Problem Set 3 for Biomath 202: Due May 17, 2017

1. For a random walker starting from the origin on a lattice with spacing 1 and spatial sites \(i\) within the set \((-n,-(n-1),...,-1,0,1,...,n-1,n)\), where \(n\) is time or the total number of steps taken since the clock started, explicitly write out the non-zero probabilities for being at all sites \(i\) for the first 6 time steps, assuming only one spatial step can be taken per time step.

2. a. Explain why the non-zero probabilities in 1 can be expressed as

\[
P(i,n) = \frac{1}{2^n} \binom{n+i}{\frac{n-i}{2}} \binom{n-i}{\frac{n-i}{2}}
\]

b. Show that \(\sum_i P(i,n) = 1\).

c. The probabilities for a random walk by definition follow the rule

\[
P(i,n) = \frac{1}{2} P(i-1,n-1) + \frac{1}{2} P(i+1,n-1).
\]
Show that our expression in 2a satisfies this rule.

3. Using results from problem 2, show that the expected value of the position, \(\langle i \rangle\), is zero, and that the variance is given by \(n\). The latter is straightforward if the distribution given in problem 2 is converted into one that we have already studied in class. Because \(n\) is the number of time steps, this shows that the square displacement goes linearly with time, a distinct signature of diffusion.

4. a. What is the mean time or expectation value for the time in the past for 3 loci to coalesce into a single common ancestor loci? The population size is \(2N\), and you should assume \(2N \gg t \gg 1\). Assume only genetic drift and thus the Wright-Fisher model. You can find the answer in the Rice chapter but carefully explain and derive why.

b. For \(k\) loci, derive that the mean time or expectation value of the time for the first coalescence of a pair of loci from the set of \(k\) loci is given by \(2N/\binom{k}{2}\), where the denominator is \(k\) choose 2.

c. Using the result from part b, iterate to find the total expected time for all \(k\) loci to coalesce into a single common ancestor. Show that it is \(4N \sum_{i=2}^{k} \frac{1}{(i-1)}\).

d. Prove that the sum in c. is equal to \(\frac{k-1}{k}\). (Hint: Use partial fractions or induction.) What is the expected time for all \(k\) loci to coalesce in the limit \(k \gg 1\)? How does this time compare to the time for just a single pair (\(k=2\)) to coalesce?
5. Show explicitly that \( P_n = P_0 \prod_{j=1}^{n-1} \frac{b_i}{d_{i+1}} \) is a solution to \( P_{n+1} d_{n+1} + P_{n-1} b_{n-1} - P_n (b_n + d_n) = 0 \) as in Volkov et al.

6. Given the three series

\[
\begin{align*}
  u &= 1 + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots \\
  v &= x + \frac{x^4}{4!} + \frac{x^7}{7!} + \cdots \\
  w &= \frac{x^2}{2!} + \frac{x^5}{5!} + \frac{x^8}{8!} + \cdots
\end{align*}
\]

show that \( u^3 + v^3 + w^3 - 3uvw = 1 \).

7.a. As will be explained in class, the fundamental signature of diffusion is that mean spatial distance from the origin increases like the square root of time, \( x \propto \sqrt{t} \). Based on this biological/physical/chemical intuition, one might guess that a solution to the diffusion equation would depend on a single parameter that encapsulates this relationship, such as the ratio \( \psi = x/\sqrt{t} \). Alternatively, instead of a ratio, one might guess it depends on the difference of these factors, such as \( x - c \sqrt{t} \), which can also be expressed as \( \sqrt{t} \left( \frac{x}{\sqrt{t}} - c \right) = \sqrt{t} g(\psi) \). Consequently, as a general guess for a solution, one might try \( f(x,t) = t^p g(\psi) \), where \( p \) is an arbitrary but constant exponent to be determined. Substitute \( f(x,t) = t^p g(\psi) \) into the diffusion equation,

\[
\frac{\partial f(x,t)}{\partial t} = D \frac{\partial^2 f(x,t)}{\partial x^2}
\]

which is a partial differential equation, and use the chain rule to show you can write this as an ordinary differential equation in the variable \( \psi \).

b. Show when \( p = -1/2 \), this differential equation can be expressed as

\[
\frac{d}{d\psi} \left[ \frac{\psi}{2} g(\psi) + D \frac{d g(\psi)}{d\psi} \right] = 0
\]

It turns out this is the only choice of \( p \) that makes the equation easily solvable.

c. Solve this equation and find the solution for the case that you have the boundary condition

\[
\frac{d g(\psi)}{d\psi} \bigg|_{\psi=0} = 0
\]
meaning that at the origin \((x=0)\) the solution is at a maximum (because the derivative is zero), and this makes sense because the origin is the most probable position for the random walker to be as time moves forward. That is, the distribution of possible locations is symmetric about the origin (because there is nothing to break this symmetry), so the origin is the expectation value or the mean value for the location of the random walker and also the most likely or maximally likely location because on average the walker keeps going back and forth from one side to the other.

d. Partial differential equations are solved using a variety of methods, including the one above, but also one known as separation of variables in which you would guess or look for a solution of the form \(f(x, t) = h(x)u(t)\) where \(h\) and \(u\) are unspecified, separable functions of only \(x\) and \(t\) respectively. Can your solution to part c. be found using this method? Or do you think a different solution could be found using this method? You do not need to actually find the solution.

8. Consider a network that is grid-like, such that nearest neighbors are all connected and each node has three edges connected to it. Starting from some central location, imagine flow of material can happen either to the “right” or “left”. When the flow reaches the next node, it cannot travel back to the node from which it came, and it must go the opposite direction (right versus left or vice versa) than it did at the previous node. As the flow of material continues to traverse the network, it must alternate directions through edges forever after that. Prove that if this pattern continues that the flow of material must eventually return to the starting node. This is equivalent to proving that the flow cannot get stuck traveling forever through a loop of edges and nodes of which the starting node is not a part. (Hint: Consider the “dual” network constructed by replacing each vessel with a node at its mid-point and connecting these new nodes only if the original edges were connected at an original node. Map the flow through the original network onto what flow means through the new “dual” network and consider what this means about loops. In the figure below, the first diagram is a path through the original network, and the second diagram is a path through the dual network.)