CSEP527, Fall 2019, Homework 1

Anthony Adkins (anadkins@uw.edu)

1.1.1 Stoichiometric matrix of the CRN

$$A = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix}$$

1.1.2 Kinetic mass vector

$K_1 = k_1 * c_A * c_b$
$K_2 = k_2 * c_A * c_c$
$K_3 = k_3 * c_b * c_c$
So,

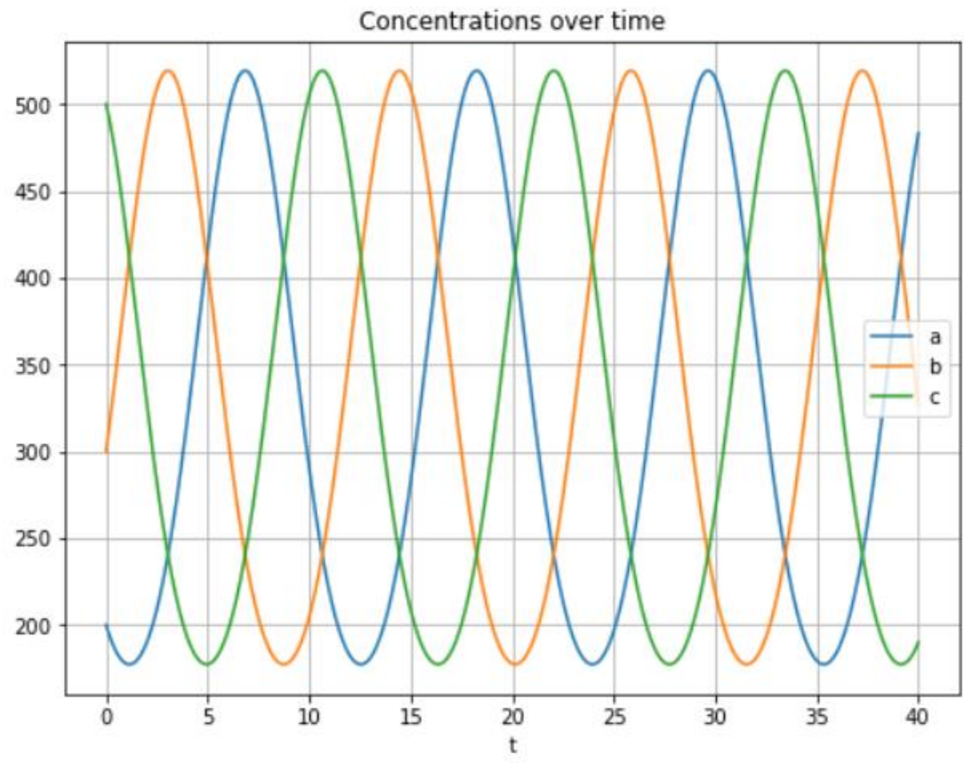
$$\vec{K} = \begin{pmatrix} k_1 * c_A * c_B \\ k_2 * c_A * c_C \\ k_3 * c_B * c_C \end{pmatrix}$$

1.1.3 ODE system

$$\nabla_t \vec{X}(t) = A * \vec{K} = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix} \times \begin{pmatrix} k_1 * c_A * c_B \\ k_2 * c_A * c_C \\ k_3 * c_B * c_C \end{pmatrix} = \begin{pmatrix} k_1 * c_A * c_B - k_2 * c_A * c_C \\ -k_1 * c_A * c_B + k_3 * c_B * c_C \\ k_2 * c_A * c_C - k_3 * c_B * c_C \end{pmatrix}$$

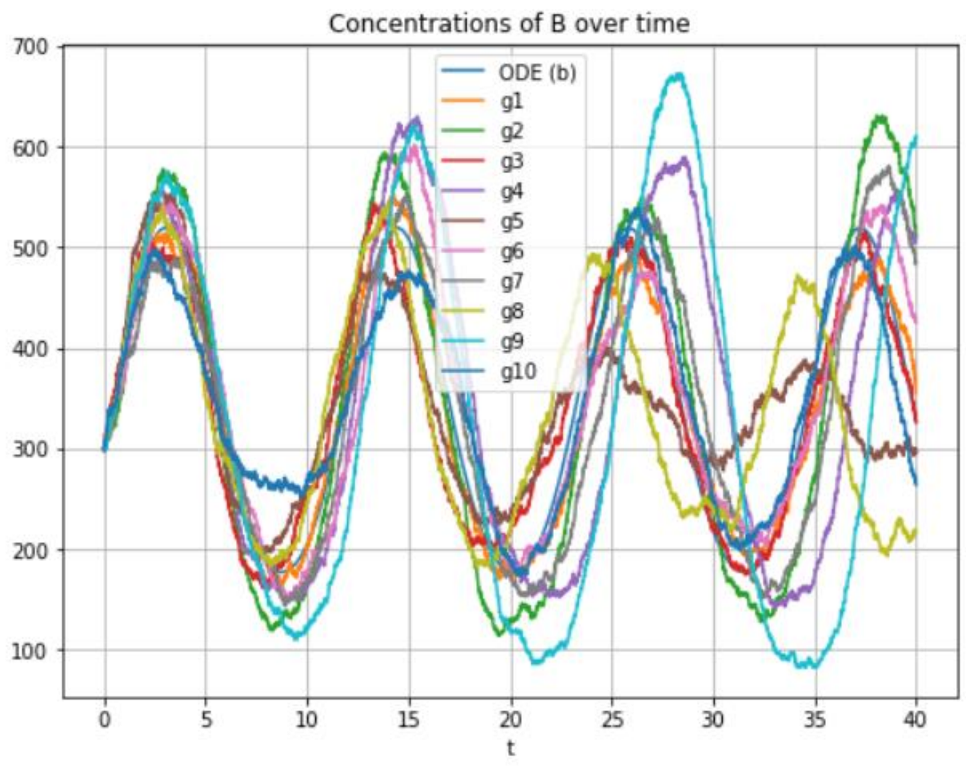
1.1.4 ODE solution

From jupyter notebook solution:



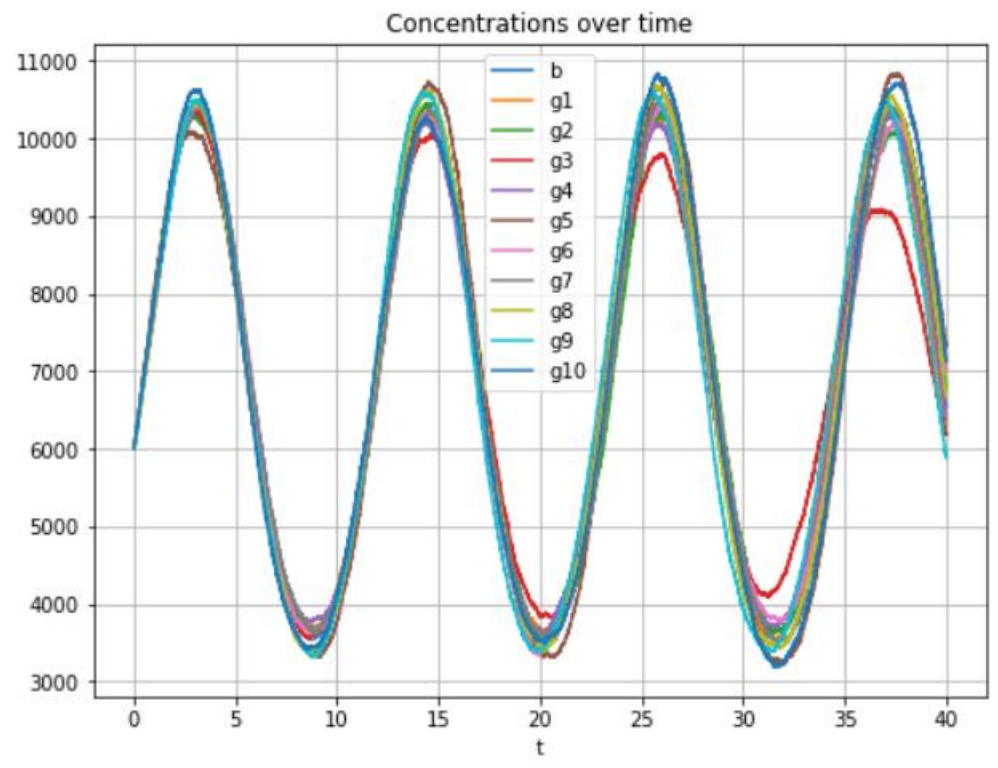
1.2.1 Plot the Gillespie simulations of substance B

From jupyter notebook solution:



1.2.2 Scale the constants

From jupyter notebook solution:



Why does the simulation take much longer time compared to problem 1.2.1?

The Monte Carlo step of Gillespie's algorithm generates much smaller time intervals when a reaction is more likely to occur. With a higher number of interactions being more likely, the average time interval for which a reaction is expected to occur decreases significantly.

Specifically, this is because a_0 increases, which inversely affects $t = \left(\frac{1}{a_0}\right) * \ln\left(\frac{1}{r_1}\right)$

Why do the trajectories seem less noisy compared to problem 1.2.1?

As explained in part 1.2.1, we are taking many more samples at shorter time intervals, which creates smoother sampling because there is less random chance that the stochastic algorithm strays from the determinative ODE integration solution. With more samples, the influence of randomness on the chemical interactions decreases.

2.1.1 Bowtie commands

```
bowtie2-build lambda_virus.fa lambda_virus
bowtie2 -x lambda_virus -U ./reads_1.fq output.sam --norc
```

2.1.2 Bowtie questions

Does the index permit search errors? If yes, how many? Why is it limited to this number?

Bowtie did not originally permit alignment gaps, but with Bowtie2 they added a second pass which allows up to 1 mismatch. The error limit is not unbounded because the amount of memory and runtime of the algorithm grows significantly in relation to the requirement to permit more errors.

If the index is limited to e errors, does that imply a read cannot be aligned if it contains > e errors?

If the read contained more than e errors which resulted in e mismatches between the read and the sample genome, then the entire read would not be successfully aligned. Sub-pieces of the read could be aligned to the genome.

2.1.3 Summary statistics

2.1.3.a What is the length of the Lambda Phage genome?

48502 nucleotides

2.1.3.b How many raw reads are there? Hint: The reads occur on Line 2, 6, 10, etc.

There are 10,000 reads

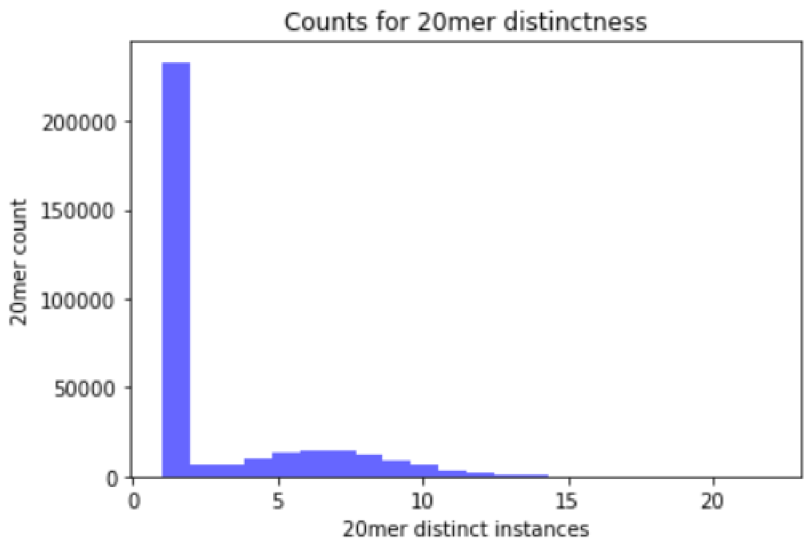
2.1.3.c What is the minimum, maximum and average read length?

min read length: 40

max read length: 354

avg read length: 108.8399

2.1.3.d Plot a histogram of the occurrence count of distinct 20-nt subsequences across all reads



2.2.1 Smith Waterman alignment

Alignment for R1

max score : 232

Start position on R1: 1

Start position on lambda virus: 18,401

```
R1: TGAATGCGAACTCCGGGACGCTCAGTAATGTGACGATAGCTGAAAACTGTACGATAAACNGTACGCTGAG
GGCAGAAAAAATCGTCGGGGACATTNTAAAGGCGGCGAGCGCGGCTTTTCCG
Lambda virus: TGAATGCGAACTCCGGGACGCTCAGTAATGTGACGATAGCTGAAAACTGTACGATAAACG
GTACGCTGAGGGCGGAAAAAATCGTCGGGGACATTGTAAAGGCGGCGAGCGCGGCTTTTCCG
```

Alignment for R2

max score : 526

Start position on R2: 5

Start position on lambda virus: 8,890

```
R2: TGATGCGGGCTTGTGGAGTTCAGCCGATCTGACTTATGTCATTACCTATGAAATGTGAGGACGCTATGCC
TGTACCAAATCCTACAATGCCGGTGAAAGGTGCCGGGATCACCCTGTGGGTTTATAAGGGGATCGGTGACCCCT
ACGCGAATCCGCTTTCAGACGTTGACTGGTCGCGTCTGGCAAAAGTTAAAGACCTGACGCCCGGCGAACTGACC
GCTGAGNCCTATGACGACAGCTATCTCGATGATGAAGATGCAGACTGGACTGC
Lambda virus: TGATGCGGGCTTGTGGAGTTCAGCCGATCTGACTTATGTCATTACCTATGAAATGTGAGG
ACGCTATGCCTGTACCAAATCCTACAATGCCGGTGAAAGGTGCCGGGACCACCCTGTGGGTTTATAAGGGGAGC
GGTGACCCTTACGCGAATCCGCTTTCAGACGTTGACTGGTCGCGTCTGGCAAAAGTTAAAGACCTGACGCCCGG
CGAACTGACCGCTGAGTCCTATGACGACAGCTATCTCGATGATGAAGATGCAGACTGGACTGC
```