Studying the Hardy-Weinberg Equation with more Realistic Assumptions

Dynamic Models in Biology (640:336)

Final Project

Nicholas DiMarzio, Anusha Nagar, Tim Nasralla, Deven Singh

Introduction - Nick

The Hardy-Weinberg rule is a principle stating that the genetic variation of a population remains constant throughout generations. This principle helps characterize certain distributions regarding genotype frequencies in specific populations. This principle also relates allele frequencies to genotype frequencies in certain populations. These specific populations are populations that follow the assumptions of the Hardy-Weinberg principle. These assumptions are as follows:

1. Mating is random
2. There is no variation in the number of progeny from parents of different genotypes.
3. Progeny have equal fitness (they are equally likely to survive)
4. There are no mutations at any step

If we follow these assumptions, we can see that gene frequencies will not change, therefore they must be constant. This is the Hardy-Weinberg law.

Looking closer at our assumptions, random mating is mating between parents where the choice of partner is not influenced by genotype. When mating is random, The Hardy-Weinberg rule predicts that the population's genotype frequencies and allele frequencies will not change and remain in equilibrium. This implies that there are no disturbing factors that would skew these probabilities.

Next, we look at variation in the number of progeny from parents of different genotypes. If there is variation, then the next generations of the population are affected by a disturbing factor, which in this case is variation of progeny. This would cause the frequencies to change which is against the Hardy-Weinberg rule.

Further, we look at populations that have equal fitness. This is the idea that the progeny of a population all have equal likelihoods of surviving. This is important as having variation would cause the frequencies of the population to change which again is directly against the Hardy-Weinberg rule.

Finally, taking a closer look at our last assumption, we must have no mutations in our population for the Hardy-Weinberg rule to hold. Mutations in populations affect allele frequencies due to the creation of new alleles produced by the mutations. If alleles mutate at different rates, allele frequencies are subject to change from generation to generation. If this occurs, then frequency would not be consistent and constant. Thus this would interfere with non-random breeding.

Using these assumptions, we are able to create tables to show the frequencies under the Hardy-Weinberg rule. First we look at the Mating table:

Genotype		AA	Aa	aa
	Frequency %	u	v	w
AA	u	u^2	uv	uw
Aa	v	uv	v^2	vw
aa	w	uw	vw	w^2

Table 3.1 from "Mathematical Models in Biology"

In the table we are able to calculate the probability that parents of specific genotypes will mate. These probabilities will remain consistent according to the Hardy-Weinberg rule. We then can summarize our information using an offspring table.

Type of Parents	Frequency	AA	aA	aa
AA x AA	u^2	u^2	0	0
AA x Aa	$2uv$	uv	uv	0
AA x aa	$2uw$	0	$2uw$	0
Aa x Aa	v^2	$\frac{1}{4}v^2$	$\frac{1}{2}v^2$	$\frac{1}{4}v^2$
Aa x aa	$2uv$	0	vw	vw
aa x aa	w^2	0	0	w^2
	Total	$u^2 + uv + \frac{1}{4}v^2$	$uv + 2uw + \frac{1}{2}v^2 + vw$	$\frac{1}{4}v^2 + vw + w^2$

Table 3.2 "Mathematical Models in Biology"

Using the Hardy-Weinberg principle and its assumptions, we now use maple to model the assumptions to get a more detailed understanding of the Hardy-Weinberg Principle.

1. Equal Fitness Assumption

1.1 Analytic Setup

To combat the Equal Fitness assumption, we consider a survival matrix, M. M[i][j] denotes the survival fraction of a type-i father and a type-j mother surviving to the next generation. In the case of the original Hardy-Weinberg Equation, this would result in M being defined as a three-by-three matrix of ones. For i,j equal to 1, the respective mother or father is of type AA; for i,j equal to 2, the respective mother or father is of type Aa, and for i,j equal to 3, the respective mother or father is of type aa. This can be seen below in Figures 1a and 1b.

$$\begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix} \quad \begin{bmatrix} AA \times AA & AA \times Aa & AA \times aa \\ Aa \times AA & Aa \times Aa & Aa \times aa \\ aa \times AA & aa \times Aa & aa \times aa \end{bmatrix}$$

Figure 1a (left): Survival Matrix setup, **Figure 1b (right):** Offspring genetic makeup for each entry of survival matrix

To further analyze how to combat the equal fitness assumption, we consider Table 1 below, which shows the frequency of each genotype. Each entry in the three-by-three table is multiplied by its corresponding entry in the survival matrix. The subsequent transformations for each variable then become as follows:

$$u \rightarrow u^2 \cdot M_{11} + \frac{1}{2}uv \cdot M_{12} + \frac{1}{2}uv \cdot M_{21} + \frac{1}{4}v^2 \cdot M_{22}$$

$$v \rightarrow \frac{1}{2}uv \cdot M_{12} + uw \cdot M_{13} + \frac{1}{2}uv \cdot M_{21} + \frac{1}{2}v^2 \cdot M_{22} + \frac{1}{2}vw \cdot M_{23} + vw \cdot M_{31} + \frac{1}{2}vw \cdot M_{32}$$

$$w \rightarrow \frac{1}{4}v^2 \cdot M_{22} + \frac{1}{2}vw \cdot M_{23} + \frac{1}{2}vw \cdot M_{32} + w^2 \cdot M_{33}$$

The transformations are further divided by the sum of each of these transformations in the Maple code in order to normalize the results.

The analytic setup work described in this section is based mainly off of Anne Somalwar's solutions to the Homework 15 problem, and subsequent Maple command HW2g in previous iterations of the in-class maple packages, and the current version, HWg, in the DMB.txt package.

	AA - Father			Aa - Father			aa - Father		
AA Mother	**AA**	**Aa**	**aa**	**AA**	**Aa**	**aa**	**AA**	**Aa**	**aa**
	u^2	0	0	$\frac{1}{2}uv$	$\frac{1}{2}uv$	0	0	uw	0
	Multiply by M_{11}			Multiply by M_{12}			Multiply by M_{13}		
Aa Mother	**AA**	**Aa**	**aa**	**AA**	**Aa**	**aa**	**AA**	**Aa**	**aa**
	$\frac{1}{2}uv$	$\frac{1}{2}uv$	0	$\frac{1}{4}v^2$	$\frac{1}{2}v^2$	$\frac{1}{4}v^2$	0	$\frac{1}{2}vw$	$\frac{1}{2}vw$
	Multiply by M_{21}			Multiply by M_{22}			Multiply by M_{23}		
aa Mother	**AA**	**Aa**	**aa**	**AA**	**Aa**	**aa**	**AA**	**Aa**	**aa**
	0	uw	0	0	$\frac{1}{2}vw$	$\frac{1}{2}vw$	0	0	w^2
	Multiply by M_{31}			Multiply by M_{32}			Multiply by M_{33}		

Table 1: Frequency of Genotypes separated by type of parent

1.2 Numerical Analysis Setup

The numerical analysis of the equal fitness assumption heavily relies on the analytic setup done by Anne Somalwar, and the HWg command in the DMB.txt maple package. For numeric analysis of this assumption, we set up a random number generator to generate values for the survival matrix between 0.0 and 1.0. This is imperative, as we cannot have a survival fraction higher than one. We held each entry of the survival matrix constant at 1 except for one entry that would be changed according to the random number generator. For example, the first iteration involved changing the value of M_{11} while all other entries were equal to one. We input these values into the HWg maple command, and used the function 'Orb' to witness the long-term orbit. For purposes of analysis, we tried three different randomly-generated values for each entry of the survival matrix. The values [0.3,0.4] were used as initial conditions in the 'Orb' function based on information in the Keshet book.

The second type of analysis was to check for fixed points and stable fixed points for each of the randomly-generated entries of the survival matrix. The command 'FP' from DMB.txt was used to evaluate the fixed points of each survival matrix, and 'Orb' was used to test initial conditions close to the fixed points in order to see whether we have a stable fixed point or not. It is important to note that this analysis of stable fixed points utilizes initial conditions very close to the fixed points found by the 'FP' function. Therefore, if the code is run again, it is very likely that the 'Orb' functions will not reflect the proper initial conditions based on the new random

values for the survival matrix. For this reason, an extra PDF of results with the specific random numbers is included along with the Maple code.

1.3 Results and Insight

Through our analytical and numerical analysis on the Equal Fitness assumption, there are a few things to note. One is the difference in the number of generations it takes to stabilize. In the original Hardy-Weinberg equation, the population seems to stabilize after just one generation. However, in our analysis, we see that it takes longer for our population to stabilize, which is more realistic. In some cases explored, we still see slight fluctuation in population after 1000 generations!

Next, we also find it important to note that when we hold one value in the survival matrix constant and change one, we see that the populations have two stable fixed points: [0,0] and [1,0]. When we were conducting part one of our numerical analysis, or just changing the values of the entries of the survival matrix, we found that sometimes our [u,v] values in the long run would approach [1,0]. When we analyzed the same functions for fixed points and stable fixed points, we found that for every case, [0,0] and [1,0] are stable fixed points. However, even in analyzing our stable fixed points, we saw that it takes thousands of generations to actually get close to the stable fixed points. We conducted our analysis between the 2000 and 2010 generations, and were still far but approaching these fixed points. One important thing to note here is that we ignored fixed points which had a negative or complex value, as these do not make sense in this real-world application.

The Maple code (.txt) and a PDF file (.pdf) with specific randomly-generated analyses are attached with this report for this section.

2. Nonrandom Mating Assumption

2.1. Modeling Nonrandom Mating

A crucial assumption of Hardy Weinberg genetics is that mating is random. In real-world genetics, traits are often selected through sexual selection. To model the genetics of a population, we must account for this real-world reality. Fortunately, correcting for this assumption requires us to perform the same mathematics we do when correcting for equal fitness.

$$\begin{bmatrix} N11 & N12 & N13 \\ N21 & N22 & N23 \\ N31 & N32 & N33 \end{bmatrix} \qquad \begin{bmatrix} AA \times AA & AA \times Aa & AA \times aa \\ Aa \times AA & Aa \times Aa & Aa \times aa \\ aa \times AA & aa \times Aa & aa \times aa \end{bmatrix}$$

Figure 2a (left): Mating Combination Matrix setup, **Figure 2b (right):** Offspring genetic makeup for each entry of survival matrix

N[i][j] denotes the mating likelihood of a type-i father and a type-j mother surviving to the next generation. In the case of the original Hardy-Weinberg Equation, this would result in N being defined as a three-by-three matrix of ones, or that the likelihood of mating is certain.

The likelihood of each mating combination, however, can range from 0 to 1. Each probability can be multiplied by the frequency of each genotype produced by all mating combinations. To illustrate this, we can look to a modified mating table.

	AA - Father			**Aa - Father**			**aa - Father**		
AA Mother	**AA**	**Aa**	**aa**	**AA**	**Aa**	**aa**	**AA**	**Aa**	**aa**
	u^2	0	0	$\frac{1}{2}uv$	$\frac{1}{2}uv$	0	0	uw	0
	Multiply by N_{11}			Multiply by N_{12}			Multiply by N_{13}		
Aa Mother	**AA**	**Aa**	**aa**	**AA**	**Aa**	**aa**	**AA**	**Aa**	**aa**
	$\frac{1}{2}uv$	$\frac{1}{2}uv$	0	$\frac{1}{4}v^2$	$\frac{1}{2}v^2$	$\frac{1}{4}v^2$	0	$\frac{1}{2}vw$	$\frac{1}{2}vw$
	Multiply by N_{21}			Multiply by N_{22}			Multiply by N_{23}		
aa Mother	**AA**	**Aa**	**aa**	**AA**	**Aa**	**aa**	**AA**	**Aa**	**aa**
	0	uw	0	0	$\frac{1}{2}vw$	$\frac{1}{2}vw$	0	0	w^2
	Multiply by N_{31}			Multiply by N_{32}			Multiply by N_{33}		

Table 2: Frequency of Genotypes separated by type of parent

Let's examine the Aa Mother and Aa Father Mating Combination. When they mate, they can produce three types of offspring: AA, Aa, and aa genotypes. The frequencies of AA, Aa, and aa offspring in the population are .25 * v^2, .5*v^2, and .25*v^2 respectively. Because we are modifying these offspring frequencies in relation to the probability of their parents mating, we will multiply

all three terms by N_{22} from the mating combination matrix (see **Figure 2a**). The subsequent transformations are as follows:

$$u \rightarrow u^2 \cdot N_{11} + \frac{1}{2}uv \cdot N_{12} + \frac{1}{2}uv \cdot N_{21} + \frac{1}{4}v^2 \cdot N_{22}$$

$$v \rightarrow \frac{1}{2}uv \cdot N_{12} + uw \cdot N_{13} + \frac{1}{2}uv \cdot N_{21} + \frac{1}{2}v^2 \cdot N_{22} + \frac{1}{2}vw \cdot N_{23} + vw \cdot N_{31} + \frac{1}{2}vw \cdot N_{32}$$

$$w \rightarrow \frac{1}{4}v^2 \cdot N_{22} + \frac{1}{2}vw \cdot N_{23} + \frac{1}{2}vw \cdot N_{32} + w^2 \cdot N_{33}$$

, where u is the frequency of homozygous dominant individuals, v is the frequency of heterozygous individuals, and w is the proportion of homozygous recessive individuals.

These transformations comprise a first-order dynamical system with 2 quantities, u and v (w is expressed in terms of u and v, 1-u-v). In addition, they are normalized to ensure all transformed genotype frequencies sum to 100% of the population. The initial condition is [0.3,0.4], which was chosen because it controls for the frequency of the dominant and recessive allele in the initial population. When any probability of mating is altered, it is reduced from 1 to 0.5.

2.2 Interpreting the Model

Although the mathematics for correcting for random mating are the same as correcting for equal fitness, the biologically-relevant scenarios simulated to test a population with nonrandom mating are different.

In Mendelian Genetics, dominant alleles completely contribute to the phenotype of the offspring. Therefore, offspring with homozygous dominant (AA) and heterozygous (Aa) genotypes display the same phenotype, whereas offspring with homozygous recessive (aa) genotypes display the alternative phenotype. Hardy-Weinberg Models assume that each individual, regardless of its genotype, is equally likely to mate with another individual, regardless of its genotype. In non-ideal scenarios, some phenotypic traits are selected over others during mating. This can be modeled in multiple ways.

Scenario 1: We can imagine that parents with the dominant phenotype (AA or Aa genotype) are more likely to mate than parents with the recessive phenotype (aa genotype). This can be modeled by reducing the probability of mating for each mating combination with a parent of a homozygous recessive phenotype from 1 to 0.5.

Results: The homozygous dominant phenotype is heavily favored. After the 1000 generation, the frequency of the homozygous dominant genotype in the population is still increasing to 100% with virtually no heterozygous or homozygous recessive individuals.

Scenario 2: We can also imagine that parents with different phenotypes (AA and aa, Aa and aa) are less likely to mate than parents with the same phenotype. This can be modeled by reducing

the probability of mating for each mating combination with a parent of a homozygous recessive genotype and a parent with either a homozygous dominant or heterozygous genotype.

Results: The homozygous dominant phenotype is still heavily favored, but the frequency of this genotype in the population after the 1000 generation is slightly less than in the previous scenario. The behavior of the system is the same.

Scenario 3: We can also imagine the opposite, that parents with different phenotypes are more likely to mate than parents with the same phenotype.

Results: The population stabilizes in this scenario. The frequency of the homozygous recessive genotype remains the majority of the population. The frequency of the heterozygous genotype is ~45% of the population, the frequency of the homozygous dominant genotype is ~5%, and the frequency of the homozygous recessive genotype is ~50%.

Scenario 4: We can imagine a population where the maternal parent with a recessive phenotype is more likely to mate.

Results: The homozygous recessive genotype is nearly 100% the entire population by the 1000th generation.

Scenario 5: Let's imagine a population where the paternal parent with a dominant phenotype is more likely to mate.

Results: The homozygous dominant genotype overtakes the population by the 1000th generation.

Scenario 6: Further, we can imagine that a certain maternal-paternal mating combination is favored. Combining the previous two scenarios, we can imagine that a maternal parent with a recessive phenotype and a paternal parent with a dominant phenotype are most likely to mate.

Results: The population stabilizes in this scenario. The frequency of the homozygous recessive genotype remains the majority of the population. The frequency of the heterozygous genotype is ~44% of the population. The homozygous dominant genotype is almost eradicated from the population.

Scenario 7: We can also imagine the opposite of Scenario 6, where a maternal parent with a dominant genotype and a paternal parent with a recessive phenotype are most likely to mate.

Results: The population stabilizes, and the frequencies of each genotype after the 1000th generation are exactly the same as those in Scenario 6.

Overall Impressions: Manipulating the likelihood of mating based on fundamental intuitions about sexual selection can dramatically change the Hardy-Weinberg equilibrium of a population. Mating combinations that favor dominant or recessive traits lead to the dominance of their corresponding genotypes. In contrast, likely mating combinations that depend on parents with different genotypes tend to increase the heterozygosity of a population over time and in turn, its phenotypic diversity.

The Maple code (.txt) and a PDF file (.pdf) with the outlined analyses are attached with this report for this section.

3. Mutation Assumption

3.1 Measuring Mutated Mating

Another interesting assumption behind Hardy Weinberg genetics is that mutations will not occur or are not allowed to occur. Obviously this goes against nature itself as mutation is an extremely common concept that occurs in humans at an astonishingly high rate. In order to try and interpret mutations in the genetics of a population, we must first understand the different variants of mutations and how they matter in the sense of Hardy Weinberg. Generally speaking there are many types of mutations that occur on the scale of DNA, which build up and can affect alleles and genotypes dramatically.

$$p = A(n) = A(n-1) + (a(n-1)\cdot g - A(n-1)\cdot h)$$
$$q = a(n) = a(n-1) + (A(n-1)\cdot h - a(n-1)\cdot g)$$
g and h represent mutation rates in opposite directions

$$\begin{bmatrix} u^2 & u\,v & u\,w \\ u\,v & v^2 & v\,w \\ u\,w & v\,w & w^2 \end{bmatrix}$$

Figure 3a (left) shows a system of equations, **Figure 3b (right)** mating probabilities based on genotypes

Some of the most common mutations are forwards and backwards mutations, in which an allele converts to the other possible allele (i.e. for a gene represented with AA Aa and aa, A—> a and a —> A). By this occurring, the allelic frequency of both alleles will be changing dramatically which will affect a potential Hardy Weinberg Equilibrium. In order to attempt mapping these changes, we would need to adapt the Hardy Weinberg law and equation to factor in a mutation rate. The frequency of Allele A and Allele a would therefore not be solely dependent on the allele frequency of the previous generation but also a mutation rate which we assumed to be

constant in between generations. These changing frequencies are then represented in Figure 3a, which shows a discrete dynamical system that changes based on the mutation rates repeatedly over generations.

- Another Perspective: Another way to attempt to map the changes in mutations would be to apply the Orb function to the system provided in Figure 3a and then compare the Matrix 3b with the original values of p and q to the new p and q values after generations of mutations. However since we have predominantly worked with u, v, and w in class altering the original Matrix provided a more interesting approach.

In order to alter the original matrix shown in 3b to adapt to the new equations for p and q, we must understand what u, v and w represent. Since u, v, and w represent the frequency of each genotype, we know that these values are dependent on the allelic frequencies directly. Based on the Hardy Weinberg Equation, we know the following equivalency applies:

$u = p^2$

$v = 2 * p * q$

$w = q^2$

Based on this equivalency, we were able to rewrite Figure 3b in terms of the p and q in 3a. By doing this, we could calculate the change in genotypic frequencies by using the allelic frequencies that were being impacted by mutation rates.

- Mutation Rate: there are relatively low rates of mutation when it comes to forward and backward mutations, and most realistic values for this are seen to range from 10^-3 to 10^-8. A mutation rate of 0.01 for example would be rather uncommon, which is helpful as this mutation rate would be much more impactful on a population.
- Forward and Reverse: The rates of mutations between the forward and reverse are most commonly not equivalent which allows for one allele to be much more prevalent from a genotypic standpoint.

After applying these changes and finding the genotypic changes in terms of p and q, we are able to find the allelic frequencies of p and q in the long term. If p and q stabilized in the long term, then we know that u, v, and w have also stabilized which means the community has achieved hardy weinberg equilibrium.

3.2 Results

Since there is a constant change in allele frequencies due to these forward and reverse mutations, the probability of a generation reaching a stable equilibrium is very low. As seen in the attached .txt and .pdf files, the result of a majority of the tested Orbits found realistic mutation rates to never exactly stabilize, with their allelic frequencies varying even after 1000 generations. Since these allelic frequencies varied so often, based on the Hardy Weinberg Equation the frequency of genotypes were also constantly changing, even if not by a major scale. The Orbit seemingly

yielded a stable equilibrium with an unrealistic mutation rate, however when we applied a more realistic mutation rate (both in rate itself and the fact that forward/reverse rates are not equivalent), the long term behavior of p and q seemed to never be exactly stable. These minor changes in allelic frequency would naturally have a dramatic effect due to the assumption that there is a large population. Therefore, the assumption of no mutations does affect the Hardy Weinberg equilibrium in a rather dramatic way and therefore is essential in the HW law, as it allows for more stability when trying to calculate population genetics, which would be extremely important for applications of the Hardy Weinberg law such as calculating variance between a non-evolving population and a constantly evolving population.

3.3 Further Research

In addition to changes amongst the frequency of alleles, another form of mutation that is worth considering is the creation of a new allele, which would have far more drastic changes onto the population. The creation of a new allele is not something uncommon, and one of the most known cases would be the creation of the allele that allows for people to have blue eyes. It is speculated that at the beginning of human evolution, everyone had brown eyes until there was a mutation that led to a recessive allele that reflects those who are green and blue eyed. Factoring this into the Hardy Weinberg Equilibrium would be much more difficult as it not only breaks the assumption of No Mutation but impacts the fact that the general Hardy Weinberg Law is applied most commonly to genes that are dependent on 1 locus (2 alleles). Breaking this law, i.e. allowing for an A+ mutation forces one to factor into many other factors that impact population genetics such as lethality, beneficial mutations that would break other Hardy Weinberg assumptions such as random mating, survival, or an extremely large population.

Conclusion - Nick

From our data, we can surmise that the assumptions of the Hardy-Weinberg are similar to the real-world assumptions we have worked to create. In our data we attempted to investigate realistic scenarios that show how generations of species are affected under specific conditions. Using maple we were able to generate long term generations for these species and were able to see how long it took for them to stabilize under our assumptions. This in turn allows us to see these assumptions from a more realistic point of view. When accounting for the fitness of the offspring and sexual selection, we conclude that the real-world corrections for the Hardy-Weinberg equilibrium are the same.

Acknowledgements

Many portions of this paper would not be possible without the work done by Anne Somalwar on the HWg command, and the work done by Dr. Doron Zeilberger on his DMB.txt Maple functions. We thank them immensely for these resources.

References

"Mathematical Models in Biology" by Leah Keshet