Homework 6. [1]

Pankaj Mehta

May 3, 2021

[1] **Due on Thursday May 6th. NO EXTENSIONS**

The goal of this HW is to think about topics we have touched open in our chapter on viral evolution.

1. **COVID-19 strain fitness**

In class, we saw that we could estimate the fitness of a variant by making a straight-line plot for an allele where the y-axis was in logit space and the x-axis was time. Here, we will use recent data from India to use this technique to estimate the fitness.

(a) Use the NextStrain database to plot the frequency of the three

strains B.1.617, B.1.1.7,B.1.351 in India from Jan 1 2021 to the beginning of April? We will make the approximation that we can think of this as a two-allele problem (the allele of interest and everything else).

(b) Use this to estimate the fitness of the strains? How big different are the fitness differences?

(c) Based on our discussions in class, which strains of COVID-19 would you estimate are the fittest globally based on the genealogical trees? You can do this qualitatively. Check your prediction at the end of the summer when hopefully the pandemic is over in the US :)– but unfortunately probably not the global South since the greedy Pharma companies are bullying the government into extending their patent rights even though this technology was largely publicaly funded.

2. **Working with sequence data**

Please do Problem 1 from Problem Set 1 in from Ben Goods class `https://bgoodlab.github.io/courses/apphys237/`.

3. **Working with Traveling Waves**

In this problem, we will gain more experience with the mathematics of traveling waves. We will analyze the one dimensional Burgers equation for a field of the form $u(x,t)$

$$\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2} - u \frac{\partial u}{\partial x}, \tag{1}$$

with boundary conditions that in the far past $u(x = -\infty, t) = 1$, $u(x = \infty, t) = 0$, and $\partial_x u(x = -\infty, t) = \partial_x u(x = \infty, t) = 0$.

(a) Look for a traveling wave solution by substituting $u(x,t) =$

$U(z) = U(x - ct)$ into the equation above. Show that the U satisfies the equation

$$-cU = \alpha U' - \frac{U^2}{2}, \tag{2}$$

where U' denotes a derivative with respect to $z = x - ct$.

(b) Use the boundary conditions to show that the wave speed must equal $c = 1/2$?

(c) Use these results to show the traveling wave solution takes the form

$$u(x,t) = \frac{U_0}{U_0 + (1 - U_0)e^{\frac{x-ct}{2\alpha}}}, \tag{3}$$

where U_0 is value of $U_0 = U(0)$. Why is there still a free parameter in this solution? Plot the solution as function of x and t for $U_0 = 1/2$.

4. **Properties of Kingman Coalescent**

We showed that the Kingman coalescent has very nice and simple properties. For example, we showed that if there are n sequences for a community with effective population size N, the probability that two of these sequences will coalesce in the previous generations is given by

$$p = \binom{n}{2} \frac{1}{N}. \tag{4}$$

(a) Let T_n be a random variable which denotes the probability that two sequences will coalesce after j generations. Note that T_n take values in the integer $j = 1, 2, \ldots$ Show that T_n follows the probability distribution

$$prob[T_n = j] = p(1-p)^j \tag{5}$$

To calculate properties of this distribution, it useful to define a generating function

$$G(z) = \sum_{j=1}^{\infty} p(1-p)^{j-1} e^{-zj} \tag{6}$$

(b) Show that the expectation of the first two moments of T_n can be written as

$$E[T_n] = -\partial_z G(z)|_{z=0} \tag{7}$$

$$E[(T_n)^2] = \partial_z^2 G(z)|_{z=0}. \tag{8}$$

and use these expressions to calculate the mean and variance of the T_n.

(c) Argue that the time T_{MCRA} to the most common ancestor of the n sequences is a random variable of the form

$$T_{\text{MCRA}} = \sum_{i=2}^{n} T_i. \tag{9}$$

Use the fact that each coalescent event is independent in neural dynamics to calculate the mean and variance of the T_{MCRA}.