

Stochastic Behaviour of Parameter Convergence in Genetic Algorithm: An Experimental Analysis

Prof. D. P. Sharma¹, Prof. Shweta²

Abstract—Parameter convergence in Genetic Algorithm (GA) is quite unpredictable since it is not tied with the number of episodes (i.e., the number of generations), convergence time, or any other parameter it requires to converge. Thus, in such a situation, trial to frame out any hypothesis to predict definite number of episodes gets disapproved; in turn, the alternate hypothesis becomes acceptable that proves that there is non-dependent relationship between various parameters with regard to number of episodes for proper convergence of a Genetic Algorithm. The reason behind is that, it is an optimization dependent technique that can take any number of episodes to converge within the given convex set until convergence criteria is satisfied. In other words, unless it converges down to adequate optima, it can go on acquiring more number of episodes. It is to note that, optimization techniques may not provide the pin-pointed target but in most of the cases it guarantee to provide optimum result, i.e., in the proximity of target. Here, an experimental study is undertaken to test the hypothesis in favour of verifying inter-relationship of parameters and its effect on convergence. The GA has been designed for the purpose of generating various parameters of interest to assist the study.

Keywords—GA, Genetic Algorithm, Convergence, GA parameters, Episodes, Hypothesis Test,

I. INTRODUCTION

Genetic Algorithm is inspired by biological phenomenon that exhibits natural selection between species. Genetic Algorithm is stochastic by nature. Developed by John Holland [1962], the GA is a largely used Artificial Intelligence algorithm that has been exploited for the problems requiring the most desirable solution set from among various choices, such as the lowest cost path in a communication network. It is a search algorithm that operates over a population of encoded candidate solutions to solve a given problem.

Genetic algorithm is basically an Optimization technique that simulates the phenomenon of natural selection, as first observed by Charles Darwin. He laid principle of evolution through natural selection in his *magnum opus*, *The Origin of Species*, published in 1959. The basic idea contained in Darwin's work can be summarized in the following points:

- a. Each individual tends to pass on its traits to its offspring.
- b. Nevertheless, nature produces individuals with different traits.
- c. The fittest individuals – those with most favourable traits, tends to have more offspring than do those with unfavourable traits, thus driving the population as a whole toward favourable traits.
- d. Over long periods, variation can accumulate, producing entirely new species whose traits make them especially suited to particular ecological niches.

From a molecular point of view, natural selection is enabled by the variation that follows from crossover and mutation producing new genes.

In natural evolution, the species search for increasingly beneficial adaptations for survival within their complex environments. The search takes place in species' chromosomes where changes and their effects are graded by the survival and reproduction of species. This makes the basis for survival of the fittest – survival and passing on these characteristics to future generations. Survival in nature is the ultimate utility function.

The genetic algorithm, instead of trying to optimize a single solution, works with a population of candidate solutions that are encoded as chromosomes. Within the chromosomes are separate genes that represent the independent variables for the problem at hand. Basic parameters are drawn from the given problem domain, and then chromosomes are created that represent further unique independent parameters. The parameters could represent binary encoded integers, bit strings, or floating point variables.

II. EXPERIMENTAL FINDINGS

In the following paragraphs, it is studied to verify and establish a definite relationship between the parameters assigned to a GA and its convergence time, that is, more precisely the number of episodes it takes to fully converge. The following chart is created to find the relationship of various input, fitness and genome parameters, their convergence under worst and best chromosome condition, convergence basis (whether the initial population, crossover, or others) and the number of episodes it acquires to converge. The following assumption is made using purist form –

Null hypothesis (H_0): there is a definite relationship between parameters of convergence with the number of episodes in GA.
Alternate Hypothesis (H_1): there is no definite relationship between parameters of convergence with the number of episodes in GA.

Stochastic Behaviour of Parameter Convergence in Genetic Algorithm an Experimental Analysis

Here, the table created, using the software prepared for this experiment, expose that the data generated for number of episodes keeping value of initial population static is not in consent with or support any relationship between itself and the parameters (e.g., population size, crossover rate, mutation rate, etc.) used for convergence of GA, rather an arbitrary variation in number of episodes is evidenced. And the difference between one to another episode number is non-trivial.

Sl. No.	y	Worst Chromosome		Made By	Best Chromosome		Made By	No. of Episodes
		Fitness	Genome		Fitness	Genome		
1	36	10.49562	99896	Initial Population	97.29729	02144	Initial Population	6
2	36	13.43283	57879	Initial Population	97.29729	42223	Initial Population	38
3	36	12.63157	97759	Initial Population	97.22223	14411	Crossover	4
4	36	10.84337	99895	Crossover	97.29729	02144	Mutation of all individuals	18

Figure – 1 Repeated convergence of same value of “y” (initial population) indicates differences in convergence pattern.

Thus, it would be better to follow a structured pattern of finding the truth through hypothesis test. So, we consider a chart of *observed values* (i.e., historical record) along with the *expected values* (i.e., the current observation) to evaluate the trustworthiness of our hypothesis.

Chart of *observed values* is created using the software.

Sl. No.	y	Worst Chromosome		Made By	Best Chromosome		Made By	No. of Episodes (Generation)
		Fitness	Genome		Fitness	Genome		
1	36	10.49562	99896	Initial Population	97.29729	02144	Initial Population	6
2	76	21.46892	98889	Initial Population	48.68421	23615	Initial Population	11
3	97	27.40112	98889	Initial Population	98.97959	56106	Initial Population	11
4	135	19.25925	32320	Crossover	99.26470	93631	Initial Population	3
5	198	4.04040	12111	Crossover	98.98989	46848	Initial Population	9
6	214	8.57142	22230	Initial Population	99.59349	92849	Crossover	1
7	237	14.34599	00334	Initial Population	99.57983	76885	Mutation of all individuals	73
8	248	11.29032	40222	Initial Population	99.59839	82698	Mutation of all individuals	134
9	256	7.03125	02132	Crossover	99.21875	83698	Crossover	16
10	276	10.869561	22.332	Crossover	95.28985	89916	Initial Population	3
ΣX		134.773851	407045.332		936.49599	648390		
\bar{X}		13.4773851	40704.5332		93.649599	64839.0		

Figure – 2 Historical Record generated using GA software designed for this purpose.

The current observation of *expected values* obtained is as below:

Sl. No.	y	Worst Chromosome		Made By	Best Chromosome		Made By	No. of Episodes
		Fitness	Genome		Fitness	Genome		
1	311	4.18006	02212	Initial Population	99.35691	68988	Initial Population	62

Figure – 3 Experimental result of new value of “y”.

Now, we consider the current chart of expected values along with the historical record to prove the assumption taken before, using statistical tool - chi square test.

Category		O (Observed Values)	E (Expected Values)	M= (O-E)	M ²	M ² /E
Worst Chromosome	Fitness	13.4773851	4.18006	9.2973251	86.440735617201	20.6793
	Genome	40704.5332	02212	38492.5332	1481675112.15310224	669835.0416
Best Chromosome	Fitness	93.649599	99.35691	-5.707311	32.573398850	0.32784
	Genome	64839.0	68988	-4149	17214201	249.5245
$\chi^2 = \sum_{i=1}^k \frac{(O - E)^2}{E}$						670105.5733

Figure – 4 Analysis of Observed and Expected Values.

Considering degree of freedom (df) = (row-1) (col-1) = (4-1)(2-1) = 3, taken at 0.05 confidence level, which is valid for most of the scientific experimentations, i.e., 7.815 considering the Chi-Square distribution table – Probability of exceeding the critical value

δ	0.10	0.05	0.025	0.01	0.001
1	2.706	3.841	5.024	6.635	10.828
2	4.605	5.991	7.378	9.210	13.816
3	6.251	7.815	9.348	11.345	16.266
4	7.779	9.488	11.143	13.277	18.467
5	9.236	11.070	12.833	15.086	20.515

Figure – 5 Upper critical values of chi-square distribution with δ degrees of freedom

Here, 670105.5733 >> 7.815,

Thus, the $\chi^2_{\text{calculated}}$ is greater than the χ^2_{critical} , thus the hypothesis H₀ is rejected.

Therefore, finally the alternate hypothesis stands forth, stated as below,

Alternate Hypothesis (H₁): there is no definite relationship between parameters of convergence with the number of episodes in GA.

III. VERIFYING REASONS FOR RANDOMNESS

Following are the reasons to turn the GA a stochastic process that presents unpredictable behavior of the algorithm in terms of estimating number of episodes for its convergence.

- i. Genetic Algorithms use probabilistic transition rules, not the deterministic rules.
- ii. Evolutionary process adopted in the Genetic Algorithms is not strictly formula bound rather it adopts the natural process of chance selection.
- iii. The fitness-proportionate selection adopted by Genetic Algorithm, in which the “expected value” of an individual is purely a chance dependent phenomenon, since the roulette wheel searching is commonly used for this purpose. Verifying the algorithm make us feel the process:

Algorithm:

1. Sum the total expected value of individuals in the population (say, sum T)
2. Repeat N times
3. Choose a random integer r between 0 and T.
 - i. Loop through the individuals in the population,
 - ii. Summing the expected values, until the sum ≥ r.

(The individual whose expected value puts the sum over this limit is the one selected)

This stochastic method statistically results in the expected number of offspring for each individual.

Note: James Baker (1987) proposed *Stochastic Universal Sampling* (SUS) method which adopts spinning roulette wheel once instead of N number of times (yet it is also stochastic).

IV. CONCLUSION

Thus, to conclude that – it is highly unlikely to get the number of episodes pre-estimated on the basis of relative values of parameters assigned to a genetic algorithm. Therefore, the number of episodes becomes free from the bindings of parameters’ convergence.

It further signifies that, the actual time of convergence of a genetic algorithm cannot be predicted. However, the techniques have evolved to make the convergence faster to a considerable extent, yet it does not guarantee the count of episodes or convergence time estimation, which are unpredictable phenomena.

Software Programs Developed for the Experiment

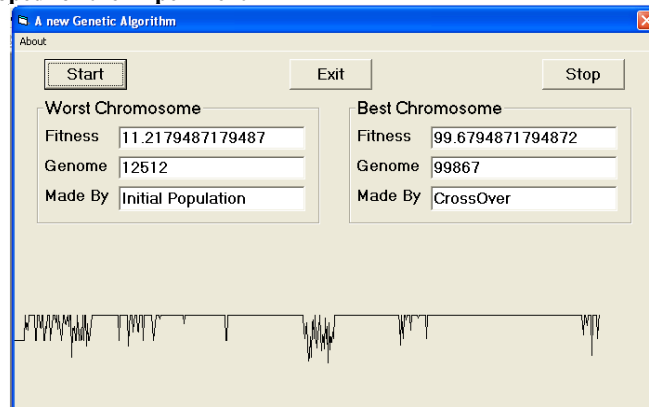


Figure – 6 (A) Snapshot of software used for experiment.

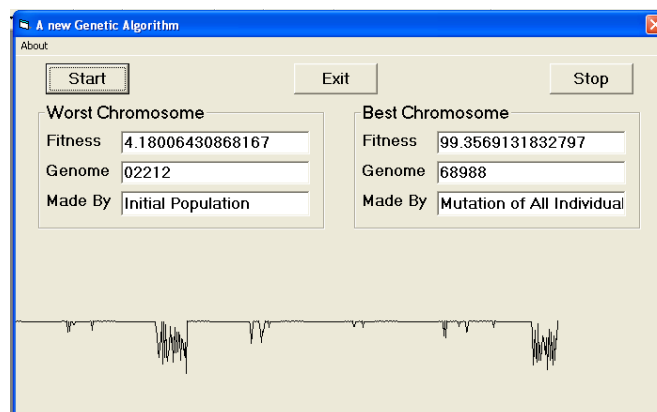


Figure – 6 (B) Snapshot of software used showing case of different episode.

REFERENCES

- [1]. Write, A. H., 1991. Genetic algorithms for real parameter optimization. In G. Rawlings, ed., *Foundation of Genetic Algorithms*, Morgan Kaufmann.
- [2]. Spears, W. M., and De Jong, K. A. 1991. On the virtues of parameterized uniform crossover. In R. K. Below and L. B. Booker, eds., *Proceedings of the Fourth International Conference on Genetic Algorithms*, Morgan Kaufmann.
- [3]. Schaffer, J. D., Eshelman, L. J., and Offut, D. 1991. Spurious correlation and premature convergence in genetic algorithms. In G. Rawlings, ed., *Foundation of Genetic Algorithms*. Morgan Kaufmann.
- [4]. O'Reilly, U. M., and Oppacher, F. 1995. The troubling aspects of a building block hypothesis for genetic programming. In L. D. Whitley and M. d. Vose, eds., *Foundation of Genetic Algorithms 3*, Morgan Kaufmann.
- [5]. Grefenstette, J. J. 1986. *Optimization of control parameters in genetic algorithms*. IEEE transactions on Systems, Man, and Cybernetics 16, no. 1:122-128.
- [6]. Bedau, M. A., and Packard, N. H. 1992. Measurement of evolutionary activity teleology, and life. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, eds. *Artificial Life II*, Addison Wesley.
- [7]. David E. Goldberg, 2006, *Genetic Algorithms in Search, Optimization & Machine Learning*. Pearson, London.
- [8]. Papoulis, A. 1984. *Probability, Random variables, and stochastic processes*, 2nd Ed. New York, McGraw Hills.
- [9]. Melanie Mitchell. 2002. *An introduction to Genetic Algorithms*. Prentice-Hall. 2nd Ed. USA.
- [10]. Holland, J. H. 1971. *Genetic Algorithms and the optimal allocation of trials*. SIAM Journal of Computing. 2(2), 88-105.
- [11]. Goldberg, D. E. 1981. *Algebraic and probabilistic properties of genetic algorithms*. 2009. Journal of Computing. 5, 28-44.