

Sharp Bounds for Genetic Drift in Estimation of Distribution Algorithms (Hot-off-the-Press Track at GECCO 2020)

Benjamin Doerr

Laboratoire d'Informatique (LIX)
École Polytechnique, CNRS
Institut Polytechnique de Paris
Palaiseau, France

Weijie Zheng*

Guangdong Provincial Key Laboratory of Brain-inspired
Intelligent Computation
Department of Computer Science and Engineering
Southern University of Science and Technology
Shenzhen, China

ABSTRACT

Estimation of distribution algorithms (EDAs) are a successful branch of evolutionary algorithms (EAs) that evolve a probabilistic model instead of a population. Analogous to genetic drift in EAs, EDAs also encounter the phenomenon that the random sampling in the model update can move the sampling frequencies to boundary values not justified by the fitness. This can result in a considerable performance loss.

This work gives the first tight quantification of this effect for three EDAs and one ant colony optimizer, namely for the univariate marginal distribution algorithm, the compact genetic algorithm, population-based incremental learning, and the max-min ant system with iteration-best update. Our results allow to choose the parameters of these algorithms in such a way that within a desired runtime, no sampling frequency approaches the boundary values without a clear indication from the objective function.

This paper for the Hot-off-the-Press track at GECCO 2020 summarizes the work "Sharp Bounds for Genetic Drift in Estimation of Distribution Algorithms" by B. Doerr and W. Zheng, which has been accepted for publication in the IEEE Transactions on Evolutionary Computation [5].

CCS CONCEPTS

• **Theory of computation** → **Theory and algorithms for application domains**; *Theory of randomized search heuristics*;

*Also with School of Computer Science and Technology, University of Science and Technology of China, Hefei, China.

This work was supported by a public grant as part of the Investissement d'avenir project, reference ANR-11-LABX-0056-LMH, LabEx LMH, in a joint call with Gaspard Monge Program for optimization, operations research and their interactions with data sciences. This work was also supported by the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (Grant No. 2017ZT07X386); Guangdong Basic and Applied Basic Research Foundation (Grant No. 2019A1515110177); Shenzhen Peacock Plan (Grant No. KQTD2016112514355531); the Program for University Key Laboratory of Guangdong Province (Grant No. 2017KSYS008); and the Science and Technology Innovation Committee Foundation of Shenzhen (Grant No. JCYJ20190809121403553).

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

GECCO '20 Companion, July 8–12, 2020, Cancún, Mexico

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-7127-8/20/07.

<https://doi.org/10.1145/3377929.3397489>

KEYWORDS

Estimation of distribution algorithms (EDAs), genetic drift, running time analysis, theory.

SUMMARY OF OUR RESULTS

Estimation of distribution algorithms (EDAs) are evolutionary algorithms (EAs) that evolve a probabilistic model instead of a population. An iteration of an EDA usually consists of three steps. (i) Based on the current probabilistic model, a population of individuals is sampled. (ii) The fitness of this population is determined. (iii) Update of the probabilistic model: Based on the fitness of this population and the probabilistic model, a new probabilistic model is computed.

Since almost all theoretical results for EDAs regard univariate models [10], this paper deals exclusively with univariate EDAs, that is, the bit positions of the probabilistic model are mutually independent. Univariate EDAs include Population-Based Incremental Learning (PBIL) with special cases Univariate Marginal Distribution Algorithm (UMDA) and Max-Min Ant System with iteration-best update, and the Compact Genetic Algorithm (cGA).

In evolutionary algorithms, it is known that the frequencies of bit values in the population are not only influenced by the contribution of the bit to the fitness, but also by random fluctuation stemming from other bits having a stronger influence on the fitness. These random fluctuations can even lead to certain bits converging to a single value different from the one in the optimal solution. This effect is called *genetic drift* [1, 14].

Genetic drift also happens in EDAs. González, Lozano, and Larrañaga [8] showed that for the 2-dimensional ONEMAX function, the sampling frequency of PBIL can converge to any search point in the search space with probability near to 1 if the initial sampling frequency goes to that search point and the learning rate goes to 1. Droste [6] noticed the possibility of the cGA getting stuck, but he only analyzed the runtime conditional on it being finite, and no analysis of genetic drift or stagnation times was given. Costa, Jones, and Kroese [2] proved that a constant smoothing parameter for the Cross Entropy (CE) algorithm (which is equivalent to a constant learning rate ρ for PBIL) results in that the probability mass function converges to a unit mass at some random candidate, but no convergence speed analysis was given. In summary, as Krejca and Witt said in [10], the genetic drift in EDAs is a general problem of martingales, that is, that a random process with zero expected change will eventually stop at the absorbing boundaries of the range. Results of Witt [16] and Lengler, Sudholt, and Witt [13] as well as the two works by Lehre and Nguyen [12] and by Doerr and

Krejca [3] showed that genetic drift can result in a considerable performance loss.

In the work [5], we quantify this effect asymptotically precisely for several EDAs and this via proven results. The few **previous works** in this direction have obtained the following results. Friedrich, Kötzing, and Krejca [7] showed that for the cGA with hypothetical population size K , the frequency of a neutral bit position is arbitrary close to the borders 0 or 1 after expected $\omega(K^2)$ generations. Though not stated in [7], from its Corollary 9, we can derive an upper bound of $O(K^2)$ for the expected time of leaving the interval $[\frac{1}{4}, \frac{3}{4}]$, and $O(K^2 \log K)$ for expected hitting time of a boundary value. For the UMDA selecting μ best individuals from λ offspring, the situation is similar [7]. After $\omega(\mu)$ iterations, the frequencies are arbitrary close to the boundaries and the expected hitting time can be shown to be $O(\mu \log \mu)$ via similar arguments as above. Sudholt and Witt [15] mentioned that the boundary hitting time of the cGA is $\Theta(K^2)$, but without a complete proof (in particular, because they did not discuss what happens once the frequency leaves the interval $[\frac{1}{6}, \frac{5}{6}]$). Although Krejca and Witt [11] focused on the lower bound of the runtime of the UMDA on ONEMAX, we can derive from it that the hitting time of the boundary 0 is at least $\Omega(\mu)$. This follows from the drift of ϕ in Lemma 9 in [11] together with the additive drift theorem [9].

Our results: While the results above give some indication on the degree of stability of PBIL and the cGA, a sharp proven result is still missing. This paper overcomes this shortage and gives precise asymptotic hitting times for PBIL (including the UMDA and the MMAS_{lb}) and the cGA. With a simultaneous analysis of the UMDA and the cGA, we prove that for the UMDA selecting μ best individuals from λ offspring on some D -dimensional problem, the expected number of iterations until the frequency of the neutral bit position is absorbed in 0 or 1 for the UMDA without margins or when the frequency hits the margins $\{1/D, 1 - 1/D\}$ for the UMDA with such margins is $\Theta(\mu)$, and the corresponding hitting time is $\Theta(K^2)$ for the cGA with hypothetical population size K . This paper also gives a precise asymptotic analysis for PBIL selecting μ best individuals from λ offspring and with a learning rate of ρ : In expectation in $\Theta(\mu/\rho^2)$ generations the sampling frequency of a neutral bit position leaves the interval $[\Theta(\rho/\mu), 1 - \Theta(\rho/\mu)]$ and then always the same value is sampled for this position.

For the lower bounds implicit in these estimates we prove an exponential tail bound.

We also extend the lower bound results to bit positions that are neutral or have a preference for some bit value. For example, we prove that for PBIL it takes an expected number of $\Omega(\mu/\rho^2)$ iterations until the sampling frequency of a position that is neutral or prefers a one (neutral or prefers a zero) reaches the interval $[0, \frac{1}{4}]$ ($[\frac{3}{4}, 1]$). The corresponding hitting time is $\Omega(K^2)$ for the cGA.

Impact: These results are useful both to interpret existing performance results and to set the parameters right in future applications of EDAs. As an example, we note that the recent work [12] shows that the UMDA with $c \log D \leq \mu = o(D)$, c a sufficiently large constant, with $\lambda \leq 71\mu$, and with the margins $1/D$ and $1 - 1/D$, has a weak performance of $\exp(\Omega(\mu))$ on the D -dimensional DECEPTIVELEADINGBLOCKS benchmark function. This runtime is at least some unspecified, but most likely large polynomial in D ; it is

super-polynomial as soon as μ is chosen super-logarithmic. From our work, we know that the expected time for the frequency of a neutral bit position to reach the boundaries is only $O(\mu)$ iterations. Since the DECEPTIVELEADINGBLOCKS function, similar to the classic LEADINGONES function, has many bit positions that for a long time behave like neutral, a value of $\mu = o(D)$ results in that a constant fraction of these currently neutral bit positions will have reached the boundaries at least once within the first D iterations. Hence also without looking at the proof of the result in [12], which indeed exploits the fact that frequencies reach the margins to show the weak performance, our results already indicate that the weak performance might be caused by the use of parameter values leading to strong genetic drift.

Our result also aides in choosing good parameter values for an EDA. If we plan to run the EDA for a given time T , then our result suggests to choose the parameters in such a way that no frequency reaches a boundary before time T due to genetic drift. This can be taken as a general rule of thumb, but can also be used to design automated ways to set the parameters as demonstrated in [4].

REFERENCES

- [1] Hideki Asoh and Heinz Mühlenbein. 1994. On the mean convergence time of evolutionary algorithms without selection and mutation. In *Parallel Problem Solving from Nature, PPSN 1994*. Springer, 88–97.
- [2] Andre Costa, Owen Dafydd Jones, and Dirk Kroese. 2007. Convergence properties of the cross-entropy method for discrete optimization. *Operations Research Letters* 35 (2007), 573–580.
- [3] Benjamin Doerr and Martin S. Krejca. 2020. The univariate marginal distribution algorithm copes well with deception and epistasis. In *Evolutionary Computation in Combinatorial Optimization, EvoCOP 2020*. Springer, 51–66.
- [4] Benjamin Doerr and Weijie Zheng. 2020. From understanding genetic drift to a smart-restart parameter-less compact genetic algorithm. In *Genetic and Evolutionary Computation Conference, GECCO 2020*. ACM. <https://doi.org/10.1145/3377930.3390163> To appear.
- [5] Benjamin Doerr and Weijie Zheng. 2020. Sharp bounds for genetic drift in estimation-of-distribution algorithms. *IEEE Transactions on Evolutionary Computation* (2020). <https://doi.org/10.1109/TEVC.2020.2987361> To appear.
- [6] Stefan Droste. 2006. A rigorous analysis of the compact genetic algorithm for linear functions. *Natural Computing* 5 (2006), 257–283.
- [7] Tobias Friedrich, Timo Kötzing, and Martin S. Krejca. 2016. EDAs cannot be balanced and stable. In *Genetic and Evolutionary Computation Conference, GECCO 2016*. ACM, 1139–1146.
- [8] Cristina González, Jose A. Lozano, and Pedro Larrañaga. 2001. The convergence behavior of the PBIL algorithm: a preliminary approach. In *International Conference on Artificial Neural Nets and Genetic Algorithms, ICANNGA 2001*. Springer, 228–231.
- [9] Jun He and Xin Yao. 2001. Drift analysis and average time complexity of evolutionary algorithms. *Artificial Intelligence* 127 (2001), 51–81.
- [10] Martin Krejca and Carsten Witt. 2020. Theory of estimation-of-distribution algorithms. In *Theory of Evolutionary Computation: Recent Developments in Discrete Optimization*, Benjamin Doerr and Frank Neumann (Eds.). Springer, 405–442. Also available at <https://arxiv.org/abs/1806.05392>.
- [11] Martin S. Krejca and Carsten Witt. 2020. Lower bounds on the run time of the Univariate Marginal Distribution Algorithm on OneMax. *Theoretical Computer Science* (2020). <https://doi.org/10.1016/j.tcs.2018.06.004> To appear.
- [12] Per Kristian Lehre and Phan Trung Hai Nguyen. 2019. On the limitations of the univariate marginal distribution algorithm to deception and where bivariate EDAs might help. In *Foundations of Genetic Algorithms, FOGA 2019*. ACM, 154–168.
- [13] Johannes Lengler, Dirk Sudholt, and Carsten Witt. 2018. Medium step sizes are harmful for the compact genetic algorithm. In *Genetic and Evolutionary Computation Conference, GECCO 2018*. ACM, 1499–1506.
- [14] Kimura Motoo. 1964. Diffusion models in population genetics. *Journal of Applied Probability* 1 (1964), 177–232.
- [15] Dirk Sudholt and Carsten Witt. 2019. On the choice of the update strength in estimation-of-distribution algorithms and ant colony optimization. *Algorithmica* 81 (2019), 1450–1489.
- [16] Carsten Witt. 2019. Upper bounds on the running time of the univariate marginal distribution algorithm on OneMax. *Algorithmica* 81 (2019), 632–667.