Lethal Recessive with Heterozygote Disadvantage

Now let's look at sickle-cell anemia, another genetic disease. Like cystic fibrosis, sickle-cell anemia occurs when an individual has two copies of a recessive allele (aa). Sickle-cell anemia causes the red blood cells to have a sickle, or hook like shape, rather than the normal round, donut like shape. Individuals with this genotype have severe and often fatal anemia and can suffer from symptoms ranging from tiredness and rheumatism to heart and kidney failure, brain damage and paralysis. Sickle-cell anemia differs from cystic fibrosis in that individuals with only one copy of the gene (Aa) also have a mild anemia and, under ordinary circumstances, have slightly lower fitness than individuals with no copies of the allele (AA). We'll talk more about "ordinary circumstances" in our next example. For now, it will be enough to assume that we are looking at a population in North America or Northern Europe.

It seems pretty likely that the "bad" allele (a) will eventually disappear from the population (as it did in the cystic fibrosis case). However, we can't predict how fast it will disappear. Will the a allele disappear faster or slower than before? Let's use the techniques we've developed to answer this example. As before we'll start with an entire population of heterozygotes (Aa) so $D_0 = R_0 = 0$, $H_0 = 1$, $p_0 = 1/2$ and $q_0 = 1/2$. According to Hardy-Weinberg, the fractions of the three genotypes in the next generation will be:

$$D_1 = p_0^2 = \left(\frac{1}{2}\right)^2 = \frac{1}{4} \qquad H_1 = 2p_0q_0 = 2\left(\frac{1}{2}\cdot\frac{1}{2}\right) = \frac{1}{2} \qquad R_1 = q_0^2 = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

As before, the aa individuals will all die before they reproduce, so we have $R_1 = 0$. In this case, the Aa individuals are less fit than the AA individuals. To match our class simulation, we'll assume that the fitness of the Aa individuals is $1/2$ and the fitness of the AA individuals is 1. Then, in the first generation of children we actually have p_0^2 homozygous AA's and $\frac{1}{2}(2p_0q_0) = p_0q_0$ heterozygous Aa's. The genotype frequencies are

$$D_1 = \frac{p_0^2}{p_0^2 + p_0q_0} = \frac{1/4}{1/4 + 1/4} = \frac{1}{2} \qquad H_1 = \frac{p_0q_0}{p_0^2 + p_0q_0} = \frac{1/4}{1/4 + 1/4} = \frac{1}{2}$$

and the new allele frequencies are

$$p_1 = D_1 + \frac{H_1}{2} = \frac{1}{2} + \frac{1}{4} = \frac{3}{4} \qquad q_1 = \frac{H_1}{2} + R_1 = \frac{1}{4}$$

We already notice that things seem to be happening faster. In our cystic fibrosis example $p_1 = 2/3$ and $p_2 = 3/4$. Now, we have $p_1 = 3/4$ in just one step. Let's see what happens after one more generation.

$$D_2 = \frac{p_1^2}{p_1^2 + p_1 q_1} = \frac{(3/4)^2}{(3/4)^2 + (3/4)(1/4)} = \frac{3}{4}$$

$$H_2 = \frac{p_1 q_1}{p_1^2 + p_1 q_1} = \frac{(3/4)(1/4)}{(3/4)^2 + (3/4)(1/4)} = \frac{1}{4}$$

$$p_2 = D_2 + \frac{H_2}{2} = \frac{3}{4} + \frac{1}{8} = \frac{7}{8} \qquad q_2 = \frac{H_2}{2} + R_2 = \frac{1}{8}$$

In this generation, the fraction of a alleles has decreased from 1/4 to 1/8 and the fraction of A alleles seems to be increasing towards 1. Will the fraction of A alleles actually go to 1? Let's go through the same analysis as before.

Let's switch to the more general case where p_0 is not specified from the start. Like before, we can use our general result to look at our class simulation by setting $p_0 = 1/2$. First, we write down a general equation for p_1 remembering that $q_0 = 1 - p_0$.

$$p_1 = D_1 + \frac{H_1}{2} = \frac{p_0^2}{p_0^2 + p_0 q_0} + \frac{p_0 q_0/2}{p_0^2 + p_0 q_0} = \frac{p_0\,(2p_0 + 1 - p_0)\,/2}{p_0\,(p_0 + 1 - p_0)} = \frac{(p_0 + 1)/2}{1} = \frac{1}{2}(p_0 + 1)$$

This formula will hold in general for any step so we can write

$$p_{n+1} = \frac{1}{2}(p_n + 1)$$

Let's solve for the equilibrium points, p^*. We have

$$p^* = \frac{1}{2}(p^* + 1) \;\Rightarrow\; 2p^* = p^* + 1 \;\Rightarrow\; p^* = 1$$

As before, the only equilibrium point is $p^* = 1$, or the case when A is the only allele in the population. Will the population actually go to this equilibrium? Our first few calculations suggested that the population will.

Let's see if we can come up with a formula for p_n in terms of p_0. Do you see a pattern in what we've calculated so far for the $p_0 = 1/2$ case? We have $p_0 = 1/2$, $p_1 = 3/4$ and $p_2 = 7/8$. A

general formula doesn't jump out quite as easily as before. Let's try looking at the first few iterations of the general case and see if that helps.

$$p_1 = \frac{1}{2}p_0 + \frac{1}{2}$$
$$p_2 = \frac{1}{2}(p_1 + 1) = \frac{1}{2}\left(\frac{1}{2}p_0 + \frac{1}{2} + 1\right) = \frac{1}{4}p_0 + \left(\frac{1}{4} + \frac{1}{2}\right)$$
$$p_3 = \frac{1}{2}(p_2 + 1) = \frac{1}{2}\left(\frac{1}{4}p_0 + \left(\frac{1}{4} + \frac{1}{2}\right) + 1\right) = \frac{1}{8}p_0 + \left(\frac{1}{8} + \frac{1}{4} + \frac{1}{2}\right)$$

It looks like

$$p_n = \frac{1}{2^n}p_0 + \left(\frac{1}{2^n} + \frac{1}{2^{n-1}} + \cdots + \frac{1}{2}\right)$$

might be what we are looking for. Remembering that

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 1 \quad \text{and} \quad \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n} = 1 - \frac{1}{2^n}$$

we have

$$p_n = \frac{1}{2^n}p_0 + \left(1 - \frac{1}{2^n}\right) = \frac{1}{2^n}(p_0 - 1) + 1 = 1 - \frac{1}{2^n}(1 - p_0)$$

Again, we can use induction to prove that this is the correct formula. First, we notice that the formula gives us the correct answer when $n = 1$.

$$p_1 = 1 - \frac{1}{2}(1 - p_0) = 1 - \frac{1}{2} + \frac{1}{2}p_0 = \frac{1}{2}(1 + p_0)$$

Let's assume it is correct for $n = k$ and show that it works for $n = k + 1$.

$$p_{k+1} = \frac{1}{2}(p_k + 1) = \frac{1}{2}\left(1 - \frac{1}{2^k}(1 - p_0) + 1\right) = \frac{1}{2} - \frac{1}{2^{k+1}}(1 - p_0) + \frac{1}{2} = 1 - \frac{1}{2^{k+1}}(1 - p_0)$$

This is exactly the form our formula predicts. Now, we can use our formula for p_n to predict what will happen after a long time has passed. As $n \to \infty$, the fraction $1/2^n \to 0$. Therefore, $p_n \to 1$ as expected.

As in the cystic fibrosis example the "bad" allele will eventually disappear, but will it disappear faster or slower? Let's go back to our specific example with $p_0 = 1/2$ and make a table of values. For comparison, we'll include our results from before.

	p_0	p_1	p_2	p_3	p_4	p_5	$\cdots$	p_{10}	p_{20}
$AA = Aa$	0.5	0.667	0.750	0.8000	0.83333	0.857143	$\cdots$	0.9167	0.95
$AA > Aa$	0.5	0.75	0.875	0.9375	0.96875	0.984375	$\cdots$	0.9990	1.0

The second example seems to be approaching one must faster. This makes sense as there is a disadvantage to having any a's at all. Let's look at q_n like we did before.

$$q_n = 1 - p_n == 1 - 1 + \frac{1}{2^n}(1 - p_0) = \frac{1}{2^n}(1 - p_0)$$

From this equation, we see that q_n (the fraction of allele a in the population) approaches zero at a rate related to 2^n which is exponential. In our cystic fibrosis example this rate was related to n. Hopefully, it is obvious that in the second example the a allele will disappear much, much faster than in the first example.

To summarize, we've found that if a recessive allele is lethal when it occurs as a homozygote (aa), it will eventually disappear from the population. If the heterozygote (Aa) is also detrimental, the rate of disappearance will be much faster.

Lethal Recessive with Heterozygote Advantage

Now, we'd like to return to an idea mentioned in the introduction to our sickle-cell anemia example. That is, we said that under ordinary circumstances, heterozygous individuals (Aa) are slightly less fit than dominant homozygous (AA) individuals. What exactly is meant by "ordinary circumstances"? It turns out that in regions of Africa that have high levels of (falciparum) malaria, the heterozygous genotype (Aa) actually has higher fitness than the normal homozygotes (AA). In this case, the slight abnormalities in the red blood cells caused by anemia actually protect a person from malaria. In a place like Africa, where large numbers of people suffer from malaria at some point in their lives, it is actually better (in terms of fitness) to suffer slightly from anemia, as it provides protection from the more lethal malaria. In general, we call this type of situation *heterozygote advantage*.

What will happen in this situation? Will the detrimental allele a still disappear, and if so, how fast? As before we'll start with an entire population of heterozygotes (Aa) so

$D_0 = R_0 = 0$, $H_0 = 1$, $p_0 = 1/2$ and $q_0 = 1/2$. Hardy-Weinberg gives us the fractions of the three genotypes in the next generation:

$$D_1 = p_0^2 = \left(\frac{1}{2}\right)^2 = \frac{1}{4} \qquad H_1 = 2p_0q_0 = 2\left(\frac{1}{2}\cdot\frac{1}{2}\right) = \frac{1}{2} \qquad R_1 = q_0^2 = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

The aa individuals will all die as before, but in this case, the AA individuals are less fit than the Aa individuals. To match our class simulation, we'll assume that the fitness of the AA individuals is $1/2$ and the fitness of the Aa individuals is 1. Then, in the first generation of children we actually have $p_0^2/2$ homozygous AA's and $2p_0q_0$ heterozygous Aa's. The genotype frequencies are

$$D_1 = \frac{p_0^2/2}{p_0^2/2 + 2p_0q_0} = \frac{1/8}{1/8 + 1/2} = \frac{1}{5} \qquad H_1 = \frac{2p_0q_0}{p_0^2/2 + 2p_0q_0} = \frac{1/2}{1/8 + 1/2} = \frac{4}{5}$$

and the new allele frequencies are

$$p_1 = D_1 + \frac{H_1}{2} = \frac{1}{5} + \frac{2}{5} = \frac{3}{5} \qquad q_1 = \frac{H_1}{2} + R_1 = \frac{2}{5}$$

The frequency of the A allele increased, but not by much. Let's see what happens after a few more generations.

$$D_2 = \frac{p_1^2/2}{p_1^2/2 + 2p_1q_1} = \frac{(3/5)^2/2}{(3/5)^2/2 + 2(3/5)(2/5)} = \frac{9}{33} = \frac{3}{11}$$
$$H_2 = \frac{2p_1q_1}{p_1^2/2 + 2p_1q_1} = \frac{2(3/5)(2/5)}{(3/5)^2/2 + 2(3/5)(2/5)} = \frac{24}{33} = \frac{8}{11}$$
$$p_2 = D_2 + \frac{H_2}{2} = \frac{3}{11} + \frac{4}{11} = \frac{7}{11} \qquad q_2 = \frac{H_2}{2} + R_2 = \frac{4}{11}$$

$$D_3 = \frac{p_2^2/2}{p_2^2/2 + 2p_2q_2} = \frac{(7/11)^2/2}{(7/11)^2/2 + 2(7/11)(4/11)} = \frac{49}{161} = \frac{7}{23}$$
$$H_3 = \frac{2p_2q_2}{p_2^2/2 + 2p_2q_2} = \frac{2(7/11)(4/11)}{(7/11)^2/2 + 2(7/11)(4/11)} = \frac{112}{161} = \frac{16}{23}$$
$$p_3 = D_3 + \frac{H_3}{2} = \frac{7}{23} + \frac{8}{23} = \frac{15}{23} \qquad q_3 = \frac{H_3}{2} + R_2 = \frac{8}{23}$$

The frequency of the A allele doesn't seem to be going to 1 in this example or, if it is, it is moving very slowly.

Let's go ahead and look at the more general case with p_0 as a variable. A general equation for p_1 is

$$p_1 = D_1 + \frac{H_1}{2} = \frac{p_0^2/2}{p_0^2/2 + 2p_0q_0} + \frac{2p_0q_0/2}{p_0^2/2 + 2p_0q_0} = \frac{p_0\left(p_0 + 2 - 2p_0\right)/2}{p_0\left(p_0 + 4 - 4p_0\right)/2} = \frac{2 - p_0}{4 - 3p_0}$$

In general, we have

$$p_{n+1} = \frac{2 - p_n}{4 - 3p_n}$$

Let's solve for the equilibrium points, p^*. We have

$$p^* = \frac{2 - p^*}{4 - 3p^*} \;\Rightarrow\; 4p^* - 3(p^*)^2 = 2 - p^* \;\Rightarrow\; 3(p^*)^2 - 5p^* + 2 = 0 \;\Rightarrow\; (3p^* - 2)(p^* - 1) = 0$$

In this case there are two equilibrium points $p^* = 2/3$ and $p^* = 1$. Which equilibrium will the population go to? Our calculations seem to be heading towards the first equilibrium, but how can we be sure? Let's see if we can write down a general formula for p_n.

$$p_1 = \frac{2 - p_0}{4 - 3p_0}$$

$$p_2 = \frac{2 - p_1}{4 - 3p_1} = \frac{2 - \frac{2-p_0}{4-3p_0}}{4 - 3\left(\frac{2-p_0}{4-3p_0}\right)} = \frac{8 - 6p_0 - 2 + p_0}{16 - 12p_0 - 6 + 3p_0} = \frac{6 - 5p_0}{10 - 9p_0}$$

$$p_3 = \frac{2 - p_2}{4 - 3p_2} = \frac{2 - \frac{6-5p_0}{10-9p_0}}{4 - 3\left(\frac{6-5p_0}{10-9p_0}\right)} = \frac{20 - 18p_0 - 6 + 5p_0}{40 - 36p_0 - 18 + 15p_0} = \frac{14 - 13p_0}{22 - 21p_0}$$

Can you figure out a general formula? In this case its pretty hard to get the right form, so we'll use another method to study the dynamics. We can plot a graph of our iterative formula

$$p_{n+1} = \frac{2 - p_n}{4 - 3p_n}$$

with p_n as the x-axis and p_{n+1} as the y-axis and use a method called cobwebbing to evaluate the long term dynamics. At the same time, we'll make plots of our previous models and see how our results compare to the cobweb diagrams.

Looking at a graph of the iteration function, we see that if we draw a vertical line starting at p_0 on the x-axis, and stop when we hit the iteration function, we will be at a height of p_1. For the next iteration, we want to start at p_1 on the x-axis. We can make this easy by drawing a second line on the graph corresponding to the function $y = x$ or $p_n = p_{n+1}$. Now, starting at the height of p_1, we draw a horizontal line until we intersect the line $y = x$. This point will correspond to a value of p_1 on the x-axis. Now, draw a vertical line from this point until we intersect the iteration function again. This will give us the next allele frequency. We can continue this process indefinitely. After a few iterations, we will notice that we seem to be moving closer and closer to a point where the iteration function intersects the line $y = x$. This intersection point is an equilibrium point.

The cobwebbing procedure for our iteration function predicts that if, at the start, there is at least one allele of each type, the A allele will eventually be present in 2/3 of the population and the a allele will eventually be present in 1/3. So, even though the a allele is pretty bad when it occurs alone, the fact that it can do some good in combination with the A allele allows it to survive in the population.

It turns out that the general formula for p_n is

$$p_n = \frac{\frac{1}{2^n} + \left(2 - \frac{3}{2^n}\right)(1 - p_0)}{\frac{1}{2^n} + 3\left(1 - \frac{1}{2^n}\right)(1 - p_0)} = \frac{1 + (2^{n+1} - 3)(1 - p_0)}{1 + 3(2^n - 1)(1 - p_0)}$$

As $n \to \infty$, the fraction $1/2^n \to 0$. Therefore,

$$p_n \to \frac{2(1 - p_0)}{3(1 - p_0)} = \frac{2}{3}$$

This matches the prediction from our cobwebbing picture.

Unlike the previous two examples, the "bad" allele does not disappear. How fast will the

population approach the equilibrium point? Let's add one more row to our table.

	p_0	p_1	p_2	p_3	p_4	p_5	$\cdots$	p_{10}	p_{20}
$AA = Aa$	0.5	0.667	0.750	0.8000	0.83333	0.857143	$\cdots$	0.9167	0.95
$AA > Aa$	0.5	0.750	0.875	0.9375	0.96875	0.984375	$\cdots$	0.9990	1.0
$AA < Aa$	0.5	0.600	0.636	0.6522	0.65957	0.663158	$\cdots$	0.6666	0.667

The good allele A seems to converge pretty fast to the 2/3 equilibrium. Let's look at q_n like we did before.

$$q_n = 1 - p_n = 1 - \frac{1 + (2^{n+1} - 3)(1 - p_0)}{1 + 3(2^n - 1)(1 - p_0)} = \frac{1 + 3(2^n - 1)(1 - p_0) - 1 - (2^{n+1} - 3)(1 - p_0)}{1 + 3(2^n - 1)(1 - p_0)}$$

$$= \frac{\left[3(2^n) - 3 - 2^{n+1} + 3\right](1 - p_0)}{1 + 3(2^n - 1)(1 - p_0)} = \frac{(3 - 2)(1 - p_0)}{\frac{1}{2^n} + 3(1 - p_0) - \frac{3}{2^n}(1 - p_0)}$$

$$= \frac{1 - p_0}{3(1 - p_0) + \frac{1}{2^n}(1 - 3(1 - p_0))} = \frac{1}{3 + \frac{1}{2^n}\left(\frac{3p_0 - 2}{1 - p_0}\right)}$$

From this equation, we see that q_n (the fraction of allele a in the population) approaches $1/3$ at a rate related to 2^n which is exponential and fairly rapid.

In summary, we've found that if a recessive allele is lethal when it occurs as a homozygote (aa), it will eventually disappear from the population. If the heterozygote (Aa) is also detrimental, the rate of disappearance will be much faster. However, if the heterozygote (Aa) is beneficial, the bad a allele will never disappear.

Extra Problems

(1) In the human population there are four blood types: A, B, O and AB. These blood groups are caused by three alleles: I^A, I^B and I^O. Individuals with genotypes I^AI^A and I^AI^O have blood type A, those with I^BI^B and I^BI^O have blood type B, those with I^AI^B have blood type AB, and those with I^OI^O have type O.

(a) If p is the fraction of I^A alleles in the population, q is the fraction of I^B alleles, and r is the fraction of I^O alleles, find the Hardy-Weinberg equilibrium proportions for the four blood types A, B, AB and O.

(b) If $p = q = r = 1/3$, what are the proportions of each blood type in the population?

(2) For the examples done in class, we assumed that fitness values were 0, 1/2 or 1. What if in our sickle-cell anemia/ malaria example, the fitness of AA individuals is only slightly less than that of Aa individuals? Will we get the same result? Let Aa individuals have a fitness of 1, AA individuals have a fitness of $(1 - s)$ and aa individuals have a fitness of 0.
(a) Find a formula for p_{n+1} as a function of p_n.
(b) Find the equilibrium points.
(c) Find a general formula for p_n as a function of p_0. It may be helpful to let $\alpha = 1 + s$ and $\beta = 1 - s$. Also, notice that

$$1 + \beta + \cdots \beta^{n-1} = \frac{1 - \beta^n}{1 - \beta}$$

(d) Make a table of p for the case $s = 0.01$. How fast does the population go to its equilibrium? Is it similar to the case when $s = 0$?

References

Davis, Tom. *Mathematical Biology*, 1999. http://www.geometer.org/mathcircles

Futuyma, Douglas. *Evolutionary Biology*, 1998. Sinauer Associates, Inc., Massachusetts.

Griffiths, A., Miller, J. Suzuki, D. Lewontin, R. and Gelbart, W. *An Introduction to Genetic Analysis, 6th Ed.*, 1996. W. H. Freeman and Company, New York.

Hartl, Daniel. *A Primer of Population Genetics, 3rd Ed.*, 2000. Sinauer Associates, Inc., Massachusetts.