
APPENDIX: STUDY CASES

A SURVEY OF NONPARAMETRIC TESTS FOR THE STATISTICAL ANALYSIS OF EVOLUTIONARY COMPUTATION EXPERIMENTS

Rafael Lahoz-Beltra , Carlos Perales Gravan

NIT 2010. International Conference Natural Information Technologies. September 28 - October 02, 2010, Madrid (Spain). <http://bioinformatica.net/tests/survey.html>

Statistical analysis of performance. Study cases

Study case 1. Simple genetic algorithm

The performance of a simple genetic algorithm was evaluated using the classical program SGA written in Pascal and described in Goldberg (1989). The simulation experiments were carried out with a population size (N) of 20 chromosomes (solutions are defined as a binary strings), using as genetic operators one-point crossover and mutation (simulated by the flip-a-bit method). In the experiment, the crossover probability was set up to 75% and the mutation probability was equal to 5%. The problem to be optimized is the one consisting in maximizing the function $f(x)=x^2$, where x is the decimal value of a binary string, thus the chromosome, composed of 12 bits. The fitness function is the power function $f(x)=(x/c)^n$, where c has been chosen equal to 1000 to normalize x , and $n=2$. Starting with a random population of chromosomes, selection, crossover and mutation were simulated (Goldberg, 1989; Davis, 1991; Lahoz-Beltra, 2004), obtaining generations of equal population size. The experiment was carried out during 20 generations registering for each generation the chromosome fitness value $f(x_i)$ as well as the minimum and maximum fitness values, thus f_t^{\min} and f_t^{\max} . A classical on-line performance measure, the average fitness, was also calculated:

$$\bar{f} = \frac{\sum_{i=1}^N f(x_i)}{N} \quad (3)$$

Study case 2. Ant population evolution

MicroAnts1.0 is an educational and entertaining software (Prata, 1993; Lahoz-Beltra, 2008) written by Steve Wright that uses genetic algorithms to simulate the evolution of an ant colony. During simulation each ant has a chromosome made up of 10 bits, being genes defined as is shown in Table II. In the present experiment we defined a target ant with chromosome 1111111010, thus an ant with the following general phenotype (features list): 'ant can see 15 squares away (gene 1 = 1111), moves in a random direction (gene 2 = 11), avoid poison (gene 3 = 1), mate if food level >500 (gene 4 = 0), only shares its food if the other ant's 9, 10 bits are equal to 10 (gene 5 = 10)'. If an ant moves onto a food spot then the ant adds 10 to its food level, otherwise it loses 1 food level. Food and poison are both randomly place. Mating may occur when two ants occupy the same location reducing each ant food level in 500, resulting a 'baby ant' after one point crossover and mutation (flip-a-bit method). The new ant starts out with a food level of 50 being randomly placed on the environment.

We performed a simulation experiment setting up the following parameters: initial number of ants=30, initial number of food=30, food grown in one year=remain constant, maximum number of food=150, number of poison=50, crossover rate=75% and mutation rate=5%. Once the simulation was started we sampled the ant population at 1, 2, 3, 4 and 5 minutes, registering the chromosome of each ant currently 'alive' in the simulation. Based on the obtained data the Hamming distance between each ant's chromosome and the target one was calculated as follows:

$$d(a,t) = \sum_{k=1}^l a[k] \text{ xor } t[k] \quad (4)$$

where $a[k]$ and $t[k]$ are the k^{th} bit values in the chromosomes of the ant a in the population and the target ant t , with l (chromosome length) equal to 10. Note how this distance is a measure of the genotypic (combination of gene values located on chromosomes) similarity between an ant in the population and the target one. Using the Hamming metric we also calculated the genetic drift (Frederick et al., 1993) in the current generation or sample time. It is a measure of the movement of the population towards a target, best adapted or fittest population member. The genetic drift was obtained as the average Hamming distance of the ants or individuals a_i in the population from the target ant t :

$$\text{Genetic drift} = \frac{\sum_{i=1}^N d(a_i, t)}{N} \quad (5)$$

Study case 3. Dawkin's biomorph evolution

This model (Prata, 1993; Lahoz-Beltra, 2008) simulates the evolution of artificial creatures known as 'biomorphs' introduced by Richard Dawkins in the book *The Blind Watchmaker* (Dawkins, 1986). The simulation illustrates how creatures with asexual reproduction can evolve by the accumulation or selection of those useful random changes or mutations. Once a progenitor is selected the offspring is obtained being each descendant similar to the progenitor excepting a mutation that occurred in its chromosome. Each biomorph has a chromosome made up of 8 integer numbers, being the genes defined as is shown in Table III. Based on the present model (Prata, 1993) we carried out a simulation experiment defining a target biomorph. The target biomorph had a chromosome 3,1,7,1,3,128,89,8, thus a biomorph with the following general phenotype (features list): 'biomorph with the offspring of size 3 (gene 1), plotted with 1 iteration (gene 2), being 7 (gene 3) and 1 (gene 4) the values used to calculate the branch length. Since gene 5 is equal to 3 (odd) then branch length is calculated as the product of gene 3 by gene 4, thus 7. The angle of 1st branch is 128 (gene 6), and the angle of 2nd branch is 89 (gene 7), being the length of the trunk (initial branch) equal to 8 (gene 8)'.

In the experiment the mutation probability is always equal to 100%, applying this genetic operator to the offspring. Once the simulation was started we registered the chromosome of each biomorph at each generation, obtaining the Euclidean distance between the biomorph b_i in the population at the current generation, from the target biomorph t_i :

$$d(b,t) = \sqrt{\sum_{i=1}^{gene} (b_i - t_i)^2} \quad (6)$$

where *gene* is the total number of genes in the chromosome, thus eight in the present model. Note how this distance is a measure of the genotypic similarity between a biomorph in the population and the target one. Finally, we also calculated the average Euclidean distance of the biomorphs in the population at the current generation, from the target biomorph:

$$\bar{d} = \sum_{j=1}^N \frac{\sqrt{\sum_{i=1}^{gene} (b_i - t_i)^2}}{N} \quad (7)$$

Note how this measure is similar to a genetic drift, thus a measure of the movement of the population towards a target, best adapted or fittest population member.

Statistical comparisons. Study cases

The statistical comparison of two or more series of computer simulation results is a common analysis in Evolutionary Computation (EC). The statistical comparison of two or more EC batch of data is accomplished by the juxtaposition of two or more boxplots together including a Mann-Whitney or Kruskal-Wallis tests, respectively.

Study case 4. Experiments with a simple genetic algorithm

In this experiment we used the program SGA described in the *study case 1*. A first, second and third simulation experiments were carried out with population sizes (N) equal to 40 chromosomes. In the first experiment the crossover and mutation probabilities were set equal to 75% and 5%, the second experiment was carried out with only crossover being its probability equal to 75%, and the third one with only mutation being its probability equal to 5%. The algorithm cycles through epochs searching for an optimum $f(x)=x^2$, where x is the decimal value of a binary chromosome of 12 bits length, until a maximum of 20 generations is reached. Starting with a random population of chromosomes, selection, crossover and/or mutation were simulated (Goldberg, 1989; Davis, 1991; Lahoz-Beltra, 2004) obtaining generations of equal population size. Once the experiments were finished we compared the first experiment with second one, as well as the three experiments together.

Study case 5. Ant colony optimization experiments

Ant Colony Optimization (ACO) (Dorigo and Gambardella, 1997) is an example of evolutionary method where a set of agents called artificial ants search for good solutions to a given optimization problem (i.e. the travelling salesman problem, TSP). Solutions search is stochastic and biased by a pheromone model. For instance, to apply ACO to the TSP we consider a graph defined by a set of cities (vertices) being the lengths of the edges proportional to the distance between two cities i and j . Once an ant selects a route from i to j , the ant releases a special chemical called pheromone that can be detected by other ants beginning to move toward the region of increasing concentration. During simulation, a given ant k is placed randomly in each city i choosing during the iteration the next city j to go according the probability:

$$p_k(i, j) = \begin{cases} \frac{\tau_{ij}^\alpha d_{ij}^\beta}{\sum_{g \in J_k(i)} \tau_{ig}^\alpha d_{ig}^\beta} & (8) \\ 0 & \end{cases}$$

where $p_k(i, j)$ is the aforementioned probability, τ_{ij} the pheromone concentration or pheromone trail value, and $J_k(i)$ the set of cities that have not yet been visited by ant k in city i . There are two parameters, α or the relative importance of the pheromone trail and β or the relative importance of the distance d_{ij} between cities. Once a tour has been completed, thus each city has been visited once by each ant, the amount of pheromone associated to each trail has to be updated by the following expression:

$$\tau(i, j) = \rho \cdot \tau(i, j) + \sum_{k=1}^n \Delta \tau(i, j) \quad (9)$$

where ρ is the evaporation constant, thus the amount of dissipated or evaporated pheromone, and $\Delta \tau(i, j)$ the amount of pheromone deposited by ant k . Finally, the length L_k of the tour covered by ant k is related with the amount of pheromone such that short tours will result in higher concentrations of deposited pheromone:

$$\Delta \tau_k = \begin{cases} \frac{1}{L_k} & , (i, j) \in \text{tour} \\ 0 & , \text{otherwise} \end{cases} \quad (10)$$

In the present simulation experiment (Sinclair, 2006) we compared three different cases of TSP, where ants had to search for a closed tour of minimal length visiting each city once and only once. The three cases consist of 10 cities with different locations in each case, with common values of the following parameters: $\alpha=1.0$, $\beta=2.0$, $\rho=0.5$ and the initial value of pheromone $\tau_{ij} = 0.1$. Once the simulation was started we registered the length of each tour until ant-cycle 100. Finally, we compared the tour lengths in the three instances at ant-cycle 100.

Study case 6. Symbolic regression

The method of least squares is by far the most widely used method in regression analysis (Draper and Smith, 1998). In symbolic regression a mathematical expression is evolved to fit a target function. However, the object of search is a symbolic description of a model, not just a set of coefficients. The goal is to find an approximation of the target function in symbolic form. These functions can be obtained combining primitives such as +, -, * and /, as well as Sin(x), Cos(x), e^x, etc., including input variables and coefficients. The fitness is scored using some measure of error (i.e. mean squared error, sum of the absolute differences) between the evolved approximated function and the target function. Using a genetic programming (Koza, 1992, 1994) algorithm a function is represented as a Lisp-like expression or tree, with crossover between branches of two different expressions and mutation in individual branches of a tree (Figure A1). The general scheme of genetic programming is similar to

the protocol of a genetic algorithm, except that chromosomes of a fixed length are substituted by variable tree structures representing mathematical or Boolean expressions. In the present simulation experiment (Gerber, 1998) we searched for an approximation of the target function $f(x)=3x^4 - 3x + 1$ under six different experimental protocols. A first set of experiments - experiment one, two and three - was carried out being the method of selection, thus the method how parents are selected when breeding a new generation, the fitness proportionate technique. In this method fitter individuals are more likely to be chosen as parents. In contrast, a second set of experiments - experiments four, five and six - was performed but choosing as method of selection the tournament method. In this case the fittest parents are selected from a pool or sample of randomly chosen trees. The crossover and mutation probabilities were respectively set up as follows: 80% and 20% (experiment one), 80% and 5% (experiment two), 5% and 20% (experiment three), 80% and 20% (experiment four), 80% and 5% (experiment five), 5% and 20% (experiment six).

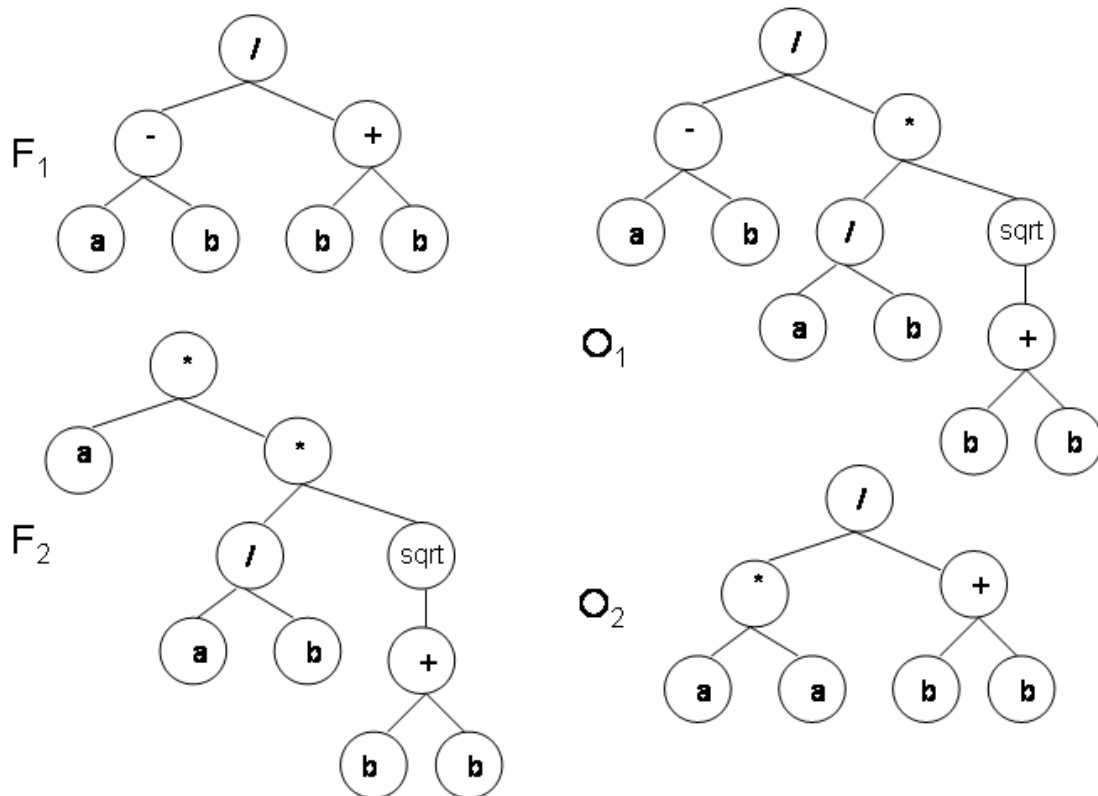


Figure A1.- Genetic programming individuals represented as parse trees. Usually crossover and mutation are simulated as follows. In crossover, the recombinant child (O_1 tree) is obtained by the exchange of two randomly chosen subtrees from the parents F_1 and F_2 . Mutation (i.e. O_2 tree) results from F_1 parental tree by the replacing of a randomly chosen subtree, $a-b$, by a randomly generated tree, $a*a$.

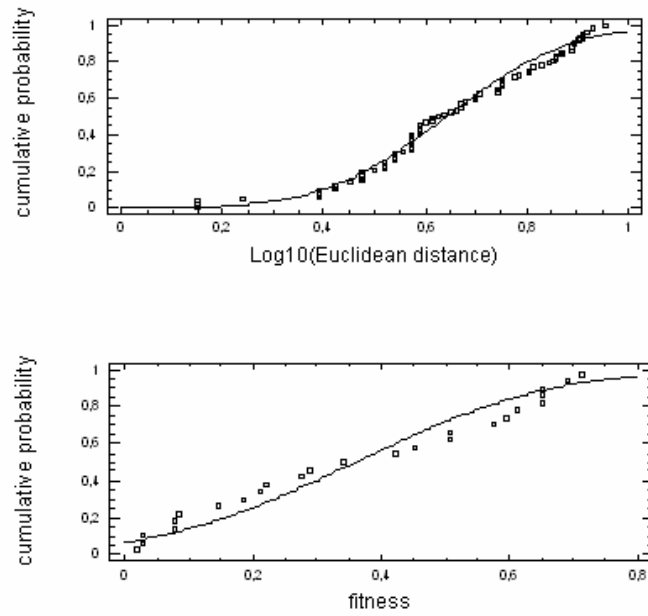


Figure A2.- Normal probability plot. Symbolic regression experiment (bottom) and transformed data in the Darwin's biomorphs experiment (top).

TABLES

Table II .- Genetic code table in MicroAnts (see Prata, 1993; Lahoz-Beltra, 2008)

GENE 1 - BITS 1,2,3,4:	BINARY CODE	FUNCTION
	0000	ant is blind
	0001	ant can see 1 square away
	0010	ant can see 2 squares away
	0011	" " " 3 " "
	:	:
	:	:
	1111	" " " 15 " "
GENE 2 - BITS 5,6:	00	ant will not move
	01	ant moves up and down
	10	ant moves right and left
	11	ant moves in a random direction
GENE 3 - BIT 7:	0	ant will not avoid poison
	1	ant will avoid poison
GENE 4 - BIT 8:	0	ant will mate if food level > 500
	1	" " " " " " > 1000
GENE 5 - BITS 9,10:	00	ant never shares its food
	01	" always " " "
	10	ant only shares its food if the other ant's 9,10 bits = 10
	11	ant only shares its food if the other ant's 9,10 bits = 01 or 11

Table III .- Dawkin's biomorph genetic code (see Prata, 1993; Lahoz-Beltra, 2008)

GENE 1:	Number of offspring.
GENE 2:	Number of iterations used when plotting a biomorph.
GENE 3:	gene for branch length
GENE 4:	gene for branch length
GENE 5:	gene for branch length, such that: if odd then branch length is GENE 3 * GENE 4 (epistatis) else branch length is GENE 3 + GENE 4 (polymorphism)
GENE 6:	Angle of 1st branch from precursor (in degrees).
GENE 7:	Angle of 2nd branch from 1st (2 branches come from each branch point).
GENE 8:	Length of initial branch or trunk.

Table IV.- Statistical summary in SGA experiment

Generation	Minimum	Maximum	Average	Median	IQR
0			2.960	0.963	3.248
1	0.076	16.483	2.748	0.801	3.589
2	0.002	14.311	3.609	3.604	5.127
3	0.097	5.569	2.631	-	-
4	2.782	6.901	4.901	4.380	1.867
5	1.960	4.177	3.203	2.796	1.527
6	4.305	6.385	4.962	4.389	1.523
7	2.534	4.080	3.686	4.008	0.556
8	4.206	6.466	4.568	4.380	0.151
9	2.443	4.177	3.820	4.008	0.160
10	4.206	6.145	4.590	4.380	0.294
11	2.621	4.177	3.706	3.976	0.876
12	4.206	6.205	4.668	4.389	0.650
13	2.572	4.177	3.780	4.038	0.572
14	4.206	5.736	4.557	4.330	0.406
15	2.944	4.177	3.924	4.064	0.216
16	4.206	5.659	4.463	4.322	0.319
17	3.504	4.161	3.981	4.064	0.263
18	4.256	4.870	4.418	4.322	0.278
19	3.579	4.129	3.978	4.064	0.287
20	4.239	4.626	4.377	4.322	0.176

Table V.- Kruskal-Wallis test in the ant population evolution experiment

Sample time	<i>n</i>	1.1.1 Average rank
0	30	68.550
1	7	57.928
2	7	66.928
3	17	50.823
4	21	47.261
5	27	44.740

Statistical value=11.7459 *p*-value=0.0384

Table VI.- Statistical summary in the ant population evolution experiment

Sample Time	Genetic drift	Median	Variance	Minimum	Maximum	Q_1	Q_3	IQR	SS ⁽¹⁾	SK ⁽²⁾
0	5.166	5.0	2.557	1.0	9.0	4.0	6.0	2.0	-0.0448	1.1711
1	4.428	5.0	2.952	1.0	6.0	4.0	6.0	2.0	-1.6060	1.4397
2	4.714	5.0	2.904	1.0	6.0	5.0	6.0	1.0	-2.3874	2.9327
3	4.176	5.0	2.029	1.0	6.0	4.0	5.0	1.0	-1.5788	0.1591
4	4.000	4.0	2.100	1.0	6.0	3.0	5.0	2.0	-1.2230	-0.6173
5	3.888	4.0	1.871	1.0	6.0	3.0	5.0	2.0	-1.1953	-1.0330

(1) Standardized skewness, $\frac{\text{skewness}}{\sqrt{\frac{6}{n}}}$ (2) Standard kurtosis, $\frac{\text{kurtosis}}{\sqrt{\frac{24}{n}}}$

Table VII.- Hamming ball of radius $r=2$ in the ant population evolution experiment

Sample Time	0	1	2	3	4	5
$d(a, t)$	1	1	2	3	5	6

Table VIII.- Statistical summary in the Dawkin's biomorph evolution experiment

Sample Time	Genetic drift	Median	Variance	Minimum	Maximum	Q₁	Q₃	IQR	SS⁽¹⁾	SK⁽²⁾
1	8.370	8.245	0.124	8.062	9.000	8.124	8.544	0.420	1.400	0.782
2	7.741	7.809	0.076	7.211	8.000	7.745	7.874	0.129	-1.799	1.892
3	7.097	7.176	0.141	6.403	7.416	7.000	7.416	0.416	-1.517	1.287
4	6.163	6.201	0.216	5.656	6.783	5.656	6.480	0.824	0.056	-0.932
5	5.478	5.567	0.173	5.000	6.082	5.000	5.656	0.656	0.029	-0.362
6	4.934	4.845	0.177	4.472	5.656	4.690	5.099	0.409	1.015	0.498
7	4.346	4.358	0.132	3.872	4.795	4.123	4.582	0.459	-0.109	-0.530
8	3.785	3.741	0.079	3.461	4.242	3.741	3.741	0.000	1.026	1.227
9	3.757	3.872	0.094	3.316	4.123	3.605	3.872	0.267	-0.496	-0.026
10	3.554	3.741	0.201	2.828	4.000	3.464	3.741	0.277	-1.196	0.878
11	3.079	3.000	0.024	3.000	3.316	3.000	3.158	0.158	1.632	1.632
12	3.457	3.464	0.055	3.162	3.741	3.313	3.602	0.289	-0.129	0.618
13	3.208	3.158	0.270	2.645	3.872	2.822	3.594	0.771	0.408	-0.009
14	2.879	2.449	0.556	2.449	3.741	2.449	3.741	1.292	1.224	-
15	2.459	2.645	0.427	1.732	3.000	1.732	3.000	1.268	-0.831	-
16	2.104	2.449	0.357	1.414	2.449	1.414	2.449	1.035	-1.224	-
17	0.942	1.414	0.666	0.000	1.414	0.000	1.414	1.414	-1.224	-

(1) Standardized skewness, $\frac{\text{skewness}}{\sqrt{\frac{6}{n}}}$ (2) Standard kurtosis, $\frac{\text{kurtosis}}{\sqrt{\frac{24}{n}}}$

Table IX.- Statistical summary in three SGA experiments

Experiment	Mean	Median	Variance	Minimum	Maximum	Q₁	Q₃	IQR	SS⁽¹⁾	SK⁽²⁾
1	2.348	2.099	1.706	0.177	6.497	1.610	2.325	0.715	5.843	6.359
2	15.183	15.358	0.768	12.616	15.737	15.295	15.609	0.314	-6.712	6.985
3	3.442	3.003	0.221	2.992	3.932	3.003	3.932	0.929	0.268	-2.706

(1) Standardized skewness, $\frac{\text{skewness}}{\sqrt{\frac{6}{n}}}$ (2) Standard kurtosis, $\frac{\text{kurtosis}}{\sqrt{\frac{24}{n}}}$

Table X.- Mann-Whitney test in two SGA experiments

Experiment	Median	1.1.2 Average rank
1	2.099	20.5
2	15.358	60.5

Statistical value=1600.0 p -value=0.0000

Table XI.- Kruskal-Wallis test in three SGA experiments

Experiment	N	1.1.3 Average rank
1	40	24.5
2	40	100.5
3	40	56.5

Statistical value=97.017 p -value=0.0000

Table XII.- Dunn's Multiple Comparison test in three SGA experiments

Experiment vs Experiment	Difference in rank sum	1.1.4 Significant $p < 0.05$
Exp 1 vs Exp 2	-68.38	Yes
Exp 1 vs Exp 3	-25.38	Yes
Exp 2 vs Exp 3	43.00	Yes

Table XIII.- Statistical summary in three ACO experiments

Experiment	Mean	Median	Variance	Minimum	Maximum	Q₁	Q₃	IQR	SS⁽¹⁾	SK⁽²⁾
1	1129.510	1088.040	7327.850	1011.650	1254.610	1088.040	1221.350	133.309	0.362	-0.981
2	641.593	641.575	0.003	641.575	641.757	641.575	641.575	0.000	4.082	6.454
3	662.431	683.118	2395.140	576.790	721.099	621.321	694.775	73.454	-0.746	-0.642

(1) Standardized skewness, $\frac{\text{skewness}}{\sqrt{\frac{6}{n}}}$ (2) Standard kurtosis, $\frac{\text{kurtosis}}{\sqrt{\frac{24}{n}}}$

Table XIV.- Kruskal-Wallis test in three ACO experiments

Experiment	N	1.1.5 Average rank
1	10	25.5
2	10	9.5
3	10	11.5

Statistical value=20.197 p -value=0.0000

Table XV.- Dunn's Multiple Comparison test in three ACO experiments

Experiment vs Experiment	Difference in rank sum	1.1.6 Significant $p < 0.05$
Exp 1 vs Exp 2	16.00	Yes
Exp 1 vs Exp 3	14.00	Yes
Exp 2 vs Exp 3	-2.00	No

Table XVI.- Statistical summary in the symbolic regression experiments

Experiment	Mean	Median	Variance	Minimum	Maximum	Q₁	Q₃	IQR	SS⁽¹⁾	SK⁽²⁾
1	0.360	0.342	0.058	0.018	0.714	0.145	0.595	0.450	-0.016	-1.599
2	0.233	0.233	0.005	0.092	0.352	0.178	0.263	0.085	0.484	-0.827
3	0.235	0.257	0.002	0.111	0.268	0.214	0.268	0.054	-2.940	1.112
4	0.162	0.025	0.053	0.001	0.576	0.011	0.200	0.189	2.478	-0.423
5	0.116	0.085	0.009	0.025	0.315	0.047	0.147	0.100	2.475	0.183
6	0.297	0.421	0.028	0.005	0.422	0.093	0.421	0.328	-1.733	-1.229

(1) Standardized skewness, $\frac{\text{skewness}}{\sqrt{\frac{6}{n}}}$ (2) Standard kurtosis, $\frac{\text{kurtosis}}{\sqrt{\frac{24}{n}}}$

Table XVII.- Kruskal-Wallis test in the symbolic regression experiments

Experiment	N	1.1.7 Average rank
1	25	98.42
2	25	80.40
3	25	83.98
4	25	48.78
5	25	48.00
6	25	93.42

Statistical value=31.9828 p -value=0.0000

Table XVIII.- Dunn's Multiple Comparison test in the symbolic regression experiments

<i>Experiment vs Experiment</i>	<i>Difference in rank sum</i>	1.1.8 Significant p<0.05
Exp 1 vs Exp 2	18.020	No
Exp 1 vs Exp 3	14.440	No
Exp 1 vs Exp 4	49.640	Yes
Exp 1 vs Exp 5	50.420	Yes
Exp 1 vs Exp 6	5.000	No
Exp 2 vs Exp 3	-3.580	No
Exp 2 vs Exp 4	31.620	No
Exp 2 vs Exp 5	32.400	No
Exp 2 vs Exp 6	-13.02	No
Exp 3 vs Exp 4	35.200	No
Exp 3 vs Exp 5	35.980	No
Exp 3 vs Exp 6	-9.440	No
Exp 4 vs Exp 5	0.780	No
Exp 4 vs Exp 6	-44.640	Yes
Exp 5 vs Exp 6	-45.420	Yes

Corresponding author

R. Lahoz-Beltra: Dept. of Applied Mathematics, Faculty of Biological Sciences, Complutense University of Madrid, 28040 Madrid, Spain ; e-mail: lahozraf@bio.ucm.es