Evolution Module

6.2 Selection (Revised)

Bob Gardner and Lev Yampolski

Integrative Biology and Statistics (BIOL 1810)

Fall 2007

FITNESS VALUES

Note. We start our quantitative exploration of selection by again studying a one locus/two alleles model. Suppose the two alleles are A_1 and A_2 and so (as usual, assuming diploid organisms) the possible genotypes are A_1A_1, A_1A_2, and A_2A_2. We start by assuming random mating (along with the other assumptions of Hardy-Weinberg, except we will have selection this time) and that zygotes are conceived at the proportions expected by the Hardy-Weinberg principle. That is, the ratio of zygote genotypes $A_1A_1 : A_1A_2 : A_2A_2$ is $p^2 : 2pq : q^2$ where p is the frequency of A_1 and $q = 1 - p$ is the frequency of A_2.

Note. We now suppose that selection plays a role and that the conceived zygotes are not equally successful in reaching a reproductive state (perhaps they do not survive to reproductive age or are sterile). To set up the computations, suppose the zygotes reach a reproductive state in the genotypic ratio $A_1A_1 : A_1A_2 : A_2A_2$ of $w_{11}p^2 : 2w_{12}pq : w_{22}q^2$. That is, we define the relative fitness values for each genotype as

Genotype	A_1A_1	A_1A_2	A_2A_2
Relative Fitness	w_{11}	w_{12}	w_{22}

In Hardy-Weinberg equilibrium, we know that the allele frequencies and genotypic frequencies remain constant. However, with the presence of selection, we expect these parameters to change with time.

Example. Suppose a population of 100 has genotypic ratio $A_1A_1 : A_1A_2 : A_2A_2$ of $3 : 5 : 2$. What are the number of individuals of each genotype.

Solution. First, sum $3 + 5 + 2 = 10$. Next, divide the population size by this sum: $100/10 = 10$. Now multiply each number in the ratio by this quotient: $10 \times 3 : 10 \times 5 : 10 \times 2$, or $30 : 50 : 20$. So the number of A_1A_1 individuals is 30, the number of A_1A_2 is 50, and the number of A_2A_2 is 20.

Note. Since we are starting with a population in which the frequency of A_1 is p and the frequency of A_2 is q, denote the frequencies of these alleles in the next generation as p' and q', respectively (we are assuming discrete generations). To calculate p' and q', we find the ratio $A_1 : A_2$ in gametes. Now an A_1A_1 individual produces 100% of its gametes with the A_1 allele, A_1A_2 individuals produce 50% of their gametes with the A_1 allele, and A_2A_2 individuals produce 0% of their gametes with the A_1 allele. Similar percentages can be calculated for gametes containing the A_2 allele. As a result, the ratio $A_1 : A_2$ in gametes which produce the next generation is

$$(w_{11}p^2 \times 100\% + 2w_{12}pq \times 50\% + w_{22}q^2 \times 0\%) : (w_{11}p^2 \times 0\% + 2w_{12}pq \times 50\% + w_{22}q^2 \times 100\%),$$

which simplifies to $(w_{11}p^2 + w_{12}pq) : (w_{12}pq + w_{22}q^2)$. This ratio is the same as $p' : q'$, and so

$$\frac{p'}{q'} = \frac{w_{11}p^2 + w_{12}pq}{w_{12}pq + w_{22}q^2} \text{ or } p' = \frac{w_{11}p^2 + w_{12}pq}{w_{12}pq + w_{22}q^2}q'.$$

Now $p' + q' = 1$, so

$$\begin{aligned} p' + q' = \left(\frac{w_{11}p^2 + w_{12}pq}{w_{12}pq + w_{22}q^2}q'\right) + q' &= 1 \\ q'\left(\frac{w_{11}p^2 + w_{12}pq}{w_{12}pq + w_{22}q^2} + 1\right) &= 1 \\ q'\left(\frac{w_{11}p^2 + w_{12}pq}{w_{12}pq + w_{22}q^2} + \frac{w_{12}pq + w_{22}q^2}{w_{12}pq + w_{22}q^2}\right) &= 1 \\ q'\left(\frac{w_{11}p^2 + 2w_{12}pq + w_{22}q^2}{w_{12}pq + w_{22}q^2}\right) &= 1 \end{aligned}$$

or

$$q' = \frac{w_{12}pq + w_{22}q^2}{w_{11}p^2 + 2w_{12}pq + w_{22}q^2}.$$

Next,

$$\begin{aligned} p' = 1 - q' &= 1 - \frac{w_{12}pq + w_{22}q^2}{w_{11}p^2 + 2w_{12}pq + w_{22}q^2} \\ &= \frac{w_{11}p^2 + 2w_{12}pq + w_{22}q^2}{w_{11}p^2 + 2w_{12}pq + w_{22}q^2} - \frac{w_{12}pq + w_{22}q^2}{w_{11}p^2 + 2w_{12}pq + w_{22}q^2} \\ &= \frac{w_{11}p^2 + w_{12}pq}{w_{11}p^2 + 2w_{12}pq + w_{22}q^2}. \end{aligned}$$

Define the *average fitness* of the population as

$$\overline{w} = w_{11}p^2 + 2w_{12}pq + w_{22}q^2.$$

Then we have

$$p' = \frac{w_{11}p^2 + w_{12}pq}{\overline{w}} \text{ and } q' = \frac{w_{12}pq + w_{22}q^2}{\overline{w}}.$$

CHANGES IN ALLELE FREQUENCIES

Note. Now, lets see how selection has affected the frequency of A_1. Computing the change in p over one generation of selection, we get

$$\begin{aligned}
p' - p &= \frac{w_{11}p^2 + w_{12}pq}{\overline{w}} - p = p\left(\frac{w_{11}p + w_{12}q}{\overline{w}} - 1\right) \\
&= p\left(\frac{w_{11}p + w_{12}q - (w_{11}p^2 + 2w_{12}pq + w_{22}q^2)}{\overline{w}}\right) \\
&= \frac{p}{\overline{w}}\left(w_{11}(p - p^2) + w_{12}(q - 2pq) - w_{22}q^2\right) \\
&= \frac{p}{\overline{w}}\left(w_{11}p(1 - p) + w_{12}q(1 - 2p) - w_{22}q^2\right) \\
&= \frac{p}{\overline{w}}\left(w_{11}pq + w_{12}q(q - p) - w_{22}q^2\right) \\
&= \frac{pq}{\overline{w}}\left(w_{11}p + w_{12}(q - p) - w_{22}q\right) \\
&= \frac{pq}{\overline{w}}\left(p(w_{11} - w_{12}) + q(w_{12} - w_{22})\right).
\end{aligned}$$

If we represent the change in p as Δp, where Δ is the Greek letter delta, then we have

$$\Delta p = \frac{pq}{\overline{w}}\left(p(w_{11} - w_{12}) + q(w_{12} - w_{22})\right.$$

Note. We now turn our attention to the long term behavior of the value of p. In particular, we are interested in *equilibrium points* at which $\Delta p = 0$. First, observe that if $p = 0$ or $q = 0$, then $\Delta p = 0$. The biological interpretations of these two equilibria are (1) 100% of the alleles are A_2 (when $p = 0$), and (2) 100% of the alleles are A_1 (when $q = 0$). In each case, the locus is at *fixation* and genetic diversity has been lost. We can also have an equilibrium when $p(w_{qq} - w_{12}) + q(w_{12} - w_{22}) = 0$, or equivalently when

$$p(w_{11} - w_{12}) + (1 - p)(w_{12} - w_{22}) = 0. \qquad (*)$$

Note. Denote the value of p in this equation as p_{eq}. Then $p_{eq} = \dfrac{w_{22} - w_{12}}{w_{11} - 2w_{12} + w_{22}}$, if $w_{11} - 2w_{12} + w_{22} \neq 0$. In order for p_{eq} to make sense, we need (in interval notation) $p_{eq} \in [0, 1]$ (or $0 \leq p_{eq} \leq 1$). We analyze this in three cases:

Case 1. Suppose $w_{11} - 2w_{12} + w_{22} = 0$. From equation $(*)$ we see that

$$\begin{aligned} p_{eq}(w_{11} - w_{12}) + (1 - p_{eq})(w_{12} - w_{22}) &= 0 \\ p_{eq}(w_{11} - 2w_{12} - w_{22}) + w_{12} - w_{22} &= 0 \\ p_{eq}(0) + w_{12} - w_{22} &= 0 \\ w_{12} &= w_{22}. \end{aligned}$$

Now substituting $w_{12} = w_{22}$ into $w_{11} - 2w_{12} + w_{22} = 0$ gives $w_{11} - 2w_{12} + w_{12} = w_{11} - w_{12} = 0$ or $w_{11} = w_{12}$. So $w_{11} = w_{12} = w_{22}$ and there is no selective difference between genotypes A_1A_1, A_1A_2, and A_2A_2. Strictly speaking, *every* value of $p \in [0, 1]$ is an equilibrium point and the population remains in Hardy-Weinberg equilibrium.

Case 2. Suppose $0 \leq p_{eq} = \dfrac{w_{22} - w_{12}}{w_{11} - 2w_{12} + w_{22}} \leq 1$ and $w_{11} - 2w_{12} + w_{22} > 0$. Then

$$\begin{aligned} 0 &< w_{22} - w_{12} < w_{11} - 2w_{12} + w_{22} \\ w_{12} &< w_{22} < w_{11} - w_{12} + w_{22} \end{aligned}$$

or $w_{12} < w_{22}$ and $0 < w_{11} - w_{12}$, or $w_{12} < w_{22}$ and $w_{12} < w_{11}$.

So the fitness of the heterozygote A_1A_2 is less than the fitness of either of the homozygotes, A_1A_1 and A_2A_2. This is called the *deleterious heterozygote model.*

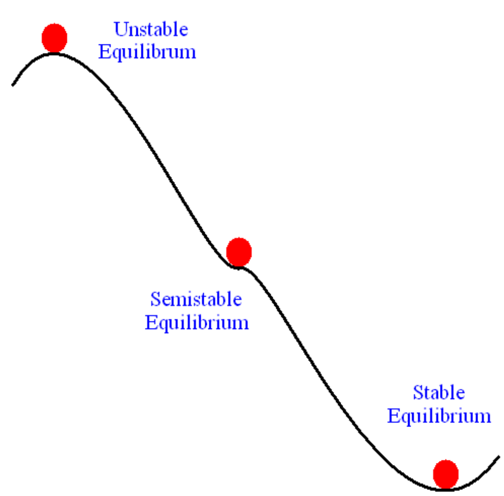
Case 3. Suppose $0 \leq p_{eq} = \dfrac{w_{22} - w_{12}}{w_{11} - 2w_{12} + w_{22}} \leq 1$ and $w_{11} - 2w_{12} + w_{22} < 0$. Then

$$\begin{aligned} 0 &\geq w_{22} - w_{12} \geq w_{11} - 2w_{12} + w_{22} \\ w_{12} &\geq w_{22} \geq w_{11} - w_{12} + w_{22} \end{aligned}$$

or $w_{12} \geq w_{22}$ and $0 \geq w_{11} - w_{12}$, or $w_{12} \geq w_{22}$ and $w_{12} \geq w_{11}$. So the fitness of the heterozygote is greater than that of either homozygote. This is called the *heterozygote advantage model.*

STABILITY

Note. In any physical system that involves change with time (in the most general case, called a *dynamical system*), two properties are of interest. The first, as explored above, is the equilibria—the states of the system which do not change with time. The second property, called *stability*, is a bit more subtle. We discuss it through analogy. Imagine a small ball which rolls around on a smooth surface. If the ball is sitting at the bottom of a trough or well, then the ball is said to be in a *stable equilibrium*. If the ball is disturbed by a small amount, then the ball will return to the equilibrium. If the ball is hit hard it can be knocked away from the bottom of the well and may go to some other equilibrium state. However, the idea of stability only involves "small" perturbations. Next, imagine the ball precariously perched at the top of a hill. In every direction, the hill slopes down, so if the ball is nudged slightly then it will not return to the equilibrium at the top of the hill, but will roll away to some other state. For this reason, such an equilibrium is called an *unstable equilibrium*. Finally, imagine a hillside which slopes downhill from the left to the right, but at one point the hill levels off (the profile of such a hill might look like the graph of the function $y = -x^3$ which is "flat" right at $x = 0$ and is sloping downhill elsewhere). In this case, if the ball is sitting right on the point where the hill is level, then the ball is in equilibrium. The behavior of the ball when perturbed depends on the direction of the perturbation. If the ball is pushed slightly downhill, then the ball will roll away (like an unstable equilibrium). If the ball is pushed slightly uphill, then the ball will roll back downhill towards the equilibrium (like a stable equilibrium). For these reasons, such an equilibrium is said to be a *semi stable equilibrium.* The following picture illustrates these three cases:



Now this is a description of the idea of stability in a very informal way. In the case of a semi stable equilibrium, you might think that displacing the ball slightly uphill will cause it to roll back towards the equilibrium, but it will gain momentum and roll right past the semi stable equilibrium point. Similarly, perturbing the ball at the stable equilibrium will cause it to roll back to the stable equilibrium point, but momentum should carry it past the equilibrium, slightly uphill where it will again roll back towards the stable equilibrium—in fact, we might think of the ball as oscillating back and forth at the bottom of the well. While we could explore this behavior in more detail, let's view these "momentum" ideas simply as artifacts of our model. Perhaps you can think of the surface as kind of "sticky" so that the ball slows down as it approaches an equilibrium and will slowly glide back to the equilibrium in the case of a stable equilibrium (or in the case of a semi stable equilibrium from one side). In continuous dynamical systems, this is the type of behavior displayed. In fact, in such systems, the ball will move slower and slower as it approaches an equilibrium and will never reach the equilibrium (in finite time) but will asymptotically approach the equilibrium—the equilibrium will be the *limit* of the system.

Note. The reason to study stability of an equilibrium is because nature is more complicated and imprecise than our models! We would expect any natural system to eventually experience some type of disturbance, and so it is highly improbable for a system to stay in an unstable or semi stable equilibrium state. The long term behavior of a natural system should be to go to some stable equilibrium state.

EXAMPLES

Example. Consider a value x which changes periodically according to the equation $\Delta x = (x-5)(x-10)^2(15-x)$. Find the equilibria and classify their stability.

Solution. To find equilibria, set $\Delta x = 0$: $\Delta x = (x-5)(x-10)^2(15-x) = 0$. So the equilibrium points are $x = 5$, $x = 10$, and $x = 15$. To analyze the stability of each equilibria, we need to know how x changes "near" the equilibria. That is, we need to know where x is increasing and where it is decreasing. This information is contained in the sign of Δx. Now Δx is either positive, zero, or negative. We have found the three points where it is 0, so this breaks the real number line into four pieces (or intervals) and the sign of Δx is the same throughout each of these pieces (since Δx is a continuous function of x—this will be spelled out in more detail later in the form of the Intermediate Value Theorem). So we choose a "test value" from each of the four pieces (intervals) and the sign of Δx at the test value gives the sign of Δx throughout the interval. We use the standard interval notation (a, b) to mean all numbers x such that $a < x < b$. Consider:

interval	$(-\infty, 5)$	$(5, 10)$	$(10, 15)$	$(15, \infty)$
test value k	0	6	11	16
Δx at $x = k$	$(-5)(-10)^2(15) = -7500$	$(1)(-4)^2(9) = 144$	$(6)(1)^2(4) = 24$	$(11)(6)^2(-2) = -792$
sign of Δx	$-$	$+$	$+$	$-$
behavior of x	decreasing $\longleftarrow$	increasing $\longrightarrow$	increasing $\longrightarrow$	decreasing $\longleftarrow$

Now consider how x changes for x near 5. If x is slightly less than 5, then x decreases (away from 5) and if x is slightly greater than 5, then x increases (away from 5). Therefore, the equilibrium $x = 5$ is an unstable equilibrium. Next, for x slightly less than equilibrium value 10, x increases (away from 10). So the equilibrium $x = 10$ is a semi stable equilibrium. Finally, for x slightly less than 15, x increases (toward 15) and for x slightly greater than 15, x decreases (toward 15). Hence the equilibrium $x = 15$ is a stable equilibrium.

Note. Now let's analyze the stability of the equilibria for some of our selection models. Consider the heterozygote advantage model in which $w_{12} > w_{11}$ and $w_{12} > w_{22}$. We have an equilibrium at

$$p_{eq} = \frac{w_{22} - w_{12}}{w_{11} - 2w_{12} - w_{22}} = \frac{w_{12} - w_{22}}{2w_{12} - w_{11} - w_{22}}$$

and

$$\Delta p = \frac{pq}{\overline{w}} \left(p(w_{11} - w_{12}) + q(w_{12} - w_{22}) \right) = \frac{pq}{\overline{w}} \left(p(w_{11} - 2w_{12} + w_{22}) + (w_{12} - w_{22}) \right).$$

In this case, p, q, and $\overline{w}$ are each positive, so the sign of Δp depends only on the sign of $p(w_{11} - 2w_{12} + w_{22}) + (w_{12} - w_{22})$. First, choose a test value of p slightly less than p_{eq}. Since $2w_{12} - w_{11} - w_{22} > 0$, we can take as our test value $p = \dfrac{k}{2w_{12} - w_{11} - w_{22}}$ where $k < w_{12} - w_{22}$. For this test value,

$$\begin{aligned} p(w_{11} - 2w_{12} + w_{22}) + (w_{12} - w_{22}) &= \frac{k}{2w_{12} - w_{11} - w_{22}}(w_{11} + 2w_{12} + w_{22}) + (w_{12} - w_{22}) \\ &= -k + (w_{12} - w_{22}) > 0. \end{aligned}$$

So for p less than p_{eq}, $\Delta p > 0$ and p increases towards p_{eq}.

Next, choose a test value of p slightly greater than p_{eq}, say $p = \dfrac{k'}{2w_{12} - w_{11} - w_{22}}$ where $k' > w_{12} - w_{22}$. For this test value,

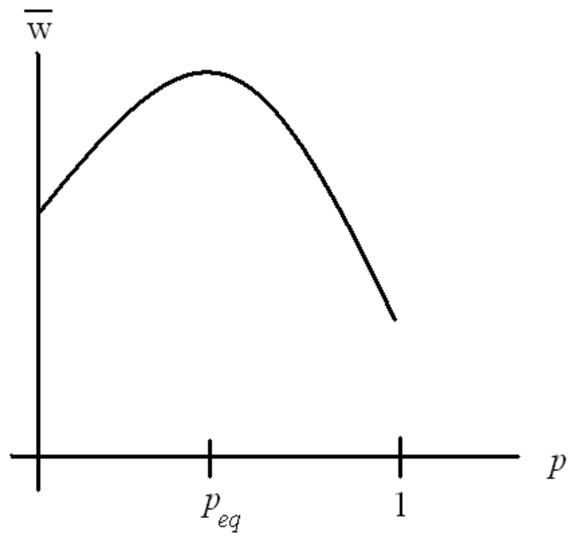
$$\begin{aligned}
p(w_{11} - 2w_{12} + w_{22}) + (w_{12} + w_{22}) &= \frac{k'}{2w_{12} - w_{11} - w_{22}}(w_{11} + 2w_{12} + w_{22}) + (w_{12} - w_{22}) \\
&= -k' + (w_{12} - w_{22}) < 0.
\end{aligned}$$

So for p greater than p_{eq}, $\Delta p < 0$ and p decreases towards p_{eq}. Therefore, with the heterozygote advantage, the equilibrium is stable. Biologically, this means that selection maintains the presence of both the A_1 and A_2 alleles (and hence polymorphism is maintained through selection). In fact, if we graph the average fitness of the population, $\overline{w}$, as a function of p, we see that it is a parabola which opens downward with formula

$$\begin{aligned}
\overline{w} = p^2 w_{11} + 2pq w_{12} + q^2 w_{22} &= p^2 w_{11} + 2p(1-p)w_{12} + (1-p)^2 w_{22} \\
&= p^2(w_{11} - 2w_{12} + w_{22}) + 2p(w_{12} - w_{22}) + w_{22}.
\end{aligned}$$

Recall that the vertex of a parabola with formula $y = ax^2 + bx + c$ has x-coordinate $\dfrac{-b}{2a}$. Therefore the vertex of the parabola determined by the graph of $\overline{w}$ occurs at the p-value

$$\frac{-2(w_{12} - w_{22})}{2(w_{11} - 2w_{12} + w_{22})} = \frac{w_{22} - w_{12}}{w_{11} - 2w_{12} + w_{22}} = p_{eq}.$$



Therefore, in the heterozygote advantage model, selection is expected to produce allele frequencies which produce a maximum of $\overline{w}$ (since the graph of $\overline{w}$ is an opening downward parabola, the vertex is the highest point on the graph). That is, in this case, selection acts to maximize average fitness of the population—not exactly "survival of the fittest," but reminiscent of it.

EXERCISES

Exercise. Show that the average fitness of a population is the expected value of the fitness of a randomly chosen individual.

Exercise. Find the equilibrium points of $\Delta x = x^2(x - 5)(7 - x)$. Find the stability of each equilibrium point. If Δx is expressed as a polynomial function in factored form, what properties of the factors determine the stability of the corresponding equilibrium point?

Exercise. Explore the stability of $p_{eq} = 1$ when $w_{11} = w_{12} > w_{22}$ and when $w_{11} = w_{12} < w_{22}$. Since this is a case of fixation, explain how *biologically* that p might be perturbed to less than 1.

Exercise. Explore the stability of p_{eq} in the deleterious heterozygote model. What is the expected long term behavior of allele frequencies of such a population? Discuss the biological significance of these results.