Math 215 Project 2: Genetics

Written by: Rebecca Strautman

Problem Description: Organisms inherit traits from their parents encoded by segments of DNA called *genes*. Genes have different forms that cause different characteristics, for example brown eyes vs. blue eyes, or straight hair vs. curly hair. In simple representations, the dominant form is symbolized by a capital letter (A), and the recessive form is symbolized by a lowercase letter (a). An organism has two forms of each gene, one from each parent. This two-letter set is known as its *genotype* for that gene. For example, offspring from parents with the genotypes AA and aa will have the genotype Aa.

We can model this inheritance as a transition from one generation to another. Each transition is given by some matrix of probabilities that the next generation will have some given properties (for example, and AA and aa produce an Aa with probability 1). One wishes to know the long term behavior so we take the matrix product over many generations. We'll see that certain genetic traits will die out over time, while others will dominate. This process of transitioning is very closely related to the concept of Markov chains or Markov processes.

Example Problem: In rats, F (the version of the gene for fur) is dominant to f (the version of the gene for hairless). Since dominant traits override recessive traits, *FF* rats have fur, *Ff* rats have fur, and *ff* rats are hairless. If $\frac{1}{3}$ of a population of rats has each of these genotypes, then the distribution of genotypes is given by the vector

$$\mathbf{x}_0 = \begin{bmatrix} FF \\ Ff \\ ff \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}.$$

Consider crossing these offspring with rats with the genotype *FF* only. For *FF* crossed with *FF*, the probabilities of offspring with *FF*, *Ff*, and *ff* are, respectively, 1, 0, and 0. For *FF* crossed with *Ff*, the probabilities of *FF*, *Ff* and *ff* are $\frac{1}{2}$, $\frac{1}{2}$, 0. For *FF* crossed with *ff*, the probabilities of *FF*, *Ff* and *ff* are 0, 1, and 0.

In the total population of offspring, the probability of the genotype of *FF* is *(1)(1/3) + (1/2)(1/3) + (0)(1/3)*. The probability of *Ff* is *(0)(1/3) + (1/2)(1/3) + (1)(1/3)*. The probability of *ff* is *(0)(1/3) + (0)(1/3) + (0)(1/3)*. So the vector $\mathbf{x}_1$, representing the probabilities of each genotype in the first generation, can be given by $\mathbf{x}_1 = A\mathbf{x}_0$, where

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

Note that our transition matrix, in the first row, gives the probability of that FF crossed with FF, Ff and ff is an FF. The second row gives the probability that FF crossed with FF, Ff and ff is an Ff and the third row gives the probability that FF crossed with FF, Ff and ff is ff. Since we want to see what happens after multiple generations, we treat this as an iteration $\mathbf{x}_k = A\mathbf{x}_{k-1}$ which makes $\mathbf{x}_k = A^k\mathbf{x_0}$. We find

$$\mathbf{x}_1 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} \frac{3}{4} \\ \frac{1}{4} \\ 0 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} \frac{7}{8} \\ \frac{1}{8} \\ 0 \end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix} \frac{15}{16} \\ \frac{1}{16} \\ 0 \end{bmatrix}, \mathbf{x}_5 = \begin{bmatrix} \frac{31}{32} \\ \frac{1}{32} \\ 0 \end{bmatrix}, \ldots$$

These are converging to the vector $\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. This means that eventually, assuming there are no mutations, the rats in this population will no longer carry the version of the gene for hairlessness.

Assigned Problem: Consider, now, a plant with a gene that codes for flower color (R for red, and r for white – R is dominant), and a gene that codes for height (T for tall, and t for short – T is dominant). These genes are on different chromosomes, and are inherited *independently* of each other. So the different combinations of genes are $RRTT$, $RRTt$, $RRtt$, $RrTT$,... etc. The probabilities of each genotype are given by the vector

$$\mathbf{x}_0 = \begin{bmatrix} RRTT \\ RRTt \\ RRtt \\ RrTT \\ RrTt \\ Rrtt \\ rrTT \\ rrTt \\ rrtt \end{bmatrix} = \frac{1}{12} \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \\ 2 \\ 1 \\ 1 \\ 2 \\ 1 \end{bmatrix}.$$

Suppose we cross these plants with only plants that have the genotype $rrTt$. What is the probability for each genotype in the 6^{th} generation? The 12^{th}? The 40^{th}? What genotype(s) does the population seem to be converging to? What do these plants look like?