Ranking Empirical Cumulative Distribution Functions using Stochastic and Pareto Dominance

Hao Wang
LIACS, Leiden University
Leiden, Netherlands
h.wang@liacs.leidenuniv.nl

Thomas Bäck
LIACS, Leiden University
Leiden, Netherlands
t.h.w.baeck@liacs.leidenuniv.nl

ABSTRACT

In this paper, we propose two new approaches to rank the frequently used empirical cumulative distribution functions (ECDFs) for performance assessment of stochastic optimization algorithms. In the first approach, the different orders of stochastic dominance among running lengths are adopted in a hierarchical manner: the first order stochastic dominance is tested and the second order is used when the first order leads to incomparable results. In the second approach, ECDFs are considered as local Pareto front of the bi-criteria decision-making problem, in which one objective is to achieve a high success rate and the other is to use as few function evaluations as possible. In this case, it is proposed to adopt the multi-objective performance indicator to handle incomparable ECDFs.

CCS CONCEPTS

• **Theory of computation → Theory of randomized search heuristics**; *Random search heuristics*;

KEYWORDS

Empirical cumulative distribution functions, Stochastic dominance, Pareto dominance

ACM Reference Format:

Hao Wang and Thomas Bäck. 2018. Ranking Empirical Cumulative Distribution Functions using Stochastic and Pareto Dominance. In *GECCO '18 Companion: Genetic and Evolutionary Computation Conference Companion, July 15–19, 2018, Kyoto, Japan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3205651.3205683

1 INTRODUCTION

Many performance metrics have been applied in benchmarking the stochastic optimizers, e.g., expected running time (ERT), fixed cost error (FCE) and *Empirical Cumulative Distribution Functions* (ECDFs). ECDFs are of particular interest due to fact that they exhibit performance information at *any-time* of the optimization process. In the work, we shall restrict the discussion to ECDFs of the running length. Let X denote the running length of *successful* runs for an stochastic optimizer under budget N and target precision Δf. In addition, let $F_X : [0, N] \rightarrow \mathbb{R}_{\geq 0}$ represent the ECDF of X.

GECCO '18 Companion, July 15–19, 2018, Kyoto, Japan

ACM ISBN 978-1-4503-5764-7/18/07…$15.00
https://doi.org/10.1145/3205651.3205683

The empirical comparison of two running lengths X and Y is considered as the task of assigning a (partial) order to the corresponding ECDFs, which are estimated from benchmarking data. For example, it is common to take the right-most point of ECDFs for the ranking. This method suffers from loosing information on any-time performance. Alternatively, the area under (above) the ECDF curve can also be used as average performance. However, this approach requires a rigorous formulation. In the paper, two ranking approaches are proposed using stochastic and Pareto dominances.

2 STOCHASTIC DOMINANCE

2.1 First Order Dominance

The first order stochastic dominance (FSD) [4, 6] is typically defined in the ascending manner: a random variable Y first order stochastically dominates random variable X ($X \leq_1 Y$), if and only if

$$\forall t \geq 0 : \quad F_X(t) \geq F_Y(t) \wedge \exists t \geq 0 : \quad F_X(t) > F_Y(t).$$

Note that if the strict inequality does not hold, then two distributions F_X and F_Y are the same. For ECDFs, this order indicates F_X is always **not below** F_Y, as shown in Fig. 1: algorithm A outperforms algorithm B at **any-time in terms of success rate** because its cumulative distribution function exhibits success rates that are always not less than that of B. When the crossing occurs, X and Y become incomparable according to this definition (thus $\leq_1$ is a partial order) and we have to resort to the second order dominance.

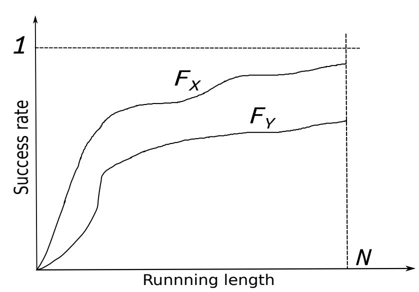

Figure 1: First order stochastic dominance: $X \leq_1 Y$.

2.2 Second Order Dominance

The second order stochastic dominance (SSD) [3, 6] strengthens the idea of comparing the area under the ECDF curves. Running length X is stochastically dominated by Y in the second order ($X \leq_2 Y$), if and only if

(1) $\forall t \geq 0, \quad \int_0^t [F_X(u) - F_Y(u)] \, \mathrm{d}u \geq 0$, and
(2) The strict inequality holds for some t.

Basically, this ordering can be interpreted as: the relation $X \leq_2 Y$ holds as long as the *cumulated* c.d.f. of X has to be greater or equal to that of Y for all points on the support $[0, N]$. Although cumulating the c.d.f. looks strange at the first glance, consider the following relation between the expectation and the c.d.f.: $\mathrm{E}[X] = \int_0^N (1 - F(x)) \, \mathrm{d}x = N - \int_0^N F(x) \, \mathrm{d}x$. In this sense, the condition $\int_0^t F_X(u) \, \mathrm{d}u \geq \int_0^t F_Y(u) \, \mathrm{d}u$ implies the *conditional expectations*

$$\forall t \in [0, N], \quad \mathrm{E}[X | X < t] \leq \mathrm{E}[Y | Y < t].$$

Or equivalently, algorithm A outperforms algorithm B **at any-time in terms of running lengths**. This order is illustrated in Fig. 2. Unfortunately, SSD is still a partial order. For instance, in Fig. 2

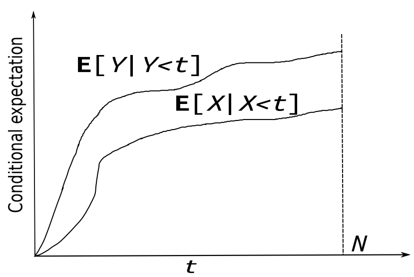

Figure 2: Second order stochastic dominance: $X \leq_2 Y$.

where the relation $X \leq_2 Y$ is assumed, if we run both algorithms a bit longer than N the c.d.f. of Y could climb even higher such that the condition is violated. In this case, it is suggested to test those algorithms in the third order stochastic dominance.

2.3 Third Order Dominance

In the third order stochastic dominance (TSD, denoted as $X \leq_3 Y$) [5], the condition is built by cumulating the area under ECDFs:

(1) $\forall t \geq 0, \quad \int_0^t \int_0^r F_X(u) \, \mathrm{d}u \, \mathrm{d}r \geq \int_0^t \int_0^r F_Y(u) \, \mathrm{d}u \, \mathrm{d}r$,
(2) $\mathrm{E}[X] \leq \mathrm{E}[Y]$ and at least one strict inequality holds.

In all, to rank the performance of algorithms, it is proposed to test a pair of ECDFs hierarchically by firstly using the FSD, testing on SSD if FSD leads to incomparable results and finally trying TSD if SSD fails. When two algorithms are incomparable even on TSD, it is suggested to stop from going up to the higher order of stochastic dominance as the computational time in calculating the multiple integrals becomes intractable in practice.

3 PARETO DOMINANCE

As an alternative to the stochastic dominance, ECDFs are formulated as a (local) *Pareto front*. The function evaluation budget can be treated as a performance metric (cost) of algorithm A as it bounds the actual running time from above. With the other performance metric: success rate, it is straightforward to formulate a **bi-criteria decision making** task: we are looking for an algorithm that gives the highest success rate r with the lowest evaluation budget setting N. We shall call each pair (r, N) as a performance point of an algorithm. In addition, Pareto dominance can be established on performance points: for $p_1 = (r_1, N_1), p_2 = (r_2, N_2)$, p_1 (Pareto) dominates p_2 if and only if $r_1 \geq r_2, N_1 \leq N_2$ and at least one inequality holds. This consideration is illustrated in Fig. 3. The ideal point of the problem is $(0, 1)$ while the nadir point is $[N, 0]$. As ECDFs

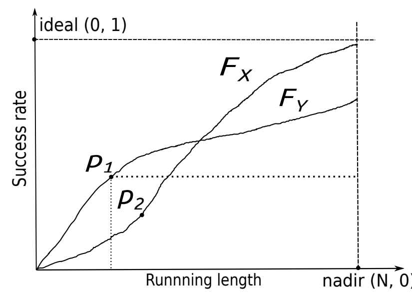

Figure 3: Pareto dominance on ECDFs.

are non-decreasing functions, each of them forms a Pareto front naturally. Consequently, the dominance relation between ECDFs can be formulated using Pareto dominance. Representing ECDFs as a collection of performance points: $F_X = \{p_1, p_2, \ldots\}$ and $F_Y = \{p'_1, p'_2, \ldots\}$, F_X Pareto dominates F_Y if and only if each p'_i is dominated by at least one point in F_X. Under this setting, if two ECDFs are incomparable in terms of Pareto dominance, it is proposed to adopt the well-known multi-objective performance indicators for the rankings, e.g., hypervolume indicator [7], $R2$ indicator [1] and inverted generational distance [2]. The choice of multi-objective performance indicators remains open. Note that, when using the hypervolume indicator, the resulting Pareto dominance compares the expectation of running length, namely $\mathrm{E}[X]$ and $\mathrm{E}[Y]$. It is not the same as SSD because SSD examines the conditional expectation $\mathrm{E}[X | X < t]$ for all t in $[0, N]$.

4 FUTURE RESEARCH

In this paper, two approaches based on stochastic and Pareto dominance are proposed to rank empirical cumulative distribution functions. For the next step, the proposed ranking method will be implemented and tested on a benchmark.

REFERENCES

[1] Dimo Brockhoff, Tobias Wagner, and Heike Trautmann. 2012. On the Properties of the R2 Indicator. In *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation (GECCO '12)*. ACM, New York, NY, USA, 465–472. https://doi.org/10.1145/2330163.2330230

[2] Carlos A Coello Coello and Margarita Reyes Sierra. 2004. A study of the parallelization of a coevolutionary multi-objective evolutionary algorithm. In *Mexican International Conference on Artificial Intelligence*. Springer, 688–697.

[3] Josef Hadar and William R Russell. 1969. Rules for ordering uncertain prospects. *The American economic review* 59, 1 (1969), 25–34.

[4] Giora Hanoch and Haim Levy. 1975. The efficiency analysis of choices involving risk. In *Stochastic Optimization Models in Finance*. Elsevier, 89–100.

[5] George A Whitmore. 1970. Third-degree stochastic dominance. *The American Economic Review* 60, 3 (1970), 457–459.

[6] Elmar Wolfstetter, U Dulleck, R Inderst, P Kuhbier, and M Lands-Berger. 1993. *Stochastic dominance: theory and applications*. Humboldt-Univ., Wirtschaftswiss. Fak.

[7] Eckart Zitzler and Lothar Thiele. 1999. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE transactions on Evolutionary Computation* 3, 4 (1999), 257–271.