BMIF 201 Lecture 3

Michelle M. Li

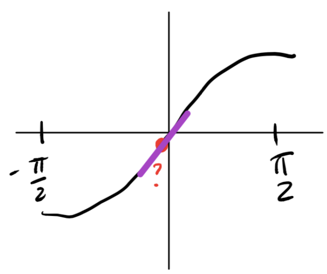
September 11, 2019

1 Review: Taylor's Expansion

Motivation: Calculating values (i.e. 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. If X is large, it is very difficult to compute the exact value. Instead, we can use Taylor's Expansion to approximate the purple function in Figure 1.

Figure 1: Approximating $\sin(x)$



Let

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

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 $T(X)$,

$$\begin{aligned} T(X) &= a_0 + a_1x + a_2x^2 + a_3x^3 \\ T'(X) &= a_1 + 2a_2x + 3a_3x^2 \\ T''(X) &= 2a_2 + 3 * 2a_3x \\ T'''(X) &= 3 * 2a_3 \end{aligned}$$

and

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

Hence, at $X = 0$,

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

Finally, we derive the Taylor's Expansion as

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

Expanding (2), we get

$$f(x_1, \dots, x_d) = \sum_{i_1=0}^{m} \cdots \sum_{i_d=0}^{m} \frac{\partial^{i_1} \dots \partial^{i_d}}{i_1! \dots i_d! \partial_{x_1}^{i_1} \dots \partial_{x_d}^{i_d}} f(a_1, \dots, a_d)(x_1 - a_1)^{i_1} \dots (x_d - a_d)^{i_d} \tag{3}$$

Using an example, where $m = 2$,

$$f(x, y) = e^x \ln(1+y) \approx y + xy - y^2 \tag{4}$$

Here is how we arrived at that approximation. The derivatives are as follows

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

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

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

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

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

$$\tag{10}$$

At $(x, y) = 0$,

$$\frac{\partial f}{\partial x} = 0 \tag{11}$$
$$\frac{\partial^2 f}{\partial^2 x} = 0 \tag{12}$$
$$\frac{\partial f}{\partial y} = 1 \tag{13}$$
$$\frac{\partial^2 f}{\partial^2 x} = -1 \tag{14}$$
$$\frac{\partial f}{\partial x \partial y} = 1 \tag{15}$$
$$\tag{16}$$

Notes:

- $\frac{\partial^2 f}{\partial x \partial y} \neq \frac{\partial^2 f}{\partial x \partial y}$ depending on the function.

2 Heterozygosity Equation

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

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

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

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

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.

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.

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

3.2 Simulations

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

3.3 Shannon Entropy

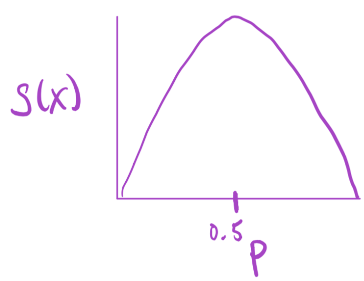
The class seems to agree that the time to fixation when $N = 20$ and $p = 0.5$ is around 27. We will now derive the results.

Firstly, the Shannon entropy is defined as

$$S(X) = -\sum p_i \log_i(p_i) \tag{20}$$

where $p_i = \dfrac{x_i}{N}$. When $p = 0.5$, the entropy is at its peak (Figure 2).

Figure 2: Shannon Entropy



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

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

The partial derivatives of $S(X)$ are

$$\begin{aligned} \frac{\partial S(X)}{x_i} &= -\frac{1}{N}(1 + \log(\frac{x_i}{N})) \\ \frac{\partial^2 S(X)}{\partial x_i \partial x_j} &= -\frac{\delta_{ij}}{N x_j} \end{aligned} \tag{21}$$

where

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \tag{22}$$

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

$$\langle j \rangle = np \tag{23}$$

and the variance is

$$\begin{aligned} Var\,\langle j \rangle &= \left\langle j^2 \right\rangle - \langle j \rangle^2 & (24) \\ &= np(1-p) & (25) \end{aligned}$$

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.

$$\begin{aligned}
\langle S(x+\Delta x) \rangle &\approx \left\langle S(x) + \sum_i^k \frac{\partial S(x)}{\partial x} \Delta x_i + \frac{1}{2} \sum_i^k \sum_i^k \frac{\partial^2 S(x)}{\partial x_i \partial x_j} \right\rangle & (26) \\
&= S(x) + \sum_i^k \frac{\partial S(x)}{\partial x_i} \langle \Delta x_i \rangle + \frac{1}{2} \sum_i^k \sum_j^k \frac{\partial^2 S(x)}{\partial^2 x_i \partial^2 x_j} \langle \Delta x_i \Delta x_j \rangle & (27) \\
&= S(x) + \frac{1}{2} \sum_i^k \sum_j^k \frac{\partial^2 S(x)}{\partial^2 x_i \partial^2 x_j} \langle \Delta x_i \Delta x_j \rangle & (28) \\
&= S(x) - \frac{1}{2} \sum_i^k \sum_j^k \frac{\delta_{ij}}{N x_j} \langle \Delta x_i \Delta x_j \rangle & (29) \\
&= S(x) - \frac{1}{2} \sum_i^k \frac{1}{N x_i} \langle \Delta x_i \Delta x_i \rangle & (30) \\
&= S(x) - \frac{1}{2} \sum_i^k \frac{1}{N x_i} \left(N(\frac{x_i}{N})(1 - \frac{x_i}{N}) \right) & (31) \\
&= S(x) - \frac{1}{2N} \sum_i^k (1 - \frac{x_i}{N}) & (32) \\
&= S(x) - \frac{k-1}{2N} & (33)
\end{aligned}$$

Remarks:

- Taking the partial derivative of x_i in (27) results in $\langle \Delta x_i \rangle = 0$.
- For (28), substitute with (21) to get (29).
- For (29), recall that $\delta_{ij} = 1$ when $i = j$ from (22).

- For (30), apply the variance equation (25) where $p = \frac{x_i}{N}$ to get (31).
- For (31), pull out the N from the summation to get (32).
- For (32), remember that $\sum_i^k 1 = k$ and $\sum_i^k x_i = N$, which simplifies (32) into (33).

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

$$\langle \Delta S(X) \rangle = \langle S(x + \Delta x) \rangle - S(x) = -\frac{k-1}{2N} \tag{34}$$

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

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

where, after substituting in (34),

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