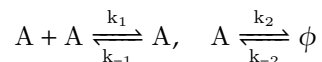


Methods for Stochastic Systems

December 7, 2023

0 The Problem

Consider a chemical species A in a container of volume v subject to the following chemical reactions



For simulations, we will use the following parameters:

$$k_1/v = 0.1, \quad k_{-1} = 1, \quad k_2 = 0.1, \quad k_{-2}v = 1$$

1 Question 1

Find $p_n(t)$ the probability master equation, denoting the probability of there being n particles at time t .

1.1 Propensity Functions

We can calculate our propensity functions as the following:

$$\alpha_1 = A(t)(A(t) - 1)k_1/v, \quad \alpha_2 = k_{-1}A(t), \quad \alpha_3 = A(t)k_2, \quad \alpha_4 = k_{-2}v \quad (1)$$

Where α_1 denotes the reaction $A + A \xrightarrow{k_1} A$, α_2 denotes $A \xrightarrow{k_{-1}} A + A$, α_3 denotes $A \xrightarrow{k_2} \phi$ and α_4 denotes $\phi \xrightarrow{k_{-2}} A$

1.2 Reaction Cases

By considering the different states that our particles could be in, before reacting into state $A(t) = n$, we can find the reactions of $p_n(t + \delta t)$ and therefore find $\frac{dp_n}{dt}$.

State of A(t)	Result of reaction
$A(t) = n$	Nothing Happens
$A(t) = n + 1$	Reaction occurs and 1 molecule of A is lost.
$A(t) = n - 1$	Reaction occurs and 1 molecule of A is gained.

Table 1: A table showing the different reactions that could occur.

1.3 Constructing $p_n(t)$

By substituting in our propensity functions for each reaction, and their corresponding values of n , we find for a infinitesimal interval $[t, t + \delta t]$:

$$\begin{aligned}
p_n(t + \delta t) = & p_n(t) \left[1 - \frac{k_1}{v} n(n-1)\delta t - k_{-1}n\delta t - k_2n\delta t - k_{-2}v\delta t \right] \\
& + p_{n+1}(t) \left[\frac{k_1}{v} (n+1)n\delta t + k_2(n+1)\delta t \right] \\
& + p_{n-1}(t) [k_{-1}(n-1)\delta t + k_{-2}v\delta t] + H.O.T
\end{aligned}$$

Note for when more than one reaction takes place, it is of order $\mathcal{O}(\delta t^2)$ and we do not have to consider the higher order terms since δt is infinitesimally small. By subtracting $p_n(t)$ from both sides and dividing through by δt , then taking the limit as $\delta t \rightarrow 0$ we get the following equation:

$$\begin{aligned}
\frac{dp_n}{dt} = & p_{n+1}(t) \left[\frac{k_1}{v} n(n+1) + k_2(n+1) \right] \\
& + p_{n-1}(t) [k_{-1}(n-1) + k_{-2}v] \\
& - p_n(t) \left[\frac{k_1}{v} n(n-1) + k_{-1}n + k_2n + k_{-2}v \right]
\end{aligned}$$

For convention, we always have $p_n = 0, \quad \forall n < 0$

2 Question 2

We will now derive an ODE which describes the evolution of the mean number of particles, M , in terms of the moments of A .

Firstly, observe that $M = \sum_{n \geq 0} nP_n$, and we can differentiate under the

summation sign since our sum converges uniformly. Therefore, $\frac{dM}{dt} = \sum_{n \geq 0} n \frac{dp_n}{dt}$.

$$\begin{aligned} \frac{dM}{dt} &= \sum_{n \geq 0} np_{n+1} \left[\frac{k_1}{v} n(n+1) + k_2(n+1) \right] \\ &\quad + \sum_{n \geq 0} np_{n-1} [k_{-1}(n-1) + k_{-2}v] \\ &\quad - \sum_{n \geq 0} np_n \left[\frac{k_1}{v} n(n-1) + k_{-1}n + k_2n + k_{-2}v \right] \end{aligned}$$

We can re-index our sums, to gain a collection of p_n terms. For the first sum, $n+1 \rightarrow n$ and $n-1 \rightarrow n$ for the second sum. We are then able to start our sums back at $n \geq 0$ due to the first terms being zero. With some algebraic manipulation, we find a simplified sum for $\frac{dM}{dt}$.

$$\begin{aligned} \frac{dM}{dt} &= \sum_{n \geq 0} np_n \left[\frac{k_1}{v} - k_2 + k_{-1} \right] - \sum_{n \geq 0} \frac{k_1}{v} n^2 p_n + \sum_{n \geq 0} p_n k_{-2}v \\ &= \left[\frac{k_1}{v} - k_2 + k_{-1} \right] \langle n \rangle - \frac{k_1}{v} \langle n^2 \rangle + k_{-2}v \end{aligned}$$

Note that $\sum_{n \geq 0} p_n = 1$ since we are summing over all probability states.

3 Question 3

We will use the mean field moment closure approximation in order to find a closed equation in terms of $\langle n \rangle$ for $\frac{dM}{dt}$. The mean field closure assumes the variance is zero, however this is not always true. For approximations with smaller error, we would have to consider closing at a higher order. The mean field closure model states

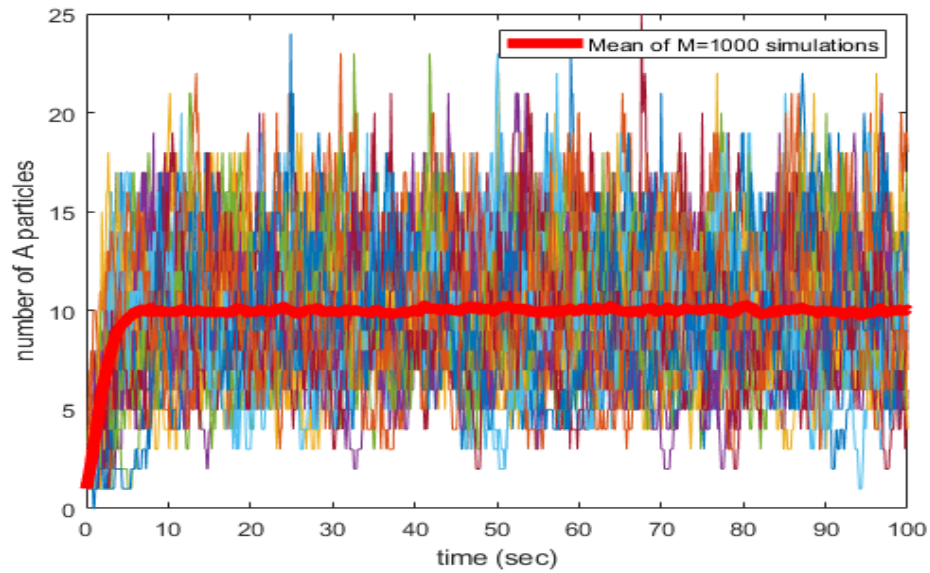
$$\langle n^2 \rangle = \langle n \rangle^2 \quad (2)$$

Applying this to our equation yields the following ODE:

$$\frac{dM}{dt} = \left[\frac{k_1}{v} - k_2 + k_{-1} \right] \langle n \rangle - \frac{k_1}{v} \langle n \rangle^2 + k_{-2}v$$

4 Question 4

We now plot the evolution of our average using MATLAB where we represent the time between reactions, τ as an exponentially distributed R.V. Explicitly, $\tau = \frac{1}{a_0} \ln(\frac{1}{r_1})$. We are using the Gillespie algorithm as our SSA. Our time step is set as 0.1 seconds, we are simulating 1000 independent runs with a time interval $[0, 100]$.



We are able to see our simulated average finds a steady state relatively quickly, after approximately 10 seconds. Let's calculate the steady state values.

5 Question 5

We now approximate the steady state values by plotting the mean and variance.

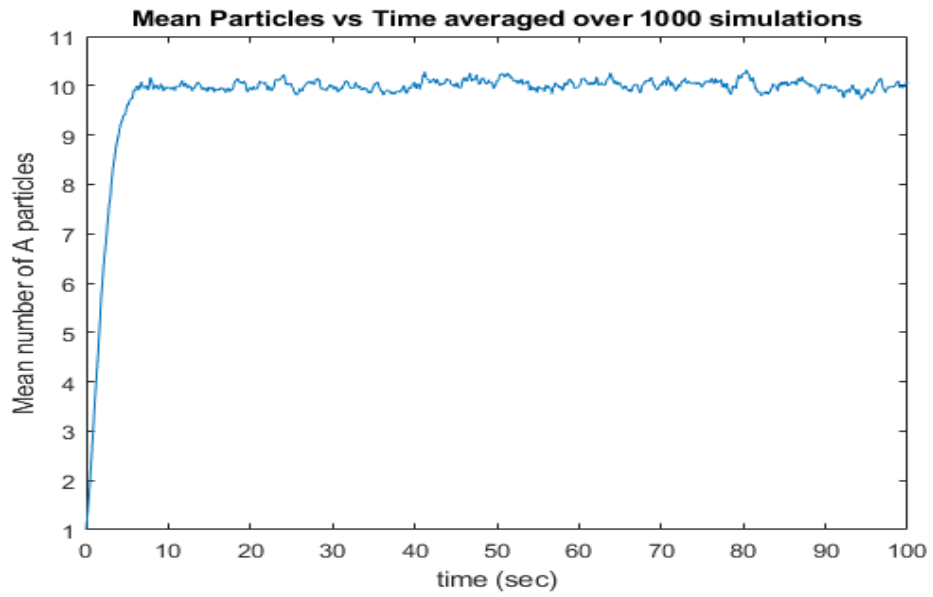


Figure 1: Simulated mean for 1000 runs

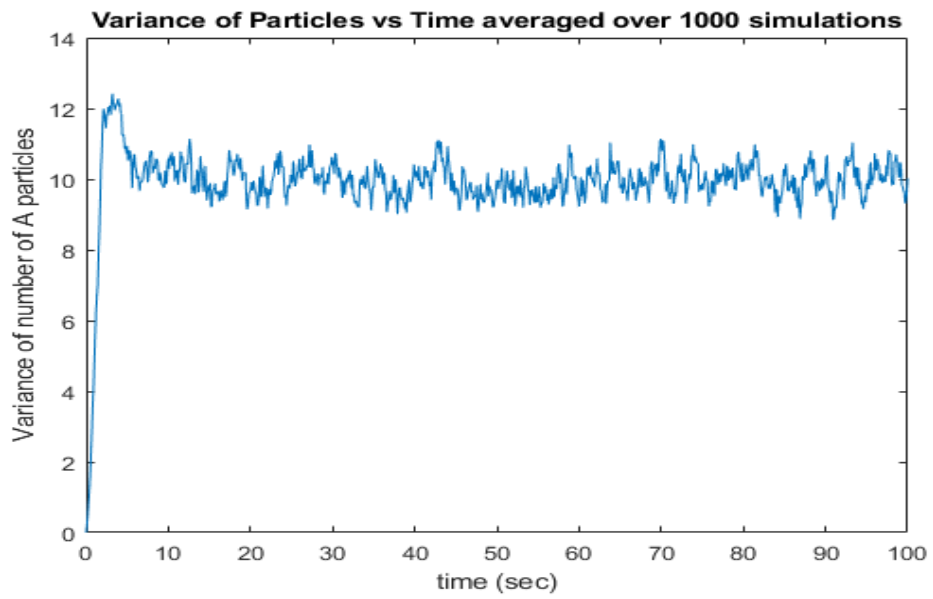


Figure 2: Simulated variance for 1000 runs

By calculating the average after both the mean and variance have stabilised,

we can approximate values for the stable steady states for the mean (M^*) and variance (V^*). We find the simulation mean stabilises after 10 seconds, and therefore we calculate our steady states as follows: $M^* = 9.9$ $V^* = 10.0$ to an accuracy of ± 0.1 .

6 Question 6

We overlay the deterministic ODE onto 4 to compare results between our mean-field approximation ODE and our stochastic simulation.

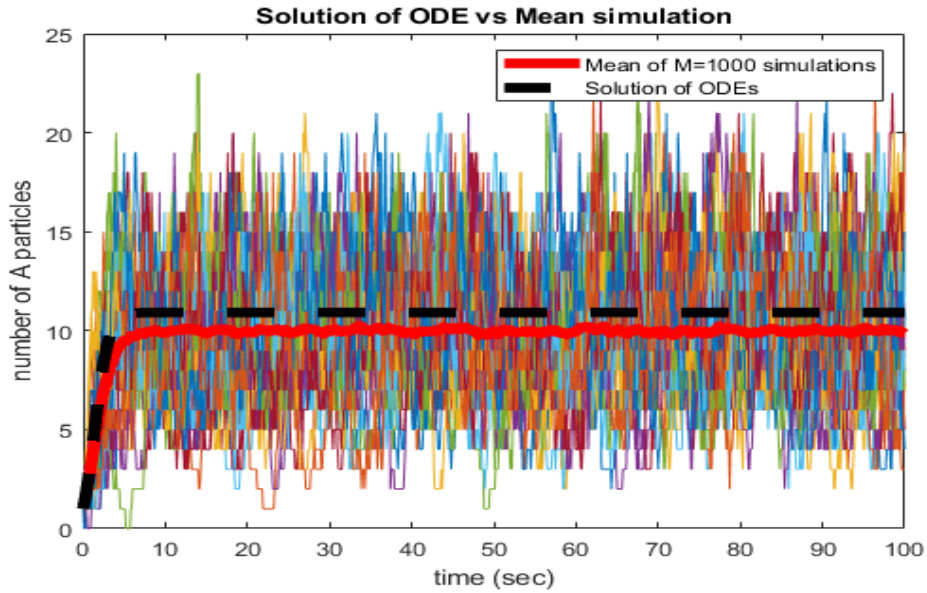


Figure 3: Simulated mean vs deterministic mean (using mean field closure)

7 Question 7

Using

$$\frac{dM}{dt} = \left[\frac{k_1}{v} - k_2 + k_{-1} \right] \langle n \rangle - \frac{k_1}{v} \langle n \rangle^2 + k_{-2}v$$

and setting $\frac{dM}{dt} = 0$ to find the steady states, M^* , we get the following quadratic equation.

$$\left[\frac{k_1}{v} - k_2 + k_{-1} \right] M^* - \frac{k_1}{v} M^{*2} + k_{-2}v = 0$$

This quadratic has roots

$$\frac{\frac{k_1}{\nu} - k_2 + k_{-1} \pm \sqrt{(\frac{k_1}{\nu} - k_2 + k_{-1})^2 + 4\frac{k_1}{\nu}k_{-2}\nu}}{2\frac{k_1}{\nu}}$$

After substituting in our constants, we find the mean to have steady states $M_1^* = 5 + \sqrt{35} \approx 10.9$ $M_2^* = 5 - \sqrt{35} < 0$. We can disregard M_2^* since it is negative, and therefore within the conditions of our model not physically possible. By conducting stability analysis on $\frac{dM}{dt}$ by taking it's derivative and substituting in our constant yields:

$$\left. \frac{d^2M}{dt^2} \right|_{M=10.9} = 1 - 0.2(10.9) < 0$$

Thus we show that M_1^* is a stable steady state.

8 Question 8

We must use the master equation to derive an equation for the evolution of the second moment of A . We are able to use a similar method to before, where we consider $\langle n^2 \rangle = \sum_{n \geq 0} n^2 p_n(t)$ and differentiate. We take care that the sum uniformly converges and therefore we are able to interchange the summation and the derivative operator to yield the following.

$$\frac{d\langle n^2 \rangle}{dt} = \sum_{n \geq 0} n^2 \frac{dp_n}{dt} \quad (3)$$

Substituting in $\frac{dp_n}{dt}$ from our master equation yields the following:

$$\begin{aligned} \frac{d\langle n^2 \rangle}{dt} &= \sum_{n \geq 0} n^2 p_{n+1} \left[\frac{k_1}{\nu} n(n+1) + k_2(n+1) \right] \\ &\quad + \sum_{n \geq 0} n^2 p_{n-1} [k_{-1}(n-1) + k_{-2}\nu] \\ &\quad - \sum_{n \geq 0} n^2 p_n \left[\frac{k_1}{\nu} n(n-1) + k_{-1}n + k_2n + k_{-2}\nu \right] \end{aligned}$$

We use the same method before, re indexing the sums so that we have a collection of p_n coefficients. For the first sum, we use $n+1 \rightarrow n$, for the second sum we use $n-1 \rightarrow n$ and we do not need to change the third sum.

$$\begin{aligned}
\frac{d\langle n^2 \rangle}{dt} &= \sum_{n \geq 1} (n-1)^2 p_n \left[\frac{k_1}{\nu} n(n-1) + k_2 n \right] \\
&\quad + \sum_{n \geq -1} (n+1)^2 p_n [k_{-1} n + k_{-2} \nu] \\
&\quad - \sum_{n \geq 0} n^2 p_n \left[\frac{k_1}{\nu} n(n-1) + k_{-1} n + k_2 n + k_{-2} \nu \right]
\end{aligned}$$

Note the summation limits change, however since in the first sum we have a collection of n 's, we are able to start the sum at zero since the first term is also zero. For the second sum, we have the convention that $p_{-1} = 0$ and therefore the term is zero and we can start at $n \geq 0$ instead.

We are able to collect all terms together, expand the brackets and simplify terms which leads to the following equation.

$$\frac{d\langle n^2 \rangle}{dt} = \sum_{n \geq 0} n^3 p_n \frac{-2k_1}{\nu} + \sum_{n \geq 0} n^2 p_n \left[\frac{3k_1}{\nu} + 2k_{-1} - 2k_2 \right] \quad (4)$$

$$+ \sum_{n \geq 0} n p_n \left[k_2 - \frac{k_1}{\nu} + k_{-1} + 2k_{-2} \nu \right] + \sum_{n \geq 0} p_n k_{-2} \nu \quad (5)$$

This simplifies to the following form

$$\frac{d\langle n^2 \rangle}{dt} = \frac{-2k_1}{\nu} \langle n^3 \rangle + \left[\frac{3k_1}{\nu} + 2k_{-1} - 2k_2 \right] \langle n^2 \rangle + \left[k_2 - \frac{k_1}{\nu} + k_{-1} + 2k_{-2} \nu \right] \langle n \rangle + k_{-2} \nu \quad (6)$$

9 Question 9

We use the normal moment closure method, meaning we set all cumulants $\kappa_i = 0 \quad \forall i \geq 3$. From the lecture notes, we see

$$\kappa_3 = \langle n^3 \rangle - 3\langle n^2 \rangle \langle n \rangle + 2\langle n \rangle^3 \quad (7)$$

Setting $\kappa_3 = 0$ and substituting 7 into 6 we find:

$$\begin{aligned}
\frac{d\langle n^2 \rangle}{dt} &= \frac{-2k_1}{\nu} [3\langle n^2 \rangle M - 2M^3] + \left[\frac{3k_1}{\nu} + 2k_{-1} - 2k_2 \right] \langle n^2 \rangle \\
&\quad + \left[k_2 - \frac{k_1}{\nu} + k_{-1} + 2k_{-2} \nu \right] M + k_{-2} \nu
\end{aligned}$$

10 Question 10

We numerically solve the coupled differential equations of $\frac{d\langle n^2 \rangle}{dt}$ and $\frac{dM}{dt}$ to analyse their steady states. We find the steady state of $\langle n^2 \rangle$ as $\langle n^2 \rangle^* = 108.8$ and for the mean, $M^* = 9.9$. We can therefore calculate a steady state value for the variance using the equation $V^* = \langle n^2 \rangle^* - (M^*)^2 = 108.8 - 9.9^2 = 10.8$. While to ± 0.1 accuracy we get the same mean, our variance is different by 0.8. This difference exists because in question 5 we used the mean field closure, which states $\langle n^2 \rangle = M^2$ where M is our mean. For the steady states, we find $(M^*)^2 = 98$ but $\langle n^2 \rangle^* = 108.8$ and therefore we need to use a higher order closure method such as the normal closure method.

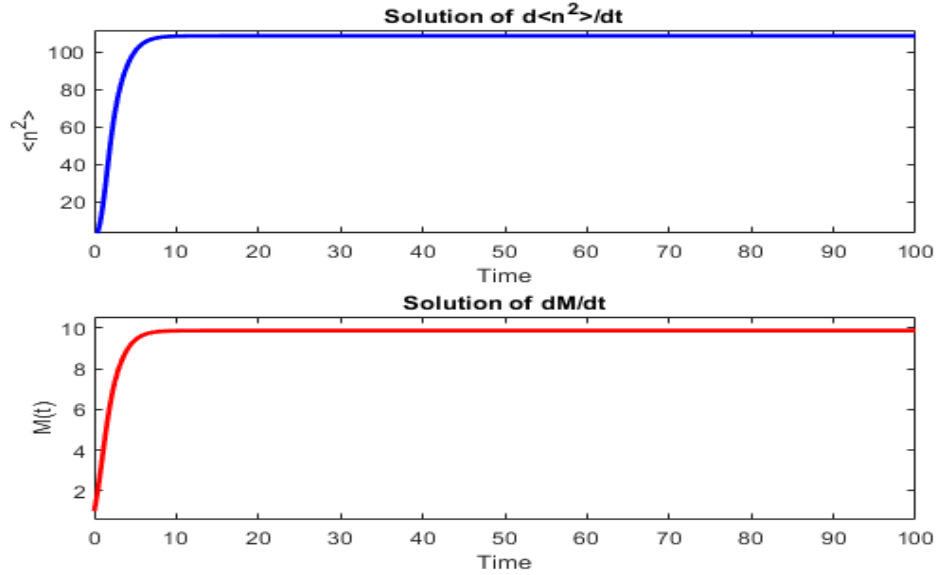


Figure 4: Numerical Solutions to the coupled ODE system.

11 Question 11

By using our previous steady state analysis from question 5, we plot our simulated stationary distribution against a normal distribution with parameters $M^* = 9.9$ $\sigma = \sqrt{V^*} = 3.2$ to yield the following comparison:

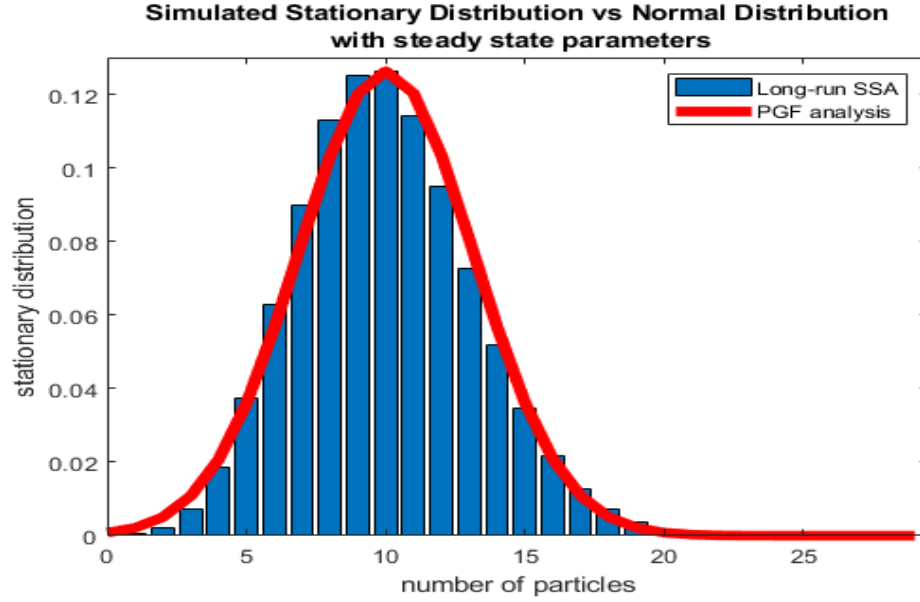


Figure 5: Simulated stationary distribution comparison

We are able to see that while our steady state normal distribution with parameters described above is a very close approximation to the simulated distribution, it is skewed slightly to the right, meaning our normal distribution overestimates the amount of particles. We do however find a much better agreement between the normal distribution approximation and the stationary distribution than we did between Question 10 and Question 5.

12 Question 12

We now consider a tau-leaping simulation using tau values of $\tau = [1, 0.5, 0.25, 0.125]$

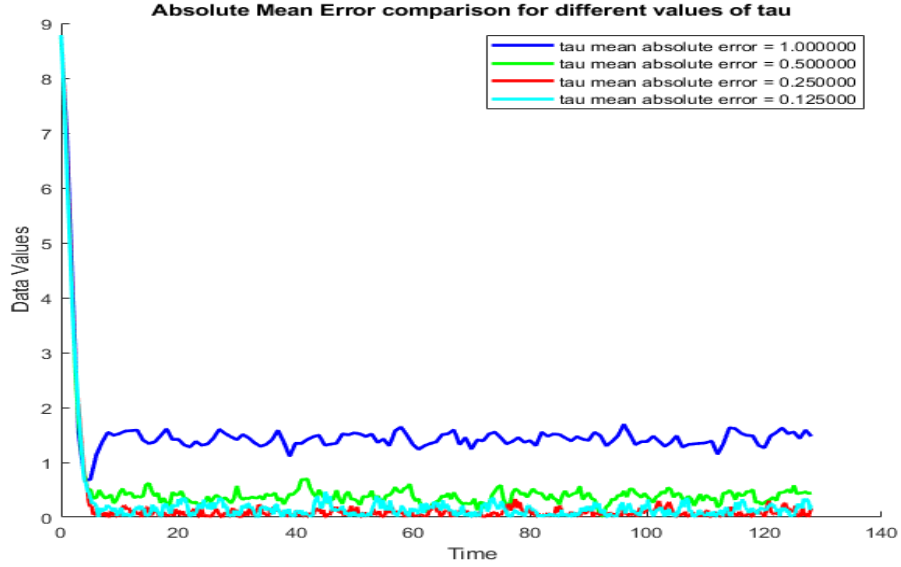


Figure 6: Absolute error between steady state of stochastic simulation and tau leaping approximation

We find the following steady states for our simulation:

```
tau=1 tau=0.5 tau=0.25 tau=0.125
paths=2426 paths=1572 paths=1188 paths=1050
mean_A=8.3108 mean_A=9.3378 mean_A=9.6195 mean_A=9.999
```

13 Question 13

To derive a probability density function for the time interval between updates of the propensity functions, we consider each reaction as an independent Poisson process. From the lecture notes, we find that these reactions occur with $PP(\alpha_i t)$ for reaction i and propensity function α_i .

Since we have a collection of independent Poisson Processes, we are able to sum over the Poisson processes for each reaction channel. Note that we can say they are independent Poisson Processes since we are not updating the propensity functions between each reaction. This yields a combined Poisson process with rate $\alpha_0 t$ where

$$\alpha_0 = \sum_{i=1}^M \alpha_i$$

where we have M possible reaction channels. We also know that the time interval between reactions given by a Poisson Process is distributed exponentially.

Explicitly $\tau \sim \text{Exp}(\frac{1}{\alpha_0})$. Therefore, for R total reactions we have a total time interval given by the sum of τ .

$$T = \sum_{n=1}^R \tau \sim \text{Gamma}(R, \frac{1}{\alpha_0}) \quad (8)$$

Since the sum of exponential distributions is a gamma distribution. Therefore the probability density function from which the time increment should be drawn is given by

$$\mathbf{P}(T = x) = \frac{x^{R-1} e^{-\frac{1}{\alpha_0} x} (\frac{1}{\alpha_0})^R}{(R-1)!}$$

Since we have integer values for R , this is an Erlang Distribution.

We want to find a way of distributing R reactions between each different reaction channel. One method to do this may be to consider the propensity functions as weights. If we say for reaction channel i , we have a total of $\frac{\alpha_i}{\alpha_0} R$ reactions occurring, this then weights on the likelihood of that reaction taking place, allowing accurate sampling. An issue with the aforementioned method may be that there is no condition of $\frac{\alpha_i}{\alpha_0} R$ being an integer, however in a physical system only integer amounts of reactions can occur. We could apply the floor function to this, however it is not then guaranteed that we will reach R total reactions in the given interval.