Stochastic Models in Biology: Computational Challenges

Krešimir Josić and Alexander Stewart.

Computational Challenge 1

Rejection sampling is a useful method for generating random samples from a distribution $f(x)$ that could be hard to sample from direclty. The idea is to generate samples from a known distribution $g(x)$, and then either to accept or reject the samples. The criterion for acceptance and rejection, and the algorithm are described here:

https://bookdown.org/rdpeng/advstatcomp/rejection-sampling.html

Use either the algorithm described in sections 6.3.1 or 6.3.3 on this webpage to do the following

Group 1 Use the uniform density on the unit interval as a candidate density to generate samples from the a beta distribution.

Group 2 Use the normal distribution as a candidate density to generate samples from the t-distribution.

Group 3 Use the Poisson distribution as a candidate distribution to generate samples from the binomial distribution.

Group 4 Use the normal distribution as a candidate distribution to generate samples from a probability distribution composed of the sum of three other normal distributions with means $\mu_1 = -1$, $\mu_2 = 0$ and $\mu_3 = 1$ and standard deviations $\sigma_1 = \sigma_2 = \sigma_3 = 0.5$.

Questions for discussion:

1. When do you think the method becomes inefficient?

2. Do you think that rejection sampling can be extended to higher dimensions? How?

Computational Challenge 2

In this computational challenge you will implement one of two Markov chain models of a biological system:

Group 1, 2: Quasi-stationary distributions in the birth-death model. In a general birth and death process, the probability that the population increases or decreases between steps can depend on the population size. Denote the size of a population at time t by X_t. The transition probabilities are then given by:

$$P(X_{t+1} = j + 1 | X_t = j) = b_j$$
$$P(X_{t+1} = j - 1 | X_t = j) = d_j$$
$$P(X_{t+1} = j | X_t = j) = 1 - (b_j + d_j)$$

where additionally $d_0 = 0$ and $b_0 = 0$, so that the state of the population having no individuals is absorbing. Assume $\sup_i(b_i + d_i) \leq 1$. Note that all states, except for 0, are transient. To keep the population finite also assume that $b_k = 0$ for all $k \geq N$ and some $N > 1$.

If you don't remember the logistic growth model covered in most calculus classes, refresh your memory by reading the relevant Wikipedia page, or consulting any other resource. For a logistic growth process we can assume $b_i - d_i = ri(1 - i/K)$. Note that any choices of b_i and d_i that satisfy this equation are consistent with logistic growth because the expected increment over a time interval is $ri(1 - i/K)$.

Assume that $b_i = r(i - i^2/2K)$ and $d_i = ri^2/2K$, so that the maximal population size is $N = 2K$ (otherwise transition probabilities become negative), and the carrying capacity of the environment is K.

- Implement computationally the Markov chain defined by these transition probabilities. Choose $r = 0.004$, $K = 50$. Plot several trajectories to get a feel for the behavior of the population. After several thousand timesteps the population size should fluctuate around 50. Run the simulation up to $t = 3000$ (or a time of your choices where the mean stabilizes) a number of times to create a histogram of the population size at that time.

- Set $r = 0.015$ and $K = 8$. Run the simulation until the population goes extinct starting at different initial population sizes. Plot the mean time to extinction.

- Discuss how the answers to the previous two questions can be consistent: The first suggests that the population reaches a stable distribution around the carrying capacity. The second suggests that the population goes extinct.

Group 3, 4: Fixation in the Moran model. The Moran model describes the frequency of different genes in a population of fixed size, N. Here we will consider a haploid population with a genotype composed of a single locus, which can be occupied by one of two alleles, a and A. We will assume that both alleles are present in the population, and that there is no selection for or against either allele.

Denote the number of individuals with allele a by j, and the number of individuals with allele A by $N - j$. The Moran model assumes that birth and death occur simultaneously, as a single event. First an individual is chosen to reproduce with mutation. A second individual is then chosen to die. This birth-death event constitutes a single step in the Moran process. Denote the number of individuals with allele a at time t by X_t. The transition probabilities for this Moran process are given by:

$$P(X_{t+1} = j + 1|X_t = j) = \left(\frac{j}{N}(1 - \mu_a) + \frac{N - j}{N}\mu_A\right) \times \left(\frac{N - j}{N}\right)$$

$$P(X_{t+1} = j - 1|X_t = j) = \left(\frac{N - j}{N}(1 - \mu_A) + \frac{j}{N}\mu_a\right) \times \left(\frac{j}{N}\right)$$

$$P(X_{t+1} = j|X_t = j) = 1 - \left(\frac{j}{N}(1 - \mu_a) + \frac{N - j}{N}\mu_A\right) \times \left(\frac{N - j}{N}\right) - \left(\frac{N - j}{N}(1 - \mu_A) + \frac{j}{N}\mu_a\right) \times \left(\frac{j}{N}\right)$$

where μ_a is the probability that allele a mutates to allele A during reproduction and μ_A is the probability that allele A mutates to allele a.

- Assume that $\mu_a = \mu_A = 0$. In this case the Moran model has two absorbing states at $j = 0$ and $j = N$, *i.e.* one state in which all individuals have allele A and one in which all individuals have allele a. Numerically estimate the probability that the Moran process initialized at $X_0 = 1$ will eventually reach fixation for a, i.e. the absorbing state in which all individuals have allele a.

- Assume again that $\mu_a = \mu_A = 0$. Conditional on a reaching fixation, estimate the time for a to become fixed for a Moran process initialized at $X_0 = 1$.

- Numerically estimate the stationary distribution for the Moran process with (i) $\mu_a = \mu_A = 0.1/N$, (ii) $\mu_a = \mu_A = 1/N$ and (iii) $\mu_a = \mu_A = 10/N$.

- Briefly discuss how adding selection for and against a might change these results.

Computational Challenge 3

You will be implementing a Yule process, where the number of births is proportional to the size of the population, and a non-homogeneous perturbation of the process modeling, *e.g.* seasonal variability in the birth rate:

1. Implement the Yule process with fixed rate $\nu = 0.4n$, where n is the size of the population.

2. Implement the Yule process with rate $\nu(t) = 0.4n + 2\sin(t)$.

Simulate the population on the time interval $[0, 100]$. In all cases show the distribution of population sizes at $t = 10, 50$ and 100. Start with $n(0) = 1$.

Group 1 Implement the two stochastic processes by discretizing the time interval $[0, 100]$ into increments Δt. Use $\nu(t)\Delta t$ as the approximate probability that a birth occurs in the interval. Once you implement the process answer the following questions:

    (a) How should you choose the increment Δt for this approach to samples that are close to those from the actual distribution defined by the Yule process? Can this sampling approach be exact?

    (b) Can you suggest an adaptive method so that the increment size changes, increasing the efficiency of the sampling procedure without sacrificing accuracy?

Group 2 Implement the two stochastic processes by sampling the time increment to the next event. In the first case the time to the next event is exponentially distributed. Is this true in the second case?

The next two groups will be implementing different approximations to this process known as τ-leaping and K-leaping. You will need to coordinate a bit with groups 1 and 2 to make some comparisons

Group 3 Implement τ-leaping [2] by doing the following: Discretize the time into intervals of length $\tau > 0$, and assume that the birth rate in each interval is constant and determined by the population size at the beginning of the interval. Thus if the population at some time $t = j\tau$ equals $n(t)$, then the number of births in the interval $[j\tau, (j + 1)\tau)$ is approximated by a Poisson distribution with parameter $\tau\nu(t)$. Note that $\nu(t)$ depends on population size at time t.

3

(a) Discuss how this sampling approach relates to the one developed by group 1. How do the approximate population distributions compare to the different values of τ?

(b) For an accurate simulation, what do you think the size of the increments τ should depend on?

Group 4 The idea of K-leaping [1] is the following: Keep the population increments, K fixed, but vary the time over which they happen. In the case of an increment by $K = 1$, the method reduces to the exact method implemented by Group 2. When $K > 1$, the time increment is the sum of exponentially distributed random variables with different parameters. To make this tractable, when doing K-leaping we approximate the distribution of a sum of K exponentially distributed random variables with different parameters with the distribution of a sum of exponentially distributed random variables with the same parameter. The easiest thing to do is to choose $\nu(t)$ where t is the endpoint of the previous time increment to determine the parameter of the variables in the sum. Remember that a sum of exponential random variables with the same parameter follows a gamma distribution.

(a) Discuss how this sampling approach relates to the one developed by group 2. How do the approximate population distributions compare to the different values of K?

(b) For an accurate simulation, is there a way to change K adaptively to make this approximation better?

Computational Challenge 4

In this challenge you will implement stochastic simulations of two synthetic biological circuits. The synthetic genetic switch, and the repressilator were described in two back–to–back papers in *Nature* in 2000 [3, 4].

In this challenge two groups will report on the deterministic version of models of these circuits, while the other two will report on the stochastic versions. The last part of each question asks that you compare the results of the simulations. Therefore groups 1 and 2, and groups 3 and 4 will need to work together on each project.

Group 1 You will analyze the deterministic version of the genetic switch system described in section III of the paper by Loinger, et al. [7]. The differential equation is given in Eq. (9).

- Construct a Petri net representation of the system based on the description in the text (you can also consider the master equation given in the same section). How does the ODE relate to this representation? Why is this called an exclusive switch?

- Solve the ODE numerically and show the different trajectories in the phase plane. Use the parameters in the paper, and coordinate with Group 2. The nullclines are the solution to the equations $A' = 0$ and $B' = 0$. As discussed in the paper by Gardner, et al. [4] the nullclines determine the behavior of the deterministic system (See Fig. 2). Perform the equivalent analysis for the system in Eq. (9).

Group 2 Consider the exclusive switch discussed in section III of the paper by Loinger, et al. [7]. The master equation for this system is given in Eq. (14)

- Construct a Petri net representation of the system based on the description in the text and the master equation.

- Set up a stochastic simulation of this system. Use the parameters discussed in the text, and change k from 0.005 to 50, as in figure 4. What happens to the switching times between the two states as k is increased? You can use the Gillespie algorithm. Show the traces for N_A and N_B in three or four cases. Discuss whether the change is as expected.
- Together with Group 1 explain whether or whether not these results can be explained using the deterministic equations.

Group 3 Consider the deterministic repressilator described via the Michaelis-Menten equations in the second paper by Loinger, et al [6].

- Construct a Petri net representation of the system for case (i) with mRNA (i.e. corresponding to Eq. 1) and case (ii) without mRNA (i.e. corresponding to Eq.2)
- Solve the ODEs for case (i) numerically with parameters $k = d_m = 1$ and $d_p = g_p$. Plot phase portraits showing the region in which the system displays oscillations, with values of g_p on the x-axis plotted against values of g_m on the y-axis, for Hill coefficients $n = [1, 2, 3]$ (i.e three different phase plots, one for each value of n). Repeat the same exercise for case (ii)

Group 4 Consider the stochastic repressilator described in Loinger, et al [6] and in Elowitz and Leibler [3].

- Using both repressilator papers as a guide, write down a Master equation describing the stochastic version of Eq. 1 of Loinger (i.e a simple repressilator with reactions as described in the Michaelis-Menten kinetics section of that paper).
- Suggest a method for detecting oscillation-like behavior in a stochastic process. Apply your method to stochastic simulations of the simple repressilator for some of the same cases considered by Group 3 above (i.e., you do not need attempt a full phase portrait, but choose a few parameter sets that correspond to those explored by group 3, and try to find at least one case where oscillations appear to be happening).

Computational Challenge 5

In this challenge two groups will report on the deterministic version of models of these circuits, while the other two will report on the stochastic versions. The last part of each question asks that you compare the results of the simulations. Therefore groups 1 and 2, and groups 3 and 4 will need to work together on each project.

Group 1 Genetic circuits that oscillate with a period of approximately 24 hours are widespread across different species. The details of the oscillators vary considerably between species, but the ones have been characterized rely on some for of negative auto-regulation. The number of molecules involved in these oscillations also varies greatly between species.

One of the most influential models of circadian clocks is the Goodwin Oscillator. It takes the form

$$
\begin{aligned}
X' &= \alpha_1 \frac{K^N}{K^N + Z^N} - \gamma_1 X \\
Y' &= \alpha_2 X - \gamma_2 Y \\
Z' &= \alpha_3 Y - \gamma_3 Z
\end{aligned}
$$

(a) Implement the oscillator by solving the ODEs for the values in the paper by Gonze and Ruoff (see Fig. 2) [5]. Show the trajectories for all the three species.

(b) Work with Group 2 to construct the corresponding Petri Net. Assume that the equations are given for concentrations. What are reasonable values for the reaction rates so that the molecule number stays within at most a few hundred molecules in the SSA (to be implemented by Group 2)?

(c) Describe the negative auto-regulation intuitively. What is the main negative auto-regulation term? Where does the delay come from and why is it important?

(d) Change the Hill coefficient, n, to a value smaller than 8. What happens to the oscillations? Why do you think that this happens.

Group 2 First work with Group 1 to understand the oscillations in this circuit, and on part 2 of their problem. Once you have the reactions and candidate rates do the following:

(a) Implement a stochastic simulation corresponding to the system above using the Gillespie Algorithm. You may need to adjust the rates to get oscillations, and keep the number of molecules at a reasonable level (more than 10 or so, and less than a few hundred).

(b) Change the Hill coefficient, n to a value smaller than 8. What happens to the oscillations? Send a message to Group 1 to see what they observed, and compare notes.

(c) Compare the frequency of oscillations with those obtained by Group 1.

Group 3 A feed-forward loop (FFL) is a 3-gene genetic circuit in which two "input" transcription factors (TFs), X and Y, regulate a third "target" Z, with TF X also regulating TF Y. Depending on the pattern of activation and repression among the three genes, an FFL can speed or slow the response time of Z to a stimulus in the form of X switching ON or OFF. Here we will explore the behavor of the so-called "incoherent" FFL.

We will assume that X is either OFF or ON such that the concentration of the product of gene X is X^* where $X^* = 1$ when the gene is ON and $X^* = 0$ when the gene is OFF. We will look at two specific cases of an incoherent FFL.

CASE I: In the first case, the concentrations of gene product Y and Z are described by the following set of ODEs:

$$\frac{dY}{dt} = B_y + \beta_y \frac{(X^*/K_{xy})^h}{1+(X^*/K_{xy})^h} - \alpha_y Y$$

$$\frac{dZ}{dt} = B_z + \beta_z \frac{(X^*/K_{xz})^h}{1+(X^*/K_{xz})^h} \times \frac{1}{1+(Y/K_{yz})^h} - \alpha_z Z$$

CASE II: In the second case, the concentrations of gene product Y and Z are described by the following set of ODEs:

$$\frac{dY}{dt} = B_y + \beta_y \frac{1}{1 + (X^*/K_{xy})^h} - \alpha_y Y$$

$$\frac{dZ}{dt} = B_z + \beta_z \frac{1}{1 + (X^*/K_{xz})^h} \times \frac{1}{1 + (Y/K_{yz})^h} - \alpha_z Z$$

(a) Draw network diagrams showing genes as nodes and regulatory interactions as edges for CASE I and CASE II. Label the edges as either activation or repression.

(b) Use the parameters $B_y = B_z = 0$, $\alpha_y = \alpha_z = 1$, $\beta_y = \beta_z = 1$, $K_{xz} = K_{xy} = 0.1$ and $h = 2$. Consider the FFL in CASE I. Compare the behavior of the FFL with $K_{yz} = 0.01$ (strong $Y - Z$ regulation) to the same FFL with $K_{yz} = 10$ (very weak $Y - Z$ regulation) in response to X switching from ON to OFF. Repeat the same exercise for the response to X switching from OFF to ON. What do you observe about the response times of Z in the two cases?

(c) Repeat the same exercise for the FFL described in CASE II.

(d) Summarize your results. When might the two different circuits be useful in a biological setting (i.e. what might an organism use them for)?

Group 4 Work with Group 3 to understand the behavior of the different circuits in response to changes in the input X.

(a) Construct a Petri Net describing a stochastic version of the two FFLs. Assuming the ODEs describe concentrations, suggest reasonable values for reaction rates that will produce at most a few hundred molecules in the products of all three genes, while retaining the regulatory relationships of the deterministic system.

(b) Implement a stochastic simulation corresponding to the two FFLs above using the Gillespie Algorithm.

(c) Estimate the ON-to-OFF and OFF-to-ON response times with strong $Y - Z$ regulation for the two FFL cases.

(d) Assuming X is ON, plot the stationary distribution for the expression of Z with strong $Y - Z$ regulation for the two FFL cases.

Computational Challenge 6

In this challenge you will examine 3 different version of the integrate–and–fire model: the leaky (QIF), the quadratic, and the exponential integrate-and-fire model. The fourth group will implement an adaptive version of the integrate and fire model.

To start, you will need a function that generates an inhomogeneous Poisson spike train. You can do this by using the method that we described in an earlier computational challenge, or simply using discrete increments. Use a resolution of at least $0.5\ ms$.

Group 1 The leaky integrate–and–fire neuron has the form

$$\tau_m u' = -u + RI(t).$$

Set $\tau_m = 10ms$ and $R = 500M\Omega$. Set the threshold at $20\ mV$. Remember that the voltage needs to be reset once the threshold is reached. You do not need to use a refractory period.

(a) Assume the input is described by a Poisson process. Compute the firing rate as a function of the input rate, assuming that the weight of each synapse is $w = 1$, with current measured in pA. Assume that each input spike results in an instant depolarization. As a result model the effect of each spike as an instantaneous increase in the membrane potential by $\Delta u = 0.5mV$. Compute the output rate by recording the taking a sufficiently long interval, after a transient, and dividing by the number of spikes in the interval.

(b) Assume that the input is composed of two independent Poisson spike trains. One corresponds to inputs with synapse weights $w = 1$ (excitatory inputs), and the other with $w = -1$ (inhibitory inputs). Denote the two input rates by ν_E and ν_I. Start with a ν_E that gave an output spike rate of $20Hz$ in the previous part. Gradually increase the ν_I, and plot the output firing rate, and the coefficient of variation for the inter-spike interval (time between spikes). Make sure to plot some output spike trains. You can represent them as raster plots.

Group 2 The exponential integrate and fire (EIF) model is given by

$$\tau_m u' = -u + \Delta_T \exp\left(\frac{u - \eta_{\rm rh}}{\Delta_T}\right) + RI(t) \tag{1}$$

Set $\tau_m = 10ms$ and $R = 500M\Omega$. Remember that the voltage needs to be reset once the threshold is reached. You do not need to use a refractory period.

(a) Set $\eta_{\rm rh} = 18mV$ and $\Delta_T = 1mV$. Plot the right hand side of Eq. (1) and vary the input current (measure the input current in pA). What happens to the roots of the equation $-u + \Delta_T \exp\left(\frac{u - \eta_{\rm rh}}{\Delta_T}\right) + RI(t) = 0$, as I is increased from 0?

(b) Simulate the EIF for different values of constant input current, as you pass the bifurcation point observed in the previous problem. You will have to set the threshold around $50mV$. Be careful since the voltage can grow explosively. Make sure to plot the membrane potential.

(c) Assume the input is described by a Poisson process. Compute the firing rate as a function of the input rate, assuming that the weight of each synapse is $w_k = 1$, with current measured in pA. Assume that each input spike results in an instant depolarization. As

a result model the effect of each spike as an instantaneous increase in the membrane potential by $\Delta u = 0.5mV$. You can compute the output rate by recording the times between consecutive spikes, and dividing by the total number of spikes. Do this for $\Delta_T = 1, 0.5$, and 0.1.

Group 3 The quadratic integrate and fire (QIF) model is given by

$$\tau_m u' = -a_0(u - u_{\text{rest}})(u - u_c) + RI(t) \tag{2}$$

Set $\tau_m = 10ms$ and $R = 500M\Omega$. Remember that the voltage needs to be reset once the threshold is reached. You do not need to use a refractory period.

(a) Set $u_{\text{rest}} = 0mV$ and $u_c = 20mV$. Plot the right hand side of Eq. (2) and vary the input current (measure the input current in pA). What happens to the roots of the equation $-a_0(u - u_{\text{rest}})(u - u_c) + RI(t) = 0$, as I is increased from 0?

(b) Simulate the QIF for different values of constant input current, as you pass the bifurcation point observed in the previous problem. You will have to set the threshold around $60mV$. Be careful since the voltage can grow quickly.

(c) Assume the input is described by a Poisson process. Compute the firing rate as a function of the input rate, assuming that the weight of each synapse is $w_k = 1$, with current measured in pA. Assume that each input spike results in an instant depolarization. As a result model the effect of each spike as an instantaneous increase in the membrane potential by $\Delta u = 0.5mV$. You can compute the output rate by recording the times between consecutive spikes, and dividing by the total number of spikes. Do this for $a_0 = 1, 2$, and 5.

Group 4 The leaky integrate and fire model with an adaptive current has the form

$$\tau_m u' = -u + RI(t) - Rv$$
$$\tau_v v' = au - v + b\tau_v \sum_{t^f} \delta(t - t^f) \tag{3}$$

The sum $\sum_{t^f} \delta(t - t^f)$ goes over all spike times (threshold crossing times) of the model. Set $\tau_v = 100ms$, $\tau_m = 10ms$, and $R = 500M\Omega$. Set the threshold at $20\ mV$. You may find Chapter 6 in the book by Gerstner, et al. helpful for this problem.

(a) Set $b = 5pA$, and vary a from negative to positive values. Plot some of the firing patters that you observe as you increase the input current I from zero to $100pA$.

(b) Assume the input is described by a Poisson process. Assume that the weight of each synapse is $w = 1$, with current measured in pA. Assume that each input spike results in an instant depolarization. As a result model the effect of each spike as an instantaneous increase in the membrane potential by $\Delta u = 0.5mV$. Take two different output patterns you observed in the previous part, and see if they are changed if the input is described by a Poisson process, with the same average that you used in part (a).

(c) Assume that the input is composed of two independent Poisson spike trains. One corresponds to inputs with synapse weights $w = 1$ (excitatory inputs), and the other with $w = -1$ (inhibitory inputs). Denote the two input rates by ν_E and ν_I. Start with a ν_E that gave an output spike rate of $20Hz$ in the previous part with an adaptive

output. Gradually increase the ν_I, and plot the output firing rate, and the coefficient of variation for the inter-spike interval (time between spikes). Make sure to plot some output spike trains. You can represent them as raster plots. Discuss with group 1 the effect of the adaptive current on the output.

Computational Challenge 7

In this challenge you will examine the behavior of the Wright-Fisher and Moran process at different population sizes

Group 1 The haploid Wright-Fisher model with a single locus and two alleles a and A has has transition probabilities

$$p_{i,j} = \binom{N}{j} \left((1-u)\frac{w_a i}{w_a i + (N-i)w_A} + v\frac{w_A(N-i)}{w_a i + (N-i)w_A} \right)^j \times$$
$$\left((1-v)\frac{w_A(N-i)}{w_a i + (N-i)w_A} + u\frac{w_a i}{w_a i + (N-i)w_A} \right)^{N-j}$$

where the terms are as defined in lecture 14 and 15.

(a) Setting $w_a = 1 + s$ and $w_A = 1$, find the stationary distribution of the Wright-Fisher process at population sizes of $N = 10$, $N = 100$, $N = 1000$ and $N = 10000$. You should work with the other three groups and choose appropriate values of u, v and s. Note these parameters should be chosen so that the values of Nu, Nv and Ns are the same for each population size [e.g. if you choose $v = 0.01$ for $N = 10$ you must use $v = 0.001$ when $v = 100$ and so on]. Compare the stationary distribution produced via simulation to that given in the lectures as the solution for the Fokker Planck equation $P^* = x^{Nv-1}(1-x)^{Nu-1}e^{Nsx}$

(b) Now set mutation rates $u = v = 0$. Beginning with a single copy of allele a, i.e. $i = 1$ use simulation to estimate the probability of fixation of a for the same population sizes and values of s as used previously. Compare the results to the analytic expression derived for the Moran model $f_1 = \frac{w_A/w_a - 1}{(w_A/w_a)^N - 1}$ and to the approximation commonly quoted for the Wright-Fisher model $f_1 = (1 - e^{-s})/(1 - e^{-Ns})$.

(c) In both cases explain any differences you see between the simulated and analytical results.

Group 2 The haploid Moran model with a single locus and two alleles a and A has has transition probabilities

$$T(n+1|n) = (1-u)\frac{w_a i}{w_a i + (N-i)w_A}\left(\frac{n}{N}\right)\left(1 - \frac{n}{N}\right) + v\frac{w_A(N-i)}{w_a i + (N-i)w_A}\left(1 - \frac{n}{N}\right)^2$$
$$T(n-1|n) = (1-v)\frac{w_A(N-i)}{w_a i + (N-i)w_A}\left(1 - \frac{n}{N}\right)\left(\frac{n}{N}\right) + u\frac{w_a i}{w_a i + (N-i)w_A}\left(\frac{n}{N}\right)^2$$

(a) Using the parameter values you agreed with group 1, repeat the same tasks, (a), (b) and (c), set for the Wright-Fisher model above for the Moran model.

Group 3 You will implement *both* the Wright Fisher and Moran models described above in a scenario in which the population is split between two islands, such that M individuals are present on island 1 and $N - M$ individuals on island 2. The fitness of alleles a and A are assumed to be the same on both islands but, the fitness of an individual is calculated relative tot eh other members of the island only. Migration between the islands is assumed to occur at rate m_{12} from island 1 to 2 and m_{21} in the other direction.

(a) Write down the transition probabilities for for the Wright Fisher and Moran models in this two island setting

(b) Using the parameter values agreed with group 1 for u, v and s, and using $N = 10000$, repeat the same tasks as set for group 1 (i.e. find the stationary distribution with mutation and fixation probability without mutation) for both models, for values of $M = 50$, $M = 500$ and $M = 5000$.

(c) Discuss how the island structure impacts evolution.

Group 4 You will implement the Moran model in a scenario in which allele a is frequency dependent, meaning $w_a = 1 + (i - 1)s/N$ and $w_A = 1 + is/N$.

(a) Write down the transition probabilities for for the Moran model with frequency dependent fitness in this two island setting described for group 3.

(b) Using the parameter values agreed with group 1 for u, v and s, and using $N = 10000$, repeat the same tasks as set for group 1 (i.e. find the stationary distribution with mutation and fixation probability without mutation) for both models, for values of $M = 0$ (i.e. only one island) and for $M = 50$, $M = 500$ and $M = 5000$.

(c) Discuss how this frequency dependence impacts evolution.

Computational Challenge 8

In this challenge you will examine games played in evolving populations.

Group 1 Consider the one-shot rock-paper-scissors game played between two players X and Y. The game has payoff matrix:

		Player Y		
		R	P	S
	R	$(0,0)$	$(-1,1)$	$(1,-1)$
Player X	P	$(1,-1)$	$(0,0)$	$(-1,1)$
	S	$(-1,1)$	$(1,-1)$	$(0,0)$

Consider a strategy $\mathbf{p} = \{p_r, p_p, p_s\}$ where p_r is the probability of playing paper etc. and $p_r + p_s + p_p = 1$.

(a) Write down the expected payoff for a pair of strategies $\mathbf{p}$ and $\mathbf{q}$ in the one-shot rock-paper-scissors game. Can you find the Nash equilibrium for this system?

(b) We are now going to model the evolution of rock-paper-scissors strategies. Assume there are only three strategies, always rock, always paper or always scissors, i.e. $p_r = 1$, $p_s = 1$ or $p_p = 1$. Write down the fitness for each of these three strategies under the Moran model.

(c) Investigate via simulation the time evolution of the strategies in a population of size $N = 1000$. Assume that strategies can mutate to one another at rate $\mu = 0.001$ with $w = 1$. Plot a time series for the frequency of different strategies in the system. Also plot a histogram showing the frequency distribution for each strategy once the system has reached equilibrium [note you will have to decide when to stop your simulation in order to do this].

Group 2 Consider the one-shot Hawk-Dove game played between two players X and Y:

$$
\begin{array}{c}
 & \text{Player } Y \\
 & \begin{array}{cc} H & D \end{array} \\
\text{Player } X \quad \begin{array}{c} H \\ D \end{array} & \begin{array}{c|c} ((B-C)/2, (B-C)/2) & (B, 0) \\ \hline (0, B) & (B/2, B/2) \end{array}
\end{array}
$$

Consider a strategy p which gives the probability of playing hawk in the one-shot game. The probability of playing dove is thus $1 - p$.

(a) Write down the expected payoff for a pair of strategies p and q in the one-shot Hawk-Dove game. Can you you find the Nash equilibrium for this system?

(b) We are going to study the evolution of hawks and doves under the Moran model. We will assume that each individual is either a hawk $p = 1$ or a dove $p = 0$. Write down the fitness of hawks and doves under the Moran model.

(c) Investigate via simulation the equilibrium frequency of hawks and doves in a population on $N = 1000$, assuming mutations between strategies at rate $\mu = 0.001$ with $w = 1$. For $B = 4$ and $C = 6$ show a histogram for the distribution of strategies at equilibrium. Show how the average proportion of hawks and doves changes as you change B.

Group 3 Consider a one-shot, two-player Public Goods game played between two players X and Y. This game is similar to the donation game discussed in class and the moves are cooperate or defect:

$$
\begin{array}{c}
 & \text{Player } Y \\
 & \begin{array}{cc} C & D \end{array} \\
\text{Player } X \quad \begin{array}{c} C \\ D \end{array} & \begin{array}{c|c} (B-C, B-C) & (B/2-C, B/2) \\ \hline (B/2, B/2-C) & (0, 0) \end{array}
\end{array}
$$

(a) When is this game a Prisoner's Dilemma? What is the Nash equilibrium for the one-shot game when it is not a Prisoner's Dilemma?

(b) Now consider the iterated version of the game, in which both players use memory-1 strategies. Consider the strategies ALLD, $\{0, 0, 0, 0\}$, TFT $\{1, 0, 1, 0\}$, ALLC $\{1, 1, 1, 1\}$ and WSLS $\{1, 0, 0, 1\}$. Assuming there is a small error rate $\epsilon = 0.001$ such that a player who attempts to play cooperate accidentally plays defect and vice versa, determine and plot the stationary distributions for the iterated game played between all pairs of strategies (ie. 10 different matchups)

(c) Using these stationary distributions calculate the expected payoffs for all pairs of strategies in the infinitely iterated game for $B = 1.5$ and $C = 1$.

Group 4 Consider the two-player public goods game described above for Group 3

(a) We will consider three iterated prisoner's dilemma strategies in the absence of errors, ALLD, $\{0, 0, 0, 0\}$, TFT $\{1, 0, 1, 0\}$ and ALLC $\{1, 1, 1, 1\}$. Assuming TFT starts off cooperating, and there are no errors, write down the payoffs received for each of the six pairwise matchups between these strategies when playing against the other in the infinitely iterated game (you should not need to simulate this)

(b) Write down the fitness for each of these strategies under the Moran model. Calculate via simulation the fixation probability for each of the possible invasions (6 in total), for a population of $N = 1000$ and $w = 1$.

(c) Now we will run simulations allowing mutations between strategies at rate $\mu = 0.01$. Plot a histogram showing the distribution of strategy frequencies once the system has reached equilibrium.

Computational Challenge 9

In this challenge you will develop different agent based models. These are representative of the models people use in research.

It will be easiest to see what is going on if you animate your simulations. While this is optional in the first three challenges, we suggest that you do it using, for example, `FuncAnimation` from `matplotlib`.

Group 1 In this challenge you will implement Schelling' model of segregation. For a description of the model, see p. 108 of the book by Easley and Kleinberg. As the book notes, you will have several options in the implementation, but the results should not differ. I suggest using a double loop to go over rows and columns of the lattice sequentially, so that you can easily terminate the simulation if no agents have moved during one entire iteration. Note that the system can also reach a dynamic equilibrium in which some of the agents will keep moving, but the pattern is largely static.

The size of the domain is determined by the width and height. The "empty ratio" is the fraction of the domain that is initially unoccupied, and the similarity threshold determines the fraction of different neighbors above which an agent will move. Set a maximal number of iterations, but also stop your simulations if your simulation reaches a stable state. For the following you can use a 50 by 50, or larger domain. The similarity ratio $r_{\text{sim}} = n_{\text{similar neighbors}}/n_{\text{neighbors}}$ is the fraction of neighbors that have the same color as the agent.

(a) Start with `empty_ratio = 0.2`, and increase the similarity threshold from 0.2 to 0.6. Run the simulation until the average similarity ratio, r_{sim} converges (it may keep oscillating a little, so add some tolerance). Does convergence take longer for smaller or larger thresholds? Explain why this could be.

(b) Start with `similarity_threshold = 0.4`, and increase the `empty_ratio` from 0.1 to 0.6 in increments of 0.1 Run the simulation until convergence. Does convergence take longer for smaller or larger empty spaces? Explain why this could be.

(c) Run your simulations for `similarity_threshold` = 0.2–0.6 in increments of 0.1, and plot r_{sim} over time. How does r_{sim} relate to the similarity threshold?

(d) Now assume that with some probability r even satisfied agent moves. That is there is a small probability r that during a time step an agent moves to a random location, regardless of its neighborhood. How does increasing r impact your observations in part (b)?

(e) BONUS: Animate your simulation.

Group 2 In this challenge you will implement a spatial Moran model of cancer, as discussed by N. Komarova, *Bulletin of Mathematical Biology* (2006) 68: 1573–1599. Assume that cells are arranged in a regular grid, at locations $i = 0, 1, 2, \ldots, N$. In this compartment the total number of cells, N, does not change, as each cell that dies is replaced by a new cell.

The simulation now follows the following steps

1. A cell is chosen for death, and is removed from the population. All cells are equally likely to die.

2. One of the two neighboring cells is chosen for reproduction. If the fitnesses of the two neighboring cells are r_{left} and r_{right}, the probability that the left will reproduce is $r_{\text{left}}/(r_{\text{left}} + r_{\text{right}})$, and the probability that the right will reproduce is $r_{\text{right}}/(r_{\text{left}} + r_{\text{right}})$.

3. The descendant of the dividing cell fills the empty spot created by the removal of the cell in step 1.

Implement this agent based model allowing the compartment size, N, to be a variable parameter. You can use a compartment of size $N = 100$ or larger for the first part of the challenge.

(a) Starting with different initial positions for a single mutant cell, compute the fixation probabilities of mutants with fitness $r = 0.98, 1$ (neutral), , $r = 1.02$, and 1.1. You can choose a reasonable subsample of the grid for the initial position (say every fourth or so), depending on your compartment size. Plot fixation probability against position. Are there edge effects, or is the probability uniform across the compartment?

(b) Compute the fixation probability as a function of compartment size starting at $N = 20$. Start with a single mutant in the center position.

(c) Compare the fixation probabilities in the previous example against fixation probabilities in a regular Moran model with no spatial organization. Remember that we have computed these fixation probabilities in class.

(d) Implement the following variation of the linear model discussed in Chapter 12 of Martin Nowak's book:

1. A cell is chosen for reproduction according to its fitness, as in the non-spatial Moran model.

2. If the reproducing cell is left of center, it pushes all cells to the left of it by one position to the left, and the leftmost cell (at position $i = 0$), is removed from the compartment. If the cell is to the right of center it pushes all cells to the right of it by one position to the right. The rightmost cell at position $i = N$ is removed.

3. The descendant of the dividing cell fills the new empty spot created in step 2 to the left or the right of the dividing cell.

14

Repeat part (a) of this challenge using this model. Explain the difference between the results.

(e) BONUS: Animate your simulation.

You can find a master equation for the first version of the model in the paper by N. Komarova. The master equation for the second version is more complicated.

Group 3 In this challenge you will implement the model described in the paper by I. Couzin, *et al* we discussed in class. To do so you will need to define vectors with agent positions, direction and speed. To update the position of the individuals, compute the direction of heading as described by equations 1-3 in the paper. Read the methods carefully: There is small noise that is added to the movement in each step. Use the parameters given in the caption of Fig. 1.

(a) Here you will examine the ability of a group to find a source of food. Start your simulation with food located at the origin, and place the group in a circle of radius 2 at distance 50-100 units from the food. You can start with a larger circle for larger populations. An informed individual in the group will move in a direction $\mathbf{g}_i$ that points to the origin. To calibrate the time, first assume that all agents are informed, and strongly attracted to the food, *ie* ω is large. You should see that the group reaches the food within about 100 time units.

(b) Next recreate the plot in Fig 1a for populations of size 25, 50, and 100, and a different number of informed individuals. Use $\omega = 0.5, 1$, and 2. Terminate the simulations when all individuals are within a radius 10 of the food (success), or the simulation ran for more than $10000s$ (failure). Increment the number of informed individuals in steps of approximately 5 %, starting at 5 %. Use a number of simulations in each condition to estimate the chance of success.

(c) Assume that the population is running away from a predator instead of finding food. The predator starts 10 units away from the centroid of the agent population, and always moves to the closest agent, regardless of distance. If the predator comes within 0.5 units of the agent, the agent is removed from the population. In this case the vector $\mathbf{g}_i$ points in the direction opposite of the predator. Assume that the speed of the predator is slightly larger than that of each agent, $s_p > s$. Fix the population size at 25, and plot the survival time of the population as a function of the fraction of informed individuals, and the strength of interaction ω.

(d) BONUS: Animate your simulation.

Group 4 In this challenge you will implement a spatial version of the rock–paper–scissors game. The agents are placed on a lattice with periodic boundary conditions, and each agent uses one of three strategies. We can thus think of the agents as belonging to one of three species. Each cell (location) in the lattice is either empty or occupied by an agent of one particular species.

Start with populating the lattice with agents of each type, and leaving some cells empty. The simulation proceeds by picking a cell uniformly at random in the lattice, and then picking a random neighboring cell. If both cells are empty, or are occupied by agents of the same species, nothing happens. Otherwise three things can happen:

1. If only one of the cells is occupied, the agent in the occupied cell will reproduce. The descendant belongs to the same species, *ie* uses the same strategy.

2. If the two cells are occupied by different species, then with probability p they fight. The agent with the loosing strategy dies, and their cell is vacated.

3. With probability $1 - p$, the agents in the two cells swap places.

For this exercise it is best to animate your output. You can use `FuncAnimation` from `matplotlib`, but there are other choices. Start with a 30 by 30 grid at least.

(a) Illustrate what happens for high values of p, and then decrease p. You should see the formation of spiral structures that increase in size as mobility is increased.

(b) Compute the time series for the frequency for each species, and plot them in time. How does this change with p?

(c) Assume that there is a higher rate of reactions in the upper left quadrant of the lattice, so that the probability of choosing a cell in this corner has probability is twice as likely as picking a cell from another quadrant. How do the results in part (a) change?

(d) BONUS: Extend this to five species by using the rules of the rock-paper-scissors-lizzard-spock game.

References

[1] X. CAI AND Z. XU, *K-leap method for accelerating stochastic simulation of coupled chemical reactions*, The Journal of chemical physics, 126 (2007), p. 074102.

[2] Y. CAO, D. T. GILLESPIE, AND L. R. PETZOLD, *Efficient step size selection for the tau-leaping simulation method*, The Journal of chemical physics, 124 (2006), p. 044109.

[3] M. B. ELOWITZ AND S. LEIBLER, *A synthetic oscillatory network of transcriptional regulators*, Nature, 403 (2000), pp. 335–338.

[4] T. S. GARDNER, C. R. CANTOR, AND J. J. COLLINS, *Construction of a genetic toggle switch in escherichia coli*, Nature, 403 (2000), pp. 339–342.

[5] D. GONZE AND P. RUOFF, *The goodwin oscillator and its legacy*, Acta biotheoretica, (2020), pp. 1–18.

[6] A. LOINGER AND O. BIHAM, *Stochastic simulations of the repressilator circuit*, Physical Review E, 76 (2007), p. 051917.

[7] A. LOINGER, A. LIPSHTAT, N. Q. BALABAN, AND O. BIHAM, *Stochastic simulations of genetic switch systems*, Physical Review E, 75 (2007), p. 021904.