Homework # 3

Due Mon. April 11, 2016

1. **Analyzing Chemical Reaction Networks**

   The theory of reaction kinetics underpins our understanding of biological and chemical systems. It is a simple and elegant formalism: chemical reactions define rules according to which reactants form products; each rule fires at a rate that is proportional to the quantities of the corresponding reactants that are present. On the computational front, there has been a wealth of research into efficient methods for simulating chemical reactions, ranging from ordinary differential equations (ODEs) to stochastic simulation. On the mathematical front, entirely new branches of theory have been developed to characterize the dynamics of chemical reaction networks.

   Most of this work is from the vantage point of analysis: a set of chemical reaction exists, designed by nature and perhaps modified by human engineers; the objective is to understand and characterize its behavior. Comparatively little work has been done at a conceptual level in tackling the inverse problem of synthesis: how can one design a set of chemical reactions that implement specific behavior?

   This homework will consider the computational power of chemical reactions from both a deductive and an inductive point of view.

   (a) A molecular system consists of a set of chemical reactions, each specifying a rule for how types of molecules combine. For instance,

   $$X_1 + X_2 \xrightarrow{\text{k}} X_3,$$

   specifies that one molecule of X_1 combines with one molecule of X_2 to produce one molecule of X_3. The rate at which the reaction fires is proportional to (1) the concentrations of the participating molecular types; and (2) a rate constant k. (This value is not constant at all; rather it is dependent on factors such as temperature and volume; however, it is independent of molecular quantities, and so called a "constant.")

Given a set of such reactions, we can model the behavior of the system in two ways:

i. In a **continuous** sense, in terms of molecular concentrations, with differential equations;

ii. In a **discrete** sense, in terms of molecular quantities, through probabilistic discrete-event simulation.

Consider the reactions:

$$\begin{aligned} R_1: &\quad 2X_1 + X_2 \rightarrow 4X_3 &\quad k_1 = 1 \\ R_2: &\quad X_1 + 2X_3 \rightarrow 3X_2 &\quad k_2 = 2 \\ R_3: &\quad X_2 + X_3 \rightarrow 2X_1 &\quad k_3 = 3 \end{aligned}$$

For a continuous model, let x_1, x_2 and x_3 denote the concentrations of X_1, X_2, and X_3, respectively. (Recall that concentration is number of molecules per unit volume.) The behavior of the system is described by the following set of differential equations:

$$\begin{aligned} \frac{dx_1}{dt} &= -x_1^2 x_2 - 2x_1 x_3^2 + 6x_2 x_3 \\ \frac{dx_2}{dt} &= -x_1^2 x_2 + 6x_1 x_3^2 - 3x_2 x_3 \\ \frac{dx_3}{dt} &= 4x_1^2 x_2 - 2x_1 x_3^2 - 3x_2 x_3 \end{aligned}$$

For the discrete model, let the state be $\mathbf{S} = [x_1, x_2, x_3]$, where x_1, x_2 and x_3 denote the numbers of molecules of types X_1, X_2, and X_3, respectively. (Here we use actual integer quantities, not concentrations.) The firing probabilities for R_1, R_2 and R_3 are computed as follows:

$$\begin{aligned} p_1(x_1, x_2, x_3) &= \frac{\frac{1}{2}x_1(x_1-1)x_2}{\frac{1}{2}x_1(x_1-1)x_2 + x_1x_3(x_3-1) + 3x_2x_3}, \\ p_2(x_1, x_2, x_3) &= \frac{x_1x_3(x_3-1)}{\frac{1}{2}x_1(x_1-1)x_2 + x_1x_3(x_3-1) + 3x_2x_3}, \\ p_3(x_1, x_2, x_3) &= \frac{3x_2x_3}{\frac{1}{2}x_1(x_1-1)x_2 + x_1x_3(x_3-1) + 3x_2x_3}. \end{aligned}$$

Suppose that $\mathbf{S} = [3, 3, 3]$. Then the firing probabilities for R_1, R_2 and R_3

are

$$\begin{aligned} p_1(3,3,3) &= \frac{9}{9+18+27} = \frac{1}{6}, \\ p_2(3,3,3) &= \frac{18}{9+18+27} = \frac{1}{3}, \\ p_3(3,3,3) &= \frac{27}{9+18+27} = \frac{1}{2}, \end{aligned}$$

respectively.

N.B. In the continuous model, the rate of change of type is proportional to x^n where x is the concentration of a reaction and n is the coefficient. In the discrete model, the probability is proportional to $\binom{x}{n}$. This is a subtle difference. See the paper by Gillespie for an explanation.

Problem

Suppose that we define the following "outcomes":

- C_1: states $\mathbf{S} = [x_1, x_2, x_3]$ with $x_1 > 7$.
- C_2: states $\mathbf{S} = [x_1, x_2, x_3]$ with $x_2 \geq 8$.
- C_3: states $\mathbf{S} = [x_1, x_2, x_3]$ with $x_3 < 3$.

Beginning from the state $\mathbf{S} = [5, 5, 5]$, compute (or estimate) $\Pr(C_1)$, $\Pr(C_2)$, and $\Pr(C_3)$.

□

(b) Now, instead of "outcomes", let's analyze probabilities in a more general sense.

Consider again the reactions:

$$\begin{aligned} R_1: &\quad 2X_1 + X_2 \rightarrow 4X_3 &\quad k_1 = 1 \\ R_2: &\quad X_1 + 2X_3 \rightarrow 3X_2 &\quad k_2 = 2 \\ R_3: &\quad X_2 + X_3 \rightarrow 2X_1 &\quad k_3 = 3 \end{aligned}$$

Let the state be $\mathbf{S} = [x_1, x_2, x_3]$, where x_1, x_2 and x_3 denote the numbers of molecules of types X_1, X_2, and X_3, respectively.

Suppose that systems begins in the state $S = [3, 3, 3]$ (with probability 1). After one step:

- it is in state $[1, 2, 7]$ with probability $\frac{1}{6}$.
- it is in state $[2, 6, 1]$ with probability $\frac{1}{3}$.
- it is in state $[5, 2, 2]$ with probability $\frac{1}{2}$.

So, considering the first type, X_1, its discrete probability distribution after one step is

- $Pr[X_1 = 1] = \frac{1}{6}$,
- $Pr[X_1 = 2] = \frac{1}{3}$,
- $Pr[X_1 = 5] = \frac{1}{2}$,

After many steps, the system can be in many different states, with different quantities of X_1. (Of course, some of these states may have the same quantity of X_1.) The probability distribution may look something like:

- $Pr[X_1 = 0] = 0.012$,
- $Pr[X_1 = 1] = 0.025$,
- $Pr[X_1 = 2] = 0.036$,
- $Pr[X_1 = 3] = 0.061$,
- $Pr[X_1 = 4] = 0.12$,
- $Pr[X_1 = 5] = 0.19$,
- $Pr[X_1 = 6] = 0.24$,
- $Pr[X_1 = 7] = 0.19$,
- $Pr[X_1 = 8] = 0.116$,
- $Pr[X_1 = 9] = 0.010$.

(Note that this is not a real calculation.)

Problem

For the set of reactions above, beginning from the state $[6, 6, 6]$ compute (or estimate) the mean and variance for the probability distributions of X_1, X_2 and X_3 – each separately – after 5 steps. □

2. **Synthesizing Chemical Reaction Networks**

The task of synthesizing a set of chemical reactions to compute a desired function is conceptually open-ended. Like most engineering problems, there are often many possible solutions.

In class, we saw a chemical reaction network that performs multiplication. What follows are some other examples of chemical reaction networks that compute simple functions. In describing the functions that the modules implement, we add subscripts to the quantities of molecular types to denote *when* these quantities exist: zero indicates that this is the initial quantity, whereas infinity indicates that it is the quantity after the module has finished.

Exponentiation:

$$\boxed{Y_\infty = 2^{X_0}}$$

This module consumes molecules of an input type one at a time, doubling the quantity of an output type for each. Its behavior is described by the following pseudocode:

```
1  ForEach x {
2    Y = 2 * Y;
3  }.
```

The reactions are:

$$\begin{aligned} x &\xrightarrow{\text{slow}} a \\ a + y &\xrightarrow{\text{faster}} a + 2y' \\ a &\xrightarrow{\text{fast}} \varnothing \\ y' &\xrightarrow{\text{medium}} y. \end{aligned}$$

Initially, Y is one and all other quantities (except X) are zero.

Logarithm:

$$\boxed{Y_\infty = log_2(X_0)}$$

This module is similar to the exponentiation module, except that instead of doubling the output, the input is forced to halve itself; each time it does so, the output is incremented by one. Its behavior is described by the following pseudocode:

```
1  While Not(X==1) {
2    X = X/2;
3    Y = Y+1;
4  }.
```

The reactions are:

$$\begin{aligned} b &\xrightarrow{\text{slow}} a + b \\ a + 2x &\xrightarrow{\text{faster}} c + x' + a \\ 2c &\xrightarrow{\text{faster}} c \\ a &\xrightarrow{\text{fast}} \varnothing \\ x' &\xrightarrow{\text{medium}} x \\ c &\xrightarrow{\text{medium}} y \end{aligned}$$

Initially, B is a small but non-zero quantity and all other quantities (except X) are zero.

Problem

Produce a chemical reaction network that computes

$$Z_\infty = X_0 \log_2 Y_0$$

Demonstrate that your solution works (either mathematically, or by simulating it continuously or stochastically).

Problem

Produce a chemical reaction network that computes

$$Y_\infty = 2^{\log_2 X_0} \tag{1}$$

(2)

(No points for noticing that $Y_\infty = X_0$. Your network must compute this as shown!) Demonstrate that your solution works (either mathematically, or by simulating it continuously or stochastically).

3. **Multiplication**
   (*no collaboration*)

   Consider the following representation of real numbers. A real value x between 0 and 1 is represented as $\frac{x_1}{x_1+x_2}$, where x_1 and x_2 are positive integers.

   Construct a set of chemical reactions that multiplies two real numbers a and b represented this way, producing a resulting number c, also represented this way. Let the quantities or concentrations of molecular types A_1 and A_2 represent a, those of B_1 and B_2 represent b, and those of C_1 and C_2 represent c. Demonstrate that your solution works (either mathematically, or by simulating it continuously or stochastically).

4. **Iterative Computation with Molecular Reactions**

   In the last couple of problems, you implemented some simple functions with molecular reactions. In this homework, you'll implement two full-fledged algorithms.

   (a) **Euclid's Algorithm**

   Euclid's algorithm is an efficient method for computing the greatest common divisor (GCD) of two integers, also known as the greatest common factor (GCF) or highest common factor (HCF). It is named after the Greek mathematician Euclid, who described it in Books VII and X of his Elements.

   In its simplest form, Euclid's algorithm starts with a pair of positive integers and forms a new pair that consists of the smaller number and the difference between the larger and smaller numbers. The process repeats until the numbers are equal. That number then is the greatest common divisor of the original pair.

   Design a set of molecular reactions that implements the procedure. Demonstrate that your code works for $x = 66$ and $y = 30$.

   (b) **Collatz Procedure**

   The *Collatz* conjecture is a famous open problem in mathematics, proposed by Lothar Collatz in 1937. Consider the following iterative procedure. For any positive integer x,

   - if $x = 1$ stop;
   - else if x is odd, let $x = 3x + 1$;
   - else let $x = x/2$.

   The conjecture is that, starting with any positive integer x, the procedure always terminates with $x = 1$. For instance, starting with $x = 5$, one follows the sequence 16, 8, 4, 2 and 1. Starting from $x = 27$, one follows the sequence 82, 41, 124, 62, 31, 94, 47, 142, 71, 214, 107, ... (keep going, you'll see that you eventually hit one.)

   Proving this is evidently difficult. Paul Erdös said about the conjecture: "Mathematics is not yet ready for such problems". He offered a monetary reward of $500 for its solution.

   You are *not* asked to prove the *Collatz* conjecture on this homework. Rather you are asked to design a set of molecular reactions that implements the procedure. The input to the system is a quantity of a type X; the system should iterate through the Collatz sequence until it hits one.

Demonstrate that your code works for $x = 27$.