PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

MAY 2020: IN SEMESTER ASSESSMENT (ISA) B.TECH. IV SEMESTER

UE18MA251- LINEAR ALGEBRA

MINI PROJECT REPORT

ON

APPLICATIONS OF LINEAR ALGEBRA IN GENETICS

(FLOWER COLOR GENOTYPE)

Submitted by

1. Name : Eshan Naik SRN : PES2201800156

2. Name : Sunad Suhas SRN : PES2201800511

Branch & Section : CSE Section - A

PROJECT EVALUATION

(For Official Use Only)

Sl.No.	Parameter	Max Marks	Marks Awarded
1	Background & Framing of the problem	4	
2	Approach and Solution	4	
3	References	4	
4	Clarity of the concepts & Creativity	4	
5	Choice of examples and understanding of the topic	4	
6	Presentation of the work	5	
	Total	25	

Name of the Course Instructor : Dr.Girish.V.R.

Signature of the Course Instructor :

Background

Genetics is a field of biology which deals with the inheritance of traits from one generation to the next.

Each gene has two alternate forms called alleles. They are denoted as A and a, ‘A’ being for a dominant gene and ‘a’ being for the recessive gene. Each individual carries a pair of alleles which can be similar or different. The combination of the alleles gives the persons genotype and how the trait is shown by the living organism. There are 3 possible genotypes for a given trait - **AA, Aa (Aa shows dominant trait) and aa .**

When two individuals breed the offspring inherits one gene from each parent, hence we can assume that the genes are equally likely to be inherited from both parents. Whenever breeding occurs the genotype distribution changes with every generation. This is where linear algebra will be used.

The genotype distribution of a generation can be determined by a vector g_n. Where,

$$\mathbf{g_n} = \begin{matrix} a\_n \\ b\_n \\ c\_n \end{matrix}$$

Let a be the portion of the population having genotype AA, b be the portion of the population having genotype Aa and c be the portion of the population having genotype aa. Since the genotype changes every generation, we can use a general expression,

$$\mathbf{g_n = Yg_{n-1}}$$

where Y is a matrix and n = 0 ,1 ,2....... By using this equation, we can find how the genotype distribution of the nth generation is related to the first generation.

$$\mathbf{g_n = Yg_{n-1} = Y^2\, g_{n-2} = \ldots\ldots = Y^n\, g_0}$$

We then diagonalize Y to get an explicit expression so that it becomes easy to calculate for large values of n.

$$\mathbf{Y^n = P\, D^n\, P^{-1}}$$

Once Y is diagonalized the Eigen values of Y are the diagonal values of the matrix D and the columns of P are the eigenvectors respectively for each eigenvalue. Hence the final expression is,

$$\mathbf{g_n = P\, D^n\, P^{-1}\, g_0}$$

This expression being an explicit expression will be easy to compute and the it only depends on the initial generation’s genotype distribution (g_0).

Question

In an experimental farm, a large population of flowers consists of all possible genotypes (AA, Aa, and aa), with an initial frequency of $a0 = .05$, $b0 = .90$, and $c0 = .05$, respectively. **Suppose that this genotype controls flower color, and that each flower is fertilized by a flower of a genotype similar to its own.** Find an expression for the genotype distribution of the population after any number of generations. Use this equation to predict the genotype distribution of the population after 4 generations, and predict what the genotype distribution of the population will be after an infinite number of generations.

Answer

The genotype distribution of the flower color can be taken as a vector given below,

$$\mathbf{g_n} = \begin{matrix} \boldsymbol{a\_n} \\ \boldsymbol{b\_n} \\ \boldsymbol{c\_n} \end{matrix}$$

Let a be the portion of the population having genotype AA, b be the portion of the population having genotype Aa and c be the portion of the population having genotype aa. **If each plant is fertilized with a plant with the same genotype, then the possible combinations are AA and AA, Aa and Aa, and aa and aa.**

The probabilities of these combinations in the genotype is given in the table below -

		Genotypes of Parents		
		AA, AA	Aa, Aa	aa, aa
Genotype of offspring	AA	1	1/4	0
	Aa	0	1/2	0
	aa	0	1/4	1

As stated earlier when two individuals breed, the offspring inherits one gene from each parent, hence we can assume that the genes are equally likely to be inherited from both parents. By using simple probability, we can determine how likely the offspring is to get either of the genotypic combinations. We can use the **Punnett Square** to predict the relative proportions of the genotype of the offspring from the genotypes of the parent. For example, if the genotype of the parents is Aa and Aa then the Punnett Square will be -

	A	a
A	AA	Aa
a	Aa	aa

This shows that (¼) of the offspring will be of genotype AA, (½) will be of genotype Aa, and (¼) will be of genotype aa. The following equations determine the frequency of each genotype as dependent on the preceding generation.

$$\mathbf{a_n = a_{n-1} + (1/4)\, b_{n-1}}$$

$$\mathbf{b_n = (1/2) bn\text{-}1} \qquad \mathbf{(n = 1,2\ldots)}$$

$$\mathbf{c_n = (1/4) b_{n-1} + c_{n-1}}$$

This can be written in the matrix notation as:

$$\mathbf{g_n = Yg_{n-1}}$$

Where,

$$\mathbf{g_n} = \begin{matrix} a\_n \\ b\_n \\ c\_n \end{matrix}$$

$$\mathbf{g_{n-1}} = \begin{matrix} a\_n - 1 \\ b\_n - 1 \\ c\_n - 1 \end{matrix}$$

$$\mathbf{Y} = \begin{matrix} 1 & 0.25 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0.25 & 1 \end{matrix}$$

As discussed earlier we now need to diagonalize Y. Hence, we need to find the eigenvalues and eigenvectors corresponding to its eigenvalues. By using a calculator, we get the Eigen values to be $\lambda_1 = 1$ and $\lambda_2 = 0.5$.

The Eigen space for $\lambda_1 = 1$ is,

$$\mathbf{g} = \begin{matrix} g\_1 \\ g\_2 \\ g\_3 \end{matrix} = \mathbf{g_1} \begin{matrix} 1 \\ 0 \\ 0 \end{matrix} + \mathbf{g_3} \begin{matrix} 0 \\ 0 \\ 1 \end{matrix}$$

A basis for this Eigen space is,

$$\begin{matrix} \mathbf{1} \\ \mathbf{0} \\ \mathbf{0} \end{matrix} \text{ and } \begin{matrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{matrix}$$

Therefore, the Eigen space corresponding to $\lambda_1 = 1$ is two-dimensional.

The Eigen space for $\lambda_2 = 0.5$ is,

$$\mathbf{g} = \begin{matrix} g\_1 \\ g\_2 \\ g\_3 \end{matrix} = \mathbf{g_3} \begin{matrix} \mathbf{1} \\ \mathbf{-2} \\ \mathbf{1} \end{matrix}$$

A basis for this Eigen space is,

$$\begin{matrix} \mathbf{1} \\ \mathbf{-2} \\ \mathbf{1} \end{matrix}$$

Therefore, the Eigen space corresponding to $\lambda_2 = 0.5$ is one-dimensional.

The diagonal matrix D = $\begin{matrix} \lambda\_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \lambda\_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \lambda 2 \end{matrix}$

$$= \begin{matrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0.5} \end{matrix}$$

P is the matrix with the eigenvectors. Hence P = $\begin{matrix} \mathbf{1} & \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{-2} \\ \mathbf{0} & \mathbf{1} & \mathbf{1} \end{matrix}$

P^{-1} can also be calculated using the formula. Hence P^{-1} = $\begin{matrix} \mathbf{1} & \mathbf{0.5} & \mathbf{0} \\ \mathbf{0} & \mathbf{0.5} & \mathbf{1} \\ \mathbf{0} & \mathbf{-0.5} & \mathbf{0} \end{matrix}$

As stated earlier,

$$\mathbf{g_n = P\, D^n\, P^{-1}\, g_0}$$

$$\mathbf{g_n} = \begin{matrix} \mathbf{1} & \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{-2} \\ \mathbf{0} & \mathbf{1} & \mathbf{1} \end{matrix} \quad \begin{matrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0.5\hat{}}n \end{matrix} \quad \begin{matrix} \mathbf{1} & \mathbf{0.5} & \mathbf{0} \\ \mathbf{0} & \mathbf{0.5} & \mathbf{1} \\ \mathbf{0} & \mathbf{-0.5} & \mathbf{0} \end{matrix}$$

$$\begin{matrix} a\_0 \\ b\_0 \\ c\_0 \end{matrix}$$

$$g_n = \begin{matrix} 1 & 0.5-0.5^{\wedge}(n+1) & 0 \\ 0 & 0.5^{\wedge}n & 0 \\ 0 & 0.5-0.5^{\wedge}(n+1) & 1 \end{matrix} \quad \begin{matrix} a\_0 \\ b\_0 \\ c\_0 \end{matrix}$$

Finally,

$$\mathbf{a_n = a_0 + [0.5 - 0.5^{n+1}]b_0}$$

$$\mathbf{b_n = 0.5^n\, b_0} \qquad \mathbf{(n = 1,2,\ldots)}$$

$$\mathbf{c_n = c_0 + [0.5 - 0.5^{n+1}]b_0}$$

These equations give the genotype distribution of the population after any number of generations.

To find the genotype distribution after 4 years (n=4) we can use the above equation. We know that $a_0 = 0.05$, $b_0 = 0.90$, and $c_0 = 0.05$.

Hence on substituting these values into the above equation we get -

$$\mathbf{a_4 = 0.47}$$

$$\mathbf{b_4 = 0.06}$$

$$\mathbf{c_4 = 0.47}$$

That is, originally (n = 0), in a population of 100 individuals, 5 would be AA, 90 would be Aa, and 5 would be aa. After 4 generations, 47 individuals would be AA, only 6 would be Aa, and 47 would be aa.

If we let n ->∞, we find that -

$$\mathbf{a_n = 0.5}$$

$$\mathbf{b_n = 0}$$

$$\mathbf{c_n = 0.5}$$

That is, in the limit as n -> ∞, 0.5 of the individuals in the population are AA, 0.5 are aa, and none are Aa.

The breeding program in this example is an extreme case of "inbreeding", which is mating between individuals with similar genotypes.

References

1. Concise Biology, I.C.S.E textbook Class – X.
2. Linear Algebra and Its Applications by Gilbert Strang 4th Edition.
3. https://sites.math.washington.edu

4. Rorres, Chris and Anton, Howard. Applications of Linear Algebra. John Wiley And Sons, 1977
5. Farr, William M. Modeling Inheritance of Genetic Traits. http://www.math.wpi.edu/Course_Materials/MA2071A98/Projects/gene/node1.html.

Clarity of concept

From the above example we understand how the genotype distribution of a gene of a particular organism can be predicted using Linear Algebra.

We use the concept of difference equation, diagonalization of a matrix, inverse of a matrix, eigenvalues and eigenvectors to solve the problem.

The basic formula we use here is Ax = B where, $x = g_n$, A = Y and B is the solution.

A will be diagonalized so it will be easier to compute for larger values of n.

Once we find A, we multiply it with x to get the answer in the form of 3 equations for the 3 possible ways the trait can be combined. Using these expressions, we can find the genotype distribution of any generation.

This is how Genetics can be considered an application of Linear Algebra.