The Stepping Stone model – Part 1 – Simulations PiE2 - Fall 2020

One of the multiple applications of Markov chains is to genetic evolution. The following model is known as *stepping stones* and was introduced by Kimura in 1953. The process is also known as the *voter model* as it can be used to simulate peer-influence on opinions among a population.

Suppose we look at a given gene in a population that can take k different values, known as alleles. For simplicity, we will assume that the values are $\{0, 1, \ldots, k-1\}$ and think about them as colours. We want to understand the behaviour of gene mutation. Given $N \geq k$, consider an $N \times N$ array, where every cell contains an individual with certain allele. We will assume that our world, represented by the array, is toroidal: that is, the first row is adjacent to the last one, and the first column is adjacent to the last one (this deletes annoying boundary effects). At a given time, a uniformly random cell is selected, its individual dies, and a new individual is born in the same cell, taking the allele of one of its four adjacent cells (right, left, top, bottom), each with probability $1/4$. You may also think as the individual at the chosen cell "changing" their state. The process stops when all the individuals have the same allele. If the final allele is c; we call it c-domination.

In Part 1, you will simulate the random model and make some conjectures. In Part 2 (forthcoming), you will formally analyse it using the stochastic process tools given in the lectures. Use your favourite coding language/mathematical software to code a stepping stones simulator. You will use it to answer the following questions. **You must justify your answers using data extracted from simulations. You must include an example of your code as an appendix**. Let $N = 10$ and $k = 2$, unless otherwise stated.

a) Suppose that initially there are 50 0s and 50 1s. Try to maximise the probability of 0-domination by choosing an initial configuration. What is the maximum probability you can obtain? which is the best configuration?

b) Repeat the experiment in a), but now start with 90 0s. What are your conclusions?

c) We call an array *sorted*, if all 0s appear before all 1s (when we read rows from top to bottom, and each row from left to right). For $i \in \{0, \ldots, N^2\}$, plot the probability of 0-domination if the initial array is sorted with i 0s. Make a conjecture about it.

d) For $i \in \{0, \ldots, N^2\}$, plot the probability of 0-domination if the initial array is randomly generated with i 0s. Make a conjecture as general as possible.

e) Let $k = 4$. Repeat the experiment in c) and d) and plot the results. What is the expected effect of increasing k to the probability of 0-domination?

f) Let $k = 2$ again. Starting with (1) a sorted and (2) a random initial array with half the cells containing 0s, plot the time until domination for different values of N (make N as large as possible). Discuss the differences, if observed. From the data observed, make a conjecture on how should the average time to domination depend on N, stating with the random array.