***Modelling Gene Expression 2***

**INVESTIGATION USING A STOCHASTIC MODEL**

In the computer room practical sessions we build up the steps necessary to carry of a stochastic analysis of a simple system (red and green balls drawn from a bag).

We can use the same steps to model other systems that we have previously encountered. These are provided as files in the stochastic examples folder.

**Coding the stochastic system**

*Read through the examples and check you understand the structure of the stochastic versions and how they relate to the deterministic models.*

The template file gene\_exp\_stoch1.py is provided to run a simulation of stochastic gene expression (without including negative feedback via repression of transcription).

*1a) Complete the code to simulate the system stochastically:*

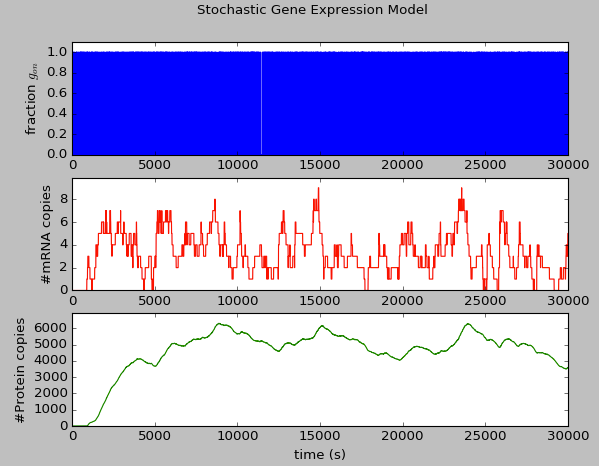
*- Fill the blanks left in the function* gillespie\_gene\_expression

*Steps 0 and 1 are already filled out.*

*Steps 2, 3 and 4 need completing.*

*- Input the correct initial conditions and parameter values.*

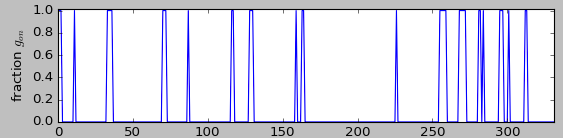
*When complete the code should plot the species concentrations over the time period 0 < t < 30000.*

*.*

*b) Contrast the observed behaviour of the stochastic simulation to the output of the deterministic model*

The value of gon now switches rapidly between 0 and 1. In the deterministic model the fractional value rapidly reached an equilibrium level ~0.17.

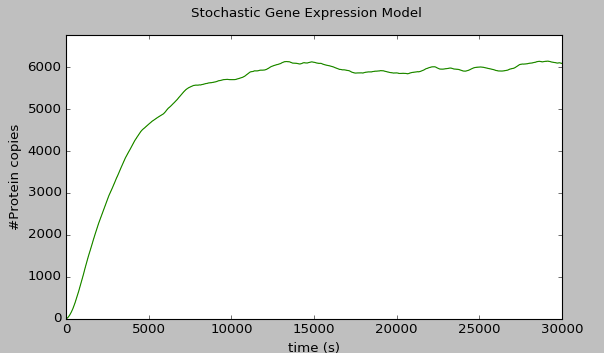
Zooming in close up we can see on average the single gene modelled in the stochastic system does spend less time in the active state in accordance with this:



In the deterministic model the value of RNA increased to a steady state value of 4 copies over a time of ~2500s. The stochastic model shows a similar behaviour but with significant fluctuations around the steady state value, sometimes falling to zero, and sometimes peaking at 9 copies.

In the deterministic model, following the rise in mRNA levels protein levels rose more slowly, reaching a steady state value of 6000 copies after ~15000s. The trajectory in the above stochastic system shows similar behaviour but agin there is noise and fluctuations around the steady state (although the fluctuations seem to be less rapid and less extreme relative to the mRNA trajectory).

*c) Add code that plots the average for the protein number over 100 simulation runs. How does this averaged behaviour compare to the deterministic model?*



Average trajectory is now in reasonable agreement with the output of the deterministic system.

**Measuring the noise expressed by the stochastic system**

*2a) Work with the results of 100 runs. Analyse the final state attained by the system at t=30000. Create a histogram showing the copy numbers of:*

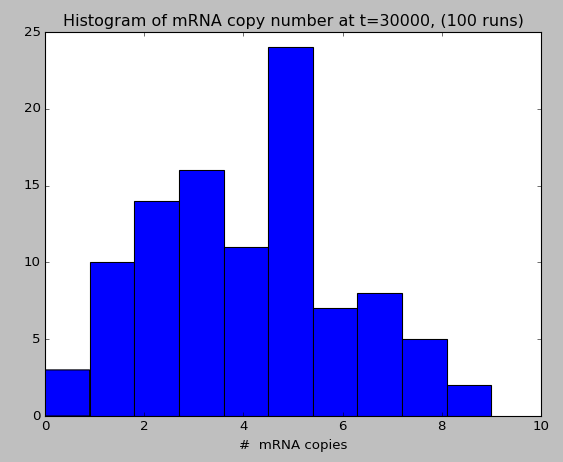
*i) mRNA*

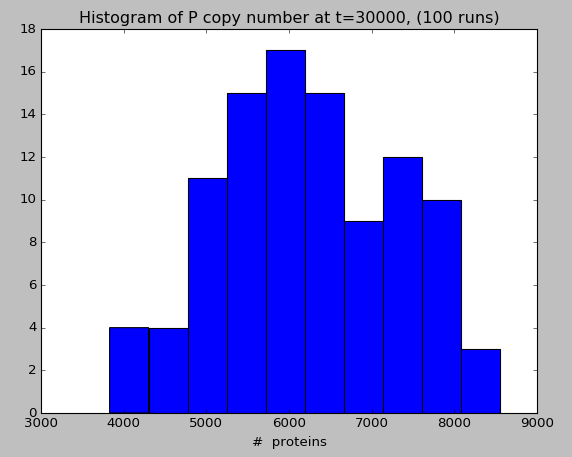
*ii) protein distribution*

**Note. For the following answers because of the relatively low number of runs we analyse the plots and values calculated may vary considerably.**

The important thing is to check you have the right techniques for generating and analysing the distributions.

(Ideally we would use at least 500 runs for the exercises but this would take a long time to run)

**

**

*Find the mean and standard deviation of these distributions and comment briefly on their shape and whether they follow the normal distribution.*

Mean mRNA copies: 4.06, stdev: 2.15

Mean Protein copies: 6240, stdev:1100

The distributions are not easily discernable to the eye – in this case 100 runs is not enough to provide a reliable guide to the shape.

Additionally because of the low copy numbers we should more appropriately try to identify whether the distribution of mRNA is consistent with a Poisson distribution.

*b) The coefficient of variation CV is a measure of the relative noise level exhibited by a system.*

*Calculate this value for the mRNA and protein distributions at time t=30000.*

*Result of code:*

Mean mRNA copies: 4.14, stdev:2.11

Mean Protein copies: 6026, stdev:1121

CV mRNA: 0.51

CV P: 0.186

*c) Adjust the model to include negative feedback via repression of transcription, using the same parameters as used in the deterministic model.*

*Repeat the calculation of CV and compare your finding to the previous result.*

*Result of code*

Mean mRNA copies: 2.39, stdev:1.5

Mean Protein copies: 3407, stdev:629.4

CV mRNA: 0.627

CV P: 0.185

*(For my results – yours may differ) The relative noise level of the mRNA has increased but the noise in the protein level has stayed approximately the same.*

**Simulating transcriptional bursting**

In the model parameters used so far we have assumed that the average switching time for the gene to transition between the active and inactive state occurs on a timescale of seconds.

Recent experimental developments mean that we can observe mRNA transcription occurring it situ and able to quantify mRNA copy numbers on an absolute basis (i.e. count the number of mRNA transcripts).

It has been observed that often mRNA is produced in bursts (on a time scale of minutes or longer) that are separated by periods of inactivity.

In our model we can simulate this by adjusting the rates that the gene switches between its active and inactive state. The original values we used are:

|  |  |
| --- | --- |
| Model parameter | Value (seconds-1) |
|  |  |
|  |  |

|  |  |
| --- | --- |
| Model parameter | Value (seconds-1) |
|  |  |
|  |  |

Let’s adjust these in to the following values:

Note: this value was originally set at 5x60

(a typo)

*3a) Repeat the analysis that estimated average time spent in the active/inactive states and the average proportion of time that the gene will spend in the active state (exercises 5a. b. and c.) Comment on your findings.*

*FROM PREVIOUS EXERCISES*

*5 a) “If the gene in its active state switches to its inactive state with rate koff the average time interval over which it remains active can be shown to be 1/koff”*

*koff* = s implies average time gene remains active before switching to inactive is 1/ *koff* = s or 6 minutes

*b)*

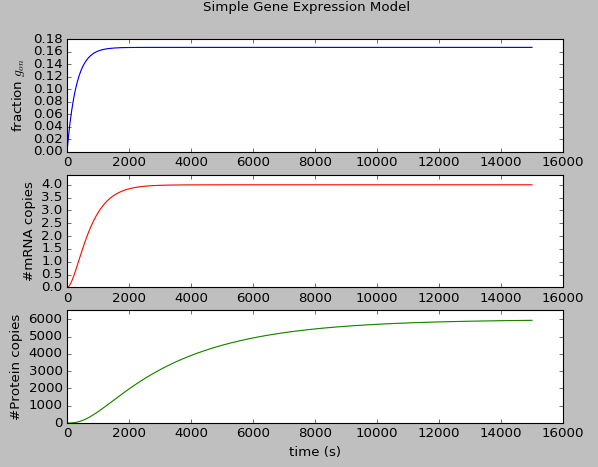
*kon* = implies average time gene remains active before switching to inactive is 1/ *knf* = s or 30 minutes

*c)*

If gene remains in active state for an average of 5min and inactive state for average time of 30min then over an “average” cycle of 36min it spends 6 min active and 30min inactive, i.e. fraction of time each gene is active averages to 6/36 = 0.167.

In this case the rate of switching has reduced so that the transition times are much longer (minutes rather than seconds). However the average fraction of time spent in the active state remains the same (this is because the parameters were both scaled by the same amount).

*b) How does the parameter change affect the behaviour of the model using a deterministic simulation? Comment this result.*

**

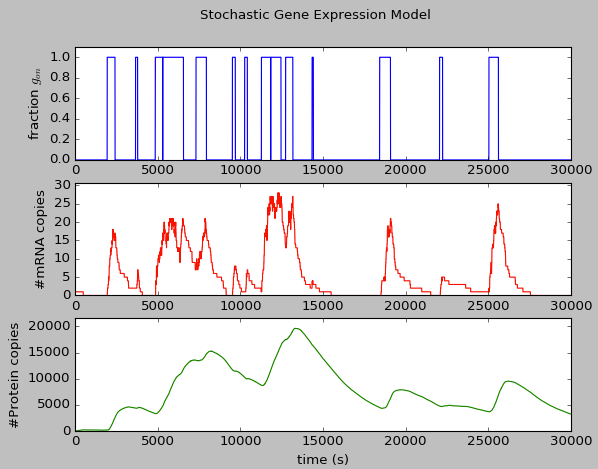
The fraction of cells with the gene in the active state gon now takes longer to reach equilibrium but reach the same steady state level.

The increase in the take taken for the gon fraction to stabilise does not seem to have a large effect on the timescales taken by mRNA and protein which remain are roughly unchanged, and reach the same steady state values.

*c) How does the parameter change affect the behaviour of the stochastic model? Repeat the analysis you completed in exercises 1 and 2 for the updated system.*

*Comment on your findings.*

*Stochastic simulation of Gene Expression with bursting (no negative feedback)*

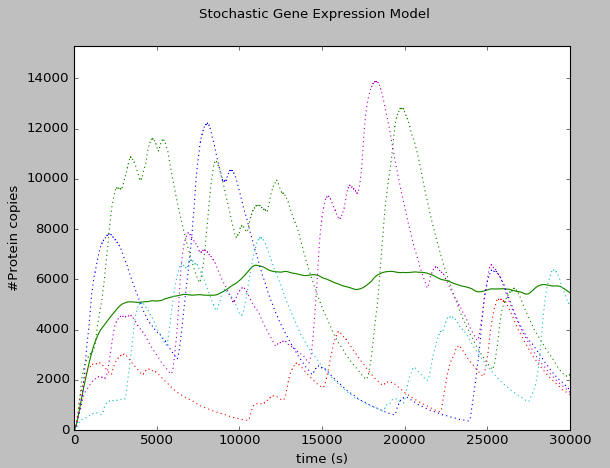
**

In the modified stochastic simulation we see very different behaviour to the deterministic model (in which the overall behaviour of the system was unchanged except for a slight decrease in the steady state levels.)

Now there are longer intervals between switching meaning:

* very long periods in the inactive state where the mRNA level falls to zero, in this state the system undergoes a prolonged fall in protein level.
* longer periods in the active state in which mRNA levels rapidly ramp up to high levels far in excess (x8) of their steady state values. When the system is in such as state it undergoes a prolonger period of increasing protein level.

This is very different behaviour and shows a clear difference in behaviour to the fast active-inactive gene switching and could be measured experimentally by observations of a single cell.

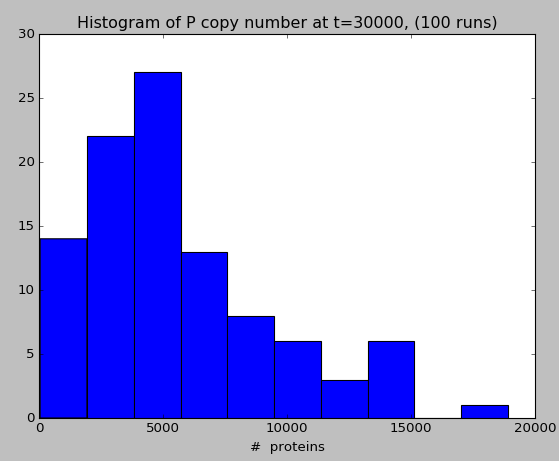


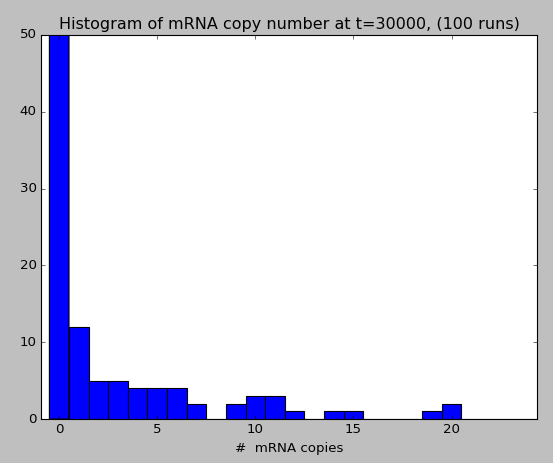
Average of 100 runs in green.

Dotted lines show examples of 5 individual runs

When we plot the average we see that although the individual cell behviour is very different to the deterministic model the averaged trajectory is fairly similar and the average protein level over a collection of cells remains similar to the case with fast gene active/inactive switching.

We might also note that there seems to be less time taken for the steady state level to be reached.





We see now that the distribution of copy numbers now looks significantly non-normal and are skewed to have a long tail on the RHS. The standard deviation is much greater indicating wider distributions for mRNA and protein, and around 50% of the runs have zero mRNA copies after t=30000.

Comparing the measured statistics:

**Results with slow gene active/inactive switching**

Mean mRNA copies: 2.92, stdev: 4.67

Mean Protein copies: 5468, stdev: 4334

CV mRNA: 1.6

CV P: 0.793

**Original Results with fast gene active/inactive switching**

Mean mRNA copies: 4.14, stdev: 2.11

Mean Protein copies: 6026, stdev: 1121

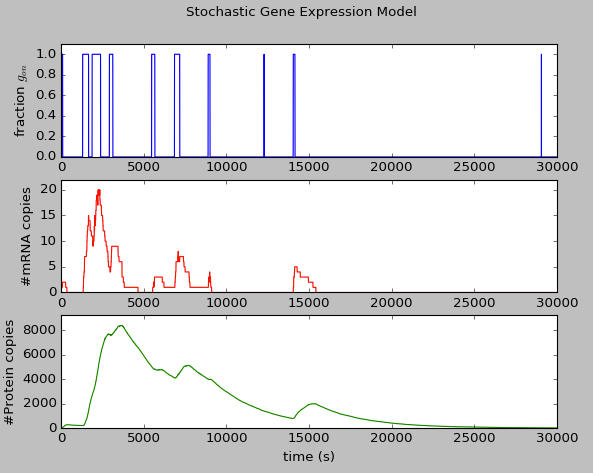
CV mRNA: 0.51

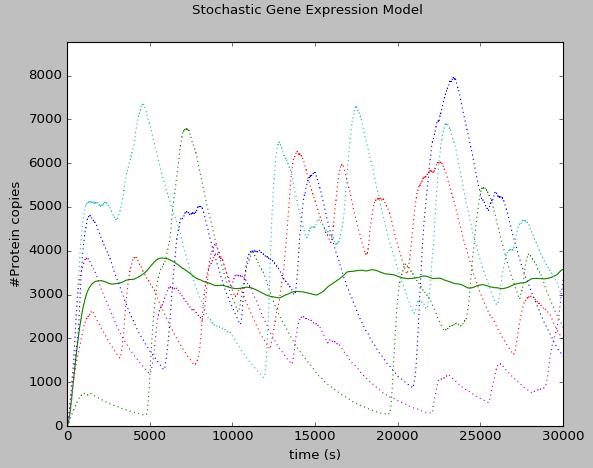
CV P: 0.186

When the switching speed is slow the mean levels of mRNA and protein drop (mRNA by ~ 25% and protein by ~10%).

The CV values that are a measurement of relative noise are both greater (mRNA noise is a factor of 3 greater, protein noise is a factor of 4 greater).

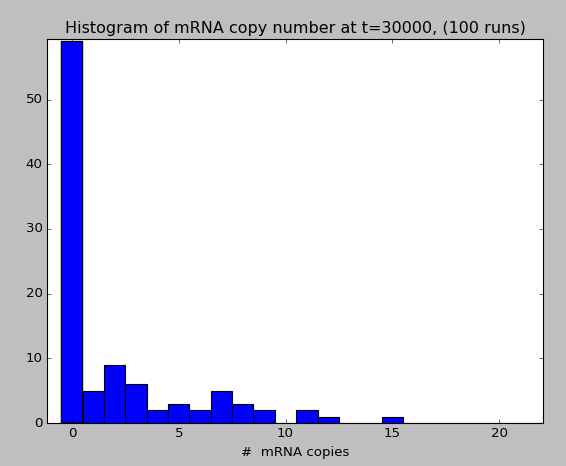
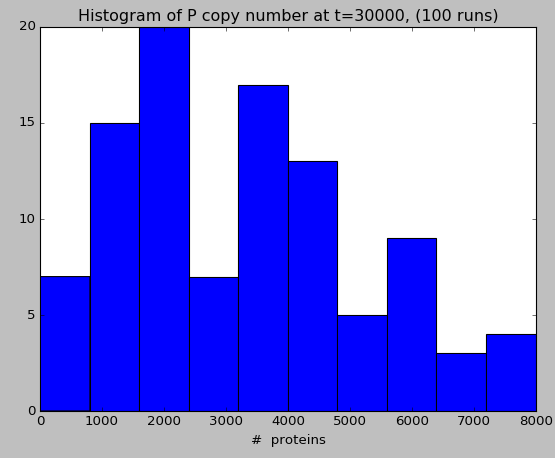
*If you are interested you can also repeat the analysis with the stochastic simulation of Gene Expression with bursting (including negative feedback)*





Average of 100 runs in green.

Dotted lines show examples of 5 individual runs.



We can again see very different behaviour, with >50% of of runs containing zero copies of mRNA. The proten distribution looks less skewed – although the low statistics make it difficult to be sure.

**Results with slow gene active/inactive switching and neg. feedback**

Mean mRNA copies: 2.02, stdev: 3.28

Mean Protein copies: 3311, stdev: 1955

CV mRNA: 1.62

CV P: 0.591

**Original results with fast gene active/inactive switching and neg. feedback**

Mean mRNA copies: 2.39, stdev: 1.5

Mean Protein copies: 3407, stdev: 629.4

CV mRNA: 0.627

CV P: 0.185

Again we see a drop (~20%) in the steady state mRNA level, and (~3%) in the protein steady-state case, but a big increase (x2.5 for mRNA and x3 for protein) in the relative noise level.