Quantile-like measures on multi-dimensional distributions of closed sets

Johann Dreo

To cite this version:
Johann Dreo. Quantile-like measures on multi-dimensional distributions of closed sets: Application in stochastic optimization. Doctoral. Innovation, Design, and Engineering, Västerås [Sweden], France. 2023. hal-04110682

HAL Id: hal-04110682
https://hal.science/hal-04110682
Submitted on 30 May 2023
Quantile-like measures on multi-dimensional distributions of closed sets

Application in stochastic optimization

• Johann Dreo • 2021-06-10
Abstract

- Algorithm to compute quantiles of 2D (and 3D) distributions.
- Well-founded statistics on top of them.
- Useful for multi-objective optimization problems.
- Example: automated tuning of stochastic optimization solvers.

Figure 3: Projections of the performance front of the D&E method, for a setting of two mutation parameters against two objectives. Right plot shows the superimposition of the 30 performances fronts, left one shows a density estimation with gaussian kernels.

[Dre09]
Summary

01
Performance estimation of optimization solvers

02
Quantiles on joint distributions

03
Implementations

04
Statistics

05
Back to Optimization

06
Supplementary material
Part 1

Performance estimation of optimization solvers
Performance(s): quality, time and probability

Optimization

\[ q_0 = q^* \]

\[ |f(\hat{x}) - f(x^*)| = q^* \]

\[ 0 = \hat{q} \]
Performance(s): quality, time and probability

Terminal qualities

\[ |f(\hat{x}) - f(x^*)| = q^* \]

\[ q \quad q^* \quad q_0 \ldots q_r \]
Performance(s): quality, time and probability

Convergence trajectories

\[ |f(\hat{x}) - f(x^*)| \leq q^* \]

\[ 0 = q \]

\[ q_0 \rightarrow q^* \]

\[ q \rightarrow t \]
Performance(s): quality, time and probability

Terminal distributions

Terminal distributions
Performance(s): quality, time and probability

Convergence distributions
Performance(s): quality, time and probability

**Expected Quality ECDF** (BBComp 2015)
Performance(s): quality, time and probability

**Expected RunTime ECDF** [HAB+16]

\[
q = q^\ast \quad |f(x) - f(x^\ast)| = q^* \quad 0 = q \\
p = q^\ast \quad P(q^\ast = q) \\
q = q^\ast \quad P(q^\ast < q) \\
qu = q^\ast \quad \text{Expected Quality} \\
q = q^\ast \quad \text{Expected RunTime}
\]
Performance(s): quality, time and probability

QT Empirical Attainment Function Levelsets [LIS14]
Performance(s): quality, time and probability

\[ |f(\hat{x}) - f(x^*)| = q^* \quad 0 = \hat{q} \]

\[ P(q^* = q) \]

\[ P(q^* < q) \]

Expected Quality

Expected RunTime

QT Empirical Attainment Function

Reference: [Kno05]

Johann Dreo

Quantile-like measures on multi-dimensional distributions of closed sets

2021-06-10
Quality-Time Empirical Attainment Function

- EQ- and ERT-ECDF are trivially computed:
  - fix a target quality (resp. time),
  - traverse all runs across time (resp. quality),
  - compute the ratio of better-than-target.

- QT-EAF requires a more complex algorithm [GdFF02].
Part 2

Quantiles on joint distributions
Properties

Closeness

- A set of optimization trajectories forms a closed-set.
- Bounded by:
  - optimal solution, \( q \in [0, \text{bound}[ \),
  - time budget, \( t \in [1, \text{budget}[ \),
  - \( P \in [0, 1] \)
Properties

Monotony

- Convergence trajectories are Pareto-optimal.
- Sequence of non-dominated points\(^a\).

\(^a\)Example from [LIS14].

---

**Fig. 6.** For each performance profile, the legend shows the classical hypervolume, and the weighted hypervolume variants \(w^{\text{opt}}\) and \(w^{\text{equal}}\).
Algorithm

Levelsets on finite sample

- Computes level sets of the distribution [GdFFH01].
- Which are essentially equivalent to quantiles.
- Because the sample is finite, there is a finite number of levelsets.
- At most $r$ level sets for $r$ input sets.
Algorithm

Example  (From [LIPS10])

Fig. 9.1: Ten independent outcomes obtained by an SLS algorithm applied to an instance of a biobjective optimization problem. In the right plot, the same outcomes are shown but points belonging to the same run are joined with a line.

Fig. 9.3: Three plots of attainment surfaces for 15 (left), 50 (middle), and 200 (right) independent runs of the same algorithm on the same problem instance.
Algorithm

2 dimensions

- (Assuming minimization on both axis).
- “Peel” level sets.
- Sweep one axis in increasing values,
- sweep the other in decreasing values.
- Essentially computes incremental Pareto-optimal archives.
- $O(m \log m + nm)$, $m$ points in $n$ runs (asymptotically optimal).
Algorithm

3 dimensions

- $O(n^2 m \log m)$, $m$ points in $n$ runs,
- a logarithmic factor worse than the upper bound.
- Output *surfaces* instead of level sets.
Part 3

Implementations
2D Empirical Attainment

EAF and EAH

- Two options:
  - EA Function (metric space)
  - EA Histogram (discrete buckets)
2D Empirical Attainment

Pros and cons

- If all level sets are computed, EAF is the true function (given the sample).
- EAH is an approximation (which converges with the discretization). . .
- . . . but can be computed on log-log scales, for better resolution.

- EAH scales better regarding the number of points. . .
- . . . but require more memory.
2D Empirical Attainment

- **C++** implementation.
- **Within IOHexperimenter**
  [https://iohprofiler.github.io/IOHexp/](https://iohprofiler.github.io/IOHexp/).
- May be ported if needed.
Statistics for bi-objective problems

Why?

- Multi-objective problems:
  - either Pareto-optimal approaches (heavy on user),
  - either objectives aggregation (not good math properties).
  - How to aggregate randomized observations?
Statistics for bi-objective problems

Why?

- Multi-objective problems:
  - either Pareto-optimal approaches (heavy on user),
  - either objectives aggregation (not good math properties).
  - How to aggregate randomized observations?

- EA[FH] is a way to aggregate Pareto-optimal fronts.

- One can compute statistics on it.
Statistics for bi-objective problems

Examples

- Orthogonal partial section statistics:
  - Area Under curves (EQ-ECDF and ERT-ECDF).
  - Attainment surface (EAF).

- Global statistics:
  - Volume under the EAF ≈ sum/mean-like.
  - Volume under a levelset ≈ quantile-like.
  - Volume under a subset of levelsets (scaling approximation).
  - Covariance [GdFFH01].
Part 5

Back to Optimization
Example of use

Automated design

- Consider the automated design of an optimization solver as a (meta) bi-objective optimization problem [Dre03].
- We want to optimize both quality and time.
- Because we have no clue about budget or target in advance.
- Maximize an average aggregate? No.
- Maximize volume under the EAF.

Figure 2: Dominance map for three instances CMA-ES on the BBOB Katsuura Lunacek Bi-Rastrigin function, dimension 50, first instance, with different population sizes (upper-left/blue: default population size, lower-left/cyan: 1/2 times the default, upper-right/brown: 2 times the default). White area shows the domain never attained by any algorithm, while gray area shows the domain attained by at least two algorithms with the same probability and colored area the domain where a single algorithm attain the higher probability.
Conclusion

Recall

- Algorithm to compute quantiles of 2D (and 3D) distributions.
- Well-founded statistics on top of them.
- Useful for multi-objective optimization problems.
- Example: automated tuning of stochastic optimization solvers.
Conclusion

Perspective

- Quantify the time/memory/loss compromises.
- Impact of the statistic choice on the optimization sub-problem.
- Use in multi-objective problems.
Time for questions

Fig. 4 Visualization of the EAFs associated to the outcomes of two algorithms (top) and the corresponding differences between the EAFs (bottom left: differences in favour of Algorithm 1; bottom right: differences in favour of Algorithm 2). In the top, the gray level encodes the value of the EAF. In the bottom, the gray level encodes the magnitude of the observed difference.
Part 6

Supplementary material
Johann Dreo.

*Adaptation de la métahéuristique des colonies de fourmis pour l’optimisation difficile en variables continues. Application en génie biologique et médical.*

Theses, Université Paris XII Val de Marne, December 2003.

Johann Dreo.

*Using performance fronts for parameter setting of stochastic metaheuristics.*


Viviane Grunert da Fonseca and Carlos M. Fonseca.

*A link between the multivariate cumulative distribution function and the hitting function for random closed sets.*

Viviane Grunert da Fonseca, Carlos M. Fonseca, and Andreia O. Hall.
Inferential Performance Assessment of Stochastic Optimisers and the Attainment Function.

Nikolaus Hansen, Anne Auger, Dimo Brockhoff, Dejan Tušar, and Tea Tušar.
COCO: Performance Assessment.
Joshua Knowles.
A summary-attainment-surface plotting method for visualizing the performance of stochastic multiobjective optimizers.
ZSCC: 0000142 ISSN: 2164-7151.

Manuel López-Ibáñez, Luís Paquete, and Thomas Stützle.
Exploratory Analysis of Stochastic Local Search Algorithms in Biobjective Optimization.
ZSCC: NoCitationData[s0].

Manuel López-Ibáñez and Thomas Stützle.
Automatically improving the anytime behaviour of optimisation algorithms.
Formal concepts

Introduction

- Optimization = find the optimum $x^*$ minimizing an objective function $f$:

$$\hat{x} = \arg \min_{x \in X} f(x)$$

- Randomized search heuristics approximates the optimum:

$$f(x_r^*) \leq f(\hat{x}) + \epsilon$$
Formal concepts

Terminal distribution

- The output distribution of a randomized search heuristic is the probability to reach a given quality target:

\[
F(q) = P[0 < Q \leq q] = \int_{0}^{q} \lim_{r \to \infty} P[\exists r \in \mathbb{N}^+ | f(x_r^*) \leq q] dq
\]
Formal concepts

Temporal convergence

- Because any solver (should) be quasi-ergodic:
  \[ P[x^* = \hat{x}] > 0 \]
- Then it (should) converge in a finite time:
  \[ \lim_{t \to \infty} P[f(x_t^*) = f(\hat{x})] = 1 \]
- Thus the probability of attaining the target should not decrease over time:
  \[ F(t) := P[f(x_t^*) = f(\hat{x})] \]
  \[ P[\exists r \in \mathbb{N}^+ | F_r(t - 1) \leq F_r(t)] > 0 \]
- Hence the trajectory in objective space of a run is monotonic.
Formal concepts

Pareto Optimality

- We say that \( u \) dominates \( v \) \((u \prec v)\) iff

\[
    u \neq v \land q(u) \leq q(v) \land t(u) \leq t(v)
\]

- The trajectory in objective space being monotonic, all its points are non-dominated, and the set is Pareto-optimal.
Formal concepts

QT-EAF

- Given a set of non-dominated sets:

\[
X_p = \{(t_1, q_1) \ldots (t_m, q_m)\} \setminus \\
\{(t_1, q_1) \ldots (t_m, q_m)| \\
\n\#(t_a, q_a), (t_a, q_a) \\
\land a \neq b \\
\land (t_a, q_a) \preceq (t_b, q_b)\}
\]

- Given the indicator function \( I(\cdot) : \mathbb{R}^2 \mapsto \{0, 1\} \)

- The Empirical Attainment Function \( EAF_r(\cdot) : \mathbb{R}^2 \mapsto [0, 1] \) is:

\[
EAF_r(z) = EAF_r(X_1, \ldots, X_r, z) := \frac{1}{r} \sum_{i=1}^{r} I(X_i \preceq z)
\]
Formal concepts

QT-EAF level set

- The $k$ level sets $QAL$ of the $EAF$:

$$Z_k(\epsilon) = \{ z \in \mathbb{R}^2 | EAF_r(z) \geq k \land EAF_r(z - \epsilon) < k \}$$

$$QAL_k(z) = \lim_{\epsilon \to 0} Z_k(\epsilon)$$