermeier, Kevin Tierney: “AC-Band: A Combinatorial Bandit-Based Approach to Algorithm Configuration”, in Proc. of the Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023, pp. 12355–12363, AAAI Press, 2023.

URL <https://doi.org/10.1609/AAAI.V37I10.26456>

In this talk, a brief introduction to the Multi-Armed Bandit and Algorithm Configuration problem was given. I presented my recent work on combinatorial bandits and the challenges during the extension of the theoretical analysis to Algorithm Configuration. In addition, the connection of the problems to EDAs was shortly outlined.

3.2 Inversion vectors: an alternative representation for permutations

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

License © Creative Commons BY 4.0 International license
© Josu Ceberio Uribe

Joint work of Josu Ceberio Uribe, Aimar Barrena, Ekhine Irurozki, Jose A. Lozano, Mikel Malagón

Many optimization algorithms use permutations to represent solutions, but despite their apparent simplicity, permutations present significant challenges for optimization methods, particularly for Global Random Search algorithms. The mutual-exclusivity constraint complicates the learning and sampling of probability distributions over the permutation space, often requiring costly numerical computations. An alternative approach involves using bijective functions on \mathbb{S}_n to transform permutation-coded solutions into integer-vector representations known as inversion vectors.

Although inversion vectors have been recognized for centuries, a comprehensive and unified framework for them is lacking. In this paper, we provide formal definitions and a unified notation for all types of inversion vector codifications. We then establish bijective transformations between these codifications, enabling the characterization of their relationships and properties. Finally, we apply this theoretical framework to explain the behavior of UMDA algorithms when different inversion vectors are used in various permutation problems.

3.3 Searching Permutations for Constructing Uniformly Distributed Point Sets

Carola Doerr (Sorbonne University – Paris, FR)

License © Creative Commons BY 4.0 International license

© Carola Doerr

Joint work of François Clément, Carola Doerr, Kathrin Klamroth, Luís Paquete

Main reference François Clément, Carola Doerr, Kathrin Klamroth, Luís Paquete: “Transforming the Challenge of Constructing Low-Discrepancy Point Sets into a Permutation Selection Problem”, CoRR, Vol. abs/2407.11533, 2024.

URL <https://doi.org/10.48550/ARXIV.2407.11533>

Uniformly distributed point sets of low discrepancy are heavily used in experimental design and across a very wide range of applications such as numerical integration, computer graphics, and finance. Recent methods based on Graph Neural Networks [1] and solver-based optimization [2] identified point sets having much lower discrepancy than previously known constructions. We show that further substantial improvements are possible by separating the construction of low-discrepancy point sets into (i) the relative position of the points, and (ii) the optimal placement respecting these relationships. Using tailored permutations, we construct point sets that are of 20% smaller discrepancy on average than those proposed by Rusch et al. In terms of inverse discrepancy, our sets reduce the number of points in dimension 2 needed to obtain a discrepancy of 0.005 from more than 500 points to less than 350. For applications where the sets are used to query time-consuming models, this is a significant reduction.

References

- 1 T.K. Rusch, N. Kirk, M.M. Bronstein, C. Lemieux, D. Rus, Message-Passing Monte Carlo: Generating low-discrepancy point sets via graph neural networks. *Proc. Natl. Acad. Sci. U.S.A.* 121, e2409913121 (2024). <https://doi.org/10.1073/pnas.2409913121> Free version available at <https://arxiv.org/abs/2405.15059>
- 2 F. Clément, C. Doerr, K. Klamroth, L. Paquete, Constructing Optimal L_∞ Star Discrepancy Sets. To appear in *Proc. of the AMS*. Preprint available at <https://arxiv.org/abs/2311.17463>

3.4 An Introduction to Search Trajectory Mining

Martin Fyvie (The Robert Gordon University – Aberdeen, GB)

License © Creative Commons BY 4.0 International license

© Martin Fyvie

Population-based metaheuristics are commonly used to solve complex optimization problems, but their stochastic nature often obscures decision-making processes. As these algorithms are increasingly deployed in critical domains requiring end-user interaction, the need for explainability has grown. This presentation introduces trajectory mining as a post-runtime analytical framework for enhancing the transparency of optimization methods including Estimation of Distribution Algorithms, Genetic Algorithms, Particle Swarm Optimization, and Differential Evolution.

Our approach is based on the use of Principal Components Analysis to project algorithm search trajectories onto lower-dimensional sub-spaces, capturing population evolution from initialization to convergence. Novel metrics such as Mean Variable Contribution identify

decision variables with highest impact on search behaviour and solution quality. These analyses, complemented by statistical testing, extract explanatory features that reveal algorithm behaviour across diverse problem landscapes, allowing search behaviour comparisons.

We show from experiments on benchmark problems and real-world applications how trajectory mining can be used to identify influential variables and reveal key learning steps that help to explain fitness outcomes. These insights can inform the design of new algorithm variants while improving end-user understanding of solution quality and algorithm behaviour, particularly in domains requiring human-algorithm collaboration.

3.5 It's Always the Step-Size


Nikolaus Hansen (INRIA Saclay – Palaiseau, FR)

License  Creative Commons BY 4.0 International license
© Nikolaus Hansen

By the semantics of the parameter K in cGA, we identify analogous places “of K ” in the CMA-ES. However, in the CMA-ES, we cannot use the same value for K in all of these places. In particular, in the (m, C) parametrization of the distribution, the C update must have a much larger K value (a smaller learning rate) than the m update, usually n^2/μ vs 1. Otherwise, CMA-ES fails to work. In the CMA-ES, genetic drift is usually only observable when the objective is noisy and we discuss how failures can be observed. Anecdotal evidence also suggests that a failure of ES or CMA-ES is often due to a failure of the step-size adaptation, in particular also when the objective is noisy where we observe the smallest eigenvalue of C to drift to zero exponentially fast.

3.6 Can EDAs be useful for Data augmentation and Imputation?

Jose Ignacio Hidalgo (Complutense University of Madrid, ES)

License  Creative Commons BY 4.0 International license
© Jose Ignacio Hidalgo

Joint work of Jose Ignacio Hidalgo, Oscar Garnica, J. Manuel Velasco, Jessica Megane

Main reference José Ignacio Hidalgo, Marta Botella, J. Manuel Velasco, Oscar Garnica, Carlos Cervigón, Remedios Martínez, Aranzazu Aramendi, Esther Maqueda, Juan Lanchares: “Glucose forecasting combining Markov chain based enrichment of data, random grammatical evolution and Bagging”, *Appl. Soft Comput.*, Vol. 88, p. 105923, 2020.

URL <https://doi.org/10.1016/J.ASOC.2019.105923>

In this talk, we present the work of our group on blood glucose prediction for individuals with diabetes, where we utilize Estimation of Distribution Algorithms in two distinct parts of the modeling process. First, we applied a UMDA to learn the weights of an ensemble model for glucose prediction. Second, we also use a UMDA to improve the data augmentation process.

I divided the talk in three parts. In the first part I quickly revised how I used a comatc GA, a basic EDA, for solving the main contribution of my PhD Thesis: Partitioning of digital circuits for Multi-FPGA implementation. I also mentioned in this part the previous work on EDAs in collaboration with UPM, Roberto Santana, Pedro Larrañaga, Concha Bielza and Alfredo Cuesta, where we proposed the use of Archimedean Copulas as the model of a Joint Distribution Function in an EDA.

The second part of the talk was devoted to explain our work on Blood Glucose prediction for people with diabetes. We have been applying Grammatical Evolution approaches for the last 12 years. In those works we use EDAs, in particular an UMDA. First, we applied a UMDA to learn the weights of an ensemble model for glucose prediction. Second, we also use a UMDA to improve the data augmentation process.

Finally, we propose a new problem for EDAs to solve in the context of the EARCO project. EARCO stands for European Alpha-1 Research Collaboration. It is a pan-European network committed to promoting clinical research and education in alpha-1 antitrypsin deficiency (AATD).

3.7 Is Deep Optimization an EDA?

Joshua D. Knowles (SLB Cambridge Research, GB)

License © Creative Commons BY 4.0 International license
© Joshua D. Knowles

Main reference Jamie Caldwell, Joshua D. Knowles, Christoph Thies, Filip Kubacki, Richard A. Watson: “Deep Optimisation: Transitioning the Scale of Evolutionary Search by Inducing and Searching in Deep Representations”, *SN Comput. Sci.*, Vol. 3(3), p. 253, 2022.

URL <https://doi.org/10.1007/S42979-022-01109-W>

Deep Optimization is an algorithm that uses deep association networks to learn epistatic structure in a problem while optimizing it, and that works recursively applying parts of itself to itself. It was introduced in 2018 by Jamie Caldwell et al, [1] and it comes out of and develops ideas from Richard A. Watson’s 25 years of research on models of symbiogenetic evolution. It appears to perform remarkably well on some problems (including learning constraints with no explicit constraint handling mechanism), however that’s presently a tentative statement because it’s not in common use. The basic form of the algorithm is an EDA but it also departs from other EDAs in some ways that I’ll explain. I have new results to show based on work I did with MSc students last year; these improve the algorithm significantly on hard multiple knapsack problem instances. Questions: Is Deep Optimization correctly classified as an EDA? Is it better than other EDAs? Why/why not? How should we measure its performance (since it has more and different overhead than other EDAs)? Is there a lighter-weight version of it that would be better for theoretical analysis? What theoretical tools would be needed to analyse Deep Optimization properly?

References

- 1 Caldwell, J.R., Watson, R.A., Thies, C. and Knowles, J.D., 2018. Deep optimisation: Solving combinatorial optimisation problems using deep neural networks. arXiv preprint arXiv:1811.00784.
- 2 Caldwell, J., Knowles, J., Thies, C., Kubacki, F. and Watson, R., 2022. Deep optimisation: Transitioning the scale of evolutionary search by inducing and searching in deep representations. *SN Computer Science*, 3(3), p.253.

3.8 Estimation-of-Distribution Algorithms for Multi-Valued Decision Variables

Martin S. Krejca (Ecole Polytechnique – Palaiseau, FR)

License  Creative Commons BY 4.0 International license

© Martin S. Krejca

Joint work of Firas Ben Jedidia, Benjamin Doerr, Martin S. Krejca


Main reference Firas Ben Jedidia, Benjamin Doerr, Martin S. Krejca: “Estimation-of-distribution algorithms for multi-valued decision variables”, *Theor. Comput. Sci.*, Vol. 1003, p. 114622, 2024.

URL <https://doi.org/10.1016/J.TCS.2024.114622>

The majority of research on estimation-of-distribution algorithms (EDAs) concentrates on pseudo-Boolean optimization and permutation problems, leaving the domain of EDAs for problems in which the decision variables can take more than two values, but which are not permutation problems, mostly unexplored. To render this domain more accessible, we propose a natural way to extend the known univariate EDAs to this setting. Different from a naive reduction to the binary case, our approach avoids additional constraints. Since understanding genetic drift is crucial for an optimal parameter choice, we extend the known quantitative analysis of genetic drift to EDAs for multi-valued, categorical variables. Roughly speaking, when the variables take r different values, the time for genetic drift to become significant is r times shorter than in the binary case. Consequently, the update strength of the probabilistic model has to be chosen r times lower now. To investigate how desired model updates take place in this framework, we undertake a mathematical runtime analysis on the r -valued LeadingOnes problem. We prove that with the right parameters, the multi-valued UMDA solves this problem efficiently in $O(r \ln(r)^2 n^2 \ln(n))$ function evaluations. Overall, our work shows that our good understanding of binary EDAs naturally extends to the multi-valued setting, and it gives advice on how to set the main parameters of multi-values EDAs.

3.9 Introduction to Theoretical Analyses of EDAs

Timo Kötzing (Hasso-Plattner-Institut, Universität Potsdam, DE)

License  Creative Commons BY 4.0 International license

© Timo Kötzing

The Theory of EDAs, conducted as a run time analysis of specific EDAs applied to test functions, has a history of close to 20 years. In this talk, I review some of the core results provided by the theory of EDAs. I give insights into main paradigms and the models used by theoreticians.

Discussing results, I present different settings pertaining to (a) hill climbing; (b) noise, rugged landscapes and local optima; (c) behavior in the absence of a signal.

3.10 Faster Optimization Through Genetic Drift

Johannes Lengler (ETH Zürich, CH)

License © Creative Commons BY 4.0 International license
© Johannes Lengler

Joint work of Cella Florescu, Marc Kaufmann, Johannes Lengler, Ulysse Schaller, Dirk Sudholt, Carsten Witt
Main reference Cella Florescu, Marc Kaufmann, Johannes Lengler, Ulysse Schaller: “Faster Optimization Through Genetic Drift”, in Proc. of the Parallel Problem Solving from Nature – PPSN XVIII – 18th International Conference, PPSN 2024, Hagenberg, Austria, September 14-18, 2024, Proceedings, Part III, Lecture Notes in Computer Science, Vol. 15150, pp. 70–85, Springer, 2024.
URL https://doi.org/10.1007/978-3-031-70071-2_5

I gave some new perspective on genetic drift. While genetic drift is often seen as something to be avoided, as it produces mistakes, I discuss a simple problem, dynamic BinVal, for which it is faster to accept those mistake and use aggressive updates that correct them quickly. This talk is part of a general discussion at Dagstuhl on when one should avoid or accept genetic drift.

References

- 1 Johannes Lengler, Dirk Sudholt, and Carsten Witt. “The complex parameter landscape of the compact genetic algorithm.” *Algorithmica* 83 (2021): 1096-1137.

3.11 Introduction to EDAs

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

License © Creative Commons BY 4.0 International license
© John McCall

In this talk I provide an introduction to Estimation of Distribution Algorithms (EDA). I begin with the general concept of an EDA: that one models the distribution of fitness over a search space and by sampling this distribution one hopes to generate high fitness solutions with a high probability.

The talk covers a range of key EDA techniques including: linkage learning; model parameterisation, model sampling; and adaptation to different representation spaces. I also present some EDA applications on test functions, theoretical benchmarks and to real world applications in biocontrol and marine scheduling.

In summary, EDAs are a mature and successful optimisation metaheuristic based on mathematically-principled statistical modelling that have been applied across all major problem representations and in the real world.

3.12 Estimation of Distribution Algorithms using generative Machine Learning models

Franz Rothlauf (Universität Mainz, DE)

License © Creative Commons BY 4.0 International license
© Franz Rothlauf

The talk discussed the development of EDAs using generative ML models over the last 15 years.

3.13 From Fitness Landscapes to Problem Factorizations: A Machine Learning Perspective

Roberto Santana (University of the Basque Country – Donostia, ES)

License © Creative Commons BY 4.0 International license

© Roberto Santana

Joint work of Roberto Santana, Arnaud Liefooghe, Bilel Derbel

Main reference Roberto Santana, Arnaud Liefooghe, Bilel Derbel: “Boomerang-shaped neural embeddings for NK landscapes”, in Proc. of the GECCO ’22: Genetic and Evolutionary Computation Conference, Boston, Massachusetts, USA, July 9 – 13, 2022, pp. 858–866, ACM, 2022.

URL <https://doi.org/10.1145/3512290.3528856>

This talk discusses the use of machine learning models – particularly neural networks – to automate various components of evolutionary search by leveraging information from the problem’s fitness landscape. A general framework is presented in which the search model comprises three elements: (1) data from the fitness landscape; (2) a neural network; and (3) a machine learning task. As a specific instance of this framework, the talk presents the problem of learning neural embeddings for NK landscapes [1]. In this scenario, the goal is to learn an embedding-based representation of an NK landscape instance, where each binary solution of the NK problem is mapped to a continuous vector. To construct this embedding, each NK model solution is treated as a word in a vocabulary. Sentences are sequences of NK solutions related by predefined semantics (e.g., walks in the fitness landscape), and a corpus is a large collection of such sentences.

The talk also proposes alternative scenarios in which fitness landscape data correspond to local optima networks (LONs) [2], represented as annotated graphs. Graph neural networks are employed to process these LONs, and the machine learning task involves algorithm parameter generation by producing a factorization of the optimization problem – one that could be exploited by a factorized distribution algorithm (FDA). The work aligns with approaches that integrate ML techniques with classical exploratory landscape analysis for automated algorithm selection [3], as well as with applications of graph neural networks in solving combinatorial optimization problems [4].

Finally, the talk addresses the question of theory choice [5] for Estimation of Distribution Algorithms, advocating for the use of factorizations as a fundamental component of EDA theory. This is justified by their potential to provide simplicity, broad applicability, fruitfulness, and accuracy – particularly due to their compatibility with widely accepted theories in the ML field.

References

- 1 R. Santana, A. Liefooghe, and B. Derbel, “Boomerang-shaped neural embeddings for nk landscapes,” in *Proceedings of the Genetic and Evolutionary Computation Conference*, 2022, pp. 858–866.
- 2 G. Ochoa, M. Tomassini, S. Vérel, and C. Darabos, “A study of nk landscapes’ basins and local optima networks,” in *Proceedings of the 10th annual conference on Genetic and evolutionary computation*, 2008, pp. 555–562.
- 3 M. Seiler, U. Škvorc, C. Doerr, and H. Trautmann, “Synergies of deep and classical exploratory landscape features for automated algorithm selection,” in *International Conference on Learning and Intelligent Optimization*. Springer, 2024, pp. 361–376.
- 4 I. Echeverria, M. Murua, and R. Santana, “Multi-assignment scheduler: A new behavioral cloning method for the job-shop scheduling problem,” in *International Conference on Learning and Intelligent Optimization*. Springer, 2024, pp. 138–152.
- 5 T. S. Kuhn, *The Essential Tension*. University of Chicago Press, 1977, ch. Objectivity, value judgment and theory choice, pp. 320–339.

3.14 The use of the Doubly Stochastic Matrix models for the Quadratic Assignment Problem

Valentino Santucci (University for Foreigners of Perugia, IT)

License © Creative Commons BY 4.0 International license

© Valentino Santucci

Joint work of Valentino Santucci, Josu Ceberio

Main reference Valentino Santucci, Josu Ceberio: “On the Use of the Doubly Stochastic Matrix Models for the Quadratic Assignment Problem”, *Evolutionary Computation*, pp. 1–33, 2025.

URL https://doi.org/10.1162/evco_a_00369

Permutation problems have captured the attention of the combinatorial optimization community for decades due to the challenge they pose. Although their solutions are naturally encoded as permutations, in each problem, the information to be used to optimize them can vary substantially. Therefore, we discussed two types of permutation problems: ordering problems and assignment problems. The Quadratic Assignment Problem (QAP) is perhaps the most prominent example of permutation problems of assignment type. In the talk we discussed how it is possible to use Doubly Stochastic Matrices (DSMs) to address QAP in the context of Estimation of Distribution Algorithms. To that end, we discussed efficient learning and sampling schemes that enable an effective iterative update of the probability model. We analyzed the effectiveness of the discussed model both on QAP and on the the Linear Ordering Problem (LOP), a well-known example of permutation problem of ordering type. The talk concluded with a description of the potential uses of DSMs for other optimization paradigms, such as genetic algorithms or model-based gradient search.

3.15 Linkage-Learning Evolutionary Algorithms: Recent Theoretical Advances

Marcus Schmidbauer (Universität Passau, DE)

License © Creative Commons BY 4.0 International license

© Marcus Schmidbauer

Joint work of Markus Schmidbauer, Dirk Sudholt

Linkage Learning (LL) aims at detecting dependencies between problem variables. We gave a short introduction into LL and, in particular, contrasted different linkage-learning techniques. Statistical Linkage Learning (SLL) uses frequencies of gene value combinations, whereas Empirical Linkage Learning (ELL) is based on investigating how perturbation influences the fitness of a solution. We proved the correctness of ELL on the Hierarchical If-and-only-if (H-IFF) problem. Further, we stated bounds on the expected runtime of the ELL-based Parameter-less Population Pyramid (P3) on H-IFF.

3.16 Estimation of Distribution Algorithms and the Ising Model

Jonathan L. Shapiro (University of Manchester, GB)


License © Creative Commons BY 4.0 International license

© Jonathan L. Shapiro

This talk discussed Estimation of Distribution Algorithms and the Ising Model.

3.17 Vine Copula-Based Estimation of Distribution Algorithms for Multivariate Continuous Optimization

Marta Rosa Soto Ortiz (CaixaBank Tech – Madrid, ES)

License  Creative Commons BY 4.0 International license
© Marta Rosa Soto Ortiz

Estimation of Distribution Algorithms (EDAs) are a class of generative models within evolutionary computation that build probabilistic models from selected solutions in order to generate new candidate solutions. In the continuous multivariate setting, traditional EDAs often assume Gaussian distributions and linear dependencies between variables. However, these assumptions are restrictive and rarely hold in complex, real-world scenarios.

To address these limitations, we adopt vine copula models, which offer a flexible framework by decoupling the modeling of marginal distributions from the dependency structure. Vine copulas capture dependencies between variables – regardless of their individual distributions – by learning them directly from the data, allowing for a broad range of pairwise relationships, including asymmetric, nonlinear, and non-monotonic interactions.

Moreover, vine copulas decompose the joint dependency structure into a product of conditional bivariate copulas, organized hierarchically in a sequence of trees known as R-vines. This architecture allows different copula families to be assigned to different variable pairs within the same decomposition, making it highly adaptable to the complex dependency patterns found in real-world data.

We illustrate the effectiveness of this approach through real-world examples, where vine copulas successfully capture diverse dependency patterns – such as tail dependence and complex variable dependencies – that Gaussian-based models fail to represent. In addition, we introduce *copulaedas*, an R package available on CRAN, which implements a complete vine copula-based EDA framework, including model learning and sampling procedures.

Our findings show that vine copulas significantly enhance the modeling capabilities of EDAs, enabling more expressive and adaptive search distributions for continuous optimization problems.

3.18 Optimal Mixing in Model-based Evolutionary Algorithms


Dirk Thierens (Utrecht University, NL)

License  Creative Commons BY 4.0 International license
© Dirk Thierens

This talk discussed Optimal Mixing in Model-based Evolutionary Algorithms.

3.19 Estimated Distributions and Warm Starts

Vanessa Volz (CWI – Amsterdam, NL)

License  Creative Commons BY 4.0 International license
© Vanessa Volz

Estimation-of-Distribution Algorithms (EDA) learn the underlying structure of a given black-box problem. But can the obtained model be reused for optimising similar problems? Below, I summarise a discussion on the topic.

In the field of evolutionary computation (EC), problems are commonly assumed to be a black box. However, in most real-world settings, solving a problem with sophisticated methods is only considered profitable if it is *recurring*, i.e. if different instances of the same problem are encountered repeatedly. For example, logistics companies need to solve comparable routing problems every day. Medical professionals need to create effective treatment plans for similar conditions. Hyperparameters of algorithms need to be adjusted for each new task. While the underlying problem structures remain unknown, black-box optimisation approaches could be iteratively tailored to the regularly encountered problem over time (across instances). Essentially: *Learning to optimise*.

Some work on warm-starting EDAs exists, for example for CMA-ES ([1] for Hyperparameter Optimisation and [2] for contextual optimisation). A structural transfer framework using EDAs has also been proposed in [3].

For some problems, such as ensemble selection of machine learning models for different data folds, the optima may be assumed to be in similar locations. In these cases, when running an EDA, warm-starting from the previously estimated distribution (with higher exploration settings) might be sufficient for improving algorithm performance. In other cases, additional domain knowledge would be required to determine how the obtained information can be transferred. However, in most settings, what kind of similarity exists between instances would need to be established first. How this can be done remains an open question.

References

- 1 M. Nomura, S. Watanabe, Y. Akimoto, Y. Ozaki, and M. Onishi, *Warm Starting CMA-ES for Hyperparameter Optimization*. In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 10, Art. no. 10, May 2021, doi: 10.1609/aaai.v35i10.17109.
- 2 Y. Sekino, K. Uchida, and S. Shirakawa, *Warm Starting of CMA-ES for Contextual Optimization Problems*. In Parallel Problem Solving from Nature – PPSN XVIII, Cham: Springer Nature Switzerland, 2024, pp. 205–220. doi: 10.1007/978-3-031-70068-2_13.
- 3 R. Santana, A. Mendiburu, and J. A. Lozano, *Structural transfer using EDAs: An application to multi-marker tagging SNP selection*. In 2012 IEEE Congress on Evolutionary Computation, Jun. 2012, doi: 10.1109/CEC.2012.6252963.

3.20 Runtime Analysis of the Multi-valued cGA

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

License © Creative Commons BY 4.0 International license

© Carsten Witt

Joint work of Sumit Adak, Carsten Witt

Main reference Sumit Adak, Carsten Witt: “Improved Runtime Analysis of a Multi-Valued Compact Genetic Algorithm on Two Generalized OneMax Problems”, CoRR, Vol. abs/2503.21439, 2025.

URL <https://doi.org/10.48550/ARXIV.2503.21439>

In this talk, we discuss recent advances in the runtime analysis of the multi-valued cGA, defined in the framework by Ben Jedidia et al. [5]. We consider an r -valued LeadingOnes function and prove a runtime bound of $O(n^2 r^2 (\log^3 n + \log^2 r))$ that holds with high probability. For an r -valued OneMax function counting the number of positions of value $r - 1$, we prove a bound of $O(nr(\log n + \log^2 r))$, which improves the previously best bound by Hamano et al. [4] by a logarithmic factor and is tight for $r = 2$. Finally, we discuss a second generalization of OneMax to the r -valued domain, which sums up the integer values of all positions and hence has a maximum value of $n(r - 1)$. For this function, the multi-valued cGA is proved to find the optimum in time $O(nr^3(\log^2 n + \log r))$ with high probability. Our proofs are

based on analyses of stochastic drift of the frequencies for value $r - 1$ and bounds on their genetic drift. Finally, we discuss why the bound for the second OneMax-function is probably not tight.


References

- 1 S. Adak and C. Witt, *Runtime Analysis of a Multi-valued Compact Genetic Algorithm on Generalized OneMax*, in Proc. of PPSN 2024, Springer, pp. 53-69.
- 2 S. Adak and C. Witt, *A Runtime Analysis of the Multi-Valued Compact Genetic Algorithm on Generalized LeadingOnes*, in Proc. of EvoCOP 2025, Springer, to appear.
- 3 S. Adak and C. Witt, *Improved Runtime Analysis of a Multi-Valued Compact Genetic Algorithm on Two Generalized OneMax Problems*, in Proc. of GECCO 2025, ACM Press, to appear
- 4 R. Hamano, K. Uchida, S. Shirakawa, D. Morinaga, and Y. Akimoto, “Tail Bounds on the Runtime of Categorical Compact Genetic Algorithm”, Evolutionary Computation, 2025, https://doi.org/10.1162/evco_a_00361, in press
- 5 F. Ben Jedidia, B. Doerr, and M. S. Krejca, *Estimation-of-Distribution Algorithms for Multi-Valued Decision Variables*, Theoretical Computer Science, vol. 1003, p. 114622, 2024.

4 Working groups

4.1 Breakout Session: Learning Distributions with Deep Learning

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

License  Creative Commons BY 4.0 International license
© Josu Ceberio Uribe

The breakout session explored applications of deep learning for learning distributions, with a focus on alternatives to traditional Maximum Likelihood Estimation (MLE) approaches. Starting point of the session was the following:

- Traditional MLE methods often require intensive computation,
- Model capacity must appropriately capture variable interactions,
- Deep learning offers computational efficiency while maintaining modeling power.

There was no common position, but rather, each participant presented a different vision of what the title of the session might mean. As a result, participants suggested various approaches:

1. Generating variation rather than modeling entire populations,
2. Moving beyond optimization to distribution estimation,
3. Finding optimal search directions,
4. Leveraging deep learning models (both custom-trained and existing LLMs) to guide search point selection,
5. Developing higher-level learning models adaptable to various problems,
6. Using fitness approximation for ranking or selection,
7. Integrating selection directly into learning processes using fitness information.

Finally, the conversation expanded to more general ideas:

- Learning movement patterns or genetic operators (crossover, mutation),
- Developing black-box iterative models that learn movement, selection, and solution representation,

- Projection techniques into embedding spaces to reduce interactions while preserving information,
- Relationships with Bayesian Optimization,
- Solution quality prediction and generation (potential GAN applications).

4.2 Breakout Session: Exploration versus Exploitation

Duc-Cuong Dang (Universität Passau, DE), Jonathan L. Shapiro (University of Manchester, GB)

License © Creative Commons BY 4.0 International license
© Duc-Cuong Dang and Jonathan L. Shapiro

Participants. Jonathan L. Shapiro, Franz Rothlauf, Carsten Witt, Vanessa Volz, Jasmin Brandt, Timo Kötzing, Martin S. Krejca, Aishwarya Radhakrishnan, Andrew M. Sutton, Marcus Schmidbauer, Duc-Cuong Dang

To get everyone on the same page, Jonathan Shapiro highlighted the importance of exploitation and exploration as in the case of Reinforcement Learning (RL): at any time one always have to decide between using the current information to exploit immediate rewards, versus exploring the alternatives that may lead to better rewards later.

The first question we discussed is whether it is a good strategy to explore first and exploit later. This is the case with traditional Evolutionary Algorithms (EAs), the algorithms often converge in the end. Some participants noted that this is not the case for EDAs like BoA. Marcus Schmidbauer added that in the case of GOMEA, a multivariate model-building algorithm, it is a combination of extreme exploitation with extreme exploration. Timo Kötzing added a comment that there is a fundamental difference between RL and EA: in EAs, the rewards are not immediate. The notion of neighbourhood is also important: EAs seem to prefer having nearby good solutions over having random good solutions that are hard to recombine. Jasmin Brandt commented that we may need a proper definition of exploitation and exploration. According to Franz Rothlauf, exploitation means finding good solutions nearby the ones already explored, while doing anything else is exploration.

We then discussed different parameters that can influence the search towards either exploitation or exploration. In the case of EAs, it appears that that the mutation rate is often settled to one on average per mutation. However, one should be careful about which EAs are considered. In the case non-elitist EAs, the balance between mutation and selection does not necessarily lead to mutation rate of $1/n$ but it depends on which selection is used or how it is set. For example, in tournament selection of size k the balance is found exactly at mutation rate $\ln(k)/n$. For EDAs, the following parameters can influence the search: the inverse of the population size $1/K$ in cGA; for UMDA, this is the ratio μ/λ when the (μ, λ) -truncation selection is used to decide which solutions to keep in order to model the next distribution; the learning rate ρ in PBIL and related algorithms like Ant Colony Optimisation. From the theory side, how to control these parameters optimally for simple problems remains an open question as the existing lower bound techniques and results are scarce.


A common set of parameters that is important for all the above-mentioned algorithms to avoid the possibility of premature convergence is the borders that limit the frequencies of bits to be strictly below 1 and strictly above 0. In practical applications, the use of these borders is essential as one may absolutely wants to avoid degenerated distributions, an example is the application Franz Rothlauf has in his talk. Related to this, Jonathan Shapiro

asked, in practices how to avoid identify that the learned model is general enough and not over-fitted. This is a hard question, and often the case this comes back to a good engineering practice, or a familiarity with the problems and algorithms. Martin S. Krejca put forwards the following question: Is there any way to adjust the borders, somehow reflecting the amount of confidence on doing the right thing at any given time? Jasmin Brandt added that in the case of multi-arm bandit problems, the confidence interval is a well-defined concept. A related question we discussed is how to detect that the entropy of the model is too low early on, or are the techniques developed for stagnate detection in EAs useful for EDAs?

In summary, we all agree that the balancing exploitation and exploration is relevant to all areas of Evolutionary Computation, including EDAs. We know how to deal with them for practical problems, however formulating them as statements with rigorous proofs remains challenging. The good news is that even in a lack of recommendations, there is a technique from theory that allow selecting the values for parameters efficiently based on power-law distributions.

4.3 Breakout Session: Univariate versus Multivariate 2

Duc-Cuong Dang (Universität Passau, DE), Jonathan L. Shapiro (University of Manchester, GB)

License  Creative Commons BY 4.0 International license
© Duc-Cuong Dang and Jonathan L. Shapiro

Participants. Jonathan L. Shapiro, John McCall, Joshua D. Knowles, Carola Doerr, Vanessa Volz, Nikolaus Hansen, Josu Ceberio Uribe, Valentino Santucci, Alistair Benford, Andrew M. Sutton, Martin S. Krejca, Roberto Santana, Duc-Cuong Dang

The breakout session began with a discussion of the trade-off between having an accurate model, at the expense of intensive computation, as in the case of building graphical models in EDAs, versus having less accurate models that are faster to learn, but at the risk of misleading or slowing down the search for optimal solutions.

We agree that in practical runs of EDAs, there may not be a clear separation between the phase of learning an accurate model and that of using that model to find optimal solutions as the model is often learned along the way. More importantly, as pointed out by Nikolaus Hansen, for solving an optimisation problem, finding an accurate model can be misleading: we only want to find the best solutions. In fact, we want models that can help the search in finding better solutions, or in finding improvements to the existing solutions, more efficiently. In other words, a good model should help guide the search towards the optimal solutions, and need not be an accurate description of those solutions.

We then discussed whether there are practical problems where there are clear interactions between variables, but the univariate model is still useful, as this is often argued by theory research (most existing theory papers only focus on univariate models). John McCall gave an example of such settings for a scheduling problem in chemotherapy. In this application, thousands of interactions between variables are detected between the variables, but univariate algorithms like PBIL/DEUM still outperform hBOA. The current hypothesis for the inefficiency of hBOA, which is stuck with medium quality solutions, is that this is due to the dominance of different terms in the objectives. It could be very interesting if there is a mathematical description that can capture this property.

As the discussion progressed, we asked the question whether the computational budget or parameter settings, like the sample size, can have some influence on the success of univariate models. This is because the univariate model is cheaper to learn, compared to models with interactions. From the practical side, practitioners have noticed that on some problems, the population or the sample size needs to be set above a certain threshold for the search to be successful. We then discussed the settings in which the computational budget is less of an issue, whether multivariate models are always the better choice. Nikolaus Hansen mentioned a counter-example of the sector-sphere function where univariate performs best. In addition, trying to adapt the step sizes on this function may worsen the progress of optimisation. Similar effects can be observed for the case of LONGPATH and SURFSAMP in discrete optimisation, and this is related to the general issue of balancing exploitation and exploration.

We then discussed whether multivariate models might limit the diversity of solutions. In other words, in multivariate EDAs, how does the entropy of the model evolve as the search progresses? Nikolaus Hansen said that it tends to increase as we learn more about the models. He added that a good indicator of a good model is that the entropy of sampled solutions within a fitness level should be high. Carola Doerr drew the parallel with black-box complexity: if we know more about the problem at hand, we want to have higher entropy. For example, if we know that we are solving a ONEMAX-type problem, the best algorithm is to sample $O(n/\log n)$ solutions uniformly at random, and then infer a consistent target string based on the acquired fitness of these random solutions.

The following questions were also discussed during the breakout session.

1. Is it always a good practice to first try univariate models first and then try the multivariate ones? There is no consensus among the participants on this question.
2. What is the current state of theory research on multivariate models? As far as we know, the only theory research on multivariate models is about linkage-learning, e.g. the papers by Dirk Sudholt and co-authors.
3. Should we model improvement moves (or directions) instead of that of solutions? After our discussion, we incline towards modelling the improvement moves.
4. Is there any benefit for building a model from scratch instead of incremental learning, and vice-versa? This is open.
5. Are the weights of detected interactions important to learn, and how to use them? This is open.
6. Is there a way to learn the model implicitly, or somehow avoid the burden of building an explicitly accurate model? This is open, but there are related ideas about projecting the search into other spaces where there may be less deception.

4.4 Breakout Session: Genetic Drift 1

Martin S. Krejca (Ecole Polytechnique – Palaiseau, FR)

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

Participants. Alistair Benford, Jasmin Brandt, Josu Ceberio Uribe, Duc-Cuong Dang, Benjamin Doerr, Carola Doerr, Martin Fyvie, Nikolaus Hansen, Joshua Knowles, Timo Kötzing, Martin S. Krejca

This breakout session discussed the phenomenon of genetic drift in estimation-of-distribution algorithms (EDAs). The discussion meandered a bit while it took place. The following summary rearranges it logically, grouping it in terms of different points that were discussed.

We started the session by Benjamin's explanations of the mathematical results for genetic drift in univariate EDAs on binary problems, which mostly concerns the cGA and the PBIL. In this setting, genetic drift refers to the random movement of frequencies in the probabilistic model without any true signal from the objective function. In the case of constant objective functions, we have strong concentration bounds that tell us how long it takes until a frequency randomly moves by at least ε . It was also mentioned that the resulting bounds can be used for a smart restart scheme, which recommends to restart an EDA whenever it is likely that a frequency could have moved too far due to genetic drift. When restarting, the step size of an update should be halved (which usually means that the population size should be doubled).

This explanation raised the question how realistic this analysis is, since real objective functions are not constant. It was clarified that genetic drift is usually superimposed by the true signal of an objective function, but that it is hard to analyze it in its generality, which is why the mathematical analysis assumes a constant objective function, which represents the full effect of pure genetic drift.

A natural question that followed was whether this phenomenon also shows up in practice and whether it is relevant. It was agreed upon that genetic drift does show up in practice, but it was less clear whether it is relevant. The consensus was ultimately that it is indeed relevant but that practitioners may not care about it due to other, more pressing issues, such as budgetary constraints or the wishes of the customer. A typical approach is to simply rely on best practices, which traditionally aim at reducing the impact of genetic drift (albeit without mathematical guarantees).

Afterward, the discussion moved to other types of search spaces. It was mentioned that genetic drift is also present for multi-valued discrete search spaces, where it is even easier to have a greater (detrimental) effect, due to generally smaller starting values of the frequencies. Afterward, the question came up to what extent genetic drift is visible in continuous search spaces. The conclusion was that it also shows up there and also poses a problem.

With genetic drift being a general problem, other questions were whether it is possible to detect genetic drift somehow and how it behaves in the presence of noise. Regarding the detection, it was mentioned that the sig-cGA operates on only updating a frequency when a statistical significance in the objective function is detected. This approach allows to directly measure (with a certain confidence) that genetic drift did *not* occur. This approach should not be largely affected by noise. Otherwise, more traditional EDAs are known to cope well with noise, but only if the step size is chosen sufficiently small (suggesting again a smart restart scheme in the general case).

These points were followed up by questions how genetic drift relates to similar phenomena in other settings, such as multi-armed bandits and reinforcement learning. For multi-armed bandits, it was not so clear how this directly translates to the setting of EDAs, since there is no real exploitation in the binary domain, as the two possible choices for a variable are either 0 or 1. However, assigning a confidence to each frequency is something that aligns well with multi-armed bandits and seems like a good strategy. As for reinforcement learning, the question was whether the approach of adding uniform samples periodically with a certain probability could prove promising to counteract genetic drift. The argument was brought forth that uniform samples are not a good choice. However, in continuous problems, it is not


uncommon to add noise on purpose to the samples and to then assess whether the selection is affected by this choice. If it is, then the learning rate is increased. Something similar could potentially prove useful for EDAs as well.

In the same vein as the connection to reinforcement learning, the question came up whether it is smart to introduce an artificial bias into the update that explicitly counteracts genetic drift. This has been done before in theoretical studies to great effect, but the paper on the sig-cGA shows that this can also easily backfire. A static approach is not useful, as the bias can be stronger than the signal of the objective function, which renders optimization unsuccessful. However, a dynamically adjusting approach could maybe work.

In conclusion, genetic drift is a real and important problem that is hard to detect. Being able to detect it could tremendously improve the situation. The sig-cGA as well as the concept of adding noise on purpose and adjusting the step size based on whether this affects selection can be promising ways to reduce the impact of genetic drift. However, these details need to be worked out in greater detail analytically.

4.5 Breakout Session: Univariate versus Multivariate 1

Timo Kötzing (Hasso-Plattner-Institut, Universität Potsdam, DE)

License  Creative Commons BY 4.0 International license
© Timo Kötzing


In this breakout session we discussed univariate versus multivariate EDAs.

We started by discussing insights dating back to 80s / 90s. In particular, we discussed a concatenated (summed) Trap_k function, and the difference to LeadingOnes-concatenated Trap_k . Univariate EDAs or uniform crossover can optimize this function if the gap in function value between local optima is large enough. It might also work if parameter scale with k sufficiently, due to averaging over many traps. If the gap in fitness is small, the function is “fully deceptive”, univariate EDAs get stuck, same with crossover. As an aside: univariate EDAs might be “more stuck” than EAs. Multivariate EDA can learn blocks; can swap all blocks. This is highly related to the work of Mühlenbein, Ochoa and others.

We then discussed how to learn the linkage. How can one do it efficiently? The direct algorithm scales quadratic in the number of variables, which can be too much. Many problems with low computation time use UMDA instead. Some modern algorithms can do better (eg GOMEA).

4.6 Breakout Session: Genetic Drift 2

Johannes Lengler (ETH Zürich, CH) and Carsten Witt (Technical University of Denmark – Lyngby, DK)

License  Creative Commons BY 4.0 International license
© Johannes Lengler and Carsten Witt

This 2nd breakout session on genetic drift considered its definition, theoretical implications, and practical considerations in various types of EDAs. The participants discussed how genetic drift affects algorithm performance, its distinction to stochastic drift, and potential strategies for mitigating its impact.

The following **key points** were debated in the session.

Definition and Characteristics of Genetic Drift

Genetic drift was described as the loss of alleles in a population due to stochastic effects. In an infinite population model, genetic drift cannot lead to a loss of alleles. To avoid too strong genetic drift, the population size of an EDA usually has to scale linearly in problem size on benchmark problems like BV (Binary Value) and LO (LeadingOnes). This was already observed in early literature.

Distinguishing Genetic Drift from Stochastic Drift

Empirical evidence suggests sharp thresholds in population size that determine genetic drift behavior. The role of minimum population size in modeling multivariate problems was debated. The impact of genetic drift on neutral versus non-neutral bits was questioned.

Genetic Drift in Specific Problem Settings

For constrained problems (e.g., knapsack), the effect of different selection operators on drift as discussed. Ising models (related to so-called checkerboard problems) were proposed as potential test cases. Whether genetic drift occurs on fitness plateaus was questioned. The Dynamic BinVal problem was suggested as another theoretical test bed.

Counteracting Genetic Drift

In theoretical studies, borders on marginal frequencies are used to avoid fixation at extreme values. Multi-factor EDAs use Bayesian priors to prevent premature fixation. Tournament selection in UMDA and its potential impact on genetic drift was explored. Hybrid approaches combining EDAs with local search were proposed to mitigate drift. Finally, theoretical studies have identified scenarios (dynamic BinVal) where genetic drift can speed up optimization.

Algorithmic and Theoretical Considerations


Studying PBIL instead of UMDA was suggested as potentially more insightful. Again, the role of population size in determining genetic drift was highlighted. The phenomenon of domino convergence was discussed in relation to problem landscapes like BinVal and OneMax.

Broader Implications and Open Questions

The applicability of genetic drift to machine learning was debated. Whether overparameterized neural networks exhibit genetic drift was questioned. Symmetry breaking in EDAs and theoretical analysis of PBIL for feature selection were suggested as future research directions.

4.7 Breakout Session: Explainability of EDAs

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 session, we approached the “explainability” of EDAs from the point of view of expert practitioners. That is to say, what can be derived empirically from the EDA process that can potentially be useful in gaining an understanding of an EDA? The key starting point

for the discussion was the observation that runs of an EDA using a Probabilistic Graphical Model (PGM) will generate sequences of PGMs. These can be considered as trajectories of the EDA through model space. What might be learned from these trajectories?

The trajectories are a function of EDA parameters (population size, selection pressure) and problem characteristics (the distribution of fitness over solutions). Generally, insights can be gained from varying algorithm or problem characteristics and deriving statistical relations.

Problems can also be decomposed using certain functional bases on solution spaces (Walsh, continuous Fourier, discrete Fourier). There is rich potential for exploiting this when the PGM explicitly distributes fitness over these basis components, also known as “problem structure.”

In black-box scenarios, where the problem structure is unknown and so there is no decomposition available, does it make sense to capture first-order structure, then second-order, and so on to higher orders? Structure detection algorithms are widely used in EDAs and so there is a potential value in mining PGM trajectories for problem structure. We wonder if there is an algorithm that tries to guess the most interesting representation to optimize a problem.

How can a decomposition of the above help when designing an algorithm? Can structural features of problem classes suggest EDAs specialized to learn how fitness is distributed across those structures? Might the behavior of such EDAs on such problems be amenable to theoretical analysis?

EDA experts in the group observed that by examining PGM trajectories you can sometimes manually identify solutions that are better than those sampled from the model by the EDA, even after parameter optimization. Can we learn more efficiently from those trajectories?

In the context of real-world problems, often similar problems need to be frequently solved. We discussed how to gain knowledge about the problem family by running EDAs on the problem and extracting recurrent model structures. Related to this, transfer learning works on EDAs, and incremental-BOA were discussed. There were cases where models learnt in certain iterations were re-used in later iterations ($i+1$, $i+2$, $i+3$, ...). We discussed aggregating trajectories of EDA executions, and then using SVD-type techniques to remove noise, and hence gain insights about recurrent structures. We discussed whether ensembles of algorithms (the same algorithm with different parameters run on the same problem) might help in problem understanding. This was related to Island models. In some real-world problems such as scheduling, such analysis can be used to identify resources that need to be carefully allocated early in the solution process, while others may be fixed later and can be flexibly allocated. This also gives a lot of information about the process. We discussed the general concept of “backbones” of important variables that, once optimally set, reduce the complexity of optimizing the remaining variables.

We discussed the relation between EDA theory and EDA explainability, with a special focus on the phase transition between divergence and convergence of algorithms and the relation to population size. A key theme arising many times in the seminar was the focus of EDA theory on genetic drift whereas researchers of large-scale and applied EDAs never consider but see premature convergence as a major problem. The process of adjusting population size appears to be analogous to the role of K in theory. Is there potential for combining theoretical and empirical analysis to understand this phase transition?

In the context of UMDA (or PBIL), we considered statistical analysis of univariate probabilities as a source of information about the problem. Does the explainability approach discussed in this group offer a bridge between the two EDA communities to build a framework

that is useful to both theory and practice? Such a framework would consider a set of problems with known and controllable structures and explore generalized concepts of K theoretically and empirically.

4.8 Breakout Session: Linkage Learning

John McCall (The Robert Gordon University – Aberdeen, GB) and Duc-Cuong Dang (Universität Passau, DE)

License  Creative Commons BY 4.0 International license
© John McCall and Duc-Cuong Dang

Participants. John McCall, Marcus Schmidbauer, Benjamin Doerr, Timo Kötzing, Martin Fyvie, Aishwarya Radhakrishnan, Andrew M. Sutton, Nikolaus Hansen, Carsten Witt, Duc-Cuong Dang.

The group considered the question “How much precision is needed in linkage learning?” More fully, for an EDA that learns linkage, either in the form of a PGM (Probabilistic Graphical Model) or more broadly as families of subsets of variables, to optimise a problem, is it necessary to incorporate in the EDA model all linkage that is detected.

The discussion focussed around a published paper comparing algorithm performance on a real world chemotherapy problem. The chemotherapy problem is based on binary strings encoding dosages of multiple drugs over a time period. The problem contains thousands of variable linkages based on tumour-size, dosage constraints and toxic side-effects on the patient. The paper compares a leading multivariate-PGM EDA, hBOA, against a GA and three univariate EDAs (UMDA, PBIL and a univariate version of DEUM).

The counter-intuitive result is that the univariate algorithms, led by UMDA, all converge to the (best-known) optimum much faster than hBOA, with far fewer fitness evaluations and massively less clock time. This is despite the fact that the univariate EDAs learn no interactions. The GA is outperformed on this problem by all of the EDAs.

There was discussion on the linkage detection method used – if there is no thresholding then spurious linkages can be detected that have no significance for the problem. hBOA uses internal linkage parameters that do use thresholding, and note that the inventor of hBOA (Martin Pelikan) collaborated on the paper. Moreover, a separate detection algorithm (LDA) was used to verify the linkages. The paper concludes that, in this example, multivariate linkage is inessential.

A contrasting example in the case of Concatenated Traps was discussed. Here it can be shown theoretically that P3 (Parameter-less Population Pyramid), a model-building EA related to EDAs, requires multivariate linkage-learning to solve the problem in polynomial time. If only univariate learning is allowed (i.e. no linkage) P3 is shown to require exponential time to solve Concatenated Traps. Therefore in this example, multivariate modelling is essential.

Some more general insight into when linkages are essential or inessential can potentially be gained through Walsh function decomposition of binary functions. More research is needed into how robust Walsh coefficients are in terms of detecting essential linkage, particularly under function scaling.

The original question on precision remains open.

5 Panel discussions

5.1 Discussion Forum: Early Insights on Design and Analysis of EDAs

Duc-Cuong Dang (Universität Passau, DE), Timo Kötzing (Hasso-Plattner-Institut, Universität Potsdam, DE), Marcus Schmidbauer (Universität Passau, DE), and Andrew M. Sutton (University of Minnesota – Duluth, US)

License © Creative Commons BY 4.0 International license
© Duc-Cuong Dang, Timo Kötzing, Marcus Schmidbauer, and Andrew M. Sutton

The first part of this session was devoted to early theoretical approaches to the analysis of *Estimation-of-Distribution Algorithms* (EDAs). The second one gave a short introduction to the *Cross-Entropy Method* (CEM).

First Part: Early Theory on EDAs

Mühlenbein et. al. [1] consider additively decomposed functions, and point out that simple genetic algorithms fail when they are deceptive. The authors present conditions for such functions on which the Boltzmann distribution factors into a product over marginal distributions, making them tractable for univariate EDAs. They also demonstrate that when these conditions do not hold, the factorization is approximate. Further, the authors consider a purely selection-based EDA using the Boltzmann distribution, called *Boltzmann EDA* (BEDA). The authors prove that, for a given ϵ , BEDA on ONEMAX finds the optimum in N_ϵ (specified in the paper) generations with probability at least $1 - \epsilon$.

Mühlenbein et. al. [2] consider EDAs with infinite populations to solve continuous maximisation problems. They define that an EDA converges *globally* if the average fitness of the population converges to the globally optimal fitness value as time goes to infinity. The authors prove that certain purely selection-based EDAs converge globally. Further, they prove that this is also the case for the *Factorized Distribution Algorithm* (FDA) with proportional selection.

Second Part: Introduction to CEM

The idea of CEM [3] when applied to an optimisation problem is to view it as a problem of estimating the probability of sampling solutions above some quality threshold; typically this is a rare event and the technique of *importance sampling* provides an efficient solution. This is equivalent to minimising the *Kullback-Leibler metrics* between a family of distributions defined by some model and the ideal sampling distribution (hence the name cross-entropy) and analytically solving this minimisation problem produces a specific algorithm for the model, like in the case of the univariate model we obtain the *Univariate Marginal Distribution Algorithm* (UMDA) and then *Population-based Incremental Learning* (PBIL) via some smoothing process.

References

- 1 Heinz Mühlenbein, Thilo Mahnig, and Alberto O. Rodriguez. 1999. Schemata, Distributions and Graphical Models in Evolutionary Optimization. *Journal of Heuristics* (1999), 215–247. <https://doi.org/10.1023/A:1009689913453>
- 2 Qingfu Zhang, and Heinz Mühlenbein. 2004. On the convergence of a class of estimation of distribution algorithms. *IEEE Transactions on Evolutionary Computation* (2004), 127–136. <https://doi.org/10.1109/TEVC.2003.820663>

- 3 Pieter-Tjerk de Boer, Dirk P. Kroese, Shie Mannor, and Reuven Y. Rubinstein. 2005. A Tutorial on the Cross-Entropy Method. *Annals of Operations Research* (2005), 19-67. <https://doi.org/10.1007/s10479-005-5724-z>

6 Open problems

6.1 Sampling of Rare Fitness Values

Duc-Cuong Dang (Universität Passau, DE)

License  Creative Commons BY 4.0 International license
© Duc-Cuong Dang

We discuss the algorithmic design problem of efficiently discovering all fitness values, thus introducing the notion of (expected) cover time instead of optimisation time. This design perspective is useful in some situations where the number of possible fitness values is relatively small and they are well connected.

A prominent algorithm that follows this idea is the so-called Fitness Frequency Assignment [2] which completely ignores the actual fitness values but focuses on the frequencies at which these values are encountered by the parent and offspring populations. We formulate this algorithm in its general form as the $(\mu+\lambda)$ -FFA algorithm using unary unbiased variation operators and show some of its invariants and a conditional upper bound on the cover time. It remains an open question how to prove unconditional bounds for some functions such as OneMax or Jump, since experiments with fixed $\lambda = 1$ suggest that they lie between $\Omega(n^2)$ and $O(n^2 \log n)$ depending on μ . Finally, we introduce the notion of the fitness accessibility graph, and show that learning this graph on the fly can improve the cover time in some settings.

As a general question to the research community on EDAs, we want to know what information and statistics are useful to learn or model, then to exploit in order to minimise the cover time. Benjamin and Timo suggest a multi-objectivization approach: run GSEMO on $(f(x), -f(x))$ and this gives a cover time of $O(n^2 \log n)$ for OneMax and Jump. In a related work, Aisha and Timo have a paper [1] where they analyse Novelty Search on OneMinMax and show a non-trivial upper bound $O(n^2)$ for the expected optimisation time.

References

- 1 D. Antipov, T. Kötzing, A. Radhakrishnan. *Greedy Versus Curious Parent Selection for Multi-objective Evolutionary Algorithms*. PPSN (3) 2024: 86-101.
- 2 T. Weise, Z. Wu, X. Li, Y. Chen, and J. Lässig. *Frequency Fitness Assignment: Optimization Without Bias for Good Solutions Can Be Efficient*. IEEE Trans. Evol. Comput. 27(4): 980-992 (2023).

Participants

- Alistair Benford
University of Birmingham, GB
- Jasmin Brandt
Universität Paderborn, DE
- Josu Ceberio Uribe
University of the Basque Country
– Donostia, ES
- Duc-Cuong Dang
Universität Passau, DE
- Benjamin Doerr
Ecole Polytechnique –
Palaiseau, FR
- Carola Doerr
Sorbonne University – Paris, FR
- Martin Fyvie
The Robert Gordon University –
Aberdeen, GB
- Nikolaus Hansen
INRIA Saclay –
Palaiseau, FR
- Jose Ignacio Hidalgo
Complutense University of
Madrid, ES
- Joshua D. Knowles
SLB Cambridge Research, GB
- Timo Kötzing
Hasso-Plattner-Institut,
Universität Potsdam, DE
- Martin S. Krejca
Ecole Polytechnique –
Palaiseau, FR
- Johannes Lengler
ETH Zürich, CH
- John McCall
The Robert Gordon University –
Aberdeen, GB
- Aishwarya Radhakrishnan
Hasso-Plattner-Institut,
Universität Potsdam, DE
- Franz Rothlauf
Universität Mainz, DE
- Roberto Santana
University of the Basque Country
– Donostia, ES
- Valentino Santucci
University for Foreigners of
Perugia, IT
- Marcus Schmidbauer
Universität Passau, DE
- Jonathan L. Shapiro
University of Manchester, GB
- Marta Rosa Soto Ortiz
CaixaBank Tech – Madrid, ES
- Andrew M. Sutton
University of Minnesota –
Duluth, US
- Dirk Thierens
Utrecht University, NL
- Vanessa Volz
CWI – Amsterdam, NL
- Carsten Witt
Technical University of Denmark
– Lyngby, DK

