Available online at www.jlls.org
JOURNAL OF LANGUAGE
AND LINGUISTIC STUDIES
ISSN: 1305-578X
Journal of Language and Linguistic Studies, 18(4), 1143-1149; 2022

Some Applications Of Linear Algebra To Genetics

Wilson Gordillo Thiriata[a], Wilson Pinzón Casallas[b], Orlando García Hurtado[c]

[b] "Universidad Distrital Francisco José De caldas", Bogotá Colombia, *wgordillot@udistrital.edu.co*
ORCID: https://orcid.org/0000-0002-3856-4691
[b] "Universidad Distrital Francisco José De caldas", Bogotá Colombia, *wjpinzonc@udistrital.edu.co*
ORCID: https://orcid.org/0000-0003-0258-6810
[b] "Universidad Distrital Francisco José De caldas", Bogotá Colombia*ogarciah@udistrital.edu.co*
ORCID: https://orcid.org/0000-0002-4155-4515

APA Citation:

Thiriata, W.G., Casallas, W.P., Hurtado, O.G., (2022). T Some Applications Of Linear Algebra To Genetics, *Journal of Language and Linguistic Studies, 18*(4), 1143-1149
Submission Date: 25/10/2022
Acceptance Date: 22/12/2022

Abstract

This work presents a linear algebra approach to provide solutions to some mathematical model's characteristic of plant genetics. With the use of the reference framework of matrix diagonalization and the elements of genetics and through a methodological analysis of case study type, situations typical of this science of modification are presented, concluding with the demonstration of cases.

Linear algebra is an essential element in the basic training of an engineer, since in addition to being a basic science subject, it has immense applications in both physics and statistics, differential equations, function modeling, circuits, computational algorithms, cryptography, etc. Therefore, it is important that engineering students develop concepts rather than techniques; of matrices, vectors, bases, subspaces, linear applications, vectors and eigenvalues, aspects that are closely linked to the solution of problems related to his profession; In addition, these concepts offer the possibility of a clear and precise advance of their applicable knowledge in subsequent courses and courses of their specific knowledge.

Keywords: Linear algebra, genetics, diagonalization, autosomal inheritance, Punnett tables.

1. **Introduction**

In 1906, William Bateson first coined the term "genetics" to refer to the study of the transmission of hereditary traits, following the work developed by George Mendel in his studies with peas and peas, which resulted in what we now call Mendel's laws that gave rise to genetic inheritance. A definition proposed by Sturtevant (1965) describes genetics as the scientific study of genes and inheritance and how certain qualities or traits are inherited from parents to offspring because of changes in the sequence of deoxyribonucleic acid (DNA).

Email id : *wgordillot@udistrital.edu.co*

A gene is a segment of DNA, which contains the instructions for making one or more molecules that help an organism function. All the genetic material of an organism, including genes and other elements that control its activity, constitutes the genome. The complete genome of an organism is found in almost all its cells. In the cells of humans, plants and animals, the genome is housed in a structure called the nucleus.

2. Theoretical Framework

For this work we will use the elements of linear algebra described by Grossman and Flores (2012) such as: Diagonalization of matrices and characteristic values, as well as the elements of genetics such as: autosomal inheritance, Punnett's Tables among others.

a. Diagonalizable matrix

A matrix *A* of *n*x*n* is diagonalizable if there exists a diagonal matrix *D* such that *A* is similar to *D.*

i. Theorem

A matrix *A* of size *nxn* is diagonalizable if and only if it has *n* linearly independent characteristic vectors. In such a case, the diagonal matrix *D* similar to *A* is given by

$$D = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}$$

where $\lambda_{1,}\ \lambda_{2,}\lambda_{3,}\ \dots, \lambda_{n,}$are the characteristic values of *A*.

If *C* is a matrix whose columns are linearly independent characteristic vectors of *A*, then $D = C^{-1}AC$

1. Corollary

If the *n*x*n* matrix *A* has *n* different characteristic values, then *A* is diagonalizable.

b. Autosomal inheritance

According to Cubero(2013), he defines autosomal inheritance as a genetic trait or condition that is passed from parents to offspring, meaning that an individual inherits one gene from each of his parents' genes, thus forming his own pair. As far as is known, it is chance that determines which of a parent's two genes is passed on to its offspring. If it were possible to classify the individuals of a population of a given species in terms of which alleles (alternative forms that the same gene can have) are formed by three types of genotypes, AA, Aa and aa (note that the genotype Aa is the same as aA), it would be possible to determine the proportions of alleles in the population. This would not be feasible if, for example, AA could not be distinguished from Aa.

For $n = 0,1,2,\ldots,$ it is established that

p_n=Proportion of the AA genotype present in the generation of order *n*
q_n=Proportion of the Aa genotype present in the generation of order *n*
r_n= Proportion of genotype aa that is in the generation of order *n*

Assuming that these proportions can be determined, note that one must have the equality, $u = p_n + q_n + r_n$
Then, the proportions u and v of the two alleles A and a in the population satisfy the following equations. $u = p_n + \frac{1}{2}q_n$ and $v = \frac{1}{2}q_n + r_n$

Here, we used the fact that the A and alleles constitute 100% of the AA genotype (with ratio p_n and 50% of the Aa genotype and similarly for the alleles. If it is assumed that the genotypes occur in the same proportions between males and females then u and v represent (in the whole population) the probabilities that the gene is A or a, respectively.

2.2.1 Hardy-Weinberg Law

Suppose that, in a large population of parents, the A and alleles of a particular gene occur in the proportions u and $v = 1 - u$.

Assuming that these proportions are the same for males and females and, furthermore, that mating is random, the first and all successive generations will consist of the three genotypes, AA, Aa and aa in the proportions $u^2, 2uv$ and v^2.

2.2.2 The Punnett Tables

A Punnett square is a chart showing all possible combinations of genes resulting from the crossing of two organisms (whose genes are known). They are named Punnett squares after the English geneticist, Reginald Punnett. He discovered some basic principles of Genetics, including sex linkage and sex determination. In addition, he worked with the feather color characteristics of chickens separately for male and female chickens.

To illustrate how a Punnett square is constructed, it should be noted that if one parent is of the Aa genotype, then it is equally likely that the offspring will inherit either the A allele or a allele from this parent. On the other hand, if one parent is of genotype aa and the other is of genotype Aa, the offspring will always receive a allele from the genotype aa parent and an A or a allele, with equal probability, from the genotype Aa parent. Thus, the offspring has the same probability of being of genotype AA or Aa.

Table 1. PROBABILITIES OF THE POSSIBLE GENOTYPES

Genotype	**PARENTS**					
SONS	AA-AA	AA-Aa	AA-aa	Aa-Aa	Aa-aa	aa-aa
AA	1	0.5	1	0.25	0	0
Aa	0	0.5	1	0.5	0.5	0
aa	0	0	0	0.4	0.5	1

Table 1 shows the probabilities of the possible genotypes of the offspring for all possible combinations of the genotypes of the parents.

Situations in which mating is not random occur frequently in controlled biological experiments. An obvious example is in racehorse breeding, where a proven winner is in great demand as a sire.

3. Methodology

This work has the characteristics of the qualitative methodology (Johnson & Onwuegbuzie, 2004), the case study with the one used in mathematical research (Gordillo & Pino-Fan 2016, Pino-Fan, et al,2018) since the interest is to use elements for the understanding of a problem to investigate. In this research, three problematic situations of genetics proposed by Jiménez (2004) are addressed, which will be solved with the elements proposed by the frame of reference and with the help of Wolfram-Alpha (2022) software to facilitate the calculations.

3.1 Situation Problem 1.

In a population, the distribution of genotypes in the n-th generation is 50% AA, 30% Aa and 20% aa. What proportions of the genes in this population are A and a?

3.2 Situation Problem 2

Pea flower color is controlled by a pair of genes. The three genotypes AA, Aa and aa are characterized by red, pink and white flowers, respectively. If a field is randomly grown with 60% red flowers and 40% white flowers, what proportions of the three genotypes will be present in the fourth generation?

3.3 Situation Problem 3

A farmer has a large population of plants with a certain distribution of the three possible genotypes, AA, Aa and aa. This man wishes to initiate a cropping program in which all plants in the population are fertilized by a plant of genotype AA. He wants to obtain the formula for the distribution of the three possible genotypes in the population after a certain number of generations.

4. Data Analysis

For the data analysis we propose the solution of each of the proposed situations in relation to the proposed theoretical framework.

4.1 Solution Problem Situation 1

To address this situation, we take $p_n = 0.5$,$q_n = 0.30$, $r_n = 0.20$ Therefore applying these values we replace the equations.

$u = p_n + \frac{1}{2}q_n = (0.50) + \frac{1}{2}(0.30) = 0.65$ and $v = \frac{1}{2}q_n + r_n = \frac{1}{2}(0.30) + (0.20) = 0.35$

That is, 65% of the gene population is of A and 35% is of a. Of frequent interest is the inverse problem of determining the proportions of the genotypes when the proportions of the alleles are known. In general, this problem has no unique solution. The system of equations given with the equations reduces to a two-unknown equation, as seen in eq. To obtain a second independent equation, we will assume random pairing. This means that the probability that a given individual mates with another individual does not depend on the genotype of the latter. In many cases, this is a correct assumption. In others it is not, for example, it is known that tall people tend to marry tall people and therefore the height characteristic in humans cannot be analyzed in this way. On the other hand, it has been shown that the random mating assumption applies to the characteristic of human blood types. Most individuals choose their spouse without concern for their blood type.

As before, suppose that u and v are the proportions of the A and a allele among males and among females. Then, if we assume that the population is large, the probability that the offspring receive the A allele from both parents is u^2. Similarly, the probabilities of the AA and aa genotypes are $2uv$ and v^2, respectively. The term $2uv$ comes from the fact that the Aa and aA alleles are the same, a fact that had already been stated. This result leads to Hardy and Weinberg's law in 1908.

4.2 Solution Problem Situation 2

For the solution we take $u = 0.6$ and $v = 0.4$. By Hardy-Weinberg's Law, the proportions of red, pink, and white flowers in the first and all subsequent generations are $u^2, 2uv$ and v^2, or 0.36, 0.48, and 0.16, respectively. Note that the random crop assumption is equivalent to the random pollination assumption.

The Hardy-Weinberg law is only valid when mating is random and when all three genotypes are equally likely. In certain cases, it is quite difficult to verify that mating is random. However, if the proportions of the genotypes remain constant over several generations and if they satisfy the Hardy-Weinberg law, this can be taken as strong evidence that mating is random. Thus, the knowledge that mating is random for human blood types, as well as for many plant and animal characteristics, was deduced from observations of the proportions of genotypes as they satisfy this law.

4.3 Solution Problem Situation 3

To solve the problem, we proceed as follows: let $p_n, q_n$and r_n be the proportions of the three genotypes in the generation. Then, for $n = 0,1,2,\ldots$, we have that

$$p_n = p_{n-1} + \frac{1}{2}q_{n-1}$$

$$q_n = \frac{1}{2}q_{n-1} + r_{n-1}$$

$$r_n = 0$$

These equations determine the distribution of genotypes in each generation from the distribution in the previous generation and were established by means of Punnett's Tables.

The system formed by these equations can be expressed in matrix notation as:

$X^{(n)} = PX^{(n-1)}$ for $n = 0,1,2,\ldots,$

$$X^{(n)} = \begin{bmatrix} p_n \\ q_n \\ r_n \end{bmatrix}, \; X^{(n-1)} = \begin{bmatrix} p_{n-1} \\ q_{n-1} \\ r_{n-1} \end{bmatrix} \text{ y } P = \begin{bmatrix} 1 & 0.5 & 0 \\ 0 & 0.5 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Note that the columns of the P matrix are equal to the first three columns of the Punnette table, thus implying:

$$X^{(1)} = PX^{(0)}$$

$$X^{(2)} = PX^{(1)} = P\left(PX^{(0)}\right) = P^2X^{(0)}$$

$$X^{(3)} = PX^{(1)} = P\left(P^2X^{(0)}\right) = P^3X^{(0)}$$

$$\vdots$$

In general

$$X^{(m)} = P^m X^{(0)}$$

Thus the proportions of the future genotypes are completely determined by the vector $X^{(0)}$ of the initial proportions and by the matrix P.

Now, it is easy to check that the eigenvalues of P are $\lambda_1 = 1$, $\lambda_2 = \frac{1}{2}$ and $\lambda_3 = 0$ with corresponding eigenvectors

$\overrightarrow{v_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\overrightarrow{v_2} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ y $\overrightarrow{v_3} = \begin{bmatrix} 0.5 \\ -1 \\ 0.5 \end{bmatrix}$

Then P will be diagonalizable by matrix $C = \begin{bmatrix} 1 & -1 & 0.5 \\ 0 & 1 & -1 \\ 0 & 0 & 0.5 \end{bmatrix}$ because

$C^{-1}PC = D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

Finally, as $P = CDC^{-1}$ you must $P^m = (CDC^{-1})^m$

This fact is not unknown because, $P^m = CD^mC^{-1}$, determine C^{-1} is obtained

$C^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix}$ and as it has already been found C y C^{-1} matrix D is the diagonal matrix containing the eigenvalues associated with P, we have that

$D^m = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5^m & 0 \\ 0 & 0 & 0 \end{bmatrix}$ now, as the $\lim_{m\to\infty} (0.5)^m$ it can be seen that D^m tends to the matrix

$\lim_{m\to\infty} D^m = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, from which you have

$$\lim_{m\to\infty} P^m = \begin{bmatrix} 1 & -1 & 0.5 \\ 0 & 1 & -1 \\ 0 & 0 & 0.5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Note that since P is a stochastic columnar matrix, then $\lim_{m\to\infty} P^m$ resulted in a matrix with all its columns equal to the probability vector corresponding to the eigenvector associated to the eigenvalue $\lambda = 1$.

On the other hand,

$$\lim_{m\to\infty} X^{(m)} = \lim_{m\to\infty} P^m X^{(0)} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_n \\ q_n \\ r_n \end{bmatrix} = \begin{bmatrix} p_0 + q_0 + r_0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

since $u = p_n + q_n + r_n$ Thus it is demonstrated that in the long run, all plants will be genotype AA.

5. Conclusions

In this work, three situations typical of genetics have been analyzed and which were solved with the use of matrix diagonalization, an element typical of linear algebra.
On the other hand, this type of situations where case analyses are carried out show the close relationship between mathematics and biology, showing a great convergence between these two sciences, contributing to the development of each one of them.

References

Cubero, J. I. (2013). Introduction to plant breeding. Madrid: Mundi-Prensa Libros.

Gordillo, W., & Pino-Fan, L. (2016). A proposed reconstruction of the holistic meaning of the antiderivative.. *Bolema: Boletim de Educação Matemática*, *30*(55), 535-558. doi: 10.1590/1980-4415v30n55a12

Grossman, S. I.,& Flores,J .J (2012). *Linear algebra*, México: McGraw Hill Educación.

Jimenez, J. A. (2004). *Class Notes Linear Algebra II With Applications in Statistics*, Bogotá: Universidad Nacional de Colombia.

Johnson, R., & Onwuegbuzie, A. (2004). Mixed methods research: a research paradigm whose time has come. Educational Research, 33(7), 14-26.

Pino-Fan, L., Font, V., Gordillo, W., Larios, V. & Breda, A. (2018). Analysis of the meanings of the antiderivate used by students of the first engineering courses. *International Journal of Science and Mathematics Education*, *16*(6), 1091-1113. doi: 10.1007/s10763-017-9826-2

Sturtevant, A. H. (1965). A History of Genetics. California: Institute of Technology

Wolfram Research, Inc., Wolfram|Alpha Notebook Edition, Champaign, IL (2022).