practical_6_stochastic_simulations_solutions

November 9, 2019

Name: Write your name here

1 SYSTEMS AND NETWORK BIOLOGY - PRACTICAL 6

2 Stochastic description of protein expression

Here we implement a stochastic description of simple protein expression processes, using a stochastic simulation algorithm know as the *next-reaction method*. We will consider a biochemical process constituted by M reactions, denoted by $\mu = 1, 2, \ldots, M$, and N biomolecular species, denoted by $i = 1, 2, \ldots, N$.

Let the rate of reaction μ be k_μ, and their corresponding reactants $x_i^{(\mu)}$. The probability that this reaction occurs in the interval $(t, t + dt)$ given the state of the system at time t is given by $a_\mu dt$, where a_μ is the *propensity* of the reaction. Following the law of mass action, and assuming that only one molecule of each reactant species is involved, the propensity of reaction μ has the form:

$$a_\mu = k_\mu \prod_{i=1}^{N} x_i^{(\mu)}$$

Another important piece of information that characterizes the reactions is the *stoichiometry matrix* $S_{\mu,i}$. The columns of this matrix are the reactions and the rows are the biomolecular species. The elements of the matrix represent how each reaction affect each species (e.g. a species that loses one molecule in a reaction has a value -1 in the corresponding matrix location, while if the reaction produces two molecules of a species, the corresponding matrix element is 2). Once the propensities and the stoichimetry matrix are known, the steps of the next-reaction stochastic simulation algorithm are:

1. Initialize the system (number of molecules and time)
2. Calculate the propensities of all reactions
3. Calculate the time τ at which the next reaction happens, using the expression:

$$\tau = \frac{1}{a_{tot}} \ln \frac{1}{z_1}$$

where $a_{tot} = \sum_{\nu=1}^{M} a_\nu$ and z_1 is a random number uniformly distributed between 0 and 1.

4. Determine the reaction that occurs at time τ by calculating the smallest integer λ such that

$$\sum_{\nu=1}^{\lambda} a_\nu > z_2 \sum_{\nu=1}^{M} a_\nu$$

where z_2 is another random number uniformly distributed between 0 and 1.

5. Update the molecule numbers according to the stoichimetry coefficients $S_{\mu,i}$.

6. Go to step 2 to compute the next reaction.

Let us implement this algorithm for a simple birth-death process representing protein expression and degradation:

$$\emptyset \overset{k_{exp}}{\rightarrow} A, \quad A \overset{k_{deg}}{\rightarrow} \emptyset$$

First, we load all necessary `Python` packages.

In [60]:

Next, we define the values of the parameters, the stoichimetry matrix and the initial conditions (step 1 above). We consider first that $k_{exp} = 100$ and $k_{deg} = 1$.

In [61]:

Now we compute the propensities (step 2 above), calculate the next reaction time (step 3), and determine the reaction that takes place (step 4). For steps 3 and 4 use the function `uniform` from the module `numpy.random`.

In [62]:

Finally, we update the molecule numbers and the time variable (step 5 above).

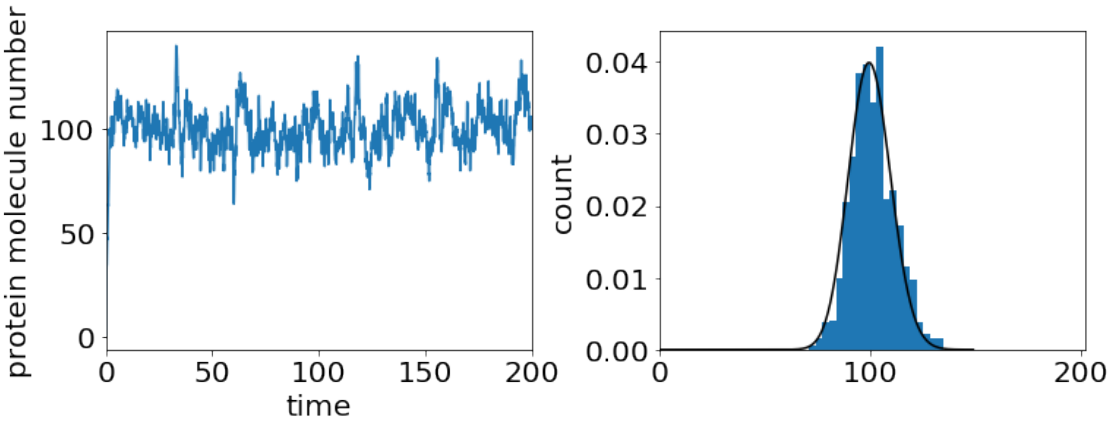
In [36]:

Now put together steps 2-5 above in a loop, and generate output every mt time units (you might need to save the same state of the system for multiple output times, if no reaction happens in a given mt interval). Run the simulation for 200 time units with sampling period $mt = 0.1$.

In [44]:

Now we plot the time series of molecule numbers, and the histogram of molecule number of all data during the second half of the simulation (stationary state). Normalize this histogram so that it has unit area, and compare it with the expression of the Poisson distribution derived in class.

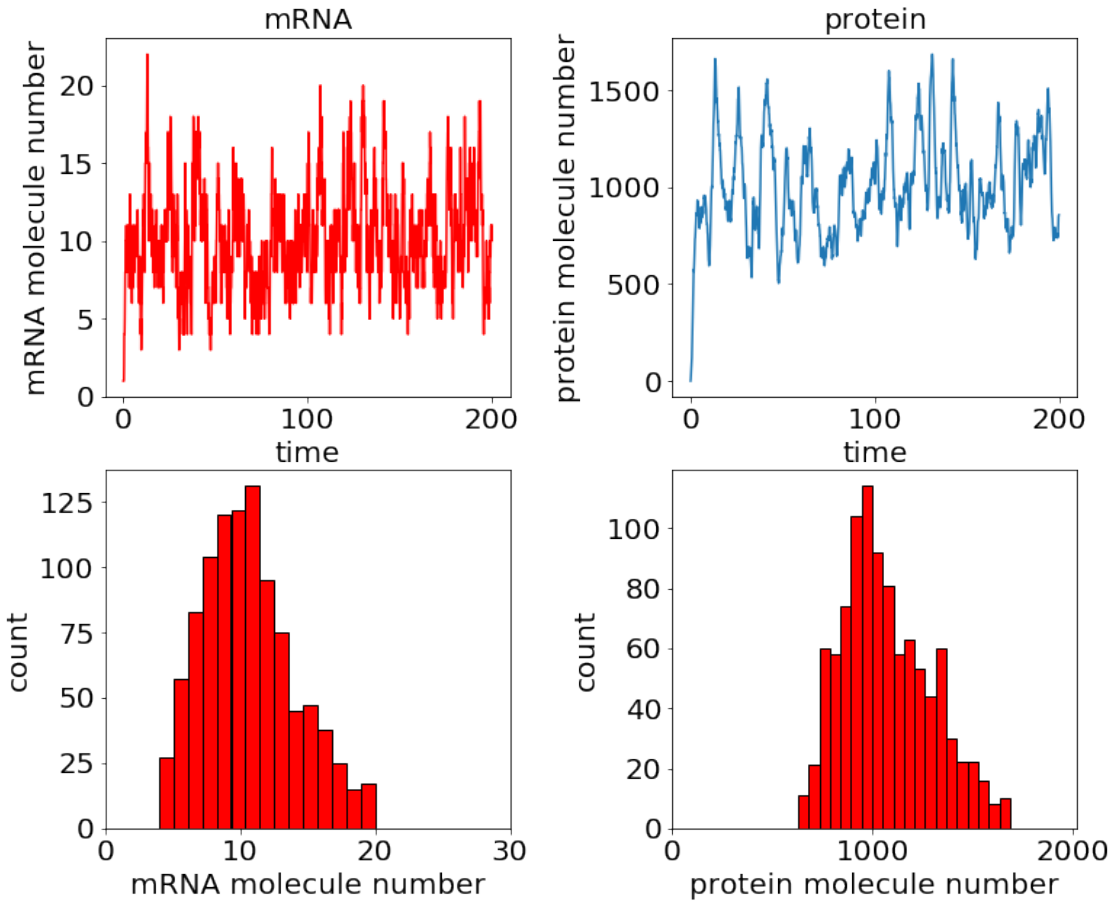
In [51]:



Comment your result here

Next, we separate the expression into transcription and translation. We thus have four reactions (transcription, translation, mRNA degradation and protein degradation) and two species (mRNA and protein). Integrate the system of reactions and plot the resulting time series and normalized histograms for the two species. Consider a transcription rate $k_{tx} = 10$, a translation rate $k_{tl} = 100$, and $k_{deg} = 1$.

In [64]:



Comment your result here

Finally, scale the transcription and translation rates by the same factor, multiplying one of the rates and dividing the other, in such a way that the product remains constant, and discuss what happens with the mean and standard deviation of the distribution of protein numbers for different values of this factor (smaller and larger than 1).

In [ ]:

Comment your result here
