## Stochastic Gene Expression and Cellular Variability

The goal of this lab is to explore stochastic gene expression, cellular variability, and the influence of feedback on noise. For your submission, include all MATLAB code, figures generated, and answers to included questions.

## Problem 1 (Demo): Simulation of stochastic gene expression dynamics

Consider the following simple birth-death process, where mRNA $M$ is produced at rate $\alpha$ and degraded at a concentration dependent rate $\gamma M$:

$$\overset{\alpha}{\rightarrow} M \overset{\gamma}{\rightarrow}$$

This process can be modeled using the following differential equation, treating the amount of $M$ as a continuous and deterministic quantity.

$$\frac{dM}{dt} = \alpha - \gamma M$$

However, in real systems, we observe differences that arise from intrinsic noise, or noise from stochastic gene expression. To account for this stochasticity, we need a different kind of approach from the continuous model described earlier. The Gillespie algorithm is a useful approach for carrying out such stochastic simulations of biological reaction networks and can be described with the following generic pseudocode:

1. **Initialization:** Specify the time range of the simulation, initialize reaction kinetics parameters, and setup initial chemical concentrations.
2. **Monte-Carlo:**
   I. Randomly find waiting time until the next event
   II. Randomly determine which event has occurred (i.e., birth or death).
3. **Update:** Update the current time based on the waiting time from step 2
4. **Repeat:** Iterate through steps 2 and 3 until stopping criteria is met (i.e., current time >= max simulation time)

A) Implement the Gillespie algorithm and simulate $M$ given $M_0 = 0, \alpha = 30\ \#\ min^{-1}, \gamma = 1\ min^{-1}, t_{max} = 720 min$. Plot the first $10 min$ of the stochastic simulation along with ODE solution for $\frac{dX}{dt} = \alpha - \gamma X$.

B) Calculate and report the mean $\mu$ and variance $\sigma^2$ for $M$ from part A). Regularize the time-course $M(t)$ to the range $0{:}\,0.1{:}\,720\ min$ using the built-in `interp1` function with the "previous" method to prevent the variable time-step of the Gillespie algorithm from skewing the results.

## Problem 2 (Demo): Modeling stochastic gene expression with SimBiology

We will use the same model from problem 1 and implement it using the SimBiology toolbox in MATLAB.

A) Implement the birth-death model using SimBiology using the same parameters as problem 1, $M_0 = 0, \alpha = 30\ \#\ min^{-1}, \gamma = 1\ min^{-1}, t_{max} = 720 min$. Create a new program with the `Simulation` method and select the `Stochastic` solver. Simulate the model with a stop time of

$720min$. Plot your results, and if using a GUI for constructing the model include a screenshot of your model in your submission.

B) Export the data from the simulation in part A) to your MATLAB workspace and calculate the mean $\mu$ and variance $\sigma^2$ of $M$.

C) Plot the histogram for $M$ and compare the distribution with a Poisson distribution, $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$, for $x = 0,1,2,3, ...$, with $\lambda = \mu$ from part B. How does the distribution of $M$ compare to the Poisson distribution?

D) Simulate 1000 realizations of the model using the "Run Ensemble Program" in the SimBiology model analyzer interface or using the `sbioensemblerun` command. Plot the distribution of the concentration of M at the final timepoint of 720 minutes.

**Problem 3: Expanding the transcriptional bursting model**

A) Let's expand on our initial model to account for multiple promoter states, $P_0$ and $P_1$. We assume that $P_1$ is the transcriptionally active promoter state. The active promoter allows for transcription of mRNA with rate $k_{tsc}$. The mRNA can be degraded with rate $\gamma$ or translated to protein X with rate $k_{tsn}$. The protein X degrades with rate $\gamma_2$. The network of reactions for this process is then

$$P_0 \underset{k_b}{\overset{k_f}{\rightleftharpoons}} P_1$$
$$P_1 \xrightarrow{\alpha} P_1 + mRNA$$
$$mRNA \xrightarrow{k_{tsn}} mRNA + X$$
$$mRNA \xrightarrow{\gamma_1} \emptyset$$
$$X \xrightarrow{\gamma_2} \emptyset$$

A) Use the SimBiology model builder to implement the transcriptional bursting model described above using the parameters $k_f = 5\ min^{-1}, k_b = 5\ min^{-1}, \alpha = 0.1\ min^{-1},\ \gamma_1 = 0.23\ min^{-1},\ \gamma_2 = 0.02\ min^{-1}, \mathrm{k_{tsn}} = 20\ \mathrm{min^{-1}}\ t_{max} = 240min$. Assume that the promoter starts in the active state Assume that the promoter starts in the active state, $P_1 = 1, P_0 = 0$ and initial $mRNA = 2\ \#, X = 2174\ \#$. If using the GUI for model development, include a screenshot of the model in your submission. Simulate your model with the stochastic solver in simbiology. How does the average value of X compare to your initial condition?

B) Simulate 1000 realizations of the model using the "Run Ensemble Program" in the SimBiology model analyzer interface or using the `sbioensemblerun` command.

C) Plot the histogram of $X$ at the final time point, $t_{max} = 240min$. Describe your observations on the distribution.

D) Calculate the standard deviation (SD) using the built-in `std` function in MATLAB and coefficient of variation (CV) of the protein concentration. The CV can be calculated as $\frac{standard\ deviation}{mean}$.

E) Repeat A-C with the following modifications of parameters (the rest of the parameters unchanged):

   a. Slower promoter switching $k_f = 0.01\ min^{-1}, k_b = 0.01\ min^{-1}$

   b. Slower translation, faster transcription: $\alpha = 1\ min^{-1}, \mathrm{k_{tsn}} = 2\ \mathrm{min^{-1}}$

Compare the histograms with one another and the one generated with C and comment on how these changes affected the noise in gene regulation and why.

**Problem 4: Feedback and regulatory noise**

Let's explore how feedback influences gene regulatory noise. Using the transcriptional bursting model from problem 3, we can add feedback by requiring protein X to bind unbound promoter $P_0$ to form $P_1$. From an unbound state mRNA is transcribed with rate $\alpha$ and from a bound state protein is transcribed with rate $f * \alpha$. For values of $f > 1$ this system contains positive feedback, and for values of $f < 1$ this system contains negative feedback.

$$P_0 + X \underset{k_b}{\overset{k_f}{\rightleftharpoons}} P_1$$
$$P_1 \xrightarrow{f*\alpha} P_1 + mRNA$$
$$P_0 \xrightarrow{\alpha} P_0 + mRNA$$

$$mRNA \xrightarrow{k_{tsn}} mRNA + X$$
$$mRNA \xrightarrow{\gamma_1} \emptyset$$
$$X \xrightarrow{\gamma_2} \emptyset$$

A) Adapt the transcriptional bursting model from problem 3 to include positive feedback using the reaction network given above and the parameter values $k_f = 0.0005\ \#^{-1}\ min^{-1}, k_b = 0.1\ min^{-1}, \alpha = 0.001\ min^{-1}, f = 100, k_{tsn} = 20\ min^{-1}, \gamma_1 = 0.23\ min^{-1}, \gamma_2 = 0.02\ min^{-1}$, and assume the promoter starts in the active state, $P_1 = 1, P_0 = 0$, and initial $mRNA = 2\ \#, X = 2174\ \#$.

B) Use the Ensemble Run program to simulate 1000 realizations of the model for 240 minutes.

C) Plot the histogram of $X$ concentration at the final time point. Does adding positive feedback change the shape of the distribution? If so, describe the differences. What are the mean, standard deviation, and CV of the final protein concentrations? How does this compare to the case without feedback in Problem 3?

D) Now we will simulate the system with negative feedback, i.e. repression. Simulate 1000 realizations of the model for 240 minutes using the parameters $k_f = 0.0005\ \#^{-1}\ min^{-1}, k_b = 0.1\ min^{-1}, \alpha = 0.1\ min^{-1}, f = 0.01, k_{tsn} = 20\ min^{-1}, \gamma_1 = 0.23\ min^{-1}, \gamma_2 = 0.02\ min^{-1}$, and assuming the promoter starts in the inactive state, $P_1 = 0, P_0 = 1$, and initial $mRNA = 2\ \#, X = 2174\ \#$.

E) Plot the histogram of X at the final time point. Does repression change the shape of the distribution? What are the mean, standard deviation, and CV of the final protein concentrations? How do these compare to the base case in Problem 3 and the positive feedback model in Problem 4A?