BMIF 201 Lecture 3

Michelle M. Li

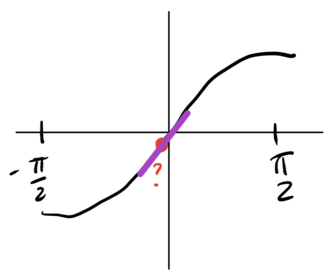
September 11, 2019

1 Review: Taylor's Expansion

Motivation: Calculating values (e.g. e^x) is difficult for humans (and often even computers)

Example: Suppose our function is $sin(x)$ and we want to calculate the value of the red dot in Figure 1. We can use Taylor's Expansion to approximate the purple function in Figure 1.

Figure 1: Approximating $\sin(x)$



Let

$$T(X = 3) = a_0 + a_1x + a_2x^2 + a_3x^3 \tag{1}$$

and set a_0, a_1, a_2, and a_3 such that

$$\begin{aligned} T(0) &= sin(0) \\ T'(0) &= sin'(0) \\ T''(0) &= sin''(0) \\ T'''(0) &= sin'''(0) \end{aligned}$$

So, for $T(3)$,

$$\begin{aligned} T(3) &= a_0 + a_1x + a_2x^2 + a_3x^3 && \approx \ sin(x) \\ T'(3) &= a_1 + 2a_2x + 3a_3x^2 && \approx \ sin'(x) = cos(x) \\ T''(3) &= 2a_2 + 3 \cdot 2a_3x && \approx \ sin''(x) = -sin(x) \\ T'''(3) &= 3 * 2a_3 && \approx \ sin'''(x) = -cos(x) \end{aligned}$$

and

$$\begin{aligned}
sin(0) &= 0 \\
sin'(0) &= cos(0) = 1 \\
sin''(0) &= -sin(0) = 0 \\
sin'''(0) &= -cos(0) = -1
\end{aligned}$$

Hence, at $x = 0$,

$$\begin{aligned}
T(0) &= a_0 \\
T'(0) &= a_1 = 1!a_1 \\
T''(0) &= 2a_2 = 2!a_2 \\
T'''(0) &= 3 * 2a_2 = 3!a_3
\end{aligned}$$

where

$$\begin{aligned}
a_0 &= 0 \\
a_1 &= 1 \\
a_2 &= 0 \\
a_3 &= -\frac{1}{3!}
\end{aligned}$$

Finally, we derive the Taylor's Expansion as

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k \tag{2}$$

Generalizing (2) for multiple variables, we get

$$f(x_1, \ldots, x_k) = \sum_{i_1=0}^{m} \cdots \sum_{i_k=0}^{m} \frac{\partial^{i_1} \ldots \partial^{i_k}}{i_1! \ldots i_k! \partial_{x_1}^{i_1} \ldots \partial_{x_k}^{i_k}} f(a_1, \ldots, a_k)(x_1 - a_1)^{i_1} \ldots (x_k - a_k)^{i_k} \tag{3}$$

Example: Expansion of $e^x(x, y) = e^x \ln(1 + y)$ at $(0, 0)$.

Let's use the Taylor Expansion on another function. So, we expand (3) where $m = 2$, $k = 2$:

$$f(x, y) = \sum_{i_1=0}^{2} \sum_{i_2=0}^{2} \frac{\partial^{i_1}\partial^{i_2}}{i_1!i_2!\partial_x^{i_1}\partial_y^{i_2}} f(a_1, a_2)(x - a_1)^{i_1}(y - a_2)^{i_2} \tag{4}$$

We first calculate the derivatives as follows:

$$\frac{\partial f(x,y)}{\partial x} = e^x \ln(1+y) \tag{5}$$

$$\frac{\partial^2 f(x,y)}{\partial^2 x} = e^x \ln(1+y) \tag{6}$$

$$\frac{\partial f(x,y)}{\partial y} = \frac{e^x}{1+y} \tag{7}$$

$$\frac{\partial^2 f(x,y)}{\partial^2 x} = e^x \left(-\frac{1}{(1+y)^2}\right) \tag{8}$$

$$\frac{\partial f(x,y)}{\partial x \partial y} = \frac{e^x}{1+y} \tag{9}$$

$$\tag{10}$$

At $(x, y) = (0, 0)$,

$$\frac{\partial f(0,0)}{\partial x} = 0 \tag{11}$$

$$\frac{\partial^2 f(0,0)}{\partial^2 x} = 0 \tag{12}$$

$$\frac{\partial f(0,0)}{\partial y} = 1 \tag{13}$$

$$\frac{\partial^2 f(0,0)}{\partial^2 x} = -1 \tag{14}$$

$$\frac{\partial f(0,0)}{\partial x \partial y} = 1 \tag{15}$$

$$\tag{16}$$

2 Heterozygosity Equation

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

$$H_t = H_0(1 - \frac{1}{N})^t \approx H_0 e^{-\frac{t}{N}} \tag{17}$$

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

$$\begin{aligned} H_{t+1} &= \frac{1}{N} * 0 + (1 - \frac{1}{N}) * H_t \\ &= (1 - \frac{1}{N}) * H_t \end{aligned} \tag{18}$$

Note that $\frac{1}{N}$ 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.

3 Wright-Fisher Model

3.1 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.

$$X \sim Bin(n, p) \tag{19}$$

The expectation for selecting j alleles with probability p is

$$\langle j \rangle = np \tag{20}$$

and the variance is

$$\begin{aligned} Var \langle j \rangle &= \left\langle j^2 \right\rangle - \langle j \rangle^2 && (21) \\ &= np(1-p) && (22) \end{aligned}$$

because the Wright-Fisher Model follows a binomial distribution.

3.2 Simulations

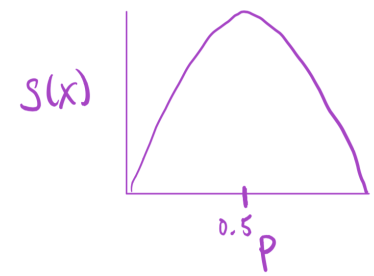
Simulation exercises can be found here as a Jupyter Notebook: `https://github.com/michellemli/BMIF201`

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 in the next section.

3.3 Shannon Entropy

Example: For two alleles with probability p and $1 - p$. When $p = 0.5$, the entropy is at its peak (Figure 2).

Figure 2: Shannon Entropy



Now, let us consider k alleles. Firstly, the Shannon entropy is defined as

$$S(X) = -\sum_{i=1}^{k} p_i \log(p_i) \tag{23}$$

where $p_i = \dfrac{x_i}{N}$.

At fixation, this value is 0 because there is only one allele in the population.

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

We will use Taylor Expansion to approximate $S(X)$ with degree 2. So, we will prepare the partial derivatives as follows

$$\begin{aligned} \frac{\partial S(X)}{x_i} &= -\frac{1}{N}(1 + \log(\frac{x_i}{N})) \\ \frac{\partial^2 S(X)}{\partial x_i \partial x_j} &= -\frac{\delta_{ij}}{N x_j} \end{aligned} \tag{24}$$

where

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \tag{25}$$

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

$$\langle k \rangle = np \tag{26}$$

and the variance is

$$Var \langle k \rangle = \left\langle k^2 \right\rangle - \langle k \rangle^2 \tag{27}$$

$$= np(1-p) \tag{28}$$

To calculate the change in Shanon entropy over one time step (i.e. from generation t to generation $t+1$):

Let us denote the change in X through one generation by ΔX. So, $x + \Delta x = t + 1$. Also, let x be the number of black alleles.

Here, we use the Taylor Expansion to approximate $\langle S(x + \Delta x)\rangle$:

$$\langle S(x+\Delta x)\rangle \approx \left\langle S(x) + \sum_i^k \frac{\partial S(x)}{\partial x}\Delta x_i + \frac{1}{2}\sum_i^k\sum_i^k \frac{\partial^2 S(x)}{\partial x_i \partial x_j} \right\rangle \tag{29}$$

$$= S(x) + \sum_i^k \frac{\partial S(x)}{\partial x_i}\langle \Delta x_i\rangle + \frac{1}{2}\sum_i^k\sum_j^k \frac{\partial^2 S(x)}{\partial^2 x_i \partial^2 x_j}\langle \Delta x_i \Delta x_j\rangle \tag{30}$$

$$= S(x) + \frac{1}{2}\sum_i^k\sum_j^k \frac{\partial^2 S(x)}{\partial^2 x_i \partial^2 x_j}\langle \Delta x_i \Delta x_j\rangle \tag{31}$$

$$= S(x) - \frac{1}{2}\sum_i^k\sum_j^k \frac{\delta_{ij}}{N x_j}\langle \Delta x_i \Delta x_j\rangle \tag{32}$$

$$= S(x) - \frac{1}{2}\sum_i^k \frac{1}{N x_i}\langle \Delta x_i \Delta x_i\rangle \tag{33}$$

$$= S(x) - \frac{1}{2}\sum_i^k \frac{1}{N x_i}\left(N(\frac{x_i}{N})(1-\frac{x_i}{N})\right) \tag{34}$$

$$= S(x) - \frac{1}{2N}\sum_i^k (1-\frac{x_i}{N}) \tag{35}$$

$$= S(x) - \frac{k-1}{2N} \tag{36}$$

Remarks:

- For (29), use Taylor Expansion with degree 2 to approximate $\langle S(x + \Delta x)\rangle$
- Applying the linearity of expectations property to get from (29) to (30)
- Taking the partial derivative of x_i in (30) results in $\langle \Delta x_i\rangle = 0$.
- For (31), substitute with (24) to get (32).
- For (32), recall that $\delta_{ij} = 1$ when $i = j$ from (25).
- For (33), apply the variance equation (28) where $p = \frac{x_i}{N}$ to get (34).
- For (34), pull out the N from the summation to get (35).
- For (35), remember that $\sum_i^k 1 = k$ and $\sum_i^k x_i = N$, which simplifies (35) into (36).

So, the expected change in Shannon entropy from generation t to generation $t + 1$ is

$$\langle \Delta S(X)\rangle = \langle S(x+\Delta x)\rangle - S(x) = -\frac{k-1}{2N} \tag{37}$$

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

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

where, after substituting in (37),

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