Numerical and Analytical Explorations of the Hardy-Weinberg System with More Realistic Assumptions

Daniyal Chaudhry, Katryn Keating, Ezra Chechik, Zachariah Ali

December 2025

Contents

1 Assumptions of the Hardy Weinberg System

The Hardy-Weinberg recurrence model works with idealized biological assumptions to calculate how the frequencies of different alleles (and genotypes) in a population change over generations. Assessing the long term behavior of the model allows one to understand the stable ratios of different genotypes in the population.

Due to the simplicity of the model and its unrealistic assumptions, the biological conclusions that can be derived from it are limited. For instance, biologists may study real-world populations and their genotype frequencies over multiple generations, comparing them with the Hardy Weinberg model, in order to determine whether evolutionary forces are or are not present. This project will aim to replace the unrealistic assumptions of the Hardy Weinberg model with more realistic assumptions, to generate new recurrence models that may give more insight into the dynamics of populations in biological reality.

The first of the Hardy-Weinberg assumptions is that the population being modeled is large enough to minimize genetic drift due to random chance. In a population that is

large enough, allele frequencies are less subject to probabilistic shifts over generations. Since this assumption eliminates the need to account for randomness and stochasticity, this project will not delve into models which eliminate or alter the assumption.

The second assumption is that mating between members of the population occurs randomly. This assumption allows for a level of simplicity that is admittedly detached from real world biology. Nonrandom mating inherently affects how genotype ratios shift over generations. This project will address what happens to the stable steady states of different genotype frequencies under systems with preferential mating.

The third assumption is that there are no natural selective forces operating on the population. Similarly, this assumption simplifies the Hardy-Weinberg system in an unrealistic way. This project will address what happens to genotype frequencies under selective pressures.

The fourth assumption is that there is no gene flow within the population. In real world biology, migration causes genetic ratios to shift within different subsections of a population. This project will address how migrant populations affect the Hardy Weinberg model overall.

The final assumption is that there is no mutation within the population. Since this assumption once again removes the need to address the stochasticity of mutations and their various potential effects on the gene pool, it will not be altered in this project. However, this project will investigate the Hardy-Weinberg model when extended to X-linked alleles.

2 Random Mating

2.1 Universal Mating Preferences

The standard Hardy-Weinberg transformation does not take into account the concept of mating biases. In nature, members of a population that exhibit a certain genotype may be preferred over others.

By making alterations to the mating matrix used to derive the Hardy-Weinberg transformation equations, it is possible to create new transformations that may have their own stable steady states for the frequency of each genotype in the population. Under these new transformations, the stable frequencies of homozygous dominant (denoted GG), heterozygous (Gg), and homozygous recessive (gg) genotypes may change.

For the sake of this project, it will be assumed that the reader is familiar with the standard Hardy-Weinberg system, where the frequencies of alleles G + g = 1; that is, $p + q = 1$ and $p^2 + 2pq + q^2 = 1$. It will also be assumed that the reader understands the role of the mating matrix, a 3x3 matrix where the rows and columns represent the GG, Gg, and gg genotypes, and each entry in the matrix represents mating between the row and column genotypes. Finally, it is also assumed that the reader is generally familiar with the method of deriving Hardy-Weinberg transformation equations for each genotype (which are denoted u (for GG), v (Gg), and w (gg)).

The main alteration to the standard Hardy-Weinberg system that will be assessed is a case **where each genotype is assigned a universal preference coefficient.** In the mating matrix, a constant multiple "a" will be assigned to the entire column corresponding to the GG genotype; similarly, "b" and "c" will be assigned to the entire column corresponding to the Gg and gg genotypes, respectively.

$$\begin{bmatrix} a(u^2) & b(uv) & c(uw) \\ a(vu) & b(v^2) & c(vw) \\ a(wu) & b(wv) & c(w^2) \end{bmatrix}$$

Figure 1: An adjusted mating matrix which integrates universal mating preference coefficients.

The standard values of a, b, and c are all 1. With this adjusted matrix, a higher value of "a" corresponds with a higher likelihood that any member of the population, regardless of their genotype, will mate with a GG genotype individual. Likewise, a higher "b" or "c" value corresponds with a higher preference for Gg and gg individuals, universally. Treating (a,b,c) as parameters, an infinite amount of new Hardy-Weinberg transformation equations can be derived from this new mating matrix. Note that, if more than one parameter (a,b,c) is adjusted, then the derived Hardy-Weinberg transformation will depend solely on the ratio between the parameters in question. That is to say, parameters of (1,3,2) lead to the same transformation equations as (2,6,4), due to how coefficients simplify.

For consistency, this altered matrix for deriving Hardy-Weinberg equations will be denoted "HWgU," such that HWgU(a,b,c) denotes a modified system of transformations given parameters a, b, and c, as defined above. There are a few distinct cases of HWgU transformations worth analyzing mathematically: particularly, the case where only one parameter is altered and the case where two parameters are altered. The third case, where all three parameters are altered, leads to redundant analyses due to previously mentioned facts.

Regarding the first case, consider a system where the parameter "a" is altered, but "b" and "c" remain at a value of 1. Such a system can be denoted HWgU(x,1,1), where the parameter a is treated as a variable "x" that can range in value. Ranging x from 2 to 30, and using the Maple program to plot the stable steady states of the genotypic frequencies of u, v, and w under these HWgU transformations, **Plot 2.1a** emerges.

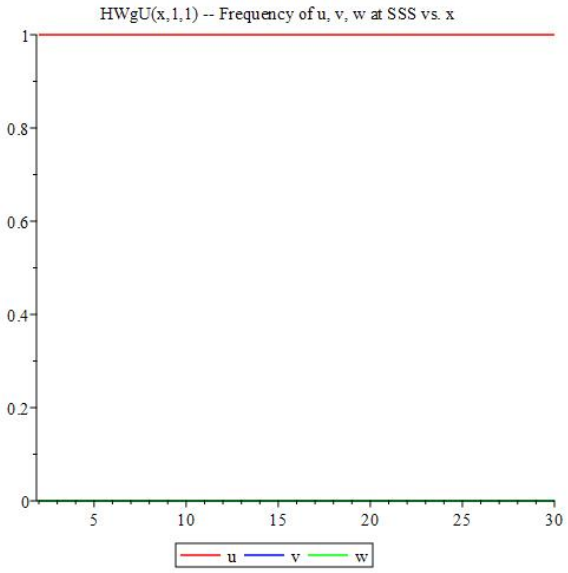

Figure 2: Plot 2.1a

The plot indicates a notable fact: in any Hardy-Weinberg transformation where the u genotype– the homozygous dominant genotype– is even remotely preferred over other genotypes, it will push out the other genotypes from the gene pool at equilibrium. This corresponds to the loss of the recessive allele (g) in the population.

Now, consider the symmetrical set of of HWgU transformations, HWgU(1,1,x). Plotting with the same methodology, **Plot 2.1b** emerges.

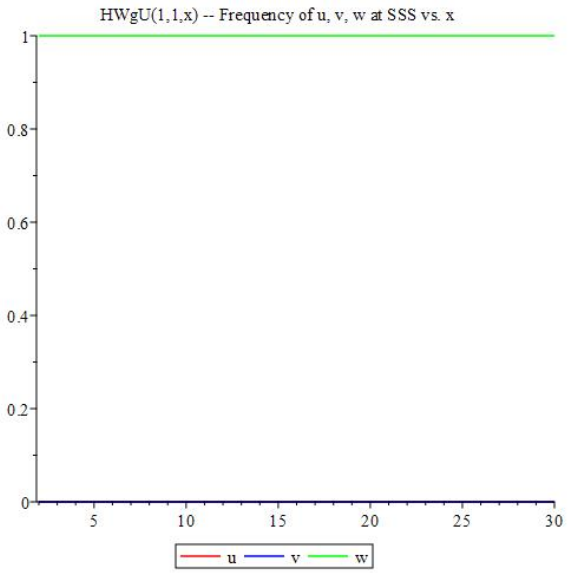

Figure 3: Plot 2.1b

Here, the plot indicates that, in any scenario where the homozygous recessive genotype is remotely preferred over other genotypes, the other genotypes will be removed from the gene pool at equilibrium, corresponding to the loss of the dominant allele (G)

in the population.

The final single-parameter change would result in the set of transformations denoted HWgU(1,x,1), which is plotted via the same method in **Plot 2.1c**.

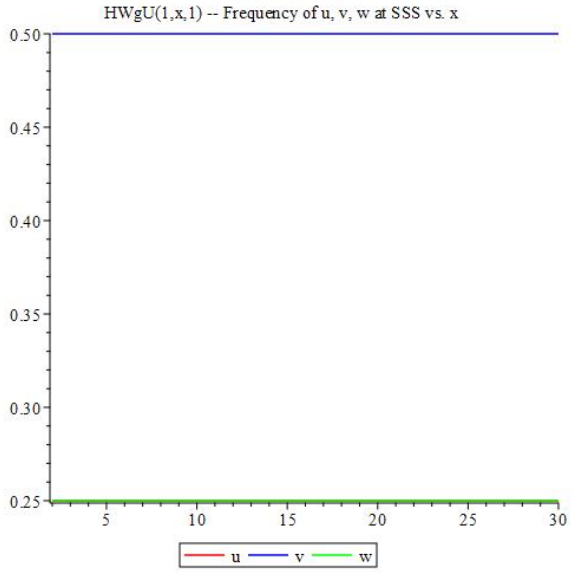

Figure 4: Plot 2.1c

Notice the range of frequencies in this plot is from 0.25 to 0.5, rather than from 0 to 1. In other words, this plot indicates a unique occurrence. When the heterozygous genotype is preferred and the other genotypes receive no preference, the equilibrium will always settle with 25% of the population being homozygous dominant, 25% homozygous recessive, and 50% heterozygous. In other words, preference for the heterozygous genotype in a population does not oust out any alleles (G or g) at equilibrium, maintaining genetic diversity.

Moving onto cases where two parameters are adjusted, consider the two following ideas: HWgU(x,y,1) and HWgU(1,x,y). In both of these scenarios, the heterozygous genotype and one homozygous genotype receive an increased level of preference in mating. These cases are fundamentally symmetrical, and thus, only HWgU(x,y,1) will be plotted, but the biological implications will be extrapolated to the HWgU(1,x,y) system as well.

The computational complexity of a 3D stable-steady-state plot (where x and y are treated as variables) is beyond the scope of the technology used to produce this project. However, by selecting sample HWgU(x,y,1) systems with fixed x values– particularly, HWgU(2,y,1), HWgU(10,y,1), and HWgU(20,y,1)– the behavior of the HWgU(x,y,1) transformations as a whole can be analyzed. The aforementioned sets of transformations are plotted in **Plot 2.1d, e, and f**, in their respective order. Note that, for the latter two, the y range used was from 5 to 100, rather than 2 to 30.

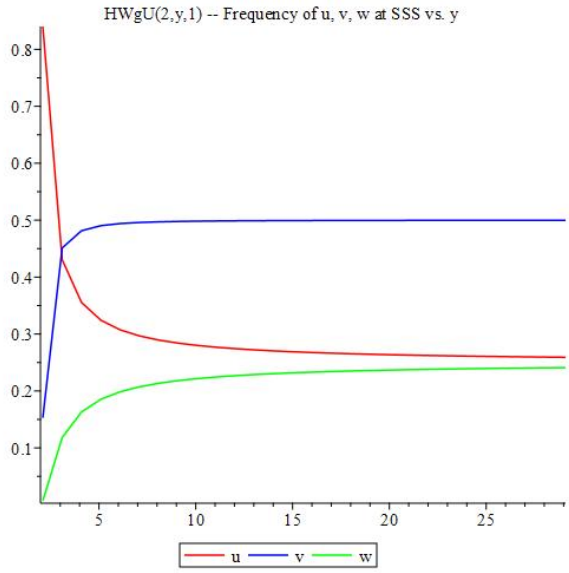

Figure 5: Plot 2.1d

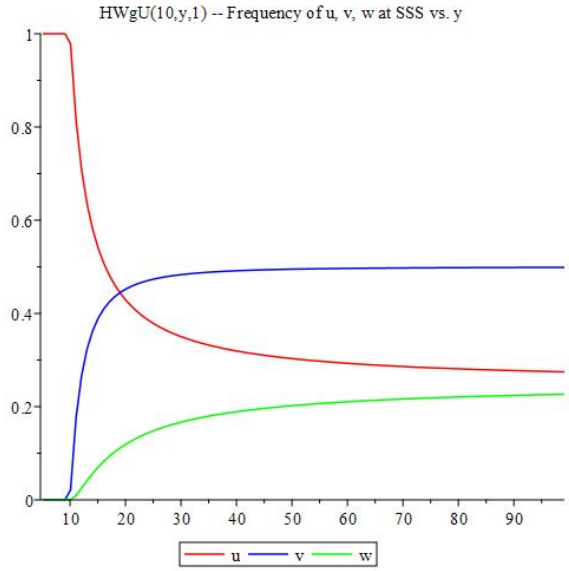

Figure 6: Plot 2.1e

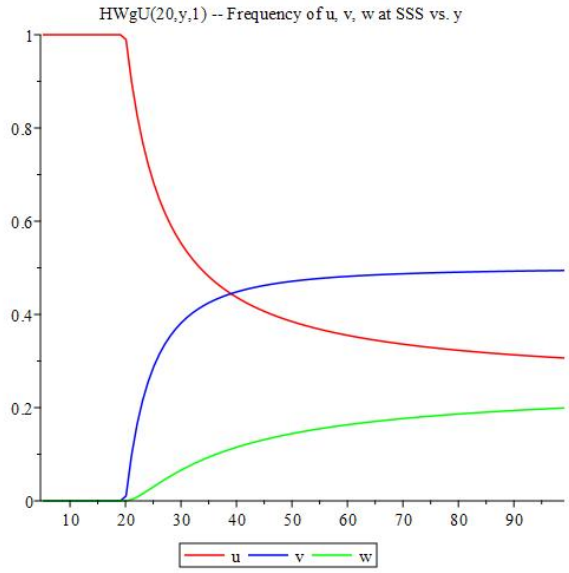

Figure 7: Plot 2.1f

The behavior of these plots leads to the following conclusions. Firstly, in any HWgU(x,y,1) transformation where y is less than x (ie. parameter "b" is less than parameter "a"), only the GG genotype remains alive at equilibrium. However, when y is greater than x (ie. parameter "b" is greater than parameter "a"), then as y increases, the stable steady state becomes exponentially closer to a ratio of 25% GG, 50% Gg, and 25% gg. In other words, **Plots 2.1d-f** mimic the stable-state behavior of **Plot 2.1a** when y is less than x, but tend towards the behavior of **Plot 2.1c** as y increases further and further away from x.

Recalling that the analysis above draws logical symmetry with HWgU(1,x,y), the following can be stated regarding those transformations. When parameter "b" is less than parameter "c," only the homozygous recessive genotype remains alive at equilibrium, mimicking the stable-state behavior of **Plot 2.1b**. When parameter "b" is greater than parameter "c," the stable steady state becomes exponentially closer to the 1:2:1 ratio in **Plot 2.1c** as "b" continues to increase.

Two distinct biological implications can be derived from the plots of these stable steady states. Firstly, even if the heterozygous genotype has a level of preference, so long as one homozygous genotype is preferred more than it (ie. has a higher preference coefficient), the homozygous genotype in question will become the only surviving genotype at equilibrium. Secondly, when the heterozygous genotype is the most preferred genotype in mating, all three genotypes will survive at equilibrium, and the equilibrium ratio will depend on how much preference is allocated to the remaining genotypes (ie. the ratio between preference coefficients), with the general pattern that higher relative preference will result in higher relative presence at equilibrium.

The final HWgU scenario to analyze is the case where parameter "b" stays at its standard value of 1, while "a" and "c" change, denoted HWgU(x,1,y). Numerical trials indicate that such systems have more than one stable steady state. In fact, they have precisely two: one stable state where only the GG genotype survives, and one stable

state where only the gg genotype survives.

The biological conclusion is that, in any HWgU(x,1,y) system, where both homozygous genotypes receive some level of mating preference, only one of these genotypes will ultimately survive. The genotype that survives depends on the initial frequencies of each genotype in the population, as well as the ratio between the preference coefficents "a" and "c" allocated to each genotype.

It is also worth noting that there is precisely one unstable steady state for each of these HWgU(x,1,y) systems. **Table 1** presents some samples of these unstable steady states.

Transformation	**Unique Unstable Steady State**			**Ratio GG:gg**
	Homozygous Dominant (GG)	**Heterozygous (Gg)**	**Homozygous Recessive (gg)**	
HWgU(2,1,3)	0.444	0.444	0.111	4:1
HWgU(3,1,3)	0.250	0.500	0.250	1:1
HWgU(4,1,3)	0.160	0.480	0.360	4:9
HWgU(5,1,3)	0.111	0.444	0.444	1:4
HWgU(6,1,3)	0.082	0.408	0.510	∼5:31

Table 1: Samples of Unstable Steady States for HWgU(x,1,y) Transformations.

Note that, for real populations that resemble the HWgU(x,1,y) systems, the initial genotypic frequencies will have a GG:gg ratio that is greater or less than the ratio at the Unstable steady state. It can be hypothesized that the homozygous genotype that will survive in the long run depends whether the actual ratio is skewed towards the GG or gg genotype when compared to the ratio at the unstable steady state, but this hypothesis requires further numerical and analytical testing.

Ultimately, through the above analysis of different HWgU transformations, a several biological implications emerge.

1. In scenarios where **a homozygous genotype** is the most preferred, then the population will reach a point where it **entirely consists** of that homozygous genotype. As an exception, if **both homozygous genotypes** receive a boosted level of preference, then either homozygous genotype may win out at equilibrium. The homozygous genotype that survives depends on the initial frequencies of genotypes in the population, as well as the preference coefficients allocated to the genotypes in question.

2. In scenarios where the **heterozygous genotype** is the most preferred, this results in the only case where the population will reach a stable steady state where all three genotypes survive. If the homozygous genotypes receive no preference, the ratio of GG:Gg:gg will be 1:2:1 at equilibrium. Otherwise, the ratio will be skewed in the direction of the homozygous genotype with a higher preference coefficient.

2.2 Non-Universal Mating Preferences

For the scope of this project, the analysis of the impact of preferential mating concludes with the HWgU set of transformations marked out in the last subsection. However, some other adjustments to the mating matrix are worth investigating in the future. Such concepts are included below:

First, a scenario where the frequency of self-type mating is controlled by a parameter "a", such that a higher "a' corresponds to a higher frequency of genotypes mating with members of the same genotype.

$$\begin{bmatrix} a(u^2) & uv & uw \\ vu & a(v^2) & vw \\ wu & wv & a(w^2) \end{bmatrix}$$

Figure 8: An adjusted mating matrix which integrates self-genotype mating preferences.

Second, an adjusted version of the last scenario, where phenotype is considered rather than genotype. Consider the idea that, generally, GG and Gg genotypes represent the same phenotypic expression. Then, the mating matrix changes into the following.

$$\begin{bmatrix} a(u^2) & a(uv) & uw \\ a(vu) & a(v^2) & vw \\ wu & wv & a(w^2) \end{bmatrix}$$

Figure 9: An adjusted mating matrix which integrates self-phenotype mating preferences.

Third, a scenario where the frequency of mating with a certain genotype is related to the current frequency of that genotype in the population. In other words, the "commonness" of a genotype alters how much it is preferred by the population. Note that, in the matrix below, "a" is a parameter that influences how strongly the current frequencies of u, v, and w affect the frequency of certain mating pairs. Also note that the system was designed very crudely and may be modified in the scenario of actual numerical exploration.

$$\begin{bmatrix} (1+au)(u^2) & (1+av)(uv) & (1+aw)(uw) \\ (1+au)(vu) & (1+av)(v^2) & (1+aw)(vw) \\ (1+au)(wu) & (1+av)(wv) & (1+aw)(w^2) \end{bmatrix}$$

Figure 10: An adjusted mating matrix which integrates frequency-dependent mating preferences.

All of the above adjustments to the mating matrix, and many more, are worth further investigation and numerical analysis, which could yield particularly interesting results and biological implications.

3 Natural Selection

Another assumption of the Hardy-Weinberg Law is that no selective forces act on the population being modeled. Knowing that natural selection is the notion that certain

traits are more likely to survive, and thus reproduce, genotypic frequencies would undoubtedly be affected by selective forces. Particularly, genotypes that produce advantageous traits (higher fitness) will prosper, while genotypes that produce less advantageous traits (lower fitness) will dwindle. [1]

Consider a set of Hardy-Weinberg equations where each genotype GG, Gg, and gg is assigned a coefficient of selection: a, b, and c, respectively. Let a, b, and c be values ranging from 0 to 1. Let it be such that, when "a" is high, selective pressures against the GG genotype are high, and thus GG has lower fitness equivalent to (1-a). Similar logical structure applies to "b" and Gg, as well as "c" and gg. Notice that, with this logic, the parameters for the standard Hardy-Weinberg equation would be 0. [2]

Let the term "gametic contribution" be defined as the surviving proportion of each genotype that contributes to making the next generation. That is, not all members of a certain genotype will survive and mate; the proportion that does is the "gametic contribution" of the genotype. Then, the "gametic contribution" of each genotype is described in **Table 2**. Notice that the total gametic contribution is fundamentally less than 1, as all fitnesses are less than 1.

	Genotype			
	GG (u)	**Gg (v)**	**gg (w)**	**Total**
Initial Frequencies	p^2	$2pq$	q^2	1
Coefficient of Selection	a	b	c	-
Fitness	$1-a$	$1-b$	$1-c$	-
Gametic Contribution	$p^2(1-a)$	$2pq(1-b)$	$q^2(1-c)$	$p^2(1-a)+2pq(1-b)+q^2(1-c)$

Table 2: Genotype Frequencies, Selection Coefficients, Fitness, and Gametic Contribution.

Understanding this, one can derive the proportion of the (n+1) generation alleles from the following manipulation of gametic contribution:

$$p_{n+1} = \frac{\text{gametic contribution contributing to p}}{\text{total gametic contribution}} = \frac{pq(1-b)+p^2(1-a)}{p^2(1-a)+2pq(1-b)+q^2(1-c)}$$

$$q_{n+1} = \frac{\text{gametic contribution contributing to q}}{\text{total gametic contribution}} = \frac{pq(1-b)+q^2(1-c)}{p^2(1-a)+2pq(1-b)+q^2(1-c)}$$

Note: In both numerators, $pq(1-b)$ is used rather than $2pq(1-b)$ because only half of the alleles from the gametic contribution of the Gg genotype contribute in both scenarios.

Then, with the understanding that $GG_{n+1} = [p_{n+1}]^2$, $Gg_{n+1} = 2\,p_{n+1}q_{n+1}$, and $gg_{n+1} = [q_{n+1}]^2$, the following equations can be calculated:

$$GG_{n+1} = \frac{p^2[p^2(1-a)^2+2pq(1-a)(1-b)+q^2(1-b)^2]}{[p^2(1-a)+2pq(1-b)+q^2(1-c)]^2}$$

$$Gg_{n+1} = \frac{2pq[(q(1-b)+p(1-a))(p(1-b)+q(1-c))]]}{[p^2(1-a)+2pq(1-b)+q^2(1-c)]^2}$$

$$gg_{n+1} = \frac{q^2[q^2(1-c)^2 + 2pq(1-c)(1-b) + p^2(1-b)^2]}{[p^2(1-a) + 2pq(1-b) + q^2(1-c)]^2}$$

The above system, along with the substitution that $p = u + \frac{v}{2}$ and $q = \frac{v}{2} + w$, describes the transformation on u, v, and w, for a Hardy-Weinberg system with selection pressures a, b, and c. The transformation in question will be denoted HWNatSel(a,b,c).

Consider the scenario where only one genotype faces a selection pressure. There are three such cases: HWNatSel(a,0,0), HWNatSel(0,b,0), and HWNatSel(0,0,c). By analyzing the stable and unstable steady states of these systems, some key patterns emerge, shown in **Table 3**.

HWNatSel Type	**Unstable Steady States**	**Stable Steady States**
HWNatSel(a,0,0), $0 < a < 1$	$(1, 0, 0)$	$(0, 0, 1)$
HWNatSel(0,b,0), $0 < b < 1$	$(0.25, 0.5, 0.25)$	$(0, 0, 1), (1, 0, 0)$
HWNatSel(0,0,c), $0 < c < 1$	$(0, 0, 1)$	$(1, 0, 0)$

Table 3: The Unstable Steady States and Steady States for Single-Parameter HWNatSel cases.

The patterns in the data suggest the following biological conclusions:

1. For a Hardy-Weinberg system where only one **homozygous genotype** is under selective pressures, it will fade out of the population at equilibrium and only the other homozygous genotype will survive.

2. For a Hardy-Weinberg system where only the **heterozygous genotype** is under selective pressure, eventually, only one homozygous genotype will survive, and the survivor depends on the initial frequency of each genotype in the population.

3. Fundamentally, in any case where only one genotype receives selective pressures, there is a loss of genetic diversity, and only one allele survives in the ultimate population.

The next case to analyze is HWNatSel scenarios where two genotypes have selective pressures applied to them. Consider HWNatSel(a,0,c), where both a and c are nonzero. In such a scenario, both the GG and gg genotypes have selective pressures acting against them.

By fixing the value of a at a sample value (ie. 0.3), and plotting the stable steady state frequencies of u, v, and w as c increases from 0.1 to 0.9, **Plot 3a** is obtained. By fixing the value of c instead, and plotting the stable steady states as a increases, **Plot 3b** is obtained. Due to logical symmetry, the graphs showcase identical patterns, where u and w are flipped.

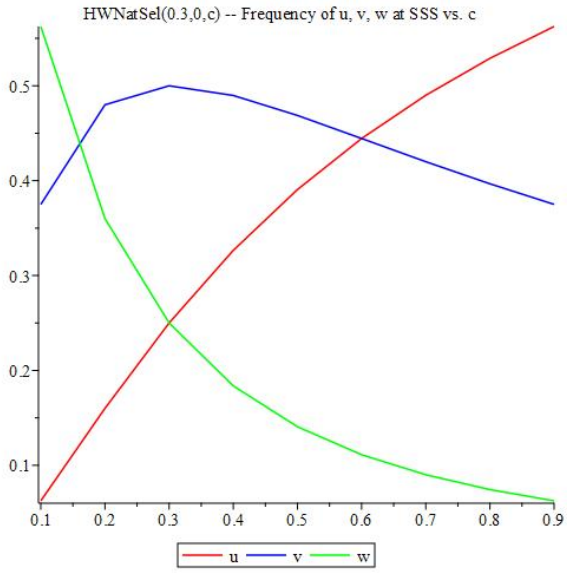

Figure 11: Plot 3a

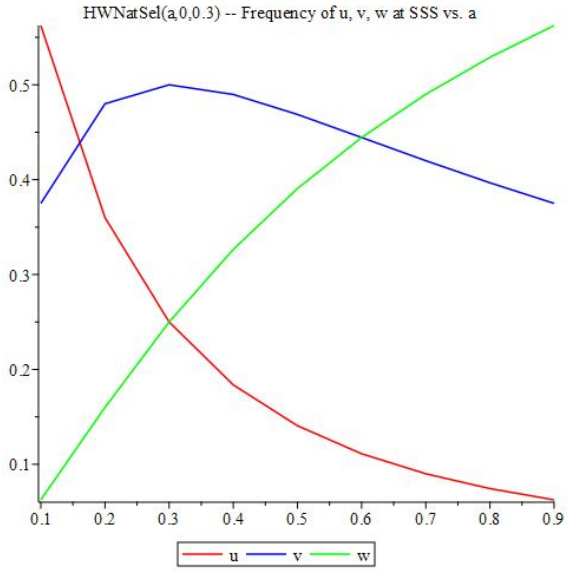

Figure 12: Plot 3b

Analyzing **Plot 3a**, a few key patterns are clear. As c increases, the frequency of the homozygous recessive (gg) genotype becomes exponentially smaller at equilibrium. This makes sense, as c is fundamentally inversely proportional to the fitness of w. Also as c increases, the frequency of the homozygous dominant (GG) genotype becomes larger at equilibrium. Finally, note that the frequency of the heterozygous genotype (Gg) increases until c = a, and then begins to decrease. In fact, it is clear that, when c = a in such HWNatSel(a,0,c) cases, the heterozygous genotype reaches its maximum possible equilibrium proportion of 0.5, and the two homozygous genotypes share the equilibrium proportion of 0.25. **Plot 3b** yields a similar analysis, swapping u with w.

The biological implications are the following. Firstly, when homozygous genotypes

have natural selective pressures against them, then the genotype with a higher selective pressure will be less prominent at equilibrium. Secondly, there is clear proportionality between the frequency of GG genotypes and the pressures against the gg genotype, and vice versa. This points to the idea that when one homozygous genotype struggles, the other thrives.

It could be interesting for future analysis to understand exactly what ratio of values for a:c yields an equilibrium point where the heterozygous population is the most prosperous. For instance, in **Plot 3b**, where c is fixed at 0.3, the heterozygous population seems to be the most prosperous at equilibrium within the rough domain of (0.15, 0.6). It is possible that these values emerge due to their relation to the value 0.3, though this hypothesis requires further testing.

The remaining scenario for a double parameter change is one where b and either a or c change, while the last parameter remains at 0. This would be denoted HWNatSel(a,b,0) or HWNatSel(0,b,c). It quickly became apparent during numerical testing that such systems always result in the homogenization of the population, with only one homozygous genotype surviving at equilibrium. The homozygous genotype that survives depends on the selective pressures a and c, as well as the initial population of GG and gg genotypes.

The final case to analyze is one where all three parameters are changed. That is, all members of the population have selective pressures applied to them. An interesting sample to analyze, and the example which has its stable states plotted in **Plot 3c**, is the case of HWNatSel(0.7,b,0.6), where a and c are fixed but unequal, and b varies.

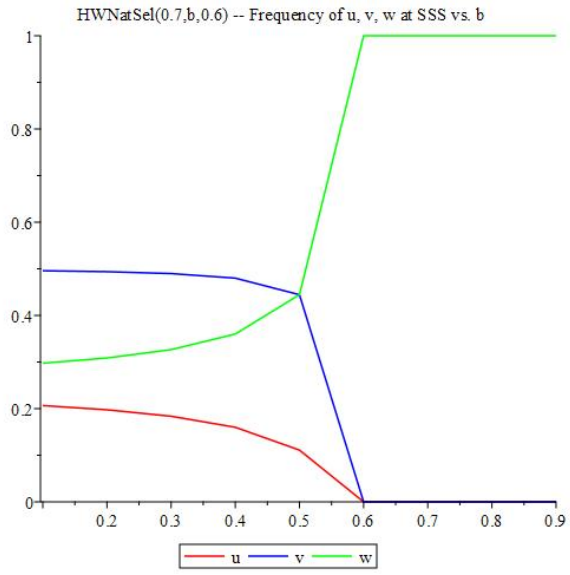

Figure 13: Plot 3c

This plot provides key insight into the behavior of HWNatSel systems where all three parameters are nonzero. Here, when b is 0.6 or greater, that is, when the selective pressures on the heterozygous population are not the least relevant in the system, only one

homozygous genotype survives at equilibrium. Importantly, although the graph shows that the homozygous recessive genotype is the survivor, and while this may seem believable due to the higher pressures on the homozygous dominant genotype, this is merely a result of the Maple procedures used to produce this graph. In reality, either the GG or gg genotypes could survive at equilibrium, depending on the initial frequencies of those genotypes in the population.

When b is less than 0.6, that is, when it is the least relevant selective pressure in the system, then all three genotypes survive at equilibrium. As b decreases, the stable point tends towards behavior that would be predicted by the HWNatSel(a,0,c) model discussed and analyzed through **Plot 3b**.

Ultimately, the biological conclusions that can be derived from this analysis of the HWNatSel set of Hardy-Weinberg transformations are listed below.

1. In any population where the **heterozygous genotype does not have the lowest magnitude of selective pressure acting against it**, only one homozygous genotype survives at equilibrium. The surviving genotype depends on the initial genotypic frequencies in the population, as well as the magnitude of selective pressures acting on the homozygous genotypes in question.

2. In any population where the **heterozygous genotype has the lowest selective pressure against it**, all three genotypes survive at equilibrium, with the frequencies depending on the relationship between the three selective pressure parameters. Generally, higher parameters correlate with lower fitness and lower presence at equilibrium.

3. In populations where only one **homozygous genotype** has selective pressures against it, then only the other homozygous genotype survives at equilibrium.

4 Migration

The next adjustment to the Hardy-Weinberg system pertains to the assumption that no migration occurs. An interesting analysis would be to see what happens to the frequency of an allele in a population after an immigration event.

Consider an isolated island population, and let the parameter p_m adjust the normal Hardy-Weinberg system by acting as the frequency of the dominant allele G in an incoming migrant population. Let m be the proportion of the island's population that are migrants in the newly created population. Then, creating a recurrence equation for the dominant allele G after a migration event (represented by p):

$$p_1 = (1 - m)p + mp_m$$

$1 - m$ can be seen as the proportion of the population that are not new migrants; multiplying it with p calculates the proportion of dominant allele G in the resulting population that comes from non-migrants. [3] Summing with mp_m, we derive the total p_1, the dominant allele's frequency after the migration event. A similar process can

derive the proportion of the recessive allele, g, in the new population; consider q_m to be the frequency of q in the migrant population.

$$q_1 = (1-m)q + mq_m$$

Considering that $GG_1 = p_1^2$, $Gg_1 = 2p_1q_1$, and $gg_1 = q_1^2$, one can derive the equations corresponding to each genotype after the migration event.

Ignoring the impacts of the reproductive step, as with random mating, allele frequencies remain unchanged due to dynamic equilibrium, one can assess what would happen to the G and g alleles in the event of a second migration event.

If we wanted to find out the next generation of the frequency of G (call it p_2), the same concept can be used:

$$p_2 = (1-m)p_1 + mp_m$$

It becomes clear that migration events cause a recurrence pattern on the frequency of alleles in the population.

$$p_n = (1-m)p_{n-1} + mp_m, \quad n \geq 1, \quad p_0 = p$$

As m and p_m do not depend on n, this is an inhomogeneous linear recurrence relation with initial conditions, which is solved as:

$$p_n = (1-m)^n p + p_m(1-(1-m)^n)$$

Thus, allele frequencies can be predicted after any amount of immigration events. An interesting concept would be to investigate what happens as the amount of immigration events tends to infinity.

$$\begin{aligned}
\lim_{n\to\infty} p_n &= \lim_{n\to\infty} \Big[(1-m)^n p + p_m\big(1-(1-m)^n\big)\Big] \\
&= \lim_{n\to\infty} (1-m)^n p + p_m\Big(1 - \lim_{n\to\infty}(1-m)^n\Big) \\
&= 0 \cdot p + p_m(1-0) \\
&= p_m
\end{aligned}$$

This analysis leads to the conclusion that, as more and more migrants make up the population, the frequency of the dominant allele G (and thus the recessive allele g) begins to hone in on the frequency within the incoming migrant population (p_m and q_m). The following plot, with $p = 0.3$, $m = 0.1$, $p_m = 0.9$, showcases this concept in action, as "generations" of immigration occur.

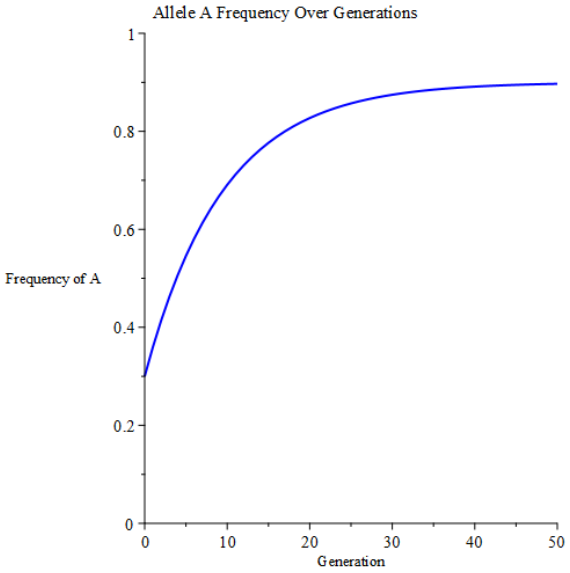

Figure 14: Frequency of Allele A with migration

Also notice that p_m is a steady state:

$$\begin{aligned} p_m &= (1-m)p_m + mp_m \\ &= p_m - mp_m + mp_m \\ &= p_m \end{aligned}$$

So it has been proven that, if the frequency of the dominant allele G in both populations is identical, then the main population will maintain an unchanging frequency of the G allele in theory.

For the migration map, treating the recurrence as a formula with its own derivative,

$$F(p) = (1-m)p + p_m,$$

$$F'(p) = 1 - m.$$

Evaluating this at the equilibrium:

$$|F'(p_m)| = |1-m|.$$

Because migration amounts satisfy $0 < m \leq 1$, we have

$$0 \leq 1 - m < 1,$$

and therefore

$$|1-m| < 1.$$

This shows that the fixed point $p = p_m$ is stable. This aligns with the understanding derived earlier, that the allele frequency in the main population moves closer to the migrant frequency each "generation" and will not diverge with any given initial conditions.

The following plots show the progression of different genotypes under a system with migration. Notice the steady shift from the initial (u,v,w) proportions of (0.36,0.48,0.16) according to the parameters m = 0.05, $p_m = 0.7$. Understanding that the red line indicates the u genotype, the blue line indicates the v genotype, and the green line

indicates the w genotype; this plot trends toward a population with a u value of $p_m^2 = 0.49$, and a w value of $q_m^2 = 0.3^2 = 0.09$, exactly as predicted.

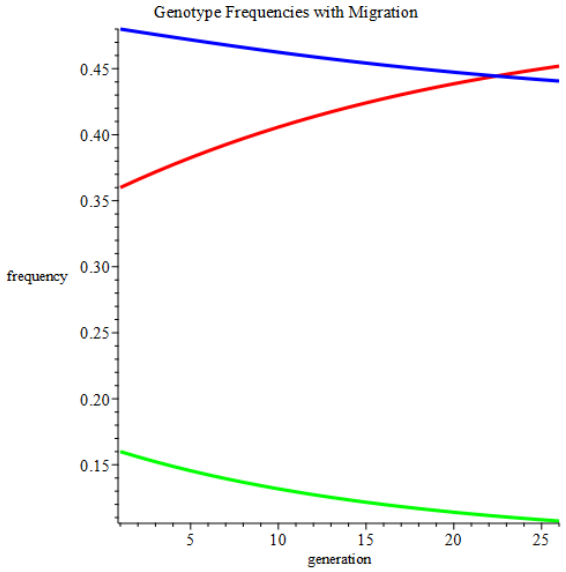

Figure 15: Frequency of alleles under multiple migrations.

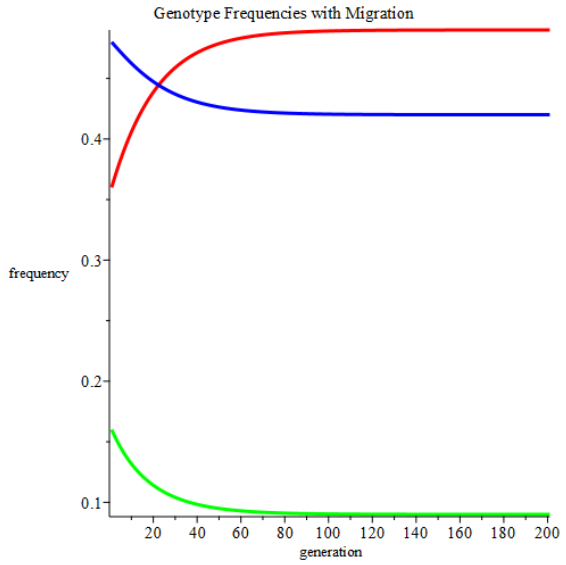

Figure 16: Extended plot.

Ultimately, this analysis of how migration events affect a population provide an interesting insight into patterns that can be missed in a standard Hardy-Weinberg model where migration is neglected. The central biological conclusion that has been reached is thus: migrant populations shift the genotype frequencies in a main population towards the frequency in their own population. Rather, immigration events cause predictable and reliable shifts to the genotype frequencies in a main population.

5 Extension to X-Linked Alleles

As a bonus investigation, this section will adapt and experiment with the Hardy-Weinberg model for an X-linked allele proposed by M. Traykov and Iv. Trenchev in a 2016 paper, Mathematical Models in Genetics. [4]

5.1 Theory and Introduction

In population genetics, Hardy-Weinberg equilibrium applies to genes located on autosomal chromosomes, where each individual carries two copies of each gene. In the usual case, an individual can be AA, Aa, or aa, where 'A' represents one allele and 'a' is another, usually the recessive allele.

Populations with heterogametic sex contain males with one X chromosome (XY) and females with two X chromosomes (XX). While females can still be homozygous and heterozygous, males are hemizygous for an X-linked gene. Each male inherits an X from the mother and a Y from the father. Each female inherits an X from the mother and father.

5.2 Traykov and Trenchev Model: Following an X-linked Allele

$$\begin{aligned} q_n &= \text{frequency of allele } \theta \text{ in females} \\ r_n &= \text{frequency of allele } \theta \text{ in males} \\ (1) r_n &= q_{n-1} \\ (2) q_n &= \frac{1}{2}q_{n-1} + \frac{1}{2}r_{n-1} \end{aligned}$$

- Males get one X-linked allele from only their mother.
- Females get two X-linked alleles from both parents, so allele frequency is the mean of the two preceding generations.

Combining (1) and (2),

$$\frac{2}{3}q_n + \frac{1}{3}r_n = \frac{2}{3}\left(\frac{1}{2}q_{n-1} + \frac{1}{2}r_{n-1}\right) + \frac{1}{3}(q_{n-1}) = \frac{2}{3}q_{n-1} + \frac{1}{3}r_{n-1}.$$

Proving there is a constant $P = \frac{2}{3}q_n + \frac{1}{3}r_n = \frac{2}{3}q_0 + \frac{1}{3}r_0$, where $n > 0$.

Explicit Formula for $q_n - P$:

As the paper presents (Traykov and Trenchev),

$$\begin{aligned} q_n - P &= q_n - \frac{3}{2}P + \frac{1}{2}P \\ &= \frac{1}{2}q_{n-1} + \frac{1}{2}r_{n-1} - \frac{3}{2}\left(\frac{2}{3}q_{n-1} + \frac{1}{3}r_{n-1}\right) + \frac{1}{2}P \\ &= -\frac{1}{2}q_{n-1} + \frac{1}{2}P \end{aligned}$$

An interesting observation of this explicit formula is that the difference reduces by half each generation:

$$|q_n - P| = \frac{1}{2}|q_{n-1} - P|$$

This is a decay with a flipping of signs. What we see is that the zigzag becomes closer and closer to the constant P, as shown in the Time Series Plot. Looking back at the original discrete system,

$$\begin{aligned} r_n &= q_{n-1} \\ q_n &= \frac{1}{2}q_{n-1} + \frac{1}{2}r_{n-1} \end{aligned}$$

we can analyze the stability by defining a Jacobian matrix of numbers

$$\begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}.$$

This is interesting compared to the Jacobian consisting of functions of the general Hardy-Weinberg transformation, which is a non-linear transformation. The eigenvalues of this system are 1 and 1/2. This represents semi-stability. The system approaches the fixed point (P, P) but oscillates around it, with decreasing difference.

5.3 Maple Numeric Experimentations & Graphs

Hypothesis: q_n and r_n quickly approach the value $P = \frac{2}{3}q_0 + \frac{1}{3}r_0$, the weighted average of the alleles since females have two X chromosomes and males have one.

To test this hypothesis and look at the characteristics of this model, we created a Maple procedure, `XalleleSIM(Q0,R0,N)`, that takes initial allele frequency in females and males, q_0 and r_0 respectively, and iterates over N generations. It gives

$$[[q_0, q_1, \ldots, q_k], [r_0, r_1, \ldots, r_k], P].$$

Values of allele frequency in females and males in consecutive generations as well as the constant P derived from the initial conditions.

For example,

```
XalleleSIM(0.33, 0.74, 8);
```

Female q_n	0	1	2	3
Values	0.33	0.5350	0.4325	0.48375
Female q_n	4	5	6	7
Values	0.4581	0.4709	0.4645	0.4677
Female q_n	8			
Values	0.4661			

Male r_n	0	1	2	3
Values	0.74	0.33	0.5350	0.4325
Male r_n	4	5	6	7
Values	0.48375	0.4581	0.4709	0.4645
Male r_n	8			
Values	0.4677			

$$P = 0.4667$$

```
XalleleSIM(0.21, 0.63, 12);
```

Female q_n	0	1	2	3
Values	0.21	0.420	0.3150	0.36750
Female q_n	4	5	6	7
Values	0.341	0.354375	0.3478125	0.35109375
Female q_n	8	9	10	11
Values	0.349453125	0.35027	0.34986	0.35006
Female q_n	12			
Values	0.349965			

Male r_n	0	1	2	3
Values	0.63	0.21	0.420	0.3150
Male r_n	4	5	6	7
Values	0.36750	0.341	0.354375	0.3478125
Male r_n	8	9	10	11
Values	0.35109375	0.349453125	0.35027	0.34986
Male r_n	12			
Values	0.35006			

$$P = 0.3500$$

Looking at the results of this numeric experiment, we can notice two important points:

1. In contrast to the general Hardy-Weinberg Formulas, equilibrium is NOT reached after one generation.
2. Allele frequency of male populations follows one generation behind the female population $\rightarrow q_0 = 0.33, r_1 = 0.33$.
3. For any generation n, the pair (q_n, r_n) satisfies the property that one of the allele frequencies exceeds the constant P whereas the other is less than P.

We can further explore these ideas by plotting both the allele frequencies of male and females as a function of the generation.

Using the same initial allele frequencies from the first numeric example, and plotting for 10 generations,

```
XalleleTimeSeriesPlot(0.33, 0.74, 10);
```

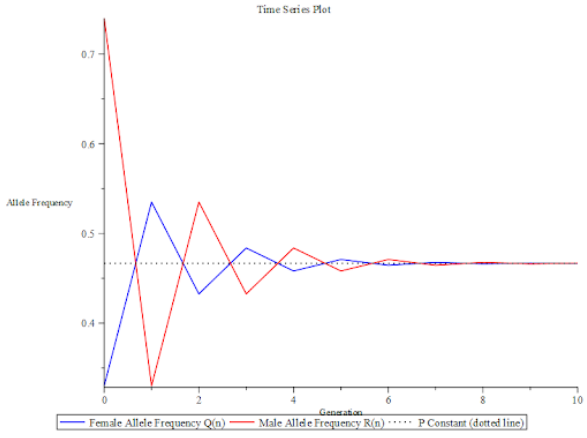

Figure 17: Oscillations of Q(n) and R(n) over 10 generations.

Here, we see how the linear recurrence system follows a zigzag pattern and the males lag one generation behind the females. We can also see how for both males and females,

the frequency of the X-linked allele will overcompensate P in one generation, and undercompensate in the next. Graphically, we can see how the allele frequencies both tend to the constant value P in only a few generations.

5.4 Adding Fitness/Survivability

In an attempt to address the generalizations of this model, we propose the following model as a starting point to look at how fitness may skew away from the General Hardy-Weinberg X-allele 'P' constant. Outlined below is a sex-specific frequency dependent fitness model.

Assumptions:

- q_n = frequency of allele θ in females
- r_n = frequency of allele θ in males
- S_f = female fitness coefficient, representing how the allele's presence affects reproductive success ($S_f > 0$ = negative fitness of females)
- S_m = male fitness coefficient, representing how the allele's presence affects male reproductive success ($S_m > 0$ = negative fitness of males)

$$(1) r_n = q_{n-1}$$
$$(2) q_n = \frac{1}{2} q_{n-1}(1 - S_f Q_{n-1}) + \frac{1}{2} r_{n-1}(1 - S_m R_{n-1})$$

Experimenting with this model reveals some interesting observations. Using the maple package, we can use the formula, `XalleleFitnessTimeSeriesPlot(Q0,R0,N,Sf,Sm)`.

If we take $S_f = 0$ and $S_m = 0.01$, a small negative effect on fitness of the male allele affects the transmission of the allele to the females of the next generation, and indirectly to the males in the $n + 2$ generation.

- `XalleleFitnessTimeSeriesPlot(0.33, 0.74, 60, 0, 0.01);`

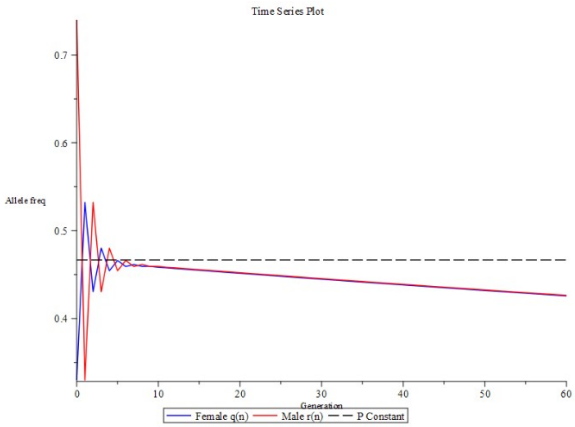

Figure 18: X-Linked Fitness Model, $S_f = 0$ and $S_m = 0.01$

- Similarly, we can plot `XalleleFitnessTimeSeriesPlot(0.33, 0.74, 60, 0.01, 0);`, where now, there is a small negative effect on fitness of females.

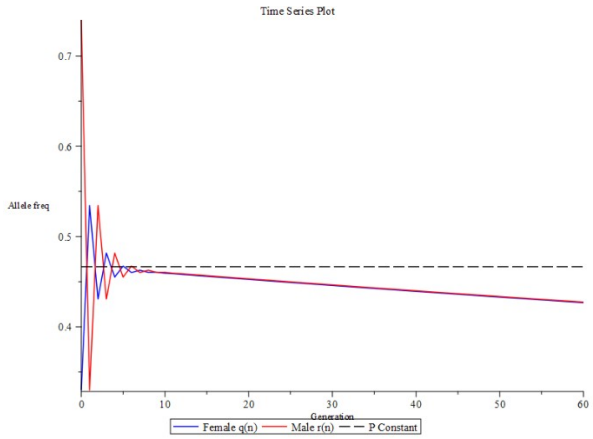

Figure 19: X-Linked Fitness Model, $S_f = 0.01$ and $S_m = 0$

There are several interesting observations about this model. First, it follows suit with the original X-linked allele model where in only a few generations, q_n and r_n converge to one another. In this case, they do not converge to a stable equilibrium but decrease linearly till the allele frequencies are 0.

Interestingly, regardless of whether fitness is imposed on females or males, we see that the behaviors converge. The same behavior can be graphed with positive fitness. In this case, due to enhanced fitness, long term behavior would show that the X-linked allele dominates over other X-linked alleles at that loci in a population. We would observe a linear equilibrium till the allele frequencies in males and females are 1.

5.5 Epistasis: Combining The Hardy-Weinberg and X-allele Transformations

Epistasis is the interaction between two different genes at independent loci. For example, a disease caused by having a specific allele on an X, sex chromosome, may depend on the genotype at another loci. In this section, we will explore the application of both the general Hardy-Weinberg transformation for autosomal genotypes and the frequency of an X-allele through the concept of epistasis.

For this model, we assume that carrying a specific X-linked allele predisposes an individual to a disease, but that the disease is expressed only when this allele is present in combination with an 'aa' genotype at another autosomal locus.

To build this combined model and plot the disease in both females or males, we note the following details. In a population where w represents the frequency of the 'aa' genotype and qn and rn represent the frequency of an X-linked allele in females and males respectively:

- Frequency of females in a population expressing the disease $\rightarrow D_F = wq_n$

- Frequency of males in a population expressing the disease $\rightarrow D_M = wr_n$

In the Hardy-Weinberg model, the frequency of individuals with the 'aa' genotype, denoted by w, stabilizes after just one generation. So if we start with some $w_0, 0 < w_0 < 1$, we will see that $w_1 = w_{n>1}$. In the X-linked allele model, the frequency of both males and females carrying an allele stabilizes to the P constant, $P = 2/3q_n + 1/3r_n = 2/3q_0 + 1/3r_0$.

This means that epistasis should look quite similar to the X-linked allele model. More importantly, after just one iteration of the Hardy-Weinberg transformation, a new equilibrium constant becomes clear. That is, the frequency of the disease in both males and females tends to a constant E:

$$E = w_1 P$$

To test this idea, we use a procedure in our maple package DiseaseTrajectoryPlot(u_0, v_0, q_0, r_0, N). This procedure uses the procedure XalleleSIM(q_0, r_0,N) from our maple package and the HW(u,v) procedure from DMB.txt to calculate and plot the points (N,D_F) and (N,D_M), where N is the generation starting at the initial conditions, N = 0.

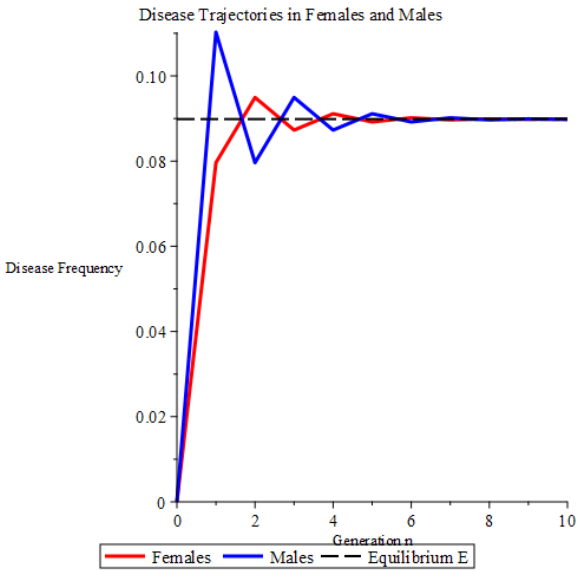

Figure 20: DiseaseTrajectoryPlot(0.3,0.7,0.9,0.4,10);

Since the population started at 'AA' (u_0) = 0.3 and 'Aa' (v_0) = 0.7, 'aa' (w_0) = 0. So, even though a large percent of both males and females carry the X-linked diseased allele, the initial disease frequency is 0 for both males and females. In the second generation, 'aa' reappears stability to a frequency of 0.1225. So now, we expect some males and females to express the disease. In the X-linked allele model, since $r_n = q_n - 1$, we notice the largest disease spike will be in the first generation as $r_1 = q_0 = 0.9$ and $D_M = (0.9)(0.1225) = 0.11025$. In the next generations, as both X-allele frequency in males and females approaches P = (2/3)(0.9) + (1/3)(0.4) and w_n, $n > 1$, is 0.1225, the frequency of the disease tends to E = 0.08085.

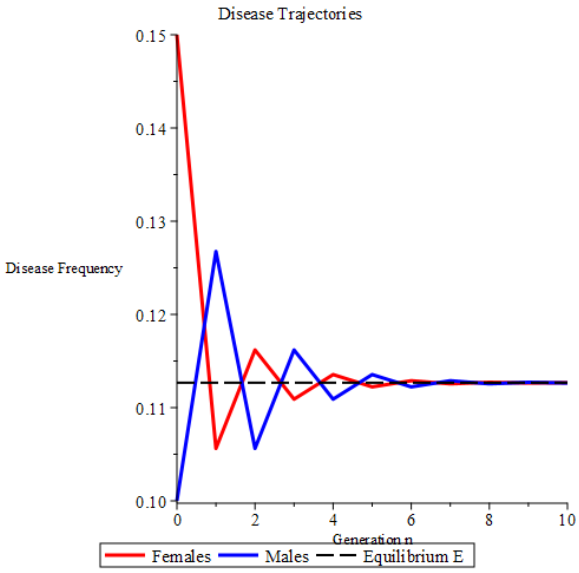

Figure 21: DiseaseTrajectoryPlot(0.2, 0.3, 0.3, 0.2, 10);

In this experimental example, we can also note another interesting observation. While in the X-linked allele model, male frequency always copied that of the female frequency in the previous generation, in the combined epistasis model, that phenomenon only occurs at the second generation and on. It is also interesting to note that similar to the X-linked allele model, equilibrium is reached very quickly.

References

[1] William Beavis et al. *Chapter 1: Gene Frequencies*. Published: 2023-10-18. 2023.

[2] *Evolution - the Theory of Natural Selection (Part 1)*. Accessed: 2025-12-11. 2025.

[3] Philip W. Hedrick. *Genetics of Populations*. 4th ed. Chapters on migration and gene flow discuss the standard discrete-generation migration model and equilibrium allele frequencies. Sudbury, MA: Jones and Bartlett Publishers, 2011.

[4] M. Traykov and Iv. Trenchev. "Mathematical Models in Genetics". In: *Russian Journal of Genetics* 52.9 (2016), pp. 958–992. DOI: 10.7868/s0016675816080130. URL: https://doi.org/10.7868/s0016675816080130.