

Predicting Genotype Frequency for Mendelian Traits at Hardy-Weinberg Equilibrium
using Linear Algebra

Introduction:

One of the many fields in which techniques of Linear Algebra can be applied is the study of Genetics. More particularly, the field of Population Genetics and Genotype Frequency.

As the name implies, Population Genetics is concerned with studying populations and their differences in their genes. These genetic differences refer to the genotype. The genotype is the actual set of genes in DNA that determines any given trait.

Here, we are interested in using Linear Algebra to predict the genotype frequency of Mendelian traits under the assumptions of Hardy-Weinberg Equilibrium. Mendelian traits are traits that are passed down via a single-locus system that involves a single dominant and single recessive allele. We want to use Mendelian traits because they illustrate how Linear Algebra can be used in Population Genetics in the easiest way to understand. To note, the assumptions of Hardy-Weinberg are: no genetic drift, a closed population, no mutations, random mating patterns, and no natural selection. The reason we want these assumptions is that they make the predictions of genotype frequency more reliable and accurate.

The model equation for Hardy-Weinberg is:

$$p^2+2pq+q^2=1$$

p^2 refers to the proportion of a population that is homozygous dominant, or AA genotype, for some given trait. $2pq$ refers to the proportion of heterozygotes, or Aa genotype. Lastly, q^2 refers to the proportions of homozygous recessive, or aa genotype.

With this knowledge, we can now model population proportions with matrices and frequency vectors.

Model Matrix:

For our model matrix, we will consider a cross between the homozygous dominant (AA) genotype with all genotypes (i.e. with AA, Aa, and aa, respectively). When we do this, we obtain the genotype frequencies in the vector form:

$$\text{Proportions} = \begin{bmatrix} AA \\ Aa \\ aa \end{bmatrix}$$

AA with AA:

$$AA \times AA = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

AA with Aa:

$$AA \times Aa = \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \end{bmatrix}$$

AA with aa:

$$AA \times aa = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Thus, we now can denote our model matrix as follows:

$$A_1 = \begin{bmatrix} AA \times AA & AA \times Aa & AA \times aa \end{bmatrix} = \begin{bmatrix} 1 & 1/2 & 0 \\ 0 & 1/2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Genotype Frequency Vector:

Using a frequency vector, we can obtain the genotype frequencies of subsequent generations for our specific genetic cross. This can be modeled by the linear equation:

$$A_1x=b$$

where x is a frequency vector of the current genotype frequencies and b is the genotype frequency vector of the subsequent generation. For simplicity, we will rewrite the equation as:

$$A_1x_n=x_{n+1}$$

where n is the number of generations after the initial generation.

For our model, we will assume that all three genotypes have the same initial frequencies such that:

$$x_0=\begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

Model Equations:

Now that we have all the information we need to proceed, say we want to find the genotype frequencies one generation after the initial generation. To do this we simply solve for x_1 via matrix multiply. This would be written as follows:

$$x_1=A_1x_0=\begin{bmatrix} 1 & 1/2 & 0 \\ 0 & 1/2 & 1 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}=\begin{bmatrix} 1/2 \\ 1/2 \\ 0 \end{bmatrix}$$

Now say that we want to find the genotype frequency for second generation after the initial. If we want to solve for the second generation's genotype frequency, we would want to use the new current genotype frequency vector. This would correspond to x_1 . Thus, we could solve for the frequency vector as follows:

$$A_1 x_1 = x_2$$

however, notice that there is another way to solve for x_2 . By substitution, we can rewrite the equation and solve for it as follows:

$$A_1 x_1 = A_1(A_1 x_0) = A_1^2 x_0 = x_2 = \begin{bmatrix} 3/4 \\ 1/4 \\ 0 \end{bmatrix}$$

With this method of solving for x_n we can much more easily solve for the genotype frequencies when Hardy-Weinberg Equilibrium is met. To do this, we would just have to look at the equation:

$$\lim_{n \rightarrow \infty} A_1^n x_0$$

We want to have n approach infinity because as each subsequent generation arises, the closer their genotype frequencies get to Hardy-Weinberg Equilibrium. Thus, we will find that, at Hardy-Weinberg Equilibrium, x_n and the genotype frequencies will be:

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

From this we can conclude that, at Hardy-Weinberg Equilibrium, our cross will result in a 100 percent frequency of the homozygous dominant genotype.

Now we can easily find the genotype frequencies for the cross between the heterozygous genotype and the cross between the homozygous recessive genotype at Hardy-Weinberg Equilibrium. Of course, we would have to construct the model matrices first. They are as follows:

$$A_2 = [AaxAA \quad AaxAa \quad Aaxaa] = \begin{bmatrix} 1/2 & 1/4 & 0 \\ 1/2 & 1/2 & 1/2 \\ 0 & 1/4 & 1/2 \end{bmatrix}$$

$$A_3 = [aaxAA \quad aaxAa \quad aaxaa] = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1/2 & 0 \\ 0 & 1/2 & 1 \end{bmatrix}$$

Using our same initial genotype frequency vector, we now can solve for the Hardy-Weinberg Equilibria using the same limit equation method as before. The Equilibria are as follows:

$$\lim_{n \rightarrow \infty} A_2^n x_0 = \begin{bmatrix} 1/4 \\ 1/2 \\ 1/4 \end{bmatrix}$$

$$\lim_{n \rightarrow \infty} A_3^n x_0 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Thus, we see that the heterozygous cross results in 25 percent of the population being homozygous dominant, 50 percent heterozygous, and 25 percent homozygous recessive at Hardy-Weinberg Equilibrium. For homozygous recessive cross, the resulting genotype frequency at Hardy-Weinberg Equilibrium is 100 percent homozygous recessive.

Multiple Traits:

The great thing about this method of solving for genotype frequencies is that it does not exclusively apply to the prediction of genotype frequency for one Mendelian trait at a time. We can use this method for finding genotype frequencies of two, three, or how ever many traits we want to look at a time. To illustrate this, we can look at two generic loci which we will creatively call loci A and loci B. This time, we will look at the cross between the genotype that is homozygous dominant for both traits (AABB) and all genotypes (i.e. with AABB, AABb, AAbb, AaBB, AaBb, Aabb, aaBB, aaBb, and aabb, respectively).

$$AB = \begin{bmatrix} 1 & 1/2 & 0 & 1/2 & 1/4 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 1 & 0 & 1/4 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/4 & 0 & 1 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/4 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$y_0 = \begin{bmatrix} 1/9 \\ 1/9 \\ 1/9 \\ 1/9 \\ 1/9 \\ 1/9 \\ 1/9 \\ 1/9 \\ 1/9 \end{bmatrix}$$

Now we can solve for Hardy-Weinberg Equilibrium the same way as before.

$$\lim_{n \rightarrow \infty} AB^n y_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Thus, we see that the resulting genotype frequency at Hardy-Weinberg Equilibrium for the two traits is 100 percent homozygous dominant for both traits.

Conclusion:

As we have demonstrated, techniques of Linear Algebra, and matrix multiply in particular, serve as useful tools in predicting genotype frequencies and Hardy-Weinberg Equilibria for Mendelian traits.

While we focused solely on simple Mendelian traits here, it is worth mentioning that these types of traits serve as a basis for more complexly inherited traits. Though we did not delve into more complex situations of genetics, such as sex-linked traits, polygenic traits, linked genes, and lethal traits, the same general tools used here can be used for these more complex genetic traits. Thus, as we can see, Linear Algebra can easily and efficiently be applied to the study of Population Genetics and Hardy-Weinberg Equilibrium at large.

References:

Becker, Andrea. "What Are the Five Parts of the Hardy-Weinberg Equilibrium?" *Why Is It Important to Have a Social Life in College? | Education - Seattle PI*, 21 Nov. 2017, education.seattlepi.com/five-parts-hardyweinberg-equilibrium-3968.html.

"Mendelian & Non-Mendelian Traits in Humans: Definition & List." *Study.com*, Study.com, 2018, study.com/academy/lesson/mendelian-non-mendelian-traits-in-humans-definition-list.html.

O'Neil, Dennis. "Modern Theories of Evolution: Hardy-Weinberg Equilibrium Model." *Early Theories of Evolution: Evidence of Evolution*, 1997, anthro.palomar.edu/synthetic/synth_2.htm.