JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

APPLICATIONS OF LINEAR ALGEBRA IN GENETICS

Vandar Kuzhali .V, Monika .I, Ishwarya .D, Saranya .A

Abstract: This paper explores how linear algebra can be applied to genetics. We will journey through the fundamental applications of linear algebra in genetics, witnessing how matrices, vectors, and linear equations play pivotal roles in modeling, calculating, and interpreting genetic phenomena. Our aim is to demonstrate how, using only the genotype distribution of the starting population, linear algebra can be used to predict the genotype distribution of a given trait in a population after any number of generations. This paper delves into how linear algebraic techniques aid in understanding hereditary patterns, population genetics, and how linear algebra is used to study Autosomal inheritance. By elucidating these applications, the abstract sheds light on the indispensable role of linear algebra in unraveling the complexities of genetics.

1.Introduction:

The confluence of mathematics and biology is explained by the title. Its foundation is the use of mathematics in biology. One area of mathematics called linear algebra is also very helpful in solving a number of real-world issues, such as those in genetics.

The area of biology known as genetics studies how genes are passed down through generations and how inherited traits differ within a species, grouping, or individual organism.

In genetics, linear algebra is essential for tasks like understanding inheritance patterns, modeling biological systems, and interpreting genetic data.

Jared Kirkham (2001) studied the genotypic distribution of the flower population in the n^{th} generation, using the present generation as the initial population.

Population genetic analysis has been made possible by the Hardy-Weinberg law. It gives mathematical approach to genetics and evolution, which can be used to predict the equilibrium rates of given alleles as well as the frequencies resultant after one generation of mating.

In order to represent genetic features, gene interactions, and population genetics, matrices are frequently used. This helps to enhance personalized medicine and genetic research. In addition to providing tools for modeling biological systems and deciphering the complexities of inheritance patterns, linear algebra offers a strong foundation for studying and interpreting complex genetic data.

2.Applications of Linear Algebra in genetics:

- Linear algebra is used to find the genotype distribution of an organism. By calculating the powers of the matrix, one can investigate the possibility that an inherited trait is passed down across the generations.
- Linear algebra is also used in the study of autosomal inheritance. By computing eigen values, eigen vectors, and matrix notation, autosomal recessive inheritance disease effectiveness can be counted.

- In population genetics, linear algebra is used in the analysis of allele frequencies and genetic diversity within populations, providing insights into evolutionary processes.

3.Genotypes distribution in n^{th} generation: Particular strategies exist for the expression of genes that we inherit from our biological parents. Most genes are found in pairs. One comes from the father, and the other from the mother. The term "genotype" describes the genetic makeup of living things. The entire collection of genes passed down from parents to children is known as a genotype. The frequency or percentage of various genetic variations, or genotypes, within a population is referred to as the genotype distribution. It symbolizes the diversity of genetic composition that exists, indicating the prevalence of different genotypes within the population.

Method: Assume that the starting genotype distribution is X_0. Let A represent the genotype transition matrix for the specified population.

After a year, the genotype of the population X_1 is determined by multiplying the initial matrix X_0 by the product of transition matrix A.

$$AX_0 = X_1$$

After two years, the population's genotype can be $X_2 = AX_1 = A(AX_0) = A^2X_0$.

After three years, the population's genotype may be $X_3=AX_2=A(A_2X_0) = A^3X_0$.

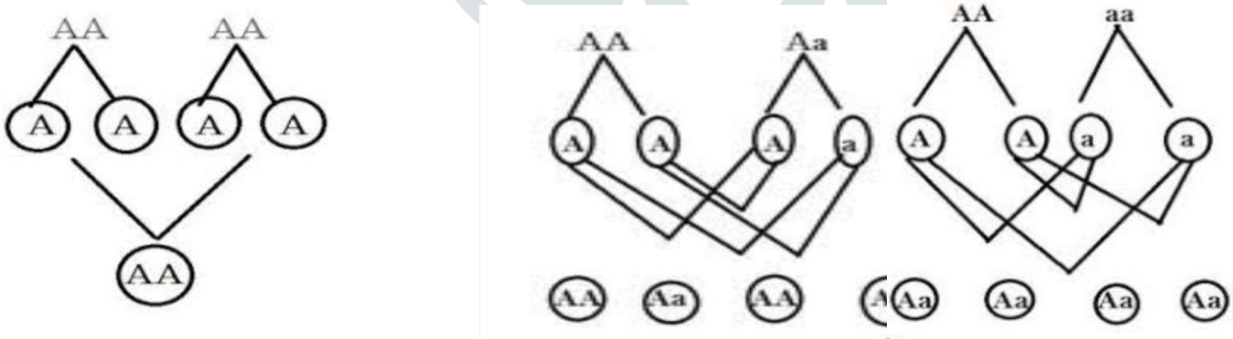
The population's genotype may be $X_n = A^nX_0$ after n years.

The diagonalization of the matrix method is an easier one to determine the n^{th} individual based on the parent genotype probability, and obtained equation $X_n = A^nX_0$.

A square matrix A_{nxn} can be said to be diagonalizable if there is matrix P which has an inverse such that $D = P^{-1}AP$ is a diagonal matrix. Matrix P is called the matrix that diagonalizes matrix A.

Problem: A farmer has a large population of plants with a certain distribution of the three possible genotypes: AA, Aa, and aa. This man wants to start a farming scheme whereby a plant with genotype AA will fertilize every plant in the community, resulting in a higher yield. He wants to know the probabilities of possible genotypes and use this as the initial population to find the genotype distribution of the population after a certain number of generations.

Solution:



AA-AA **AA-Aa** **AA-aa**

Probabilities of the possible genotype

Genotypes	AA-AA	AA-Aa	AA-aa
AA	1	1/2	0
Aa	0	1/2	1
aa	0	0	0

Assume A to be the starting population, $A = \begin{bmatrix} 1 & 1/2 & 0 \\ 0 & 1/2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

Assume that each genotype was present in equal proportion in the initial plant population.

$X_0 = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$ is the initial distribution vector, hence it may be found this way.

The distribution is $X_1 = AX_0 = \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \end{bmatrix}$ after a year.

Then $X_2 = \begin{bmatrix} 3/4 \\ 1/4 \\ 0 \end{bmatrix}$, $X_3 = \begin{bmatrix} \frac{7}{8} \\ \frac{1}{8} \\ 0 \end{bmatrix}$, $X_4 = \begin{bmatrix} \frac{15}{16} \\ \frac{1}{16} \\ 0 \end{bmatrix}$,….

The genotype distribution of n^{th} generation will be found by using Diagonalization method.

$$A = P^{-1}DP$$

For n years, $X_n = A^nX_0$, where A^n can be written as $A^n = PD^nP^{-1}$, n=1, 2, 3, …

For the matrix A, the eigen values are $\lambda_1 = 1$, $\lambda_2 = 1/2$, $\lambda_3 = 0$ and the corresponding eigen vectors are

$$V_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad V_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad V_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

Therefore, the diagonal matrix $D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and matrix $P = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$

Then the inverse matrix $P^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$

$$X_n = P\, D^n\, P^{-1}\, X_0$$

$$X^n = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1^n & 0 & 0 \\ 0 & \left(\frac{1}{2}\right)^n & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} + \frac{1}{3} - \frac{1}{3}\left(\frac{1}{2}\right)^n + \frac{1}{3} - \frac{1}{3}\left(\frac{1}{2}\right)^{n-1} \\ \frac{1}{3}\left(\frac{1}{2}\right)^n + \frac{1}{3}\left(\frac{1}{2}\right)^{n-1} \\ 0 \end{bmatrix}$$

where n = 1, 2, 3, …….

Therefore, under a crossbreeding program, the proportion of the population was determined for the n^{th} generation.

4.Autosomal Inheritance: The pattern of inheritance of genes found on autosomes—non-sex chromosomes other than X and Y—is referred to as autosomal inheritance. Autosomes in humans range in number from 1 to 22. Traits that are autosomal dominant or autosomal recessive are examples of autosomal inheritance patterns.

- ***Autosomal Dominant Inheritance*:** Only one copy of the dominant allele from either parent is needed for the trait to be expressed.
- ***Autosomal Recessive Inheritance*:** Both copies of the gene (alleles) must be recessive for the trait to be expressed.

These inheritance patterns are fundamental to understanding how traits are passed from one generation to the next for genes located on autosomes.
(*Allele*: one of two or more possible forms of a gene that are found at the same place on a chromosome).

Problem: If one of the parents has Huntington's disease (autosomal dominant) and the other does not, what is the probability that their child will inherit the disease?

Solution:

Let Hh allele be one of the parents with Huntington's disease, where H is the dominant allele for the disease, and let hh allele be the other one without disease.

Parent 1: Hh can be represented as the vector $\begin{bmatrix}1\\1\end{bmatrix}$

The first element (1) represents the presence of Huntington's allele (H), and the second element (1) represents the normal allele (h).

Parent 2: hh can be represented as the vector $\begin{bmatrix}0\\1\end{bmatrix}$

The first element (0) represents the absence of Huntington's allele (H), and the second element (1) represents the normal allele (h).

The matrix $\begin{bmatrix}1 & 1\\0 & 1\end{bmatrix}$ is the combination matrix. (by simple Mendelian inheritance)

Now we multiply the combination matrix by the allele of the parent 1, $\begin{bmatrix}1 & 1\\0 & 1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}=\begin{bmatrix}2\\1\end{bmatrix}$

$\therefore$ The resulting vector $\begin{bmatrix}2\\1\end{bmatrix}$ represents the allele in the offspring.

The first element (2) indicates that there are two copies of Huntington's allele (H) in the offspring and the second element (1) indicates that there is one copy of the normal allele (h) in the offspring.
Therefore, the offspring will inherit the Huntington's disease (Hh) based on this genetic combination. The probability of inheriting the disease is 2/3 (67%), and the probability of not inheriting is 1/3 (33%).

5.Population genetics: A population is a collection of individuals from a species that are reproducing together and inhabiting a specific area at a given time. Population genetics is the study of how phenotypic features are inherited within a particular population. This is based on an idea that was independently put forth by German physician Weinberg and British mathematician Hardy in 1908.
The simple formula $p + q = 1$ represents the Hardy-Weinberg law.
where q is the frequency of a recessive gene and p is the frequency of a dominant gene.
The frequency of dominant and recessive genes in a population can be determined using this formula.

We can calculate the frequency of homozygotes and heterozygotes in a population using an additional Hardy-Weinberg formula: $(p + q)^2 = p^2 + 2pq + q^2$
where p is the dominant gene's frequency.

q = Recessive gene frequency
p^2 = Dominant homozygote frequency
2pq = Heterozygote Frequency
q^2 = Recessive homozygote frequency

Problem: The frequency of blood types in a sample of 300 persons is as follows:

Type M = 43.3%

Type N =10%

Type MN = 46.7%

Does this fit the assumptions of segregation and random mating?

Solution:

The frequency of Type M = 43.3%

i.e., $p^2 = 43.3\% = 0.433$

$p = 0.658$

By Hardy-Weinberg law, $p + q = 1$

$q = 1 - p$

$q = 1 - 0.658 = 0.342$

$q^2 = 0.1169$ or 11.69%

i.e., The frequency of Type N = 11.69%

The frequency of Type MN = 2pq = 2 (0.658) (0.342) = 0.4501 or 45.01%

	Expected	**Observed**
Type M(p^2)	43.3%	43.3%
Type N(q^2)	11.69%	10%
Type MN(2pq)	45.01%	46.7%

On comparison it seems that these observed results fit the assumptions of segregation and random mating.

6.Conclusion: This paper attempted to discuss the application of linear algebra in genetics. One of the major applications is to find genotype distribution in the n^{th} generation, which was found with the help of the diagonalization method and by computing the highest power of the matrix. This work also explained how linear algebra is used to study autosomal inheritance and population genetics, which provided different methods to solve the problems. This type of problem, where case analyses are carried out, shows the close relationship between mathematics and biology, showing a great convergence between these two sciences and contributing to the development of each one of them.

7.Reference:

1. David. C. Lay (2000), Linear algebra and its application, Addison Wesley, New York.
2. Dr. R. P. Meyyan (2013), Genetics, Saras Publication.
3. Jared Kirkham (2001), Linear Algebra in genetics, math 308 project.
4. John Gallego, Sunit Kambli, Daniel Lee, Linear algebra and genetics, Math 247.
5. M.Naga Maruthi kumari, Kavya. H.U and Krupashree.Y.N (2018), Applications of linear algebra in genetics, Journal Of Harmonized Research (JOHR), Journal of Harmonised Research in Applied Sciences 6(4), 236-240.
6. Robert H. Tamarin, Principles of genetics.