Homework Assignments for a Graduate Introduction to Stochastic Processes in Biology

This is the collection of all four homework assignments in the course. Because students also had to work on computational challenges, we assigned less homework that we would in a traditional class. Comments about the homework that were not included in our assignments are provided in *italics*.

Homework 1

1. Show that in a pure birth process $\sum_{n=0}^{\infty} P_n(t) = 1$ for all t if and only if $\sum_{n=0}^{\infty} 1/\lambda_n = \infty$. Come up with an "explosive" process, that is a birth process which cannot be normalized for all finite times. Write a program to simulate this process. You will have to be careful in the implementation as realizations can (and typically will) "blow up" in finite time.

2. In the case of a simple birth and death process we have $\lambda_n = \lambda n$ and $\mu_n = \mu n$ for some constants λ and μ. We can add immigration to this process, by assuming that during each interval of length h, an individual moves into the population with probability νh. Here ν is constant independent of the population size. The deterministic model of the system has the form
$$\frac{dn}{dt} = (\lambda - \mu)n + \nu$$
Show that the mean of the population size in the corresponding Master Equation, $\bar{n}(t)$, is equal to the solution of this differential equation, with an appropriate initial condition. Give an example to show that this is not true for more general birth and death processes with immigration, and illustrate the example numerically.

   Comment: *We discussed how to obtain an equation for the mean of a stochastic process from the Master Equation in class.*

3. The transition matrix for a four-state Markov chain is
$$P = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/4 & 0 & 3/4 & 0 \\ 0 & 3/4 & 0 & 1/4 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}$$

   (a) Draw the directed graph corresponding to this chain.
   (b) Show that the chain is irreducible, positive recurrent, and periodic. What is the period?
   (c) Find the unique stationary probability distribution.

4. Assume that the arrival of proteins to region of DNA near a promoter can be described as a Poisson process with rate λ. If the region is not occupied, it will bind the protein. The time the proteins remain bound are independent random variables with mean μ. A protein that arrives when the promoter is occupied diffuses away. Show that the long time fraction of time that the promoter is unoccupied is $1/(1 + \lambda\mu)$.

   Comment: *This problem can be extended to provide a more realistic model of this process by, for example, including a probability that the protein will bind to the region of DNA.*

5. Assume that the sequence of action potentials (APs) fired by a neuron can be described as a Poisson process. For a fixed time t, let $T(t)$ be the time to the nearest AP in time. This could be an AP preceding or subsequent to the time t. What is the mean of $T(t)$? What is the probability density function of $T(t)$?

Homework 2

1. A stochastic Lotka-Volterra (predator-prey) model is given by the following equations for the abundance of two species, Y_1 and Y_2,

$$Y_1 \xrightarrow{c_1} 2Y_1$$
$$Y_1 + Y_2 \xrightarrow{c_2} 2Y_2$$
$$Y_2 \xrightarrow{c_3} \emptyset.$$
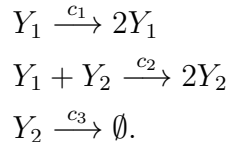
- Explain why this is called a predator-prey model. What is the corresponding ODE model? Given this interpretation, explain whether species 1 or species 2 can exist in isolation, i.e. in the absence of the other species.

- Use the Gillespie algorithm to simulate trajectories from the model. Use $c_1 = 1, c_2 = 0.005, c_3 = 0.6$.

- Changing which parameter will make it more likely for species 1 to go extinct? What about species 2? Change these parameters gradually to verify your hypothesis in simulations.

Comment: *In class we described the relation between ODE and stochastic models in some detail.*

2. Consider the following chemical reaction

$$A \xrightarrow{k} X \xrightarrow{\alpha_1} \emptyset$$
$$B \xrightarrow{k} Y \xrightarrow{\alpha_2} \emptyset$$
$$X + Y \xrightarrow{k_a} C.$$
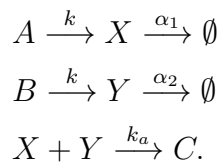
The deterministic set of equations describing this reaction has the form

$$\frac{d[X]}{dt} = k - \alpha_1[X] - k_a[X][Y]$$
$$\frac{d[Y]}{dt} = k - \alpha_2[Y] - k_a[X][Y],$$

where $[\,\cdot\,]$ denotes concentration.

- Find the fixed points of the deterministic system. Show that for the values

$k = 10$	$\alpha_1 = 10^{-6}$	$\alpha_2 = 10^{-5}$	$k_a = 10^{-5}$
$k = 10^3$	$\alpha_1 = 10^{-4}$	$\alpha_2 = 10^{-3}$	$k_a = 10^{-3}$

the fixed points are the same.

2

- Run the Gillespie algorithm and show that the behavior is very different in the two cases. Compute the stationary distributions in the two cases.

- Can you give an argument why the behavior in the two cases is so different?

3. Consider a gene whose product regulates its own transcription (autoregulation). The deterministic set of equations describing mRNA (r) and and protein (p) expression level are given by:

$$
\begin{aligned}
\frac{dr}{dt} &= k_l + \phi(p) - \gamma_r r, \\
\frac{dp}{dt} &= r k_p - \gamma_p p,
\end{aligned}
$$

(1)

where $\phi(p)$ is a function that describes how the rate of mRNA transcription depends on the protein concentration p.

- What is the transition matrix for the Markov process describing this system?

- Consider the case of *positive autoregulation*:

$$
\phi(p) = \frac{k_0 \left(\frac{p}{K}\right)^n}{1 + \left(\frac{p}{K}\right)^n}
$$

Setting $k_l = 0$, $\gamma_p = \gamma_r = k_p = k_0 = 1$ and $K = 0.5$ in Eq. 1 above, numerically determine the number of fixed points for $n = 1$ and $n = 10$. Determine the stability of the fixed points.

- Consider the case of *negative autoregulation*:

$$
\phi(p) = \frac{k_0}{1 + \left(\frac{p}{K}\right)^n}
$$

Use the Gillespie algorithm to find and plot the stationary distribution of protein for the stochastic process you wrote down above. Use transition rates $k_l = 0.001 s^{-1}$, $k_p = 0.17 s^{-1}$, $k_0 = 0.01 s^{-1}$ and $\gamma_p = 0.00028 s^{-1}$, $\gamma_r = 0.0083 s^{-1}$ with $n = 10$. Do this for the case of (i) strong regulation, $K = 100$ and (ii) weak regulation, $K = 10000$. Calculate the ratio of the standard deviation to the mean of the distribution for both cases.

Homework 3

1. In the course, we have outlined how to derive the Fokker–Planck equation for the distribution of the membrane potential under excitatory input (equation (8.41) in the chapter posted on the course website).

- Complete the missing steps in this derivation, starting from Eq. (8.37) and ending with Eq. (8.41).

- Verify that Eq. (8.45) is a solution of the equation with free boundary conditions by direct substitution.

- Simulate an integrate and fire neuron with Poisson input *without threshold* to verify that after sufficient time the membrane potential follows the distribution given by Eq.(8.46). You can do this either by running a long simulation, and sampling the potential at different times, or by running many simulations.

Comment: *This question refers to chapter 8 of the book* Neuronal Dynamics *by W. Gerstner, et al. The chapter can be found online here* $https: //neuronaldynamics. epfl. ch/ online/Ch8. html$.

2. Use the derivation in Lecture 16, and the notes from the book by Gardiner to compute the expected time for a leaky integrate and fire neuron to fire. Assume a threshold θ, resting potential set to 0, and $\tau_m = 1$. Assume that the input RI_0 is constant, and that the variance $\sigma^2(t) = \sigma^2$ is constant as well.

Simulate an integrate and fire neuron *with threshold* to show that this expected time is a good approximation. In particular, you will need to choose a synaptic weight $w_k = w$ (all equal), and input rate $\nu_k(t) = \nu$ (all equal and constant) to define the parameters RI_0 and σ^2. Note that you can use the code from the challenges for this simulation.

Comment: *This question refers the discussion of first passage time distributions in the book* Stochastic Methods: A Handbook for the Natural and Social Science *by C. Gardiner.*

3. Suppose that the times between spikes are distributed according to a gamma distribution, $\Gamma(\alpha, \beta)$.

- Compute the coefficient of variation of the inter-spike interval.

- A special case of the gamma distribution is the Erlang distribution. This is the distribution of the sum of k iid exponentially distributed random variables with parameter λ. Find the coefficient of variation for an inter-event time that follows the Erlang distribution with fixed λ and increasing k. Interpret your results.

4. Two Moran process with two alleles, a and A, with mutation but without selection has transition rates

$$
\begin{aligned}
T(n+1|n) &= (1-u)\left(\tfrac{n}{N}\right)\left(1-\tfrac{n}{N}\right) + v\left(1-\tfrac{n}{N}\right)^2 \\
T(n-1|n) &= (1-v)\left(1-\tfrac{n}{N}\right)\left(\tfrac{n}{N}\right) + u\left(\tfrac{n}{N}\right)^2
\end{aligned}
$$

where $P(n,t)$ is the probability that there are n copies of allele a at time t, u is the rate of mutation from a to A and v the rate of mutation from A to a. The master equation for this process is

$$
\frac{dP(n,t)}{dt} = -P(n,t)(T(n+1|n)+T(n-1|n))+P(n+1,t)T(n|n+1)+P(n-1,t)T(n|n-1)
$$

Setting $x = n/N$, Taylor expand the master equation retaining only terms up to order $1/N^2$ to derive the Fokker Planck equation for this system.

5. In this question you will estimate the hitting time via simulation for the two-step adaptation process (Lecture 17). Initialize the population such that all individuals have genotype 1. Assume that genotype 1 mutates to genotype 2 at rate μ_{12} and genotype 2 mutates to genotype 3 at rate mu_{23}. By simulating a haploid Moran process, estimate the expected time for the population to produce an individual with genotype 3, for mutation rates (i) $\mu_{12} = \mu_{23} = 0.1/N$, (ii) $\mu_{12} = \mu_{23} = 1/N$ and (iii) $\mu_{12} = \mu_{23} = 10/N$, using $N = 1000$ in all cases. How do the results compare to the naive expectation that the waiting time should be $\sim 1/\mu$ (which as we saw in class is the weak mutation limit which holds when $\mu \ll 1/N^2$)?

Homework 4

1. A man has two sons, and he would like them to cooperate. Therefore, in his will it states that the two sons must each specify a sum of money, s_1 and s_2, that they are willing to accept for inheritance when he dies. The value of his estate is 1,000,000 USD.

   If $s_1 + s_2 \leq 1,000,000$, then each son gets the sum he asked for and the remainder (if any) goes to charity. If $s_1 + s_2 > 1,000,000$, then neither son receives any money and all the money goes to charity.

   Find all the pure strategy Nash equilibria of the game assuming that the two men only care about their share of inheritance, they do not talk to each other, and can only ask for integer amounts.

2. This is a problem about the iterated Prisoner's Dilemma, so that the game is played between two players over multiple rounds. First look at https://ncase.me/trust/ to see a nice illustration of different strategies. We will assume that the payoff matrix on a single round is given by R to each player for mutual cooperation, P to each player for mutual defection. In the case only one player cheats then the reward is T to the cheating player, and S to the cooperating player. We assume $T > R > P > S$. Note that this is a bit more general than the payoff matrix discussed by Dr. Stewart. We will extend this payoff matrix to multiple rounds.

   (a) First consider two strategies GRIM, and ALLD. GRIM cooperates on the first move and then cooperates as long as the opponent does not defect (cheat). Once the opponent defects, GRIM will always defect afterward. Suppose that two opponents play the game repeatedly for m rounds. Compute the payoff matrix of GRIM vs ALLD after m rounds (for example you should find that the payoff to ALLD playing against GRIM after m rounds is $T + (m - 1)P$.

   (b) Based on your answer show that GRIM is stable against invasion by ALLD if

   $$m > \frac{T - P}{R - P}$$

   (c) If you know the number of rounds that you play, and you know your opponent will cooperate on the last round, then it is best for you to defect on that round. Therefore, let's introduce a new strategy GRIM* which is the same as GRIM, but always defects on the last round. Compute the payoff matrix for GRIM* vs GRIM, and conclude that GRIM* dominates GRIM.

(d) Once everyone uses GRIM*, then everyone is bound to get P on the last round. However, now there is no reason to cooperate on the second to last round. Use the same argument as above to find a strategy that dominates against GRIM*.

(e) If you continue this argument what strategy will you arrive at?

3. In some cases a more realistic scenario than a repeated game that occurs exactly m times, with m known to both players, is a scenario in which there is a probability, δ, of the game continuing after each round.

(a) Show that the expected number of rounds played in this way is $\langle m \rangle = 1 + \frac{\delta}{1-\delta}$

(b) Now, sticking with the Prisoner's Dilemma form question 2, consider a matchup between GRIM and ALLD. Write down the expected payoffs for GRIM vs ALLD, GRIM vs GRIM and ALLD vs ALLD.

(c) Show that GRIM is stable against ALLD provided

$$\delta > \frac{T - R}{T - P}$$

4. Now consider an infinitely repeated game in which players make execution errors each round with probability ϵ. Sticking with the Prisoner's Dilemma described in question 2, we will look at the behavior of three strategies: (i) Tit-for-Tat (TFT) which has strategy $\mathbf{p}_T = \{1-\epsilon, \epsilon, 1-\epsilon, \epsilon\}$, (ii) GRIM which has strategy $\mathbf{p}_G = \{1-\epsilon, \epsilon, \epsilon, \epsilon\}$ and (iii) Always cooperate (ALLC) which has strategy $\mathbf{p}_C = \{1 - \epsilon, 1 - \epsilon, 1 - \epsilon, 1 - \epsilon\}$.

(a) Write down the the 6 transition matrices for the 6 pairwise matchups between these three strategies.

(b) Calculate the stationary distribution for each of the 6 chains in terms of ϵ. Explain what happens in each case in the limit $\epsilon \to 0$.

(c) Use the stationary distribution in the limit $\epsilon \to 0$ to calculate the payoffs for each strategy in each matchup.

(d) Are there any Nash equilibria for the Prisoner's Dilemma if we limit players to only these three strategies and errors are vanishingly small, i.e. $\epsilon \to 0$. Explain your answer.