Darwin, genetic algorithms, and mathematics

ALESSIO DRIVET
GeoGebra Institute of Turin
ITALY

Abstract: - The text considers the relationship between Darwinian selection, genetic algorithms, and mathematics. The initial idea introduces some examples that can be presented to upper secondary school students. In this sense, the solution proposals differ from the traditional ones found in the literature and use IT supports already known to the students.

Key-Words: - Genetic selection, genetic algorithms, optimization problems, computational procedures, maths relations, probability, logic.

Received: July 8, 2022. Revised: March 19, 2023. Accepted: April 21, 2023. Published: May 22, 2023.

1 Introduction

Charles Darwin (1809-1882) argued that species evolve through natural selection, i.e., the survival of the fittest in their surroundings and the death of the least fit. This process of natural selection relies on the genetic variation among individuals in a population, which descendants can inherit. Genetic variations that confer a survival advantage are passed on, while disadvantageous ones are eliminated.

Recently, the theme has been taken up again regarding genetic algorithms [1][2]. These replicate the process of natural selection by using a population of possible solutions to a given problem. Solutions with the best characteristics are selected to generate a new population of solutions. This process is repeated for several iterations until an optimal solution to the problem is reached [3].

These algorithms allow one to solve combinatorial optimization problems, in which one tries to find the best solution among many possible solutions.

For example, a classical combinatorial optimization problem is known as the travelling salesman problem; it seeks to find the shortest path a travelling salesman must take to visit a set of cities only once. One can use genetic algorithms to find an approximate solution to this problem [4].

A second example refers to a combinatorial optimization problem in which one tries to find the combination of items to put in a backpack so that the total weight is less than a given threshold and the total value of the items is maximized.

A myriad of problems, some particularly complex, have been treated with this technique: engineering problems such as the design of aircraft parts, modelling of financial markets [5], testing of economic alternatives, the study of functions used for nonlinear optimization algorithms [6], design of neural networks [7], etc.

At the school level, most research has focused on the so-called timetabling problem [8] or the analysis of student performance [9]. In contrast, proposing a simplified approach to test how a Darwinian procedure can be applied to classical mathematical problems in the student's curriculum would be helpful.

2 Problem Formulation

As this is a much-explored problem, a 'corpus' of concepts, terms and consequent applications has been consolidated and can be found in the literature [10][11][12].

The aim is to show simple examples for high school students that link mathematics to the Darwinian model using flexible tools such as GeoGebra and Excel.

The four proposals allow grasping the concept of the genetic algorithm without resorting to binary vector encodings or crossover operators but establishing intuitive selection criteria.

The procedure is summarised in Fig. 1:

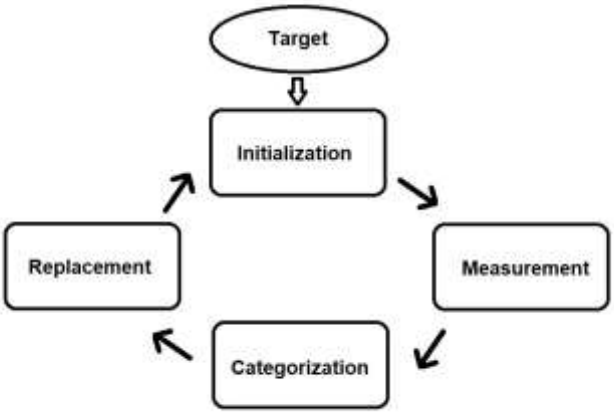

Fig. 1

2.1 Darwin's Parabola

Let us consider the parabola:

$y= -1/2\ x^2 +5\ x- 5/2$

whose graph is easily obtained with GeoGebra, as shown in Fig.2:

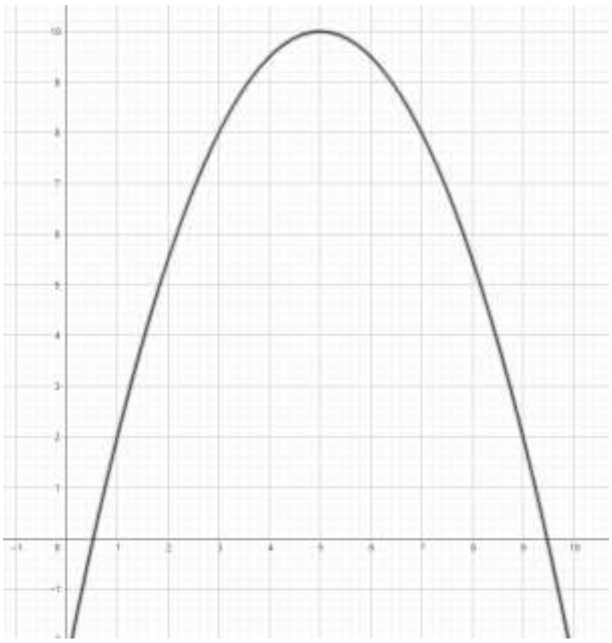

Fig. 2

In this case, it is evident that for x = 5, we obtain the maximum y = 10.

The study of derivatives can be used to determine this maximum, but the aim is to use a different procedure.

First, Excel generates thirty values of x in the interval [-5; 5]. Then we calculate the y-values of the function and the average of these. We proceed with successive tests to eliminate the values below these averages.

After a few passes, the averages coincide, and only the largest values can be retained. The maximum has been found, and the corresponding x can be associated with it (Fig.3).

a=	-0,5	b=	5	c=	-2,5		min	-5	max	5
n	x	y	test1	test2	test3	test4	result y	result x		
1	-5	-40	0	0	0	0	0	0		
2	0	-2,5	-2,5	0	0	0	0	0		
3	-5	-40	0	0	0	0	0	0		
4	5	10	10	10	10	10	10	5		
5	1	2	2	2	0	0	0	0		
6	-4	-30,5	0	0	0	0	0	0		
7	-2	-14,5	0	0	0	0	0	0		
8	-3	-22	0	0	0	0	0	0		
9	1	2	2	2	0	0	0	0		
10	5	10	10	10	10	10	10	5		
11	-2	-14,5	0	0	0	0	0	0		
12	-2	-14,5	0	0	0	0	0	0		
13	0	-2,5	-2,5	0	0	0	0	0		
14	0	-2,5	-2,5	0	0	0	0	0		
15	2	5,5	5,5	5,5	5,5	5,5	0	0		
16	-4	-30,5	0	0	0	0	0	0		
17	3	8	8	8	8	8	0	0		
18	-5	-40	0	0	0	0	0	0		
19	2	5,5	5,5	5,5	5,5	5,5	0	0		
20	-3	-22	0	0	0	0	0	0		
21	3	8	8	8	8	8	0	0		
22	-5	-40	0	0	0	0	0	0		
23	4	9,5	9,5	9,5	9,5	9,5	0	0		
24	-4	-30,5	0	0	0	0	0	0		
25	-3	-22	0	0	0	0	0	0		
26	-4	-30,5	0	0	0	0	0	0		
27	-1	-8	-8	0	0	0	0	0		
28	-4	-30,5	0	0	0	0	0	0		
29	1	2	2	2	0	0	0	0		
30	-2	-14,5	0	0	0	0	0	0		
31	0	-2,5	-2,5	0	0	0	0	0		
32	2	5,5	5,5	5,5	5,5	5,5	0	0		
	media	-12,98	1,57	2,08	1,88	1,88				

Fig. 3

Following a line closer to the concept of selection, a procedure known as *tournament selection* can be used [13][14]: a macro can order the solutions, evaluated with this algorithm:

1. choose the highest value. The probability p=1/n is associated with it.

2. assign the probability $p\ (1\text{-}p)^i$ to the following elements.

3. calculate the average of the probabilities of half of the elements.

4. only the *n* data showing a higher probability are kept.

5. repeat the operation from point 1.

2.1 Darwin's jars

In a mathematical problem, one wants to find one or more unknown data from a statement, starting from the available data. For example, knowing the radius and height, it is possible to determine the volume of the corresponding cylinder using the formula $V = \pi r^2 h$.

More complex is to answer the question: for what values of r and h (r, h ∈N) do we obtain a given volume?

Let us take a concrete example: we want to construct a jar with a volume of 1000 cm^3 with 5≤r≤9 and 8≤h≤12. The tolerated error is 20 cm^3.

A mathematician would rely on an analytical approach. Considering the 25 possible pairs, one can see which function passes through one of these pairs respecting the constraint 980<=V<=1020. In this case, we have the optimum for $V=\pi 6^2 9 = 1017.88$, as can be seen from the graph in Fig. 4 obtained with GeoGebra:

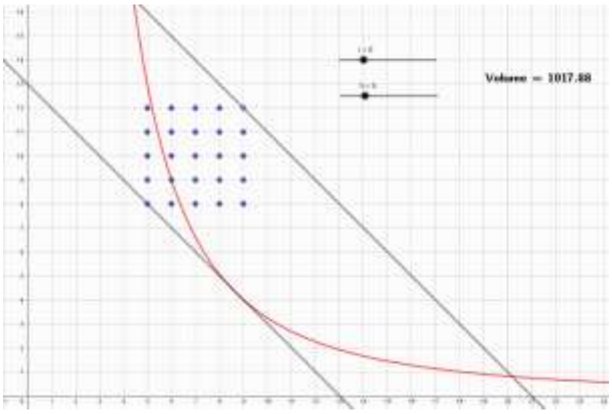

Fig. 4

However, the intention is to use a different approach. In this case, it is interesting to note how one can use a Darwinian genetic algorithm.

First, it is possible to generate *n* random jars and calculate the corresponding volumes, then the results are categorised, e.g.:

category 1 = 1000 ± 20

category 2 = 1000 ± 150

category 3 = 1000 ± 250

category 4 = 1000 ± 500

Substantial errors were deliberately chosen to show the effectiveness of the evolutionary algorithm.

At this point, the highest category is, according to Darwinian theory, destined to be replaced by a more appropriate category (e.g., the first). The process continues with new measurements and categorisation so that solutions that do not meet the proposed maximum error constraint are progressively eliminated.

The process can be implemented in a spreadsheet, as shown in Fig. 5:

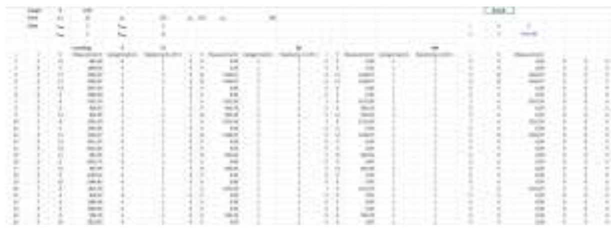

Fig. 5

Starting with 100 jars, after only three steps, the solution V = 1017.88 for r = 6 and h = 9 is found.

2.2 Darwin's four coins

The Darwinian model can be applied to a probabilistic problem. For example, assuming the toss of a fair coin, the probability of tails is 0.5. If there are four coins, we can estimate that the probability of obtaining four crosses is $p(4C)= 0.5^4 = 0.0625$.

However, the intention is to use a different approach. In this case, it is interesting to note how one can use a Darwinian genetic algorithm.

First, n random tosses are simulated (0 stands for tails and 1 for heads), and the results are added up. Obviously, the 4C event corresponds to 0. Then the results are categorized, for example:

category 0 = 0

category 1 = 1

…

category 4 = 4

At this point, the highest category is, according to Darwinian theory, destined to be replaced by a good category (for example, the first). The process continues with new measurements and categorization to progressively eliminate the category 2, 3, and 4 solutions.

Starting from 100 launches, after three passages, a solution is found which indicates the p(4C) as the ratio between the elements of category zero and those of category 1. The result of a random event requires a sufficiently consistent number of trials, and then one can use Excel Macros to repeat the experiment several times. With 2000 trials, a result like the expected one is reached, as shown in Fig. 6:

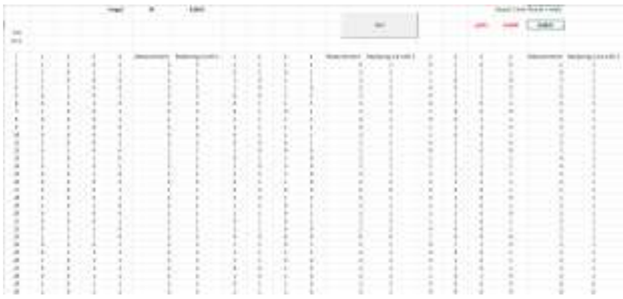

Fig. 6

2.3 Darwin's Logic

Logic is the study of reasoning and argumentation. One of the most interesting problems is checking the consistency of a set of propositions. For example, "At least one of Alberto and Bruno lives in Turin; at least one of Bruno and Alberto is a doctor; Bruno is not a doctor and does not live in Turin. Under what conditions is the set of propositions consistent?"

The problem can be easily solved by defining elementary propositions and connecting them with connectives $\vee$ (or) or $\wedge$ (and). From: A= Alberto lives in Turin; B= Bruno lives in Turin; C= Bruno is a doctor; D= Alberto is a doctor we get $(A \vee B) \wedge (C \vee D) \wedge (\neg C \wedge \neg B)$.

The tools available are the truth table or semantic tableaux. In both cases, the set appears consistent when A and D are true, and C and D are false.

However, the intention is to use a different approach. In this case, it is interesting to note how a Darwinian genetic algorithm can be used.

First, one generates n random sequences with possible truth values 0 (FALSE) or 1 (TRUE) for the four propositions and determines the result. New quadruplets are extracted if the result generates FALSE; otherwise, the previous values are retained. The process continues with a new measurement to progressively eliminate unacceptable solutions.

The process can be implemented in a spreadsheet, as shown in Figs. 7a and 7b, which show the same spreadsheet divided into two parts for ease of reading:

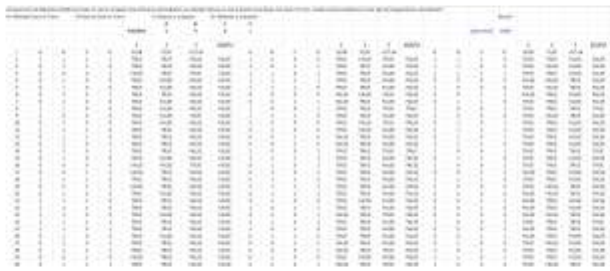

Fig. 7a

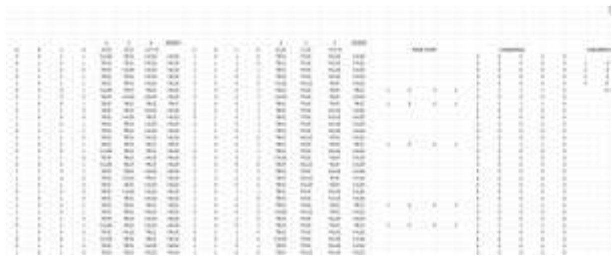

Fig. 7b

Starting from 100 quadruplets, after a few passages, only the solutions corresponding to those determined with the classical methods remain.

4 Conclusion

The examples given in the text are a simple indication of how it is possible to show the solution of a problem from a different angle than the traditional mathematical approach.
This point of view based on a concrete operational approach represents one of the most interesting aspects of teaching [15] [16].
Among other things, they lead to a stimulating relationship between the mathematics teacher and the science teacher.

References:

[1] Holland, JH., *Adaptation in natural and artificial systems*, Ann Arbor: University of Michigan Press, 1975.

[2] Golberg, DE., *Genetic algorithms in search, optimization, and machine learning*, Addison Wesley, 1989.

[3] Mitchell, M., *An introduction to genetic algorithms*, MIT press, 1998.

[4] Han, Shifen; Xiao, Li. An improved adaptive genetic algorithm. In: *SHS Web of Conferences*. EDP Sciences, 2022. p. 01044.

[5] Kumar, Gourav; Jain, Sanjeev; Singh, Uday Pratap. Stock market forecasting using computational intelligence: A survey. *Archives of computational methods in engineering*, 2021, 28: 1069-1101.

[6] Chen, Yuntian, et al. Symbolic genetic algorithm for discovering open-form partial differential equations (SGA-PDE). *Physical Review Research*, 2022, 4.2: 023174.

[7] Nigam, AkshatKumar, et al. Augmenting genetic algorithms with deep neural networks for exploring the chemical space. *arXiv preprint arXiv:1909.11655*, 2019.

[8] Tan, Joo Siang, et al. A survey of the state-of-the-art of optimisation methodologies in school timetabling problems. *Expert Systems with Applications*, 2021, 165: 113943.

[9] Lakshmi, T. Miranda; Martin, A.; Venkatesan, V. Prasanna. An analysis of students performance using genetic algorithm. *Journal of Computer Sciences and Applications*, 2013, 1.4: 75-79.

[10] Malik, Ashima. A study of genetic algorithm and crossover techniques. *International Journal of Computer Science and Mobile Computing*, 2019, 8.3: 335-344.

[11] Mirjalili, Seyedali, et al. Genetic algorithm: Theory, literature review, and application in image reconstruction. *Nature-Inspired Optimizers: Theories, Literature Reviews and Applications*, 2020, 69-85.

[12] Zhi, Hui; Liu, Sanyang. Face recognition based on genetic algorithm. *Journal of Visual Communication and Image Representation*, 2019, 58: 495-502.

[13] Prayudani, S., et al. Analysis effect of tournament selection on genetic algorithm performance in traveling salesman problem (TSP). In: *Journal of Physics: Conference Series*. IOP Publishing, 2020. p. 012131.

[14] Katoch, Sourabh; Chauhan, Sumit Singh; Kumar, Vijay. A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, 2021, 80: 8091-8126.

[16] Kaiser, G., Mathematical modelling and applications in education. *Encyclopedia of mathematics education*, 2020, 553-561.

[16] Drivet, A., *Examining an Operational Approach to Teaching Probability*, IGI Global, 2021