INDEX

ABSTRACT

Linear algebra is a branch of mathematics that is related to the study of lines and planes, vector spaces and mappings that are required for linear transformation. Since it is the mathematics of data, the tools of linear algebra are used in many domains such as machine learning, robotics , biology and so on. Linear algebra also plays a major role in genetics. Genetics is a branch of biology concerned with the study of genes, genetic variation, and heredity in organisms. With the help of linear algebra, the genotype of the existing individual can be found out.

Since the distribution varies over generations, matrices may be used to calculate the probability of each trait in each generation. To evaluate these characteristics, we can use matrix transformations, which is a part of linear algebra. Each row of the matrix represents the potential offspring genotypes, while each column of the matrix represents a mixture of parents' genotypes. This method makes it simple to decide the genotype of the nth generation. Since manual calculation is extremely difficult and time consuming, the diagonalization approach has been used to determine the genotype.

INTRODUCTION AND BACKGROUND

Genetics is the study of inheritance, or the transmission of traits from one generation to the next.
There are two types of genetic disorders:

- Sex linked
- Autosomal

Sex linked disorders are those disorders that are caused due to mutations present in the sex chromosomes which are X chromosomes and Y chromosomes of women and men respectively.
Description of a sex linked inheritance with algebras involves overcoming the obstacle of asymmetry in the genetic inheritance rules. The main problem for a given algebra of a sex linked population is to carefully examine how the basic algebraic model must be altered in order to compensate for this lack of symmetry in the genetic inheritance system.
E.g.- Color blindness, Hemophilia etc.

In autosomal inheritance, each heritable trait is assumed to be governed by a single gene on a chromosome. These disorders are caused due to mutations in genes on the autosomes, or numbered chromosomes. There are typically two different forms, or alleles of a gene (denoted by A and a). Each individual in a population carries a pair of alleles, which may be similar or different. These pairs are called an individual's genotype, and there are three possible genotypes for a particular trait: AA, Aa, or aa.
E.g.- Sickle cell anemia, Thalassemia, Phenylketonuria etc.

Autosomal disorders can be further classified as dominant or recessive. “Dominant" means that a single copy of the disease-associated mutation is enough to cause the disease. This is in contrast to a recessive disorder, where two copies of the mutation are needed to cause the disease.

It is the genotype that determines how the trait controlled by the gene is manifested in the individual. For example, in humans, eye coloration is controlled through autosomal inheritance. In such a case, the A allele is dominant over the allele, or that the allele is recessive to the A allele, since genotype Aa has the same outward trait as genotype AA.
In each generation the likelihood of each trait can be determined through matrices, since the distribution changes through generations.
We will be using transformation of matrix to determine these traits, which is a part of Linear algebra.
Linear algebra is the branch of mathematics concerning vector spaces and linear mappings between such spaces.
Each column of the matrix represents a combination of parents’ genotypes. Each row of the matrix will then represent the possible offspring genotypes.

The general genotype combinations are:

AA x AA		
	A	A
A	AA	AA
A	AA	AA

Here the cross is between the genotype ‘AA’ and the genotype ‘AA’, all the products obtained will also be of the genotype AA.

AA x Aa		
	A	A
A	AA	AA
a	Aa	Aa

Here the cross is between the genotype ‘AA’ and the genotype ‘Aa’, all the products obtained will also be of the genotype AA or Aa.

Aa x aa		
	A	a
a	Aa	aa
a	Aa	aa

Here the cross is between the genotype ‘Aa’ and the genotype ‘aa’, all the products obtained will also be of the genotype Aa or aa.

PROBLEMS FACED WITH THE MANUAL METHOD

The manual method for finding the genotypes is very time consuming and leads to lengthy calculations which in turn may lead to errors and difficulty in solving, hence we use the Diagonalization method to solve for the genotypes.

MATRIX DIAGONALIZATION

Matrix diagonalization is the process of taking a square matrix and converting it into a special type of matrix, a so-called diagonal matrix that shares the same fundamental properties of the underlying matrix.
A matrix is diagonalizable if and only if for each eigenvalue the dimension of the eigenspace is equal to the multiplicity of the eigenvalue.
Diagonalization can be used to determine the information after n amount of years. If the matrix A can be written as a product of an invertible matrix P and a diagonal matrix D and an inverse of P,
$A = PDP^{-1}$, then the computation is much simpler.

AUTOSOMAL INHERITANCE

By calculating the eigenvalues, eigenvectors and matrix rotation autosomal recessive inheritance effectiveness could be counted.

	AA x AA	AA x Aa	AA x aa	Aa xAa	Aa x aa	aa x aa
AA	1	$\frac{1}{2}$	0	$\frac{1}{4}$	0	0
Aa	0	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	0
aa	0	0	0	$\frac{1}{4}$	$\frac{1}{2}$	1

The table above shows the possible distribution of genotypes AA, Aa and aa. From this we can derive an expression for the distribution of the three possible genotypes in the population after any number of generations.

For n = 0, 1, 2, …, lets us consider

a_n = fraction of genotype AA in n^{th} generation

b_n = fraction of genotype Aa in n^{th} generation

c_n = fraction of genotype aa in n^{th} generation

Thus a_0, b_0 and c_0 specify the initial distribution of the genotypes. We also have that

$$a_n + b_n + c_n = 1, \text{ for } n = 0, 1, 2, \ldots$$

From that table we can determine the genotype distribution of each generation from the genotype distribution of the preceding generation by the following equations:

$a_n = a_{n-1} + {}^1/_2\, b_{n-1}$

$b_n = c_{n-a} + {}^1/_2\, b_{n-1}$ $\qquad n = 1, 2, \ldots$

$c_n = 0$

This equation can be written in matrix notation as:

$x^{(n)} = Mx^{(n-1)}$, $\qquad n = 1, 2, \ldots$

Where,

$$x^{(n)} = \begin{bmatrix} a_n \\ b_n \\ c_n \end{bmatrix}$$

$$x^{(n=1)} = \begin{bmatrix} a_{n-1} \\ b_{n-1} \\ c_{n-1} \end{bmatrix}$$

$$M = \begin{bmatrix} 1 & 1/2 & 0 \\ 0 & 1/2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

From the above equation,

$$x^{(n)} = Mx^{(n-1)} = M^2x^{(n-2)} = \ldots = M^nx^{(0)}$$

Consequently we can find an explicit expression for M^n, we can use the above equation to obtain an explicit expression for x^n. To find an explicit expression for M^n, we first diagonalize M. That is, we fine an invertible matrix P and diagonal matrix D such that

$$M = PDP^{-1}$$

Where,

$$D^n = \begin{bmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ 0 & 0 & 0 & \ldots & \lambda_k \end{bmatrix} = \begin{bmatrix} \lambda_1^n & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2^n & 0 & \ldots & 0 \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ 0 & 0 & 0 & \ldots & \lambda_k^n \end{bmatrix}$$

The diagonalization of M is accomplished by finding its eigenvalues and corresponding eigenvectors. These are as follows:

Eigenvalues -

$$\lambda_1 = 1, \lambda_2 = ½, \lambda_3 = 0$$

Corresponding eigenvectors -

$$V1= \begin{bmatrix}1\\0\\0\end{bmatrix}, \quad V2= \begin{bmatrix}1\\-1\\0\end{bmatrix}, \quad V3= \begin{bmatrix}1\\-2\\1\end{bmatrix}$$

$$D = \begin{bmatrix}\lambda_1 & 0 & 0\\0 & \lambda_2 & 0\\0 & 0 & \lambda_3\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\0 & \frac{1}{2} & 0\\0 & 0 & 0\end{bmatrix}$$

and

$$P = [\, v_1 \mid v_2 \mid v_3 \,] = \begin{bmatrix}1 & 1 & 1\\0 & -1 & -2\\0 & 0 & 1\end{bmatrix}$$

Therefore,

$$X^{(n)} = PD^nP^{-1}X^{(0)}$$

$$= \begin{bmatrix}1 & 1 & 1\\0 & -1 & -2\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\0 & (\frac{1}{2})^n & 0\\0 & 0 & 0\end{bmatrix}\begin{bmatrix}1 & 1 & 1\\0 & -1 & -2\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}a_0\\b_0\\c_0\end{bmatrix}$$

$$x^{(n)} = \begin{bmatrix}a_n\\b_n\\c_n\end{bmatrix} = \begin{bmatrix}1 & 1-(\frac{1}{2})^n & 1-(\frac{1}{2})^{n-1}\\0 & (\frac{1}{2})^n & (\frac{1}{2})^{n-1}\\0 & 0 & 0\end{bmatrix}\begin{bmatrix}a_0\\b_0\\c_n\end{bmatrix}$$

$$x^{(n)} = \begin{bmatrix} a_n \\ b_n \\ c_n \end{bmatrix} = \begin{bmatrix} 1 & 1-(\frac{1}{2})^n & 1-(\frac{1}{2})^{n-1} \\ 0 & (\frac{1}{2})^n & (\frac{1}{2})^{n-1} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_0 \\ b_0 \\ c_0 \end{bmatrix}$$

$$= \begin{bmatrix} a_0 + b_0 + c_0 - (\frac{1}{2})^n b_0 - (\frac{1}{2})^n c_0 \\ (\frac{1}{2})^n b_0 - (\frac{1}{2})^{n-1} c_0 \\ 0 \end{bmatrix}$$

Using the fact that $a_0 + b_0 + c_0 = 1$, we thus have

$a_n = 1 - (\frac{1}{2})^n b_0 - (\frac{1}{2})^n c_0$

$b_n = (\frac{1}{2})^n b_0 - (\frac{1}{2})^{n-1} c_0$ $\qquad$ n = 1, 2, . . .

$c_n = 0$

These are explicit formulas for the fractions of the three genotypes in the nth generation.

Because $(\frac{1}{2})^n$ tends to zero as n approaches infinity, it follows from these equations that

$a_n \longrightarrow 1$

$b_n \longrightarrow 0$

$c_n \longrightarrow 0$

as n approaches infinity. That is, in the limit, all the genotypes obtained will be AA.

So, genotype AA will be expressed while genotype Aa and aa will not be expressed

- Sickle cell anemia

Sickle cell anemia is an autosomal recessive disease therefore it can be transmitted from the parents to the offspring when both male and female are carriers of the gene. A person who is completely diseased would die before attaining maturity. The disease is controlled by a single pair of allele, HB^A and HB^S. Thus, three genotypes are possible in the population:

- $HB^A\ HB^A$ (normal, homozygous)
- $HB^A\ HB^S$ (normal, carrier)
- $HB^S\ HB^S$ (diseased, die before attaining maturity)

HB^AHB^A x HB^AHB^A		
	HB^A	HB^A
HB^A	HB^AHB^A	HB^AHB^A
HB^A	HB^AHB^A	HB^AHB^A

Here the cross is between the genotype '$HB^A\ HB^A$' and the genotype '$HB^A\ HB^A$', all the products obtained will also be of the genotype $HB^A\ HB^A$.
The probability of '$HB^A\ HB^A$' is 1 while '$HB^A\ HB^s$' and '$HB^S\ HB^S$' have a probability of 0

HB^AHB^A x HB^AHB^S		
	HB^A	HB^A
HB^A	HB^AHB^A	HB^AHB^A
HB^S	HB^AHB^S	HB^AHB^S

Here the cross is between the genotype '$HB^A\ HB^A$' and the genotype '$HB^A\ HB^S$' all the products obtained will also be of the genotype '$HB^A\ HB^A$' or '$HB^A\ HB^S$'
The probability of '$HB^A\ HB^A$' as well as '$HB^A\ HB^s$' is ½ while '$HB^S\ HB^S$' has a probability of 0

$HB^S HB^S$	x	$HB^A HB^A$
	HB^A	HB^A
HB^S	$HB^A HB^S$	$HB^S HB^A$
HB^S	$HB^A HB^S$	$HB^S HB^A$

Here the cross is between the genotype 'HB^A HB^A' and the genotype 'HB^S HB^S', all the products obtained will also be of the genotype 'HB^A HB^S'.
The probability of 'HB^A HB^A' as well as 'HB^s HB^s' is 0 while 'HB^S HB^S' has a probability of 1

Matrix formulation for Sickle cell anemia:

$$A = \begin{bmatrix} 1 & 1/2 & 0 \\ 0 & 1/2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$X = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$$

$$X_1 = AX = A = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix}$$

For any positive integer n, n years later

$$x_n = A^n x_0 = \begin{bmatrix} 1 & \sum_{k=1}^{n} \frac{1}{2^k} & \sum_{k=1}^{n-1} \frac{1}{2^k} \\ 0 & \frac{1}{2^n} & \frac{1}{2^{n-1}} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 - \frac{1}{2^{n+1}} & 1 - \frac{1}{2^{n+1}} \\ 0 & \frac{1}{2^n} & \frac{1}{2^{n-1}} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$$

As n gets larger and larger, the matrix A^n approaches to

$$= \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Therefore, $x_n = A^n x_0$ will approach to

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Since the genotype distribution changes over time, we can represent the succession of genotype distributions from one generation to the next in the form of a difference equation:

$$Xn = A^n X_{(n-1)}$$

CONCLUSION

Sickle cell anemia is an autosomal recessive disease that results in change in shape of the RBCs present in the blood from biconcave to elongated sickle like shaped. This causes problems in the transportation of o_2 in the blood and hence leading to the death of the individual who is diseased. Those who are carriers of the gene is considered normal but a single cross with another carrier may result in a progeny who is completely diseased and would die before attaining maturity.

With the help of linear algebra, we were able to show that after n^{th} generations, the only one who survives will be the complete healthy individual and those who are suffering and is a carrier of the disease would eventually die. This also supports Darwin's theory of "survival of the fittest" where only the most fittest individual survives and the others would extinct with time. The use of diagonalization method rather then manual calculation to depict the flow of inheritance of the gene up to n^{th} generation would make the work more easy and less time consuming.

LITERATURE SURVEY AND BIBLIOGRAPHY

- A. Labra, M. Ladra, U.A. Rozikov, “An evolution algebra in population genetics”, 21 May 2014 **https://www.math.uci.edu/~brusso/LadraEtAl2014.pdf**

- **https://www.slideshare.net/Shababmehebub/linear-algebra-to-solve-autosomal-inheritance**

- **http://web.csulb.edu/~jchang9/m247/m247_sp10_Danial_John_Sunny.pdf**

- **https://sites.math.washington.edu/~king/coursedir/m308a01/Projects/m308a01-pdf/kirkham.pdf**