Model of Frequency of Phenotype and its Equilibrium

Meixin Liu

February 26, 2017

Abstract

Introduction: Hardy-Weinberg law determine in what condition that the allele frequencies will not change in a give population over time as individuals within that population mate and reproduce. According to Karlin (1972), there is a number of factors can affect the path of evolution. Karlin (1972) states "The three most familiar are mutation, migration and selection. The first two are self-explanatory. For selection, it can be perform through variation in viability and fertility" (p.702). In class, we analyze the model in the case that different genotypes have different finesses. And we found three equilibrium of that model. This paper will be mainly concerned about the model of allele frequencies in the case that there is variation in the number of progeny from parents of different genotypes. Find the equilibrium of this model and compare it to the equilibrium we got in class.

Hardy-Weinberg Law. *Assume in a parent population, a particular gene has two alleles A and a, and the initial proportion of allele A is p_0 and the initial proportion of allele a is q_0. In addition, assume (i) mating is random, (ii) there is no variation in the number of progeny from parents of different genotypes, (iii) all genotypes have equal survival probability, (iv) there is no immigration nor emigration, (v) there are no mutations and (vi) generations are nonoverlapping. Then, in generation t, the allele frequencies do not change, $p_t = p_0$. In addition, the genotypic frequencies do not change from the second generation onwards, $p_{AA} = p_0^2$, $p_{Aa} = 2p_0q_0$, and $p_{aa} = q_0^2$.*

According to the Hardy-Weinberg law, the recessive trait will not die out but remain in the population at a fixed proportion. If the assumptions in Hardy-Weinberg law are violated, then the proportions change(Allen, 2007, p.103). Assume that population with non-overlapping generation is randomly mixed. The population has 2 allele types A and a, and the females of genotypes AA, Aa, and aa produce r_{AA}, r_{Aa} and r_{aa} number of eggs, respectively. The survival probability of individuals with each genotype.

To determine the Model of the fractions of individuals with each genotype. The following assumption are needed. Let N be the total population size. F be the total female amount and M be the total male amount. And define F_{AA} be the number of female with AA genotype, M_{AA} be the number of male with AA genotype, and N_{AA} be the number of total population with AA genotype, and Similar definition for F_{Aa}, F_{aa}, M_{Aa}, M_{aa}, N_{Aa}, and N_{aa}. And we assume female and male has a constant fraction f, that is $F/M = f$. Then, we can define

$$\begin{aligned}
\frac{F_{AA}}{F} &= \frac{M_{AA}}{M} = \frac{N_{AA}}{N} = p_{AA} \\
\frac{F_{Aa}}{F} &= \frac{M_{Aa}}{M} = \frac{N_{Aa}}{N} = p_{Aa} \\
\frac{F_{aa}}{F} &= \frac{M_{aa}}{M} = \frac{N_{aa}}{N} = p_{aa}
\end{aligned}$$

Thus, in the t-generation the frequency of A alleles is

$$p_t = p_{AA} + \frac{p_{Aa}}{2}$$

Since $p_{AA} + p_{Aa} + p_{aa} = 1$, we can get

$$p_t = \frac{1}{2} + \frac{p_{AA}}{2} - \frac{p_{aa}}{2}$$

To determine what happens after one generation of mating, it is necessary to consider all possible matings, their frequency, and all possible offspring and their frequency.

Mating	Offspring Fraction
F_{AA} x M_{AA}	AA
F_{AA} x M_{Aa}	$\frac{1}{2}AA$ $\frac{1}{2}Aa$
F_{AA} x M_{aa}	Aa
F_{Aa} x M_{AA}	$\frac{1}{2}AA$ $\frac{1}{2}Aa$
F_{Aa} x M_{Aa}	$\frac{1}{4}AA$ $\frac{1}{2}Aa$ $\frac{1}{4}aa$
F_{Aa} x M_{aa}	$\frac{1}{2}Aa$ $\frac{1}{2}aa$
F_{aa} x M_{AA}	Aa
F_{aa} x M_{Aa}	$\frac{1}{2}Aa$ $\frac{1}{2}aa$
F_{aa} x M_{aa}	aa

Table 1: Mating and offspring table.

Let $N_{AA}(t+1)$, $N_{Aa}(t+1)$, and $N_{aa}(t+1)$ denote the number of offspring of each genotype in the next generation. $p_{AA}(t+1)$, $p_{Aa}(t+1)$, and $p_{aa}(t+1)$ denote the genotype frequencies in the next generation. Then, applying the result from Table 1,

$$\begin{aligned}
N_{AA}(t+1) &= \{r_{AA}F_{AA}(p_{AA} + \frac{1}{2}p_{Aa}) + r_{Aa}F_{Aa}(\frac{1}{2}p_{AA} + \frac{1}{4}p_{Aa})\}\frac{1}{1+aN_t} \\
N_{aa}(t+1) &= \{r_{aa}F_{aa}(p_{aa} + \frac{1}{2}p_{Aa}) + r_{Aa}F_{Aa}(\frac{1}{2}p_{aa} + \frac{1}{4}p_{Aa})\}\frac{1}{1+aN_t} \\
N_{t+1} &= (r_{AA}F_{AA} + r_{Aa}F_{Aa} + r_{aa}F_{aa})\frac{1}{1+aN_t}
\end{aligned}$$

And we divide $N_{AA}(t+1)$ and $N_{aa}(t+1)$ by N_{t+1}, we can get the genotype frequencies in t+1-generation,

$$\begin{aligned}
p_{AA}(t+1) = \frac{N_{AA}(t+1)}{N_{t+1}} &= \frac{\{r_{AA}F_{AA}(p_{AA}+\frac{1}{2}p_{Aa})+r_{Aa}F_{Aa}(\frac{1}{2}p_{AA}+\frac{1}{4}p_{Aa})\}\frac{1}{1+aN_t}}{(r_{AA}F_{AA}+r_{Aa}F_{Aa}+r_{aa}F_{aa})\frac{1}{1+aN_t}} \\
&= \frac{r_{AA}F_{AA}(p_{AA}+\frac{1}{2}p_{Aa})+r_{Aa}F_{Aa}(\frac{1}{2}p_{AA}+\frac{1}{4}p_{Aa})}{r_{AA}F_{AA}+r_{Aa}F_{Aa}+r_{aa}F_{aa}} \\
&= \frac{\{r_{AA}F_{AA}(p_{AA}+\frac{1}{2}p_{Aa})+r_{Aa}F_{Aa}(\frac{1}{2}p_{AA}+\frac{1}{4}p_{Aa})\}/F}{(r_{AA}F_{AA}+r_{Aa}F_{Aa}+r_{aa}F_{aa})/F} \\
&= \frac{r_{AA}\frac{F_{AA}}{F}(p_{AA}+\frac{1}{2}p_{Aa})+r_{Aa}\frac{F_{Aa}}{F}(\frac{1}{2}p_{AA}+\frac{1}{4}p_{Aa})}{r_{AA}\frac{F_{AA}}{F}+r_{Aa}\frac{F_{Aa}}{F}+r_{aa}\frac{F_{aa}}{F}} \\
&= \frac{r_{AA}p_{AA}(p_{AA}+\frac{1}{2}p_{Aa})+r_{Aa}p_{Aa}(\frac{1}{2}p_{AA}+\frac{1}{4}p_{Aa})}{r_{AA}p_{AA}+r_{Aa}p_{Aa}+r_{aa}p_{aa}} \\
&= \frac{r_{AA}p_{AA}^2+\frac{1}{2}p_{Aa}p_{AA}(r_{Aa}+r_{AA})+\frac{1}{4}r_{Aa}p_{Aa}^2}{r_{AA}p_{AA}+r_{Aa}p_{Aa}+r_{aa}p_{aa}} \\
&= \frac{(2r_{AA}p_{AA}+r_{Aa}p_{Aa})(2p_{AA}+p_{Aa})}{4(r_{AA}p_{AA}+r_{Aa}p_{Aa}+r_{aa}p_{aa})}
\end{aligned}$$

and we know $p_{Aa} = 1 - p_{AA} - p_{aa}$, substitute this equation. Then,

$$\begin{aligned}
p_{AA}(t+1) &= \frac{(2r_{AA}p_{AA}+r_{Aa}(1-p_{AA}-p_{aa}))(2p_{AA}+1-p_{AA}-p_{aa})}{4(r_{AA}p_{AA}+r_{Aa}(1-p_{AA}-p_{aa})+r_{aa}p_{aa})} \\
&= \frac{\{(2r_{AA}-r_{Aa})p_{AA}+r_{Aa}-r_{Aa}p_{aa}\}(p_{AA}-p_{aa}+1)}{4\{(r_{AA}-r_{Aa})p_{AA}+(r_{aa}-r_{Aa})p_{aa}+r_{Aa}\}}
\end{aligned}$$

Similarly, we can get $p_{aa}(t+1)$,

$$p_{aa}(t+1) = \frac{\{(2r_{aa}-r_{Aa})p_{aa}+r_{Aa}-r_{Aa}p_{AA}\}(p_{aa}-p_{AA}+1)}{4\{(r_{AA}-r_{Aa})p_{AA}+(r_{aa}-r_{Aa})p_{aa}+r_{Aa}\}}$$

These results can be used to find the allele frequencies in the t+1-generation, p_{t+1}. And to simplify the equation we introduce two variables, define $\frac{r_{AA}}{r_{Aa}} = 1-w$ and $\frac{r_{aa}}{r_{Aa}} = 1-v$. Then,

$$\begin{aligned}
p_{t+1} &= \frac{1}{2} + \frac{p_{AA}(t+1)}{2} - \frac{p_{aa}(t+1)}{2} \\
&= \frac{1}{2} + \frac{\{(2r_{AA}-r_{Aa})p_{AA}+r_{Aa}-r_{Aa}p_{aa}\}(p_{AA}-p_{aa}+1)}{8\{(r_{AA}-r_{Aa})p_{AA}+(r_{aa}-r_{Aa})p_{aa}+r_{Aa}\}} \\
&\quad - \frac{\{(2r_{aa}-r_{Aa})p_{aa}+r_{Aa}-r_{Aa}p_{AA}\}(p_{aa}-p_{AA}+1)}{8\{(r_{AA}-r_{Aa})p_{AA}+(r_{aa}-r_{Aa})p_{aa}+r_{Aa}\}} \\
&= \frac{1}{2} + \frac{\{(2(1-w)-1)p_{AA}+1-p_{aa}\}(p_{AA}-p_{aa}+1)}{8\{(1-w-1)p_{AA}+(1-v-1)p_{aa}+1\}} \\
&\quad - \frac{\{(2(1-v)-1)p_{aa}+1-p_{AA}\}(p_{aa}-p_{AA}+1)}{8\{(1-w-1)p_{AA}+(1-v-1)p_{aa}+1\}} \\
&= \frac{1}{2} + \frac{\{(1-2w)p_{AA}+1-p_{aa}\}(p_{AA}-p_{aa}+1)-\{(1-2v)p_{aa}+1-p_{AA}\}(p_{aa}-p_{AA}+1)}{8(-wp_{AA}-vp_{aa}+1)} \\
&= \frac{1}{2} + \frac{(1-wp_{AA}-vp_{aa})(p_{AA}-p_{aa})+(1-w)p_{AA}+(v-1)p_{aa}}{4(-wp_{AA}-vp_{aa}+1)}
\end{aligned}$$

We want to determine the equilibrium point, let $p_t = p_{t+1}$,

$$\begin{aligned}
p_t &= \frac{1}{2} + \frac{p_{AA}}{2} - \frac{p_{aa}}{2} \\
p_{t+1} &= \frac{1}{2} + \frac{(1 - wp_{AA} - vp_{aa})(p_{AA} - p_{aa}) + (1-w)p_{AA} + (v-1)p_{aa}}{4(-wp_{AA} - vp_{aa} + 1)} \\
p_t &= p_{t+1} \\
&\Downarrow \\
p_{AA} - p_{aa} &= \frac{(1 - wp_{AA} - vp_{aa})(p_{AA} - p_{aa}) + (1-w)p_{AA} + (v-1)p_{aa}}{2(-wp_{AA} - vp_{aa} + 1)} \\
\frac{p_{AA} - p_{aa}}{2} &= \frac{(1-w)p_{AA} + (v-1)p_{aa}}{2(1 - wp_{AA} - vp_{aa})} \\
p_{AA} - p_{aa} &= \frac{(1-w)p_{AA} + (v-1)p_{aa}}{1 - wp_{AA} - vp_{aa}} \\
wp_{AA} - vp_{aa} &= (p_{AA} - p_{aa})(wp_{AA} + vp_{aa})
\end{aligned}$$

The equality is always satisfied when $w = v = 0$, that is

$$\left.\begin{aligned} \frac{r_{AA}}{r_{Aa}} &= 1 \\ \frac{r_{aa}}{r_{Aa}} &= 1 \end{aligned}\right\} \Rightarrow r_{AA} = r_{Aa} = r_{aa}$$

$r_{AA} = r_{Aa} = r_{aa}$ means that there is no variation in the number of progeny from parents of different genotypes. When this condition holds, all assumption of Hardy-Weinberg law are satisfied in this model. The equality $p_{t+1} = p_t$ is always satisfied, which means frequencies remain constant from generation to generation. Hence, $p_t = p_0$. The Hardy-Weinberg law has been verified.

What if $r_{AA} = r_{Aa} = r_{aa}$ does not hold, is there still an equilibrium for the model? In class we consider a violation of assumption (iii) in Hardy-Weinberg law. In that case, different genotypes have different fitnesses. We define ω_{AA}, ω_{Aa} and ω_{aa} be the survival rates of genotypes AA, Aa and aa. Let $\frac{\omega_{AA}}{\omega_{Aa}} = 1 - s$ and $\frac{\omega_{aa}}{\omega_{Aa}} = 1 - r$, and we find three equilibrium 0, 1 and $p^* = \frac{r}{r+s}$.

For our model we assume a violation of assumption (ii) in Hardy-Weinberg law. And we define the different number of progeny from parents of AA, Aa, and aa, r_{AA}, r_{Aa} and r_{aa} respectively. Let $\frac{r_{AA}}{r_{Aa}} = 1 - w$ and $\frac{r_{aa}}{r_{Aa}} = 1 - v$. Is 0 and 1 is an equilibrium for this model? Let $p_t = p_{t+1} = 0$, then

$p_{AA} = 0$ and $p_{aa} = 1$, plug into the equility $wp_{AA} - vp_{aa} = (p_{AA} - p_{aa})(wp_{AA} + vp_{aa})$. Then,

$$LHS = 0 \cdot w - 1 \cdot v = -v$$

$$RHS = (0 - 1) \cdot (w \cdot 0 + v \cdot 1) = -v$$

$$LHS = RHS$$

Let $p_t = p_{t+1} = 1$, then $p_{AA} = 1$ and $p_{aa} = 0$, plug into the equility $wp_{AA} - vp_{aa} = (p_{AA} - p_{aa})(wp_{AA} + vp_{aa})$. Then,

$$LHS = 1 \cdot w - 0 \cdot v = w$$

$$RHS = (1 - 0) \cdot (w \cdot 1 + v \cdot 0) = w$$

$$LHS = RHS$$

Hence, both 0 and 1 are equilibrium of the model. Is $\frac{w}{w+v}$ or $\frac{v}{w+v}$ an equilibrium for our model? Let $p_t = p_{t+1} = \frac{w}{w+v}$, then $p_{AA} = (\frac{w}{w+v})^2$ and $p_{aa} = (1 - \frac{w}{w+v})^2$, plug into the equility $wp_{AA} - vp_{aa} = (p_{AA} - p_{aa})(wp_{AA} + vp_{aa})$. We get,

$$LHS = w \cdot (\frac{w}{w+v})^2 - v \cdot (1 - \frac{w}{w+v})^2 = \frac{w^3 - v^3}{(w+v)^2} = \frac{(w-v)(w^2 + wv + v^2)}{(w+v)^2}$$

$$\begin{aligned} RHS &= \{(\frac{w}{w+v})^2 - (1 - \frac{w}{w+v})^2\}(w \cdot (\frac{w}{w+v})^2 + v \cdot (1 - \frac{w}{w+v})^2) \\ &= \frac{w^2 - v^2}{(w+v)^2} \cdot \frac{w^3 + v^3}{(w+v)^2} = \frac{(w-v)(w^2 - wv + v^2)}{(w+v)^2} \end{aligned}$$

$$LHS \neq RHS$$

Thus, $p^* = \frac{w}{w+v}$ is not an equilibrium of the model.

Let $p_t = p_{t+1} = \frac{v}{w+v}$, then $p_{AA} = (\frac{v}{w+v})^2$ and $p_{aa} = (1 - \frac{v}{w+v})^2$, plug into the equility. We get,

$$LHS = w \cdot (\frac{v}{w+v})^2 - v \cdot (1 - \frac{v}{w+v})^2 = \frac{wv(v-w)}{(w+v)^2}$$

$$\begin{aligned} RHS &= \{(\frac{v}{w+v})^2 - (1 - \frac{v}{w+v})^2\}(w \cdot (\frac{v}{w+v})^2 + v \cdot (1 - \frac{v}{w+v})^2) \\ &= \frac{(v+w)(v-w)}{(w+v)^2} \cdot \frac{vw(v+w)}{(w+v)^2} = \frac{vw(v-w)}{(w+v)^2} \end{aligned}$$

$$LHS = RHS$$

Thus, $p^* = \frac{v}{w+v}$ is an equilibrium of the model.

Conclusion The model of preceding paper are examined the allele frequencies under the condition that different genotypes have different fertility. There is three equilibria for the model 0, 1 and $\frac{v}{w+v}$. And those equilibria are very similar to what we get in class under the condition that different genotypes have different fitness. And for this model the allele frequencies changes unless $r_{AA} = r_{Aa} = r_{aa}$ holds. That is, if all conditions of Hardy-Weinberg Law are satisfied, the allele frequencies do not change. Thus, the Hardy-Weinberg Law has been verified.

References

[1] Allen, L.(2007). An *Introduction to Mathematical Biology.* Upper saddle River, NJ:Pearson Prentice Hall.

[2] Karlin, S.(1972, September). Some Mathematical Models of Population Genetics. *The American Mathematical Monthly, 79*(7), 699-739.