CSM2024: Homework 4

1. **Simulating the dynamics of a stochastic system**

   Suppose that we're interesting modeling the stochastic dynamics of an mRNA (M) as it is synthesized and degraded. We'll specify the birth/death process describing the system as follows:

$$M \xrightarrow{\lambda_1} M + 1$$

$$M \xrightarrow{\beta_1 M} M - 1$$

   Write an implementation of the Gillespie algorithm to stimulate this process with $\lambda_1 = 10$ and $\beta_1 = 1$. Present the following results for your simulation:

   (a) A plot of a simulated trajectory from t = 0 to t = 100 time units. Hint: Remember when plotting that jumps are instantaneous; the M can never take on non-integer values.

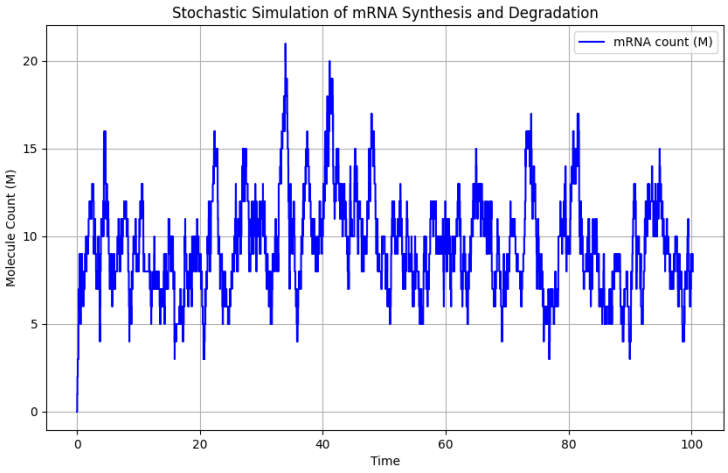

Figure 1: Gillespie algorithm simulation of stochastic mRNA dynamics

   (b) Determine the mean and variance of the process from your simulation. Hint 1: Remember our discussion in class! Naively averaging the series of M values from your simulation will give you the wrong answer! Your approach must account for the amount of time spent in each state. Hint 2: To get accurate answers, make sure your simulations run for many mRNA lifetimes or run many replicate simulations.

Mean M value: 9.9831, Variance: 10.0505 [!htbp]

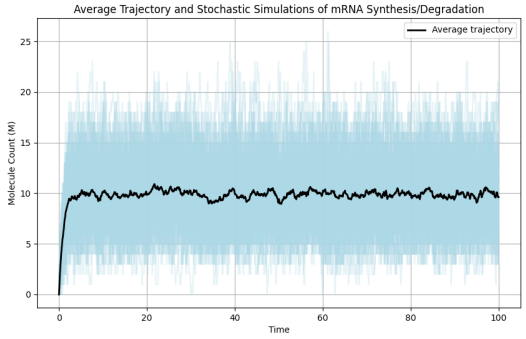

Figure 2: Average trajectory for n = 100 simulations

(c) Revisit your simulations from (b) and identify every instance where the system resides in the state M = 5. Plot a histogram of the waiting times (i.e. the duration of each visit to the state). Generate the same plot for all visits to the state M = 15.

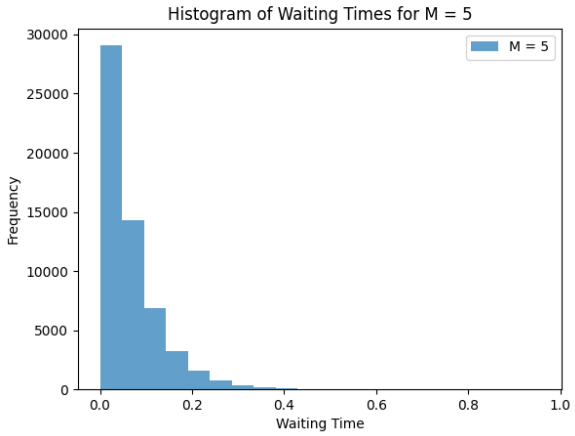

Figure 3: Histogram of waiting times for state $M = 5$

1. What type of distributions are these?
   These are long-tail gamma-like distributions, meaning that most of the values fall toward one side of the distribution.
2. In what way do they differ from one another?
   The distributions of waiting times for the states $M = 5$ and $M = 15$ show that the waiting times for $M = 15$ are shorter than the waiting times for $M = 5$.
3. What is the reason for this difference?
   The waiting times for $M = 15$ are shorter than the waiting times for $M = 5$, which indicates that the system tends to leave the $M = 15$ state faster. This is most likely due to the fact that the degradation rate is proportional to M. Thus, at higher M values, the degradation rate is increased, which increases the probability that the system will quickly leave the $M = 15$ state due to degradation. This is in contrast to the $M = 5$ state, which would have a 3x smaller rate of degradation than the $M = 15$ state. In conclusion, the waiting times for $M = 15$ are shorter than the waiting times for $M = 5$ due to the increased rate of degradation with increasing M,

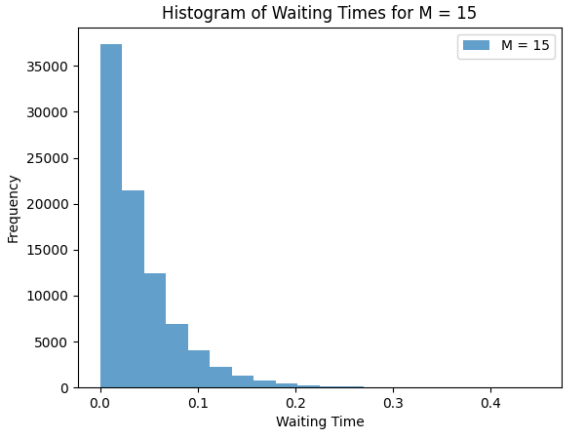

Figure 4: Histogram of waiting times for state $M = 15$

causing the system to quickly leave the $M = 15$ state due to increased degradation compared to the $M = 5$ state.

2. **Modeling multi-step reactions**

In this problem and the next one, we'll examine some of the assumptions we've been making in our models.

So far, we've been modeling all birth events as single elementary steps with exponentially distributed waiting times between events. However, many events that we might want to model are not so simple, and instead result from the accumulation of progressive steps. The time it takes to complete a given synthesis event is therefore more likely to reflect the sum of several independent exponential steps instead of a simple memoryless (i.e. exponentially distributed) waiting time. Transcriptional elongation of mRNA is a great example. It's not a single chemical step that we might model as in problem 1, but rather a series of such steps, each of which adds a base to the growing chain.

Let's explore what happens when we model these intermediate steps explicitly. We'll examine a simplified model of mRNA elongation. Imagine that we have an elongating mRNA chain of final length 3. M_i refers to the copy number of the species with length i. We'll assume that only the fully elongated species can be degraded. We can sketch out our elongation process like this:

$$\emptyset \xrightarrow{\lambda_1} M_1 \xrightarrow{\lambda_1} M_2 \xrightarrow{\lambda_1} M_3 \xrightarrow{\beta_1 M_3} \emptyset$$

And we can write out the corresponding birth-death process as follows:

$$(M_1, M_2, M_3) \xrightarrow{\lambda_1} (M_1 + 1, M_2, M_3)$$

$$(M_1, M_2, M_3) \xrightarrow{\lambda_1 M_1} (M_1 - 1, M_2 + 1, M_3)$$

$$(M_1, M_2, M_3) \xrightarrow{\lambda_1 M_2} (M_1, M_2 - 1, M_3 + 1)$$

$$(M_1, M_2, M_3) \xrightarrow{\beta_1 M_3} (M_1, M_2, M_3 - 1)$$

Write a Gillespie simulation of this process with the same parameter values as in the previous problem (i.e. with $\lambda_1 = 10$ and $\beta_1 = 1$) and answer the following:

(a) First a conceptual question: why—in molecular terms— do we simultaneously add one copy of a species and remove one of another in the middle reactions?

The reactions in the middle of the process can be interpreted as the elongation of an mRNA chain by one base at a time. When each elongation step occurs, one species is converted into the next, longer species. This is why one copy of a shorter intermediate species is removed while one copy of the next, longer species is added. Each new step consumes the previous, shorter form and produces the next one.

(b) Report the mean and variance for the fully elongated species, M_3. Is the variance larger, smaller or equal to what you saw in problem 1b?
Mean M_3 value: 9.8740, Variance: 10.3921
The variance is larger than the variance in problem 1b.

(c) Compile the distribution of waiting time between successive M_3 synthesis events from your simulations. Plot the histogram of waiting times. What kind of distribution does this appear to be? Does the result surprise you? Why or why not?

HINT: This is a question about timing between synthesis events specifically, not between any event that changes M_3.
The histogram shows a long tail, Gamma-like distribution with more wait times clustered toward the beginning of the distribution, toward the origin. Shown in the following figure:

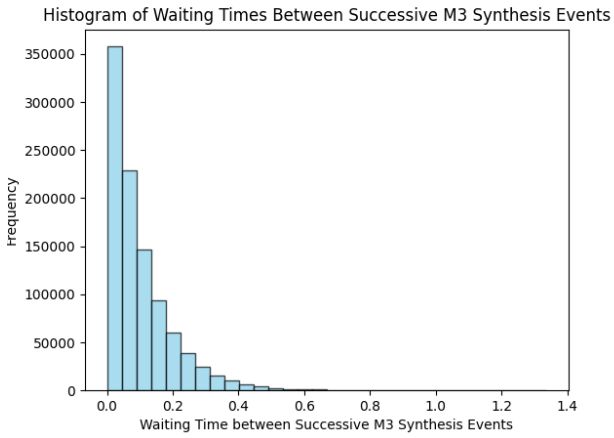

Figure 5: Histogram of waiting times for synthesis events of M_3

(d) Given your results from 2b and 2c, how do you feel about modeling mRNA transcription as a single elementary event? Does it appear to be a reasonable or unreasonable approximation?

The results of 2b and 2c indicate that modeling mRNA transcription as a single elementary event may be a valid simplification due to the similarity between the variances of a single step vs. multi step model. Additionally, the waiting time distributions for both models are extremely similar. However, the variance for the multi-step process is larger, which could indicate that the single-step process may be an oversimplification that decreases the complexity and variance of the mRNA transcription process.

3. **Modeling bursty production**

Most of the chemical reactions we've modeled so far have step sizes of one; that is, they add or remove molecules from the system one at a time. Recent work has suggested that many biomolecules are instead synthesized in large 'bursts' where many molecules appear nearly simultaneously. A well-studied example of this is mRNA transcription. Many genes—in systems ranging from bacteria to humans— appear to have through 'hot' and 'cold' periods of transcription. They'll go through long periods of transcriptional inactivity punctuated by short, intense transcriptional bursts in which many mRNAs are produced.

We could model this process microscopically— taking account of distinct promoter states, RNA polymerase binding, transcriptional elongation, etc.— but we can distill some useful intuition from a much simpler model. Consider the following simplified model of bursty transcription:

$$M \xrightarrow{\lambda_1} M + b$$

$$M \xrightarrow{\beta_1 M} M - 1$$

The only modification from the system in problem 1 is a step size of b molecules in the synthesis reaction. The burst size, b, is simply an integer ≥ 1.

Write a Gillespie simulation of this process and answer the following questions.

(a) Run a series of simulations for different combinations of b and λ_1. Specifically, try the following:

- $b = 1, \lambda_1 = \{1, 5, 10, 20, 50\}$
- $b = 5, \lambda_1 = \{1, 5, 10, 20, 50\}$
- $b = 10, \lambda_1 = \{1, 5, 10, 20, 50\}$

For each set of simulations, plot the measured average M on the x-axis against the normalized variance $\left(\frac{\sigma_M^2}{<M>^2}\right)$ on the y-axis. On your plot, you'll have 3 distinct curves that illustrate the relationship between mean and variability one for each series of simulations. Present your plot and label each curve.

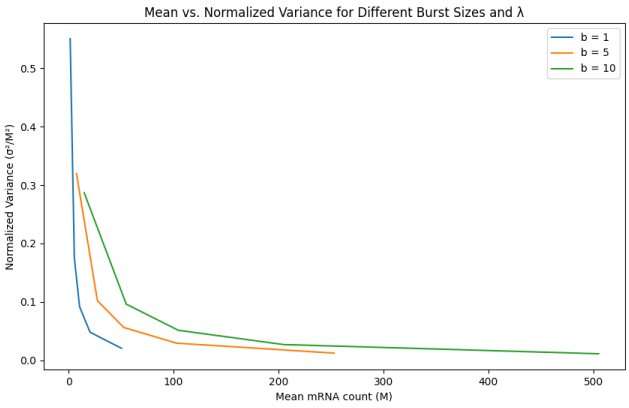

Figure 6: Mean vs. Normalized Variance for Different Burst Sizes

(b) How does the normalized variance change as the mean increases in each simulation set? Can you provide an intuitive explanation for this relationship?
Based on the plot, we observe that as the mean increases, the normalized variance decreases in all three curves ($b = 1, 5, 10$). At small M (left side of the plot), the variance is larger, and the system is more stochastic. This is because there are fewer molecules present, so each burst of synthesis significantly affects the system's state. However, As M increases, the system moves towards a more

stable state in which the effect of each burst becomes less significant, reducing the variance. So, as the system reaches larger M, the fluctuations become relatively less pronounced, and the normalized variance decreases.

(c) How does increasing the burst size affect the observed variability (as measured by the normalized variance)? That is, for a given average M does increasing the burst size lead to an increase, decrease or no change in the normalized variance? Can you provide an intuitive explanation for this trend?

Increasing burst size leads to an increase in the observed variability for a given average M. This is most likely due to the fact that larger bursts introduce greater fluctuations in the state, thus increasing variability. So, increasing the burst size leads to an increase in the normalized variance because each of the synthesis events has a larger impact, translating to greater fluctuations in mRNA count around the mean.

(d) If you were 'designing' a gene to have the lowest possible noise in M, how would you select b and λ_1?

To design a gene having the lowest possible noise in M, we would want to minimize the normalized variance. As we saw in the plot from part a, for a given average M, a larger burst size translates to a higher variability. Thus, for burst size b, we would want to choose a small value, e.g., $b = 1$. For selecting λ_1 to minimize the normalized variance, we would want a large λ_1 since the mean mRNA count M increases with the burst rate λ_1, and the normalized variance decreases as M increases as shown in the figure from part a. Thus, to achieve the lowest possible noise in M, we want to choose b and λ_1 to minimize the normalized variance; this would result in choosing a small value for b, e.g., $b = 1$, and a large value for λ_1, e.g., $\lambda_1 = 50$.