Free Lunch for Optimisation under the Universal Distribution

Tom Everitt[1] Tor Lattimore[2] Marcus Hutter[3]

[1]Stockholm University, Stockholm, Sweden

[2]University of Alberta, Edmonton, Canada

[3]Australian National University, Canberra, Australia

July 7, 2014

Outline

Are universal optimisation algorithms possible?

- Background: Finite Black-box Optimisation (FBBO) and the NFL theorems
- The Universal Distribution
- Our results
- Conclusions and Outlook

Finite Black-box Optimisation

FBBO is a formal setting for Simulated Annealing, Genetic Algorithms, etc.

It is characterized by:

- Finite search space X, finite range Y, unknown $f\colon X \to Y$.
- An optimisation algorithm repeatedly chooses points $x_i \in X$ to evaluate.
- Goal: Minimimize probes-till-max (Optimisation Time).
- Distribution P over the finite set $\{f : X \to Y\} = Y^X$.
- P-expected Optimisation Time:

$$\text{Perf}^P(a) = \mathbb{E}_P[\text{probes-till-max}(a)]$$

P affects bounds on optimisation performance.

The NFL (No Free Lunch) theorems

Definition

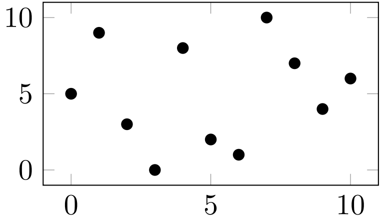
There is NFL for P if $\mathrm{Perf}^P(a) = \mathrm{Perf}^P(b)$.

Theorem (Original NFL (Wolpert&Macready, 1997))

P uniform $\implies$ NFL for P.

$\implies$ so no universal optimisation?

Uniform = unbiased? Uniform means random noise.



Our suggestion to avoid NFL: *The Universal Distribution* (not new).

The Universal Distribution – Background

Kolmogorov complexity: $K(x) := \min_p\{\ell(p) : p \text{ prints } x\}$

Universal distribution: $\mathbf{m}(x) := 2^{-K(x)}$

Example:

	000000000	0101001101
K	Low	High
$\mathbf{m}$	High	Low

- Agrees with Occam's razor with "simplicity bias"
- Dominates all (semi-)computable (semi-)measures
- Essentially regrouping invariant

Offers mathematical solution to the induction problem (Solomonoff induction). Successfully used in Reinforcement Learning (Hutter, 2005), and for general clustering algorithm (Cilibrasi&Vitanyi, 2003)

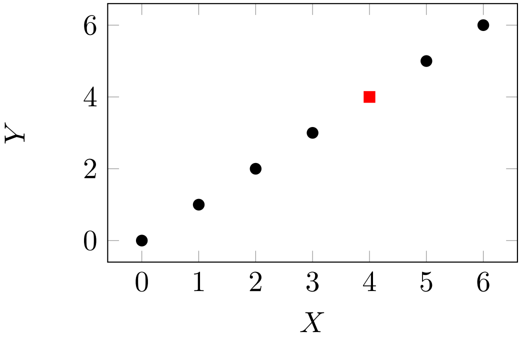
The Universal Distribution in FBBO

May equivalently be defined in two ways:

$$\mathbf{m}_{XY}(f) := 2^{-K(f|X,Y)} \tag{1}$$

$$\approx \text{“the probability that a ‘random’ program acts like } f\text{”} \tag{2}$$

(1) shows bias towards simplicity



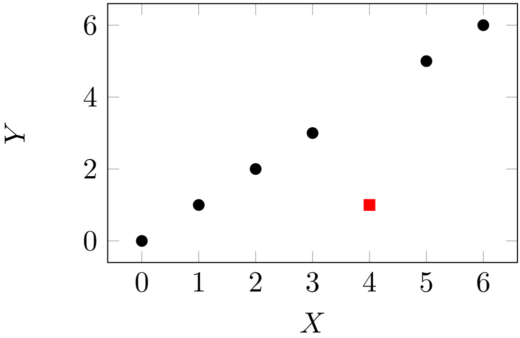
The Universal Distribution in FBBO

May equivalently be defined in two ways:

$$\mathbf{m}_{XY}(f) := 2^{-K(f|X,Y)} \quad (1)$$

$$\approx \text{"the probability that a 'random' program acts like } f\text{"} \quad (2)$$

(1) shows bias towards simplicity



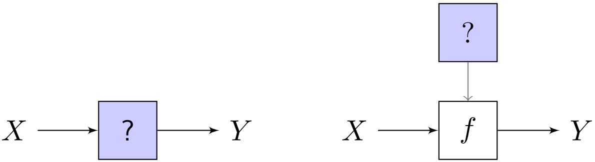
The Universal Distribution in FBBO

May equivalently be defined in two ways:

$$\mathbf{m}_{XY}(f) := 2^{-K(f|X,Y)} \quad (1)$$

$$\approx \text{“the probability that a ‘random’ program acts like } f\text{”} \quad (2)$$

(2) shows the wide applicability of the universal distribution.



Uniform distribution Universal distribution

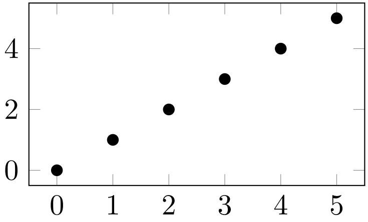
The uncertainty pertains to the *system behind* the mapping.

The universal distribution permits free lunch

Theorem (Universal Free Lunch)

There is free lunch under the universal distribution for all sufficiently large search spaces.

Follows from simplicity bias:



Results – Bad News

Unfortunately, the universal distribution does not permit *sublinear* maximum finding

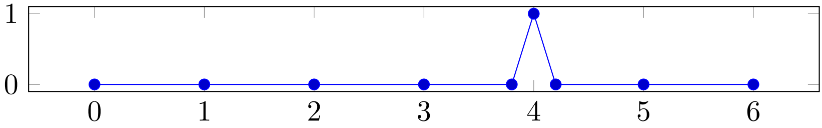
Theorem (Asymptotic bounds)

Expected optimisation time increases linearly with the size of the search space.

Optimisation is a hard problem.

Degenerate functions impede performance (NIAH-functions and "adversarial" functions).

Needle-in-a-haystack function:



Conclusions and Outlook

The universal distribution is a philosophically justified prior for finite black-box optimisation.

It offers free lunch, but not sublinear maximum finding. So meta-heuristics with different universal performance exist, but the difference is limited.

Future research: Minimal condition enabling sublinear maximum finding.

References

Rudi Cilibrasi and Paul M B Vitanyi.
Clustering by compression.
IEEE Transactions on Information Theory, 51(4):27, 2003.

Marcus Hutter.
Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability.
Lecture Notes in Artificial Intelligence (LNAI 2167). Springer, 2005.

David H Wolpert and William G Macready.
No free lunch theorems for optimization.
IEEE Transactions on Evolutionary Computation, 1(1):270–283, 1997.