Imperial College London

MRes Systems and Synthetic Biology coursework

Stochastic Gene expression and Gillespie algorithm

1. Consider an mRNA that is produced with a constant rate $k_0 = 0.2s^{-1}$ and decays with rate $k_1 = 0.01s^{-1}$. 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 matlab (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. Use these estimates to calculate the Fano factor (defined as variance over mean). Can you justify the value you obtain?

4. Now consider expression of the same mRNA $(k_0 = 0.2s^{-1}$ and $k_1 = 0.01s^{-1})$ in a cell that is dividing once every $T_{cellcycle} = 20$ minutes. Assume mRNA molecules are binomially partitioned between the daughter cells at cell division. By modifying the Gillespie algorithm above simulate mRNA numbers as a function of time as you follow a single lineage of cells in time (follow randomly one of the daughter cells after each cell division). Estimate mRNA Fano factor using both the overall mRNA trajectory and the mRNA numbers in the newly born cells. Discuss the results obtained and contrast it with results obtained in the previous case (please attach your code for this part).

5. We have done some single cell measurements of a given mRNA numbers in a bacteria and have obtained mean = variance = 10 for this mRNA. We want to use the Approximate Bayesian Computation (ABC) Scheme (see the notes about this) to decide if the appropriate model that describes the data is the model introduced in part 1 (with decay but no cell cycle) or the model in part 4 (without decay but cell cycle). For the ABC similarity distance measure between model output and data use the sum of square difference between the mean of the data and mean of simulated data and square difference between variance of the data and simulated data. Generate 100 simulations from each of the model using random parameter values k_0 and k_1 over the prior range $[0, 0.2]s^{-1}$ and k_1 over the prior range $[0, 0.02]s^{-1}$ and estimate the values of the similarity measure. For the second model keep the $T_{cellcycle} = 20$ minutes constant. Make a histogram of the similarities obtained in each model. Based on this, which model is more likely to explain the data?