Applied Stochastic Processes

Arturo del Cerro (1593930)

January 31, 2021

The model that follows the dynamics of a gene regulatory circuit is described by two variables X_1, X_2 such that

X_1 is the number of transcription factor molecules

X_2 is the bound of bound promoter sites in the gene promoter region

and the dynamics are modelled by the transition rates W_1, W_2, W_3 and W_4, and the r_1, r_2, r_3 and r_4 values related with the events systhesis of the transcription factor, degradation of the transcription factor, dimer binding to the gene promoter region and unbinding from the gene promoter region respectively.

1. Formulate the dynamics of the stochastic system described in Table 2 in terms of the corresponding Poisson representation: $X(t) = X(0) + \sum_{i=1}^{4} r_i Z_i(t)$

Letting $Z_i(\tau)$ to be the number of times process i has occurred in the interval $(t, t+\tau)$ and being $Z_i(\tau)$ distributed according to a Poisson distribution with parameter $\lambda_i(t) = \int_0^t W_i X(s) ds$ then we have that $Z_i(\tau) = \text{Poisson}(\lambda_i(\tau)) \equiv Y_i(\lambda_i(\tau))$.
$X(t)$ can be expressed in the Poisson representation as

$$X_1(t) = X_1(0) + Y\bigg(\int_0^t \hat{R} + k_1 X_2(t) dt\bigg) - Y\bigg(\int_0^t k_2 X_1(t) dt\bigg) -$$

$$-2Y\bigg(\int_0^t b_{11} X_1(t)\big(X_1(t)-1\big)\big(E - X_2(t)\big) dt\bigg) + 2Y\bigg(\int_0^t u_{11} X_2(t) dt\bigg)$$

$$X_2(t) = X_2(0) + Y\bigg(\int_0^t b_{11} X_1(t)\big(X_1(t)-1\big)\big(E - X_2(t)\big) dt\bigg) - Y\bigg(\int_0^t u_{11} X_2(t) dt\bigg)$$

2. Consider the following re-scalling of variables: $n_1^s = X_1/S, n_2^E = X_2/E$ **and** $\tau = b_{11} S E t$ **with** $\epsilon = E/S \ll 1$**, and perform this re-scalling in the Poisson representation equations. Recall that this must result in the explicit separation of the dynamics in slow** n_1^S **and fast variables** n_2^E**. You should also obtain the expression of the re-scaled parameters** $R, \kappa_1, \kappa_2, \beta_{11}$ **and** υ_{11} **as a function of the bare parameters** $\hat{R}, k_1, k_2, b_{11}$ **and** u_{11}**, respectively and** E **and** S**.**

Assuming the Haldane-Briggs analysis over the Michaelis-Menten kinectics that assumes that the concentration of the intermediate complex does not change on the time-scale of product formation, then n_1^S and n_2^E have characteristic scales S and E respectively and the Poisson representation is as follows:

$$n_1^S = n_1^S(0) + S^{-1} Y\bigg(\int_0^t \hat{R} + k_1 n_2^E E dt\bigg) - S^{-1} Y\bigg(\int_0^t k_2 n_1^S S dt\bigg) -$$

$$-2S^{-1}Y\Big(\int_0^t b_{11}Sn_1^S\big(Sn_1^S-1\big)\big(E-En_2^E\big)dt\Big)+2S^{-1}Y\Big(\int_0^t u_{11}En_2^E dt\Big)=$$

$$=n_1^S(0)+S^{-1}Y\Big(S\int_0^\tau R+\kappa_1 n_2^E d\tau\Big)-S^{-1}Y\Big(S\int_0^\tau \kappa_2 n_1^S d\tau\Big)-$$

$$-2S^{-1}Y\Big(S\int_0^\tau (n_1^S)^2\big(1-n_2^E\big)d\tau\Big)+2S^{-1}Y\Big(S\int_0^\tau \kappa_1 n_2^E d\tau\Big)$$

$$n_2^E=n_2^E(0)+E^{-1}Y\Big(\int_0^t b_{11}Sn_1^S\big(Sn_1^S-1\big)\big(E-En_2^E\big)dt\Big)-E^{-1}Y\Big(\int_0^t u_{11}En_2^E dt\Big)=$$

$$=n_2^E(0)+E^{-1}Y\Big(E\epsilon^{-1}\int_0^\tau (n_1^S)^2\big(1-n_2^E\big)d\tau\Big)-E^{-1}Y\Big(E\epsilon^{-1}\int_0^\tau \upsilon_{11}n_2^E d\tau\Big)$$

Now, making use of the Law of Large Numbers for Poisson processes, that states that

$$\Omega^{-1}Y(\Omega\bar{\lambda})\overset{\Omega\to\infty}{\longrightarrow}\bar{\lambda}$$

then working with the mean-field limit by letting $S,E\to\infty$ as $\epsilon=E/S=\text{cte}\ll 1$ it can be shown that $n_1^S,n_2^E\overset{S,E\to\infty}{\longrightarrow}X_1,X_2$ respectively and that for the equivalences between the re-scaled and original parameters the limits also are well defined. Here is shown then this relationships:

$$\beta_{11}=1,\quad \upsilon_{11}=\frac{u_{11}}{b_{11}S^2},\quad R=\frac{\hat{R}}{b_{11}S^2E},\quad \kappa_1=\frac{k_1}{b_{11}S^2},\quad \kappa_2=\frac{k_2}{b_{11}SE}$$

3. Carry out a second re-scaling $T=\epsilon^{-1}\tau$. If step 2 was done correctly you will obtain that the evolution of the fast variable occurs under a frozen (cte) value of the slow variable. Conditioned to a constant value of n_1^S find the (quasi)steady state of distribution of X_2.

As the statement says, under this re-scaling we obtain that the re-scaled rates associated to the slower reaction n_1^S fire up at a way lower rate than the rates related with the faster reaction n_2^E. This may cause that the faster variable will reach a partial equilibrium state while the slower variable remains frozen expressed as a constant. Conditioned to this constant value of n_1^S we make a second re-scaling $T=\epsilon^{-1}\tau$. Then we have $dT=\epsilon^{-1}d\tau$ and the representation of the equations now is in the form

$$n_1^S=n_1^S(0)+S^{-1}Y\Big(S\epsilon\int_0^T R+\kappa_1 n_2^E dT\Big)-S^{-1}Y\Big(S\epsilon\int_0^T \kappa_2 n_1^S dT\Big)$$

$$-2S^{-1}Y\Big(S\epsilon\int_0^T (n_1^S)^2\big(1-n_2^E\big)dT\Big)+2S^{-1}Y\Big(S\epsilon\int_0^T \kappa_1 n_2^E dT\Big)$$

$$n_2^E=n_2^E(0)+E^{-1}Y\Big(E\int_0^T (n_1^S)^2\big(1-n_2^E\big)dT\Big)-E^{-1}Y\Big(E\int_0^T \upsilon_{11}n_2^E dT\Big)$$

So on, the transition rates of the quasi-steady state of the distribution of X_2 are

$$W_3=(n_1^S)^2\big(E-X_2\big)\text{ and }W_4=\upsilon_{11}X_2$$

and due to the r values then the Master equation is described as

$$\frac{\partial P(X_2,T)}{dT} = (n_1^S)^2\big(E-X_2+1\big)P(X_2-1)+\upsilon_{11}\big(X_2+1\big)P(X_2+1)-\Big((n_1^S)^2\big(E-X_2\big)+\upsilon_{11}X_2\Big)P(X_2)$$

Using the generating function $G(q,T)=\sum_{X_2} P(X_2,T)q^{X_2}$ then we get the PDE for $G(q,T)$

$$\frac{\partial G}{dT} = \big(q-1\big)\left[(n_1^S)^2EG - \big((n_1^S)^2q+\upsilon_{11}\big)\frac{\partial G}{dq}\right]$$

with $G(1,T)=1 \ \ \forall T$. This PDE can be solved by the characteristics method taking the characteristics equations as

$$\frac{dT(s,r)}{ds}=1, \quad \frac{dq(s,r)}{ds}=(1-q)\big((n_1^S)^2q+\upsilon_{11}\big), \quad \frac{dG(s,r)}{ds}=(q-1)(n_1^S)^2EG(s,r)$$

with boundary conditions

$$T(0,r)=0; \quad q(0,r)=r; \quad G(0,r)=G(q(0,r),T(0,r))=g(r)$$

From two first equations we get $q(s,r)=e^{-s\left((n_1^S)^2+\upsilon_{11}\right)}$ and

$$-\frac{dG/ds}{dq/ds}=\frac{dG}{dq}=-\frac{(n_1^S)^2E}{(n_1^S)^2+\upsilon_{11}}G(q,t)$$

So on, as $G(q=1,T)=1, g(r=1)=1$ and it can be shown that as $r \overset{T\to\infty}{\longrightarrow} 1$, then the PDE at the quasi-steady state can be solved considering $\frac{dG}{dT}=0$. The solution for the PDE at quasi-equilibrium is in the form

$$\lim_{T\to\infty} G(q,T)=G(q)=\left(\frac{q(n_1^S)^2+\upsilon_{11}}{(n_1^S)^2+\upsilon_{11}}\right)^E$$

which means that, at quasi-equilibrium X_2 follows a binomial distribution

$$Bin(p,E) \text{ with } p=\left(\frac{(n_1^S)^2}{(n_1^S)^2+\upsilon_{11}}\right)$$

4. Perform simulations of the dynamics of the system using both the Gillespie algorithm and the QSSA Gillespie algorithm. Run the system until it reaches its steady and take the steady state statistics over a long realisation of the system and use boxplots to compare the simulation results obtained with both methods.

Following the algorithms described in [1] for the Gillespie and QSSA Gillespie algorithms the system has been run once for each algorithm and here are presented the results of the statistics of the simulation over a long simulation when the system has reached the steady state.

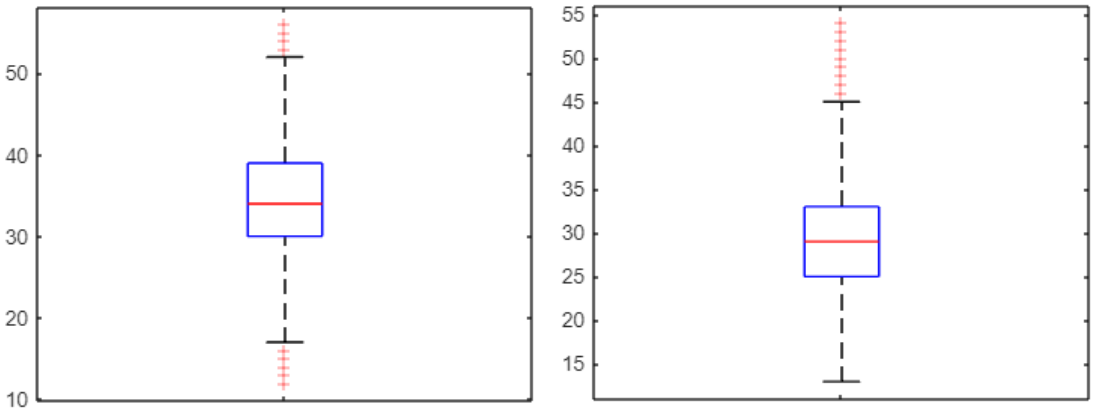

Figure 1: Statistics for Gillespie (left) and QSSA Gillespie (right) algorithms

where the data outside the whiskers represent the outlier values, the bottom and top boundaries of the box represent the first and third quartile and the red line the median value.
Both results came from a large enough ($t = 100$) realisation of the system and comparing them we can see that there are notable differences between them as the QSSA Gillespie algorithm has a median value of 29 for the X_1 variable (the one we are studying) while the Gillespie algorithm's median value is 33. An other main difference is the mean of this Gillespie and QSSA simulations: 33.58 and 28.69 respectively. So we can imagine now that this two realisations for the different methods have enormous differences and we can expect the Kolmogorov-Smirnov test result. Continuing the analysis for the variable that we are studying, also we can see how this variable changes along the iterations by looking at the standard deviation. By doing this we can see that for Gillespie and QSSA algorithms they both have a large deviation of 5.88 and 5.2 but they are close between themselves.

To compare how this two algorithms gives different results we can look at the Kolmogorov-Smirnov test to accept or reject the null hypothesis

H_0: *Gillespie and QSSA Gillespie algorithm results come from the same distribution.*

For the levels of of significance $\alpha = 0.01$ and 0.05 this test rejects the null hypothesis with a tiny p-value (lower than 10^{-20}). We can study if this hypothesis remains rejected when changing the maximum time number or if this p-value starts increasing. For max times values of 10,100 and 1000 the null hypothesis remains being rejected and the same for the p-value, with no clear evidence of this increasing p-value, maybe due to the high deviation of the data.
An other analysis of this algorithms could be made by looking at the histograms for the waiting time distribution as it affects on a relevance way each algorithm.

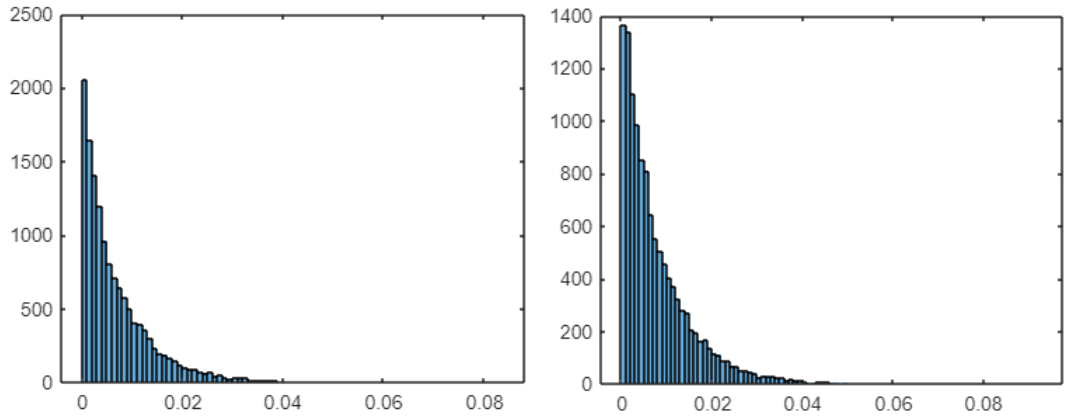

Figure 2: Histogram of the waiting time for Gillespie (left) and QSSA Gillespie (right) algorithms

These both waiting time distributions for the different algorithms seems to be in a Poisson distribution form as both have mean and standard deviation values almost equal: 0.0072 and 0.0075 for Gillespie algorithm and 0.0079 and 0.0079 for QSSA Gillespie algorithm. This fits with a Poisson(λ) as it is verified that also the standard deviation is λ.

Going further on the study of the dynamics of this system and algorithms for solving them, we have implemented the τ-leap algorithm described in [2]. Below are presented some results in order to compare them with the other two methods.

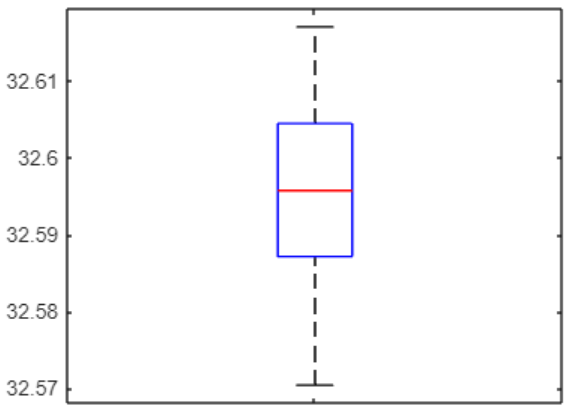

Figure 3: Statistics for τ-leap algorithm

For the steady state of the system we have that this τ-leap algorithm is way more precise on dealing with the dynamics of the system as the values for the studied variable are more bounded, with a mean, median and standard deviation values of 32.59, 32.59 and 0.0099 respectively. Also the waiting time distribution of this τ-leap algorithm is purely distributed at low values of time, with a 0.05 average and 0.62 standard deviation values. Finally we can check that neither for

this τ-leap algorithm the null hypothesis under the Kolmogorov-Smirnov test is verified, as we should expect.

5. Conclussions

In this work it has been studied both theoretically and by simulations the dynamics of a gene regulatory circuit. First a re-scaling of the variables involved in the system have been made in order to make more explicit the separation between the different scales. Then we have given a relationship between the re-scaled and original parameters using the Law of Large Numbers for Poisson processes and the mean-field properties. Later we studied the quasi steady distribution of the slow variable in order to later do simulations while using this, ending in a binomial distribution used for the QSSA algorithm. And finally the simulations for the Gillespie and QSSA Gillespie algorithm has been performed. Taking a look at the first variable of the system we arrived at the conclussion that this two algorithms gives quite large different results within the dynamics of the system, and to confirm it we made a set of Kolmogorov-Smirnov tests that confirmed this hypothesis. Ultimately and for continuing exploring the different algorithms that could have been performed we showed the results for the simulations of the τ-leap algorithm showing that is a less variance algorithm that could give more exact solutions.

References

[1] Numerical methods. Multi-scale SSA. Cao, Gillespie, Petzold pp. 395-411. (2005)

[2] Numerical Monte Carlo methods for stochastic systems. Tomas Alarcon. ICREA & Centre de Recerca Matematica

[3] MATLAB. (2010). version 7.10.0 (R2010a). Natick, Massachusetts: The MathWorks Inc.