Imperial College London

November 13, 2023

MRes Systems and Synthetic Biology coursework

Stochastic Gene expression and Gillespie algorithm

Please explain your results clearly, include figures and your code as supplement.

1. Consider an mRNA that is produced with a constant rate $k_0 = 0.2s^{-1}$ and decays with rate $k_1 = 0.01s^{-1}$. The reaction model has the following form: $\varnothing \xrightarrow{k_0}$ mRNA, mRNA $\xrightarrow{k_1} \varnothing$. You can assume cell volume is set to 1 in some units. What is the mean expression level of this mRNA at steady-state?

2. Implement the Gillespie algorithm (Stochastic Simulation Algorithm) for the mRNA birth-death process as described above in python (or any other programming language you prefer). Starting from no mRNAs at time zero, plot simulation results from 5 independent simulations for number of mRNAs as a function of time $(0 - 1000s)$.

3. Use the Gillespie simulations to estimate mean and variance of mRNA numbers at steady-state (by removing the initial transient). Use these estimates to calculate the Fano factor (defined as variance over mean). Can you justify the value you obtain?

4. Modifying your Gillespie algorithm to include protein dynamics. The extended model of stochastic gene expression reads now:

$$\begin{aligned}
&\varnothing \xrightarrow{k_0} \text{mRNA},\\
&\text{mRNA} \xrightarrow{k_1} \varnothing,\\
&\text{mRNA} \xrightarrow{k_2} \text{mRNA} + \text{Protein},\\
&\text{Protein} \xrightarrow{k_3} \varnothing.
\end{aligned}$$

   Assume that each mRNA molecule is used as template for translation of a protein with a rate of $k_2 = 5s^{-1}$ and decay rate $k_3 = 1s^{-1}$ and visualise several representative sample paths. Estimate mean and Fano factors for the protein numbers.

5. Now consider the transcription rate is a static random variable varying several orders of magnitude from the nominal value used above for each gene. Sample randomly 500 transcription rates and, for each parameter, record the mean, variance, and coefficients of variations of mRNA and protein simulations. Visualise the distributions of transcription rates, mRNA and protein expression levels. Produce scatter plots of noise versus mean relationships and interpret the result with respect to genome-wide studies of molecular noise.