# Cluster 5 (Biophysics) Challenge Problems (2018)

***Overview***

The challenge problems are intended as fun exercises for those students who finish the programming exercises given in class and wnat to try something a bit harder. Some of these problems are very hard, so don't get discouraged if you need help or run out of time. We will probably not have time to discuss the detailed solutions in class, but will share solutions with those who attempt them and ask.

***Challenge Problem in probability***

Samuel Pepys sent a letter to Isaac Newton (11/22/1693) with the following query:

"I have a Question concerning the Doctrine of determining the true proportions of the Hazards incident to this or that given Chance or Lot :

A- has 6 dice in a Box, with which he is to fling one 6

B- has in another Box 12 Dice, with which he is to fling two Sixes

C- has in another Box 18 Dice, with which he is to fling three Sixes

Q. whether B & C have not as easy a Taske as A, at even luck?"

How did Newton answer?

Challenge: Write a python program to calculate the probability for each outcome and answer Samuel Pepys' question

***Challenge Problem for gambler's ruin***

A crude model of economics is this: You have a collection of N people. Each has a certain number of dollars from $1, up to M dollars.

Pairs of people meet randomly. Let's say person 1 has M1 dollars and person 2 has M2 dollars and that M1 is less than M2. They flip a fair coin and the winner takes 1, 2, ... up to M1 dollars from the loser. That is to say, the amount in each exchange is randomly selected from $1 up to $M1. Note that if a person reaches(has) zero dollars, he/she is no longer able to enter exchanges and cannot participate in games.

Question 1: What is the ultimate distribution of money in the game? Question 2: What sets the time scale to reach that distribution?

Challenge: Write a python program which explores and answers Question 1 and 2 above.

***Challenge for RNA_to_protein translation***

Use/modify your RNA_translation program from class to access a public DNA databases:

The following link will generate a random DNA strand of any length that you may copy into your code and test:

`http://www.bioinformatics.org/sms2/random_coding_dna.html`

Generate several 1000+ length samples and test with your protein translation code. Feel free to explore this site for other types of samples.

The next site is a government sponsored database of gene (nucleotide) sequences from different life forms. Feel free to copy one of these gene sequences and test out on your program. Be careful- don't pick something too large- keep it to less than 5000 base pairs.

`https://www.ncbi.nlm.nih.gov/nuccore/?term=eme1`

***Challenge problem for 2D Random Walk***

Modify the randwalk2d_mod.py module to allow the shape of the confining boundary and/ or the target to be a circle or cross (overlapping rectangles).

Currently only a 'box' is supported(rectangle).

Allow user to set up and draw the shapes as needed in the calling program (randwalk2d_target.py)

***Challenge for DNA-HIV Mutations/ Probability***

1. An HIV genome is $N = 10^4$ base long. A mutation is any change in the genome. A one-base mutation, also called single mutation, is when exactly one base is mutated. A two-base mutation is one where exactly two bases are mutated. A three-base mutation is one where exactly three bases are mutated.

a. Find the number of different one base mutations of the HIV genome. b. Find number of different two-base mutations of the HIV genome. c. Find number of different three-base mutation of the HIV genome.

2. Now, imagine that in asymptomatic HIV patient, $10^{10}$ new virus particles are formed each day. The process of copying DNA has an error rate of one in $3 * 10^4$.

a. How many new virus particles are produced everyday with one-base mutation? b. How many new virus particles are produced everyday with two-base mutation? c. How many new virus particles are produced everyday with three-base mutation?

3. About one percent of the virus particles produced each day go on to infect new white blood cells.

a. Find the expected number of white blood cells that get one-base mutation. a. Find the expected number of white blood cells that get two-base mutation. a. Find the expected number of white blood cells that get three-base mutation.

4. Suppose an antiviral drug attacks some part of HIV but that the virus can evade a drugs effect by making one particular, single-base mutation. Show that the drug will very quickly stumble upon the right mutation and the drug isnt effective for very long.

5. Why do you suppose an effective HIV therapy involves a combination of three different antiviral drugs administered simultaneously?

Challenge: Use probaility and write python programs to answers thes questions

***Challenge for triangle monte carlo/logic***

Consider a circle of radius 1. Imagine choosing three points $A, B, C$ randomly on this circle (see the figure below). The object of this problem is to determine the probability that the triangle $ABC$ contains the center of the circle $O$. There are at least two good ways to approach this problem:

1. Carry out this task with a computer by using a Monte Carlo-style approach. This involves writing a program which repeatedly generates the triangle $ABC$, determines whether or not $ABC$ contains $O$, and averaging the results.

2. Prove your result from (1) using any math available to you. Basic probability and logic should be sufficient, although there are many possible approaches.

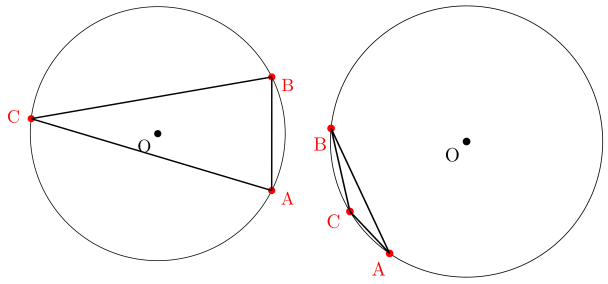


FIG. 1: Left: An example of a triangle containing $O$. Right: An example of a triangle which does not contain $O$.