# Assignment 6:

## Instructions

Complete the following problem set showing your work. Problems may be worked out "by hand" or in "python" or with the assistance of other analytical software (e.g., Mathematica, MatLab). You may use chatGPT to assist in coding.

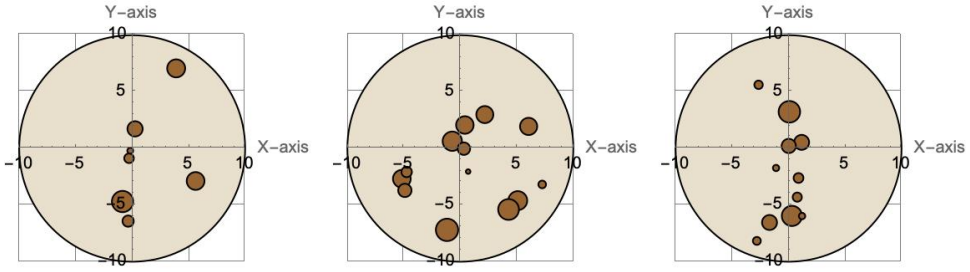
Solutions must be type written (e.g., in Jupyter, markdown, or latex). Upload PDF solution by question to crowdmark (link will be emailed to you) by **11:59pm** on the **Sunday** of the corresponding week (see syllabus). If you have issues with Crowdmark submission please email solutions to Rebekah Hall (rah11@sfu.ca).

All problems are equally weighted within an assignment. Students in 468 may or may not choose to attempt the challenge question for a bonus pts. Students in 795 are required to complete the challenge question.

## Problem Set

### 1.Modelling Microbial Colony Growth with Cellular Automata

Imagine you are helping a colleague in microbiology understand the growth of their bacterial colonies on a petri dishes. Your colleague is growing a bacteria that can either divide, die, and enter/exit a dormant state/ Your goal is to create CA simulation that provides insights into the rates at which these different events occur. The results of three replicates of their experiment are shown below.




**Part A** Define the grid and cells for your cellular automata. Specify what each cell in the grid represents in your model and the possible states a cell can have. Include states of "Active," "Dead," "Empty," and "Dormant". Are you going to use a *Von Neumann* or *Moore* neighbourhood.

**Part B** Establish rules that govern the behaviour of the microbial cells consistent with the meaning of the states above. Consider factors such as nutrient availability and cell density.

**Part C** Discuss how you would initialize the simulation, including the initial configuration of cells. Additionally, address the boundary conditions of your CA. Discuss how you would pick the time span/stopping conditions of your simulation?

**Part D Challenge for 795:** Develop the CA you outlined above and show the result of 1 simulation replicate.

### 2. Individual based predator-prey simulation.

You are helping a wildlife ecologist understand the dynamics of the caribou population of the Northwest Territories. Within this population, caribou live and travel in herds. Their major predators are wolves who themselves move and hunt in packs. Caribou reproduce annually once they reach reproductive maturity. Juvenile caribou are much more likely to be caught by predators than adults. Individual wolves must consume a certain amount of prey 'energy' to survive. Packs are dominated by an alpha male, subordinate beta males, and females. A wolfs ability to mate and access to food are determined by sex and social hierarchy within the herd.

**Part A.** Sketch the structure of an individual based simulation of caribou/wolf population dynamics. What would you use as your classes? What variables (and their data types) would you prescribe to each class? What are some class functions you would need? What global functions will you need?

**Part B** Sketch some plots of the type data you would like to create with your IBS.

### 3. Likelihood and Bayesian Inference

A GC island, also known as CpG island, is a genomic region characterized by a higher-than-expected frequency of the DNA bases guanine (G) and cytosine (C) pairs (e.g., sequences with repeats of the form GCGC). You have data from DNA sequencing reads, 200-300 base pair long sequences. Your goal is to infer the proportion of the genome enriched for CpG islands.

Suppose you model the genome with the following stochastic process:

**Part A:** Write down the transition probability matrix for the stochastic process shown above.

**Part B** You are given a read with $n$ base pairs as denoted by the sequence $x_i \quad i \in \{1, 2, \ldots n\}$ where $x_i \in \{A = 1, C = 2, G = 3, T = 4$. Express the likelihood of observing a given read under the stochastic model shown above. What is the "Data", $\mathcal{D}$, in this case? What is the "model", $\Theta$?

**Part C:** Explain why it might be best to use the log-likelihood in model fitting for this problem.

**Part D:** If using a Bayesian inference framework what priors and hyper-priors might you want to use for this problem? Hint: consider what you think the effect of CpG island should be on the transition rates. Also we have disucssed several models of molecular evolution in this class, how might you incorporate one of these with a hyperprior?