

Simulation of a Three-Tier Ecosystem and Observation of the Subsequent Population Curve and Genetic Dispersion

***Pritam Kayal**

*Don Bosco School, Park Circus, Kolkata, India

ABSTRACT

Darwin's theory of natural selection states that individuals with genetic traits better suited to the environment thrive well. This could mathematically be stated as with time, the degree of dispersion/scattering of genetic characteristics decreases as more and more of the living population tends to progress towards the fitter traits, while individuals with weaker traits are wiped out. We simulate a three tier ecosystem (autotrophs, herbivorous primary consumers, and carnivorous secondary consumers), and study the population curve and the corresponding change in the entropy, mean-absolute deviation, median-absolute deviation, and standard deviation of the genetic characteristics. This helps to validate and visualize Darwin's theory using simple artificial intelligence. The initial population is randomly assorted with traits and the successive offspring get a mixture of traits of the parents. Simulations have been run and the results have been presented.

1. INTRODUCTION

In 1859, Charles Darwin proposed "the theory of evolution by natural selection" in his book "On the Origin of Species" [1]. The theory states that individuals in a population are naturally variable, meaning that some individuals have traits better suited to the environment than others. These fitter individuals are more likely to survive and reproduce. These adaptive traits are then inherited by the offspring. Gradually, the 'fitter traits' become more and more common in the population.

Despite the general acceptance of the theory, substantiated by fossil records, some people still question the validity of the theory and school boards debate whether the theory should be taught. This motivates the visualization of a general ecosystem and observation of the population and corresponding genetic traits with time. Hence, using simple artificial intelligence, and mathematically defined animals, landforms, and genetic traits, we simulate an ecosystem with different values of initial configurations and study the population's genetic traits over time. To measure the degree of dispersion of genetic traits in each iteration, we used entropy [2, 3, 4] as the primary method. However other common methods have also been implemented. Hence, the relationship between the population curve and the dispersion of genetic traits [5] has been studied to visualize and validate Darwin's theory. Functions to use different statistical methods are also provided in the github repository [6].

The paper has been arranged as follows: In section 2, we have given a brief algorithmic overview of the process. In section 3, the mathematical definitions of the components of the ecosystem is provided. In section 4, the statistical methods used to measure the degree of dispersion of genetic traits has been given. The main algorithm has been described in section 5. For the test simulation given in the paper, the initial

configuration of the population is provided in section 6. The test results have been provided in section 7 and its corresponding conclusions are given in section 8. The references and links are given in section 9.

2. ALGORITHMIC OVERVIEW

The three-tier ecosystem consists of rabbits as the herbivorous, primary consumers and foxes as the carnivorous, secondary consumers. Initially we distribute a small number of animals uniformly randomly over the map and provide them with genetic traits (numbers) uniformly randomly from a specific domain. Each animal is assumed to have full knowledge of the geography of the map, and performs its biological tasks in the order of most important to least important. The biological needs are food, water, and urge to mate. In every iteration, an animal prioritises its need and takes a small step towards a landform (say, a pond to quench thirst.) After taking the small step, it again prioritises its needs and the process continues. Every iteration, it also checks if it is in a feasible location i.e near a pond for drinking water or grassland for eating food, and drinks or eats if required. The landforms have a small regeneration rate for their resources (due to assumed rainfall, regrowth of plants, etc.)

3. MATHEMATICAL DEFINITIONS

Definition 1: Animal

We define an *Animal* as a six-tuple (A, H, T, M, p, G) where:

A = age

H = hunger

T = thirst

M = urge to mate

p = position in the food chain

G = genetics (traits)

For a particular animal only p and G remains constant through life.

Definition 2: Rabbit and Fox

A *Rabbit* is defined as an *Animal* with $p = 2$. A *Rabbit* is prey to the animal at position $p = 1$ in the food chain. A *Fox* is defined as an *Animal* with $p = 1$.

Definition 3: Genetics

Each *Animal* has *Genetics* G defined by the six-tuple (M, S, P, Hr, Tr, V) for rabbits and (M, S, Hs, Hr, Tr, V) for foxes where:

M = mating requirement (real number)

S = size of steps the animal takes/ stride (natural number)

Hs = hunting skill (only applicable for Fox) (real number)

Hr = resistance to hunger (real number)

Tr = resistance to thirst (real number)

P = fear of predator (only applicable for Rabbit) (real number)

V = vision radius (natural number)

Definition 4: Landform

We define a *Landform* as the triplet (L, F, W) where:

L = set of the coordinates of its locations on the map
F = current food availability
W = current water availability

Landforms are of six types:

Grassland: Source of food for rabbits
Forest: Source of food for rabbits
Lake: Source of water for all animals
Pond: Source of water for all animals
Quagmire: Uninhabitable and fatal region for animals
Rugged Land: Uninhabitable and fatal region for animals

4. MEASURING THE DEGREE OF DISPERSION OF GENETIC CHARACTERISTICS

Here we will use two methods to measure the degree of dispersion: entropy and standard deviation. Entropy disfavours clustering of traits in certain domains, hence is a better method for determining the degree of “spreading out” of data. Standard deviation has no such advantage but is still used in our experiment (in the repository) because it is a commonly used statistic. The repository on github has methods for calculating other degrees of dispersion like median absolute deviation, mean absolute deviation, and coefficient of variation.

Method 1: Entropy

For a particular species a genetic character g has a domain $[m, n)$. We divide the interval into k sub-intervals of equal width. The sub-intervals are named $S_1, S_2, S_3, \dots, S_k$.

We define p_i as the proportion of animals of a particular species (say, rabbits) having a genetic trait lying in the interval S_i .

For a given current collection of rabbits R_c , the entropy of the character g is defined as:

$$\epsilon_g = - \sum_{i=1}^k (p_i) \log_2(p_i)$$

Using this method, entropy is calculated for all six of the genetic characters of R_c in each iteration. Same applies for any given current collection of foxes F_c .

Note: In the implementation shown in the paper, the value of k is chosen as 20. Different values of k can, however, be chosen.

Method 2: Standard Deviation

For a particular species, let us take into consideration, the genetic character g . Let μ_g denote the mean of the traits of the species for character g .

For a given current collection of rabbits R_c , the standard deviation of the character g is defined as:

$$\sigma = \sqrt{\frac{\sum (g_i - \mu_g)^2}{N}}$$

- g_i is the genetic trait of rabbit $r_i \in R_c$
- N is the cardinality of R_c

5. THE ALGORITHM

-**set** Gr := The set of Grasslands in the map
 -**set** F := The set of Forests in the map
 -**set** P := The set of Ponds in the map
 -**set** L := The set of Lakes in the map
 -**set** R := The set of Rugged lands in the map
 -**set** Q := The set of Quagmires in the map

-**set** $Rabbits$:= The set of rabbits with randomly assorted genetics
 -**set** $Foxes$:= The set of foxes with randomly assorted genetics

-**master function** $run_simulation(n)$:

-**for** specified number of simulations n :

-**for** each $rabbit \in Rabbits$:

-find current need N of $rabbit$ which is $\max\{H, T, M\}$ where $H, T, M \in rabbit$

-**take a small step of size** $S \in Genetics\ G$ of $rabbit$ towards nearest place (nearest rabbit for mating purposes, grassland for food, lakes or ponds for water) for satisfying need N

-**check_for_feasibility_of_current_location**($rabbit$)

-**check_if_dead**($rabbit$)

-**update_self**($rabbit$)

-**for** each $fox \in Foxes$:

-find current need N of fox which is $\max\{H, T, M\}$ where $H, T, M \in fox$

-**take a small step of size** $S \in Genetics\ G$ of fox towards nearest place (nearest fox for mating purposes, rabbit for food, lakes or ponds for water) for satisfying need N

-**check_for_feasibility_of_current_location**(fox)

-**check_if_dead**(fox)

-**update_self**(fox)

function $check_for_feasibility_of_current_location(animal)$:

-**let** $location$:= current location of animal

-**if** $animal$ is a Fox

- if *location* is **near**₁ a rabbit *r* **and** *animal.H* > some minimum value:
 - animal* eats the rabbit
 - animal.H* := *animal.H* + some value
 - Rabbits* := *Rabbits* - {*r*}
- if *location* is **near**₂ a lake $\in L$ or pond $\in P$ **and** *animal.T* > some minimum value:
 - animal* drinks water from nearest lake/ pond
 - animal.T* := *animal.H* + some value
- if *location* is **near**₂ a fox $\in Foxes$:
 - if mating is successful with nearest fox *f*:
 - f.M* := 0
 - animal.M* := -1
 - define a new Fox *offspring* with Genetics obtained from parent by the defined algorithm
 - Foxes* := *Foxes* \cup *offspring*
- if *location* is **near**₂ a quagmire $\in Q$:
 - animal.T* := *animal.T* - some value
 - animal.H* := *animal.H* - some value
- if *location* is **near**₂ a rugged land $\in R$:
 - animal.T* := *animal.T* - some value
 - animal.H* := *animal.H* - some value

Here, **near** the location of an entity means:

- location is within distance *animal.G.Hs* of the *animal* (1)
- location is within distance *animal.G.V* of the *animal* (2)

-if *animal* is a Rabbit

- if *location* is **near** a grassland $\in Gr$ **and** *animal.H* > some minimum value:
 - animal* eats the grass from the nearest grassland
 - animal.H* := *animal.H* + some value
- if *location* is **near** a lake $\in L$ or pond $\in P$ **and** *animal.T* > some minimum value:
 - animal* drinks water from nearest lake/ pond
 - animal.T* := *animal.H* + some value
- if *location* is **near** a rabbit $\in Rabbits$:
 - if mating is successful with nearest rabbit *r*:
 - r.M* := 0
 - animal.M* := -1
 - define a new Rabbit *offspring* with Genetics obtained from parent by the defined algorithm
 - Rabbits* := *Rabbits* \cup *offspring*
- if *location* is **near**₂ a quagmire $\in Q$:
 - animal.T* := *animal.T* - some value
 - animal.H* := *animal.H* - some value
- if *location* is **near**₂ a rugged land $\in R$:
 - animal.T* := *animal.T* - some value
 - animal.H* := *animal.H* - some value

Here, **near** the location of an entity means:

-location is within distance $animal.G.V$ of the *animal*

function check_if_dead(*animal*):

-if $animal.age > MAX_AGE_{species\ of\ animal}$
-remove *animal* from the set of its species
-else if $animal.T > MAX_THIRST_{species\ of\ animal}$
-remove *animal* from the set of its species
-else if $animal.H > MAX_HUNGER_{species\ of\ animal}$
-remove *animal* from the set of its species

function update_self(*animal*):

- $animal.H := animal.H + animal.G.Hr$
- $animal.T := animal.T + animal.G.Tr$
- $animal.M := animal.M + \text{some small value}$

Mating Algorithm:

Say, animal A_1 and animal A_2 are taken under consideration

Mating will be successful if and only if $A_1.M$ and $A_2.M$ are both greater than some value

Say, the offspring is animal O .

$$\forall g_i \in O.G, O.G.g_i = (A_1.G.g_i) \uparrow (A_2.G.g_i)$$

Here the $c = a \uparrow b$ means c is assigned the value a with a 50% chance and value b with a 50% chance.

6. INITIAL SETUP

The initial population is distributed randomly over the map. The initial genetic configuration is as follows:

In the initial generation the rabbits are assigned uniformly random genetic traits in the following domains:

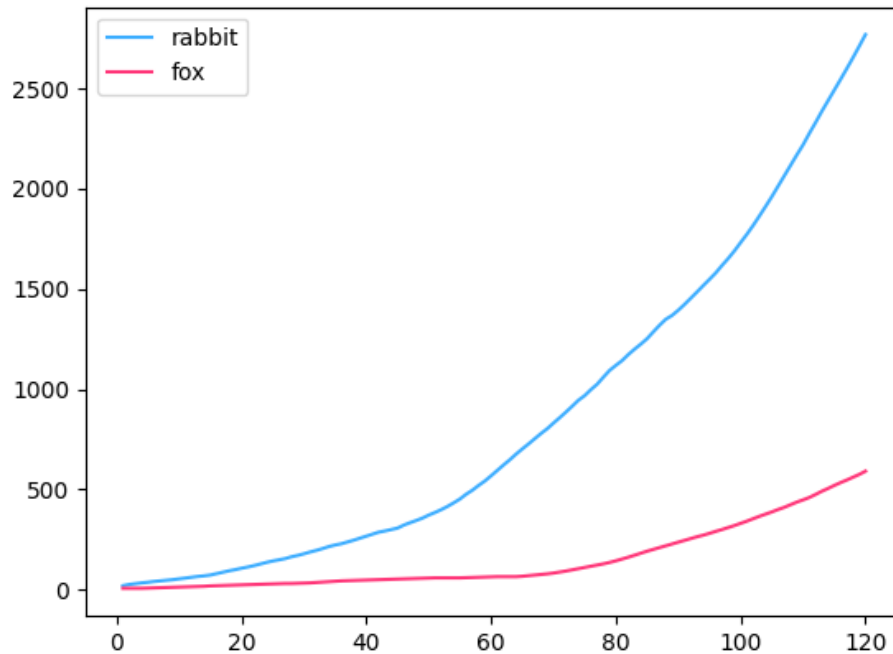
- $M : [0, 4)$
- $Hr : [0, 1)$
- $Tr : [0, 1)$
- $P : [0, 4)$
- $V : [80, 200)$
- $S : [15, 40)$

In the initial generation the foxes are assigned uniformly random genetic traits in the following domains:

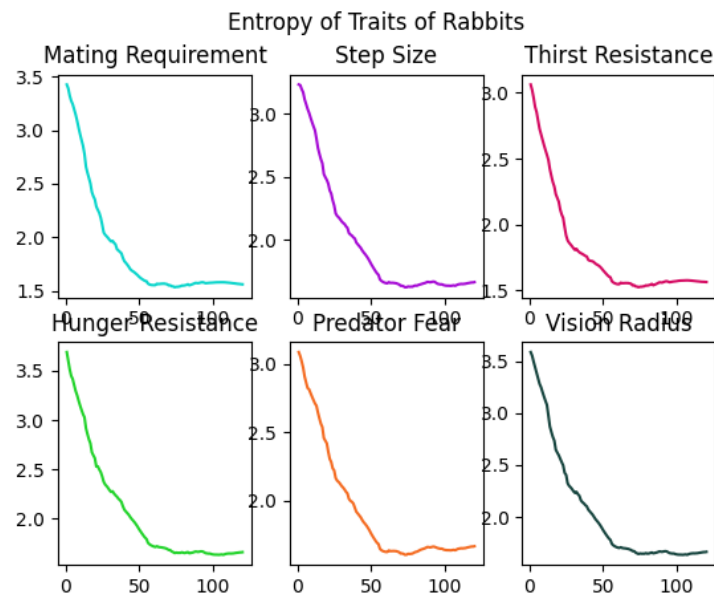
- $M : [0, 5)$
- $Hr : [0, 0.5)$
- $Tr : [0, 0.5)$
- $Hs : [130, 200)$
- $V : [100, 280)$
- $S : [20, 25)$

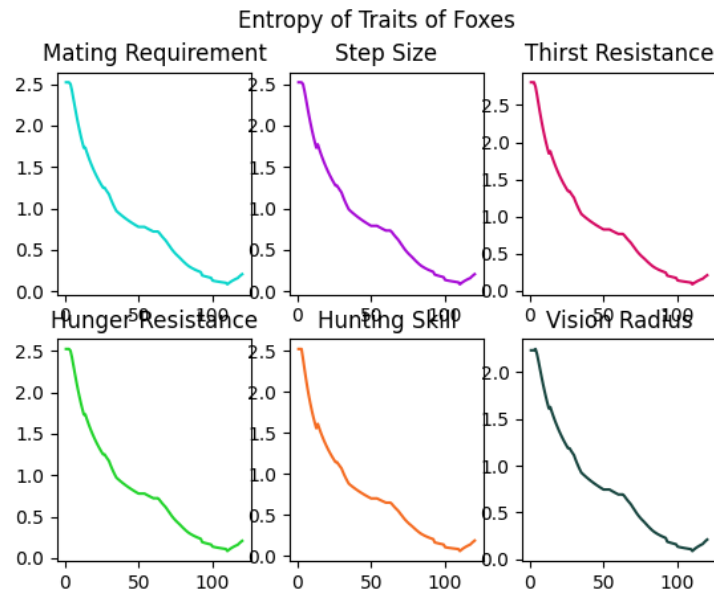
7. TEST SIMULATION

We run the simulation for 120 iterations. The initial population is set to 20 rabbits and 7 foxes. The population curve is shown below: (the X-axis shows the iteration number and Y-axis shows the population i.e number of foxes (in red) and rabbits (in blue) which are alive)



The corresponding degree of dispersion (measured by entropy only) in the genetic traits of rabbits and foxes are shown below:





The above graphs show that with increase in population, the genetic dispersion decreases, that is, the species tend to progress towards the fitter trait. The genetically inferior animals die out while the superior ones persist. This leads to a lesser degree of dispersion of genetic traits.

8. CONCLUSIONS

The performed experiment shows that Darwin's Theory of Evolution is valid. As the population increases with time, the species tend to evolve towards the fitter trait. That is, the weaker individuals die out whereas the fitter one survive. This causes a decrease in the overall degree of dispersion of the genetic characters of both the species. The above results were obtained by taking specific values of constants, specific initial configurations and mating mechanisms, and specific domains for distribution of initial genetic traits. The program, available on github <https://github.com/Kaval314/Habitat-Simulation> can be used with different values of these constants and configurations. For this specific model of the simulation, the genetic traits do progress to the fittest trait in accordance with Darwin's Theory of Natural Selection. There is also a mechanism in the program for studying deaths (not shown here) which can be used to further study any model. The program on github contains an abstract class 'Animal' which can be extended to create other animals and make the ecosystem more complex. The aforementioned algorithms can be implemented to these animals with cosmetic changes. Further, methods to study the strain on natural resources of the landscape (not shown here) are also provided. A live visualization of the ecosystem (with the movement of the animals) is also provided.

9. REFERENCES

- [1] Charles Darwin, On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life, 1859
- [2] Cover, T.M.; Thomas, J.A. Elements of Information Theory; Wiley: New York, NY, USA, 1991
- [3] Han, Te Sun & Kobayashi, Kingo, "Mathematics of Information and Coding", American Mathematical Society. ISBN 978-0-8218-4256-0, 2002

- [4] Borda, Monica, "Fundamentals in Information Theory and Coding", Springer. ISBN 978-3-642-20346-6., 2011
- [5] David J. C. MacKay, "Information Theory, Inference and Learning Algorithms", 2003
- [6] <https://github.com/Kayal314/Habitat-Simulation>