Jared Kirkham
Math 308 Project
Autumn 2001

Abstract

I have chosen to explore how linear algebra can be applied to genetics. More specifically, I will focus entirely on the phenomena of autosomal inheritance. My goal is to show how linear algebra can be used to predict the genotype distribution of a particular trait in a population after any number of generations from only the genotype distribution of the initial population. In order to perform such an analysis, I will use several critical concepts from linear algebra, including the following: Difference equations, diagonalization of a matrix, inverse of a matrix, eigenvalues, and eigenvectors.

Explanation of the application

Background:

Genetics is the study of inheritance, or the transmission of traits from one generation to the next. There are several modes of inheritance; however, this project will focus only on autosomal inheritance.

In autosomal inheritance, each heritable trait is assumed to be governed by a single gene on a chromosome. 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.

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. Genotypes AA and Aa have brown eyes, and genotype aa has blue eyes. 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.

1

The relative proportion of each of the three possible genotypes for a particular trait in a population is called the genotype distribution of that particular trait in the population. When two individuals breed, offspring inherit one allele from each of the parents. We will assume that either of a parent's genes is equally likely to be inherited. Therefore, simple probability determines how likely the offspring each to have each of the possible three genotypes. Whenever breeding exists in a population, the genotype distribution of a particular trait in a population will change with the passage of generations. The following section will give an overview of how linear algebra can be used to predict the genotype distribution of a particular trait in a population after any number of generations from only the genotype distribution in the initial population.

General overview of the method:

The genotype distribution of a particular trait in a population in the n-th generation can be represented by a genotype vector $\mathbf{x}_n = \begin{matrix} a\_n \\ b\_n \\ c\_n \end{matrix}$ where a_n, b_n, and c_n denote the portion of the population with genotype AA, Aa, aa, respectively in the n-th generation. 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,

$$\mathbf{x}_n = M\mathbf{x}_{n-1} \tag{1}$$

for a suitable matrix M and n = 1,2,3,…. (Lay 90). We seek an explicit description of $\mathbf{x}_n$ whose formula for each $\mathbf{x}_n$ does not depend on A or on the preceding terms in the sequence other than the initial term $\mathbf{x}_0$ (the initial genotype distribution in the population) (Lay 302). From equation (1), we have,

$$\mathbf{x}_n = M\mathbf{x}_{n-1} = M^2\mathbf{x}_{n-2} = \ldots.= M^n\mathbf{x}_0 \tag{2}$$

Therefore, if we can find an explicit expression for M^n, we can use equation (1) to obtain an explicit expression for $\mathbf{x}_n$ in terms of $\mathbf{x}_0$. We will proceed to diagonalize M (that is, we will seek to find an invertible matrix P and a diagonal matrix D such that $M = PDP^{-1}$). Since

$$M^n = PD^nP^{-1} \tag{3}$$

(as derived in example 2, p. 314 of Lay), and powers of D are trivial to compute (example 1, p. 313 of Lay), this diagonalization will enable us to compute M^n quickly for large values of n. By theorem 5 on p. 314 of Lay, if M is diagonalizable (that is, if M has n linearly independent eigenvectors), then the diagonal entries of D are the eigenvalues of M, and the columns of P are n linearly independent eigenvectors corresponding, respectively, to each of the eigenvalues of M. Once P and D have been specified, we will have,

2

$$\mathbf{x}_n = PD^nP^{-1}\mathbf{x}_0 \tag{4}$$

an explicit description of $\mathbf{x}_n$ that is relatively easy to compute, and is based upon $\mathbf{x}_0$.

Solved problem

Problem:

In an experimental farm, a large population of flowers consists of all possible genotypes (AA, Aa, and aa), with an initial frequency of $a_0 = .05$, $b_0 = .90$, and $c_0 = .05$, respectively. Suppose that this genotype controls flower color, and that each flower is fertilized by a flower of a genotype similar to its own (this is equivalent to an "inbreeding" program). 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 (note: this problem statement adapted from Farr).

Solution:

The genotype distribution of flower color in a population in the n-th generation can be represented

by a genotype vector $\mathbf{x}_n = \begin{matrix} a\_n \\ b\_n \\ c\_n \end{matrix}$ where a_n, b_n, and c_n denote the portion of the population with

genotype AA, Aa, aa, respectively in the n-th generation.

If each plant is fertilized by a plant with a genotype similar to its own, then the possible combinations of the genotypes of the parents are AA and AA, Aa and Aa, and aa and aa. The probabilities of the possible genotypes of the offspring corresponding to these combinations are shown in the table below.

		Genotypes of Parents		
		AA, AA	Aa, Aa	aa, aa
Genotype	AA	1	.25	0
of	Aa	0	.5	0
Offspring	aa	0	.25	1

Table 1-1

Perhaps an explanation of how the columns of the table were derived is necessary. As mentioned in the introduction, when two individuals breed, the offspring inherit one allele from each of the parents. We will assume that either of a parent's genes is equally likely to be inherited. Therefore, simple probability determines how likely the offspring each to have each of the possible three genotypes. We can use a device called a Punett square (Cambell et al. 242) to predict the relative proportions of the genotypes of the offspring from the genotypes of the parents. For example, if the parents are of genotype Aa and Aa, then the Punett square is as follows:

3

	A	**a**
A	AA	Aa
a	Aa	aa

This shows that _ of the offspring will be of genotype AA, _ will be of genotype Aa, and _ will be of genotype aa. In a similar manner, the other columns of the given table can be verified.

The following equations determine the frequency of each genotype as dependent on the preceding generation. These arise directly from consideration of table 1-1.

$$a_n = a_{n-1} + .25\ b_{n-1}$$

$$b_n = .5 b_{n-1} \qquad\qquad n = 1,2,\dots \qquad (1\text{-}1)$$

$$c_n = .25 b_{n-1} + c_{n-1}$$

Equations (1-1) can be written in matrix notation as:

$$\mathbf{x}_n = \mathbf{M}\mathbf{x}_{n-1} \qquad\qquad n = 1,2,\dots \qquad (1\text{-}2)$$

where

$$\mathbf{x}_n = \begin{array}{c} a\_n \\ b\_n \\ c\_n \end{array} , \ \mathbf{x}_{n-1} = \begin{array}{c} a\_(n-1) \\ b\_(n-1) \\ c\_(n-1) \end{array} , \text{ and } \mathbf{M} = \begin{array}{ccc} 1 & .25 & 0 \\ 0 & .5 & 0 \\ 0 & .25 & 1 \end{array} .$$

Note that this matrix corresponds exactly to the columns of Table 1-1.

Our next task is to diagonalize M (that is, we seek an invertible matrix P and a diagonal matrix D such that $M = PDP^{-1}$).

To perform this diagonalization, we need to find the eigenvalues of M and an eigenvector corresponding to each eigenvalue.

By computation by a TI-85 graphing calculator, the eigenvalues are $\_1 = 1$ (multiplicity of two), and $\_2 = .5$ (multiplicity of one).

The eigenspace corresponding to $\_1 = 1$ is $\mathbf{x} = \begin{array}{c} x\_1 \\ x\_2 \\ x\_3 \end{array} = x_1 \begin{array}{c} 1 \\ 0 \\ 0 \end{array} + x_3 \begin{array}{c} 0 \\ 0 \\ 1 \end{array}$. A basis for this

eigenspace is: $\begin{array}{c} 1 \\ 0 \\ 0 \end{array}$ and $\begin{array}{c} 0 \\ 0 \\ 1 \end{array}$. Therefore, the eigenspace corresponding to $\_1 = 1$ is two-dimensional.

4

The eigenspace corresponding to $\lambda_2 = .5$ is $\mathbf{x} = \begin{array}{c} x\_1 \\ x\_2 \\ x\_3 \end{array} = x_3 \begin{array}{c} 1 \\ -2 \\ 1 \end{array}$. A basis for this eigenspace is

$\begin{array}{c} 1 \\ -2 \\ 1 \end{array}$. Therefore the eigenspace corresponding to $\lambda_2 = .5$ is one-dimensional.

Since the sum of the dimensions of the distinct eigenspaces equals 3 (and M is 3x3), M is indeed diagonalizable (Theorem 7, p. 318 of Lay). That is we can find 3 linearly independent eigenvectors of M. By theorem 7, p. 318 of Lay, these 3 linearly independent eigenvectors of M are the set of vectors that were identified above to form the basis for each of the eigenspaces corresponding to each eigenvalue.

Therefore, $D = \begin{array}{ccc} lambda\_1 & 0 & 0 \\ 0 & Lambda\_1 & 0 \\ 0 & 0 & Lambda\_3 \end{array} = \begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & .5 \end{array}$.

$P = \begin{array}{ccc} 1 & 0 & 1 \\ 0 & 0 & -2 \\ 0 & 1 & 1 \end{array}$ and $P^{-1} = \begin{array}{ccc} 1 & .5 & 0 \\ 0 & .5 & 1 \\ 0 & -.5 & 0 \end{array}$ (by TI-85 graphing calculator)

Thus, $\mathbf{x}_n = \begin{array}{c} a\_n \\ b\_n \\ c\_n \end{array} = PD^n P^{-1} \mathbf{x}_0$

$= \begin{array}{ccc} 1 & 0 & 1 \\ 0 & 0 & -2 \\ 0 & 1 & 1 \end{array} \begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & .5^n \end{array} \begin{array}{ccc} 1 & .5 & 0 \\ 0 & .5 & 1 \\ 0 & -.5 & 0 \end{array} \begin{array}{c} a\_0 \\ b\_0 \\ c\_0 \end{array}$

$= \begin{array}{ccc} 1 & .5 - .5^{\wedge}(n+1) & 0 \\ 0 & .5^{\wedge} n & 0 \\ 0 & .5 - .5^{\wedge}(n+1) & 1 \end{array} \begin{array}{c} a\_0 \\ b\_0 \\ c\_0 \end{array}$.

Finally,

$$a_n = a_0 + [.5 - .5^{n+1}]b_0$$

$$b_n = .5^n b_0 \qquad\qquad n = 1,2,\ldots \qquad (1\text{-}3)$$

$$c_n = c_0 + [.5 - .5^{n+1}]b_0$$

Equations 1-3 give us the genotype distribution of the population after any number of generations of the breeding scheme discussed in the introduction. Notice that for $n = 0$, $a_n = a_0$, $b_n = b_0$, and $c_n = c_0$, as we expect.

After 4 generations ($n = 4$), with $a_0 = .05$, $b_0 = .90$, and $c_0 = .05$, we predict that:

$a_4 = .47$, $b_4 = .06$, and $c_4 = .47$

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 $n \rightarrow$, we find that:

$a_n = .5$, $b_n = 0$, and $c_n = .5$

That is, in the limit as $n \rightarrow$, _ of the individuals in the population are AA, _ 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. A well-known example of inbreeding is mating between brothers and sisters. Such breeding schemes were used by the royal families of England, in hopes of keeping "the royal lines pure." (Rorres and Anton 160). However, many genetic diseases are autosomal recessive, which means that the disease is carried on the recessive allele (a). While AA and Aa genotypes will not exhibit the recessive disease, aa will. As this example shows, inbreeding increases the proportion of AA and aa genotypes in the population while reducing the number of Aa genotypes. Therefore, under an inbreeding program, the proportion of the population that is infected with the disease (aa) will increase.

References

Campbell, Neil A., Reece, Jane B., and Mitchell, Lawrence G. Biology. Addison Wesley, 1999.

Farr, William M. Modeling Inheritance of Genetic Traits.
http://www.math.wpi.edu/Course_Materials/MA2071A98/Projects/gene/node1.html.

Lay, David C. Linear Algebra And Its Applications. New York: Addison Wesley, 2000.

Rorres, Chris and Anton, Howard. Applications of Linear Algebra. John Wiley And Sons, 1977.

Williams, Gordon. Genetics. http://www.math.washington.edu/%7Egwilliam/308/genetics.pdf