BMIF 201 Lecture 3

Michelle M. Li

September 11, 2019

1 Heterozygosity Equation

Recall: CZ posed the question, "Why does the Moran and Wright-Fisher models fixate when the heterozygosity equation (1) suggests otherwise?"

$$\begin{aligned} H_t &= H_0(1 - \frac{1}{N})^t \\ &\approx H_0 e^{-\frac{t}{N}} \end{aligned} \tag{1}$$

For some time t, the expected heterozygosity for two alleles is as follows

$$\begin{aligned} H_{t+1} &= \frac{1}{N} * 0 + (1 - \frac{1}{N}) * H_t \\ &= (1 - \frac{1}{N}) * H_t \end{aligned} \tag{2}$$

where $\frac{1}{N}$ is the probability two offspring share the same ancestor, and $(1 - \frac{1}{N})$ is the probability that they do not.

Explanation: We are calculating the expected value of the next time step. The probability of reaching, thereby fixating at, 0 or N is non-zero.

2 Wright-Fisher Model

2.1 Background

In the Wright-Fisher Model (a.k.a. Fisher-Wright Model), you choose n alleles (white or black) from the population of a set size N.

The probability of choosing a black allele is $p = \frac{b}{N}$, where b is the number of black alleles in the population at time t.

This is a binomial distribution problem, where you are doing n "coin flips" of probability p to get the number of black alleles, X.

$$X \sim Bin(n, p) \tag{3}$$

2.2 Simulations

Simulation exercises can be found here as a Jupyter Notebook: `https://github.com/michellemli/BMIF201`

2.3 Shannon Entropy

The class seems to agree that the time to fixation when $N = 20$ and $p = 0.5$ is around 27. We will now derive the results.

Firstly, the Shannon entropy is defined as

$$S(X) = -\sum p_i \log_i(p_i) \tag{4}$$

where $p_i = \dfrac{x_i}{N}$. In words, the Shannon diversity index quantifies the entropy. At fixation, this value is 0 because there is only one allele in the population.

$$S(X) = 0 \Longleftrightarrow \text{Fixation}$$

The partial derivatives of $S(X)$ are

$$\begin{aligned} \frac{\partial S(X)}{x_i} &= -\frac{1}{N}(1 + \log(\frac{x_i}{N})) \\ \frac{\partial^2 S(X)}{\partial x_i \partial x_j} &= -\frac{\delta_{ij}}{N x_j} \end{aligned} \tag{5}$$

where

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \tag{6}$$

Recall that the expectation for selecting j alleles with probability p is

$$\langle j \rangle = np \tag{7}$$

and the variance is

$$Var \langle j \rangle = \langle j^2 \rangle - \langle j \rangle^2 \tag{8}$$

$$= np(1-p) \tag{9}$$

because the Wright-Fisher Model follows a binomial distribution.

To calculate the change in Shanon entropy over one time step:

Let k be the number of alleles, and x be the number of black alleles.

$$\langle S(x+\Delta x)\rangle \approx \left\langle S(x) + \sum_i^k \frac{\partial S(x)}{\partial x}\Delta x_i + \frac{1}{2}\sum_i^k\sum_i^k \frac{\partial^2 S(x)}{\partial x_i \partial x_j} \right\rangle \tag{10}$$

$$= S(x) + \sum_i^k \frac{\partial S(x)}{\partial x_i}\langle \Delta x_i\rangle + \frac{1}{2}\sum_i^k\sum_j^k \frac{\partial^2 S(x)}{\partial^2 x_i \partial^2 x_j}\langle \Delta x_i \Delta x_j\rangle \tag{11}$$

$$= S(x) + \frac{1}{2}\sum_i^k\sum_j^k \frac{\partial^2 S(x)}{\partial^2 x_i \partial^2 x_j}\langle \Delta x_i \Delta x_j\rangle \tag{12}$$

$$= S(x) - \frac{1}{2}\sum_i^k\sum_j^k \frac{\delta_{ij}}{N x_j}\langle \Delta x_i \Delta x_j\rangle \tag{13}$$

$$= S(x) - \frac{1}{2}\sum_i^k \frac{1}{N x_i}\langle \Delta x_i \Delta x_i\rangle \tag{14}$$

$$= S(x) - \frac{1}{2}\sum_i^k \frac{1}{N x_i}\left(N(\frac{x_i}{N})(1-\frac{x_i}{N})\right) \tag{15}$$

$$= S(x) - \frac{1}{2N}\sum_i^k (1-\frac{x_i}{N}) \tag{16}$$

$$= S(x) - \frac{k-1}{2N} \tag{17}$$

Remarks:

- Taking the partial derivative of x_i in (11) results in $\langle \Delta x_i\rangle = 0$.
- For (12), substitute with (5) to get (13).
- For (13), recall that $\delta_{ij} = 1$ when $i = j$ from (6).
- For (14), apply the variance equation (9) where $p = \frac{x_i}{N}$ to get (15).
- For (15), pull out the N from the summation to get (16).
- For (16), remember that $\sum_i^k 1 = k$ and $\sum_i^k x_i = N$, which simplifies (16) into (17).

So, the expected change in Shannon entropy from generation to generation is

$$\langle \Delta S(X)\rangle = \langle S(x+\Delta x)\rangle - S(x) = -\frac{k-1}{2N} \tag{18}$$

To calculate the fixation time T, we start by stating that

$$S(X) - T\langle \Delta S(X)\rangle = 0$$

where, after substituting in (18),

$$T = \frac{S(X)}{\langle \Delta S(X)\rangle} = \frac{2NS(X)}{k-1}$$