

Bi/Ge105: Evolution

Homework 4

Due Date: Wednesday, February 7, 2024

“We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.”

–Pierre Simon Laplace, A Philosophical Essay on Probabilities

1. Population Genetics for Haploid Organisms

In this problem, we build on what was done in class, but now focusing on the case of haploid organisms rather than diploid organisms. Once again, we think about the one-locus, two-allele (A_1 and A_2) abstraction that we have already been using in class.

(A) In the presence of selection, with fitnesses w_1 and w_2 for our two genotypes, write down the expressions for the new values of p and q after a generation of selection. If you think of the urn idea we considered in class, then the number of A_1 alleles is N_1 and the number of A_2 alleles is N_2 , and there are a total of $N = N_1 + N_2$ alleles in our urn. This means that we assign $p = N_1/N$ and $q = N_2/N$ to the probability of drawing an A_1 and an A_2 , respectively. What we want is p' and q' which is the frequencies after a single generation of selection. Make sure you explain your notation and that you give an appropriate definition of the mean fitness in the context of this simplified haploid example.

(B) Now, consider the case in which the relative fitnesses are $w_1 = 1$ and $w_2 = 1 - s$, where s is the so-called selection coefficient. Work out an expression for Δp as a function of p , q and s . Now, assuming that the initial frequency of $p = 0.01$ and that $s = 0.1$, make a plot of Δp as a function of the generation number. Explain in what sense the system is evolving.

(C) Now imitate the concept we did in class in order to work out the mutation-selection balance. Just as we did in class for the diploid case, work out an expression for p'' (i.e. after both selection and mutation in a single generation) in terms of p , q , s and μ (the rate of mutation from A_1 to A_2). Remember that in thinking about this mutation-selection balance, we are imagining that the mutation from A_2 to A_1 is exceedingly rare compared with the rate from A_1 to A_2 (i.e. it is easier to destroy than to build!) Now, by insisting that the allele frequencies no longer change (i.e. you find the steady state), find the value of p^* and q^* in the steady-state limit. Make sure you give an intuitive explanation of the result and their dependence on the parameters such as s and μ .

2. Population Genetics for Diploid Organisms

In class we talked about the expression for Δp in the case of selection. In this problem we are going to revisit those ideas and elaborate on them further.

(A) Rederive the expression we worked out for Δp , namely,

$$\Delta p = \frac{p}{\bar{w}} [p(w_{11} - \bar{w}) + q(w_{12} - \bar{w})]. \quad (1)$$

By using the definition of \bar{w} , demonstrate that this can also be written as

$$\Delta p = \frac{pq}{\bar{w}} [p(w_{11} - w_{12}) + q(w_{12} - w_{22})]. \quad (2)$$

Make a plot of Δp vs p for the case of overdominance in which $w_{12} > w_{11} > w_{22}$. Make sure you explain what this graph demonstrates. Essentially, this is a phase portrait that shows how the allele frequency changes from one generation to the next.

(B) In class we claimed that in the case of overdominance, there is a fixed point p^* . Using the expression

$$\Delta p = \frac{pq}{\bar{w}} [p(w_{11} - w_{12}) + q(w_{12} - w_{22})], \quad (3)$$

find p^* by solving for the case in which $\Delta p = 0$. Your expression for p^* will be a simple function of w_{11} , w_{12} and w_{22} .

(C) Now repurpose your calculations to the case of the case of heterozygous advantage and choose our fitnesses symmetrically as $w_{11} = w_{22} = 0.1$ and $w_{12} = 1.0$. Once again, make a plot of Δp vs p and explain what the resulting graph shows us.

3. Loss of Heterozygosity

Our discussion of genetic drift argued that one of the ways that people examine the extent of drift is by monitoring the heterozygosity. Recall that we looked at several examples including the classic experiment of Buri on eye color in flies, the lava lizards of the Galapagos where we saw that the heterozygosity was smaller on smaller islands and the case of the snapper fish in New Zealand where the hypothesis of overfishing has been advanced. To explore the idea of loss of heterozygosity in more detail, we are going to work with our usual one-locus, two-allele model and the Wright-Fisher model. The goal will be to compute the loss of heterozygosity on a per generation basis.

(A) Given that the initial frequency of allele 1 is p_0 , work out an expression for the initial heterozygosity ($H(0) = H_0$) in terms of p_0 . Remember that in the most general case, heterozygosity is given by

$$H(t) = 1 - \sum_i p_i(t)^2, \quad (4)$$

where t refers to the generation of interest. The idea of this expression is to subtract off the probability of all of the homozygotes. Your job is to exploit this definition but in the simpler case where there are only three genotypes, two of which are homozygotes.

(B) Now let's find the heterozygosity one generation later. From the previous result, we need

$$\langle H_1 \rangle = \langle 2p_1(1 - p_1) \rangle = 2\langle p_1 \rangle - 2\langle p_1^2 \rangle. \quad (5)$$

To make sense of this, note that there are a total of N alleles and we want to find the probability that we have n of allele 1. Write the probability of getting n A_1 alleles, given that you are drawing from an urn with $p_0 N$ A_1 alleles. We need to compute

$$\langle p_1 \rangle = \left\langle \frac{n}{N} \right\rangle \quad (6)$$

and

$$\langle p_1^2 \rangle = \langle \frac{n^2}{N^2} \rangle. \quad (7)$$

Since N is a constant, all this really means is that you need to figure out $\langle n \rangle$ and $\langle n^2 \rangle$. Recall that these are defined as

$$\langle n \rangle = \sum_{n=0}^N np(n, N) \quad (8)$$

and

$$\langle n^2 \rangle = \sum_{n=0}^N n^2 p(n, N), \quad (9)$$

where $p(n, N)$ is the probability that we pull n copies of A_1 out of our urn given that the probability of getting an A_1 is p_0 . Note that to evaluate these averages, this is the moment to use the trick that was showed while discussing the Luria-Delbruck experiment, namely, differentiation with respect to a parameter. For example, if we have

$$\langle n \rangle = \sum_{n=0}^N n \frac{N!}{n!(N-n)!} p^n q^{N-n}, \quad (10)$$

the critical observation is that we can get this as

$$\langle n \rangle = p \frac{\partial}{\partial p} \sum_{n=0}^N \frac{N!}{n!(N-n)!} p^n q^{N-n}. \quad (11)$$

Note that from the binomial theorem,

$$\sum_{n=0}^N \frac{N!}{n!(N-n)!} p^n q^{N-n} = (p+q)^N. \quad (12)$$

This trick empowers you to evaluate all of the averages required above.

To be concrete, your goal is to show that

$$\langle H_1 \rangle = 2p_0(1-p_0)(1 - \frac{1}{N}) = H_0(1 - \frac{1}{N}). \quad (13)$$

Once you have demonstrated this result, it implies in turn that if we run the same argument over and over again for t generations, we have

$$\langle H_t \rangle = H_0 \left(1 - \frac{1}{N}\right)^t \approx H_0 e^{-\frac{t}{N}}. \quad (14)$$

Fill in the missing steps of that argument. The emergence of the exponential is based upon the observation that $e^{-x} \approx 1 - x$ for small x . Make sure you include all the steps leading to the claims offered above and that you explain what this final result tells you about the relationship between the time to fixation of alleles and the effective population size.