Genetic Switches between two population with regards to mRNA and proteins applying Markov Chain Stochastic Model Check.

Qin He1;2;3;4 RubinWang5  XiaochuanPan5 Jukka Peltola2;3 Susanna Narkilahti3 Jari Hytinnen4

1.Signal Processing Department,Tampere Universities

2.Medicine and Health Technology department,Tampere Universities, 3.TAYs

Tampere,Finland

4.Neural-Dyamics Lab, East China University of Science and Technology

Shanghai, China

5. Biophysics Department,Tampere Universities

Abstract

As Arc, one virus-like gene, crucial for learning and memory, was dis-covered by researchers in neurological disorders fields, Arc mRNA’s single directed path and allowing protein binding regional restric-tively is a potential investigation on helping shuttle toxic proteins responsible for some diseases related to memory deficiency. To study especially the transform between mRNA and proteins, the switching function of the phenotypes, ’normals’ multiplying populations and ’persisters’, resilient to stress instead of multiplying is of our interest. Mean time to switching (MTS) is calculated explicitly quantifying the switching process in statistical methods combining Hamiltonian Markov Chain(HMC).The model derived from predator and prey with type

1. functional response studies the mechanism of normals with intrin-sic rate of increase and the persisters with the instantaneous discovery rate and converting coefficients. During solving the results, since the numeric method is applied for the 2D approximation of Hamil-tonion with intrinsic noise induced switching combining geometric minimum action method. In the application of Hamiltonian Markov Chain, the behavior of the convertion (between mRNA and proteins through 6 states from off to on ) is described with probabilistic conditional logic formula and the final concentration is computed with both Continuous and Discret Time Markov Chain(CTMC/DTMC) through Embedding and Switching Diffusion.The MTS, trajectories and Hamiltonian dy-namics demonstrate the practical and robust advantages of our model on interpreting the switching process of genes (IGFs, Hax Arcs and etc.) with respects to memory deficiency in aging process which can be useful in further drug efficiency test and disease curing.

1 Introduction

In cell biology, non-equilibrium stochastic process is of interest since the observation of experimental results are becoming of higher res-olution, studying the molecules both with imaging and expression data are often conducted in both single and population (thousand) order, which basically described in stochastic process whether on a discrete or continuous scale with status changes either genotypi-cally or phenotypically. Many problems are thus studied related to status switching, including cell regulatory networks[1], signal re-sponse on excitability and inhibition[2], (convinced by translational and transcriptional burst of expression for instances.), metastability among populations, (binding of ligands and proteins, forming of polymerases and etc.).In this paper, we focus on the interaction among genes, mRNA, proteins and etc. To be more specific, while the switching problem among molecules can be studied on geno-type, including sequencing for single RNA, alignments and binding considering condons

and etc, we stay on the switching with expres-sion (concentration) only, which is simplified as modified population problem using Lotka-Volterra equations[3] of two populations only. Thus, rather than the competitor model(for instances, cell bifurcations.), we applied simulation of switching on predator model. The model is based on the following basic

assumptions: Prey population (promoters) is fed with enough food all the time while the predator population of the predator(the persisters) depends on the size of prey(promoters).

In our paper, we mainly study the interaction of DNA and its interaction with the associated proteins.(Clinical data of Hax1 and HS1 is downloaded from Ensmbl gene database[4]). On one hand, the switching model is calculated under the large deviation theory(LDT)[5] combining the least actions. The Markov chain[6] consider the states of the 2D coordinates (x; y) of mRNA numbers and protein numbers referencing the distribution of x, which follows the order O(1) while PX follows the time scale on O(1/e) and guaran-teeing the variant of LDT hold with the transform of the expressions in single population. Only considering the process of diffusion case, we study the binding of hax1 with simple switching between on and off status under its interaction with HS1 seen as in the constant environment, i.e. the closed system at mean field. The dimer which can be cancelled out connect the binding between two single population. On the other hand, one numeric method is applied to solve the problem, making compare with the stochastic process[7] on ap-proximation equation of the mean switching time(MST) with the transform between two status (we studied the switching time with four situations, both multiplicative and asymptotic of single population and the binding and degradation between two population.) Again, this method is also calculated based on the Hamiltonians. We give out the MST with respect to N/Nc denoting N as the population number of interest and Nc as the threshold of certain status(either of that population or the other population). Since our study only based on data in the process of transforming in the constant environment, extinction is not considered in this paper. To study both intrinsic and extrinsic noise with the exciting and inhibiting bursts is the potential topic in the future. In the following contents, the first chapter is the proposition of the model, based on least action with LDT and MTS approximation with one stochastic differential equation (SDE) [8]separately; And the second chapter gives numeric experiments based on Hamilton Markov Chain[9] computation of the expression data of hax1 and HS1; In the last chapter, the model is described in the normal logic formula with both probabilistic condition model[10] and the results are analysed with both hamiltonian, realization size, convergence, the rewards computation taking the CTMC as Poisson process[11] and the reachability computation with the transfer kernel of switching diffusion[12] through DTMC. In the appendix, there also includes the complete proof of model with action S based on Hamilton not only based on the explicit equation in this paper. Some descriptive Statistics and pre-computation based on the data can be

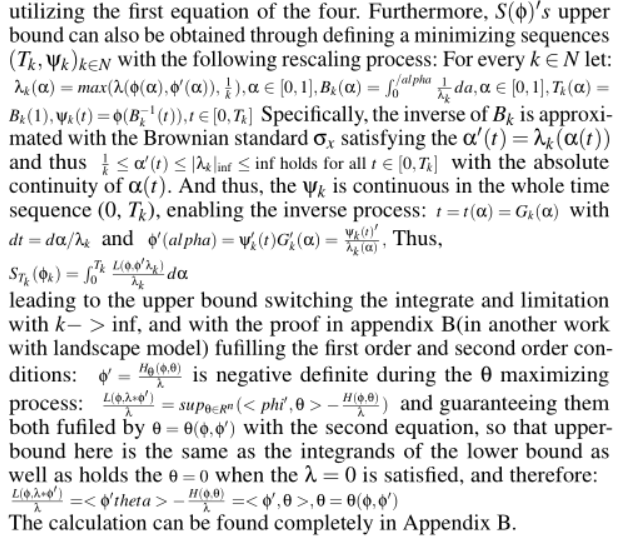
accessed through link in availability. As the process related to motor coordination and func-tion, the Hax’s function in regulation, B cell’s signal transduction can be further studied with more data considering its excitability and metastability functions with stimulation of drugs for instance in the future as well. And one computation applying DTMC with

linear regression on previous work is made as the further extension of the model.

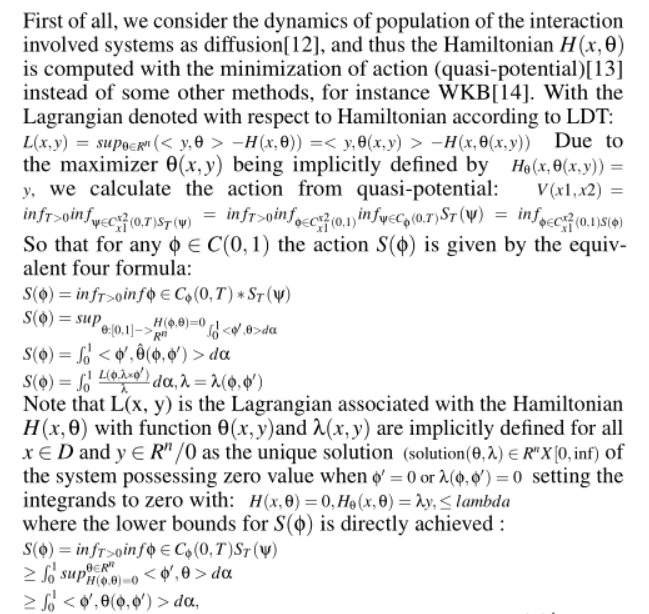
2. Proposed Model

Molecular interactions are studied on phenotypic data of the mRNA

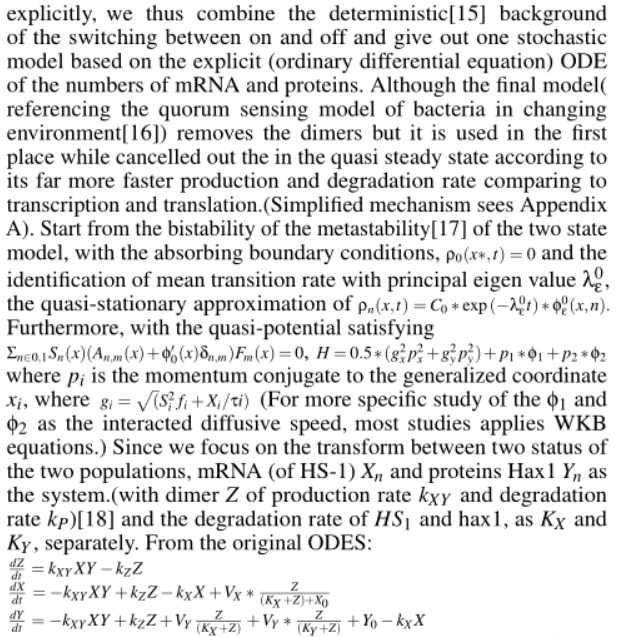
and its associated protein in this paper, especially the trajectory of the production of hax1 and HS1 with interaction with each other through least action method combining diffusion process[10]. Furthermore, in solving the equation, one stochastic differentiation equation approximates the analytic solution and calculation of MST[11] based on converging with Hamiltonian quantities, finding three convergence points through eigenvalue of position quantities as well as satisfying H = 0 and Hq = 0 where q(PX ; PY ) are momentum quantities. In the 3rd subsection, the transition is illustrated with belief graph first and then convert ratio are utilized in computing the discrete embedding of the continuous temporal logic. As comparison, the third subsection compute the discretized time markov chain as the approximation considering it as a hybrid systems.

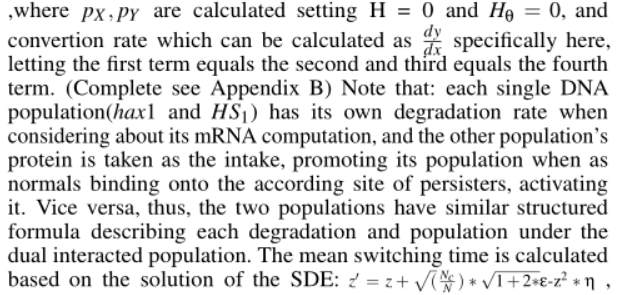
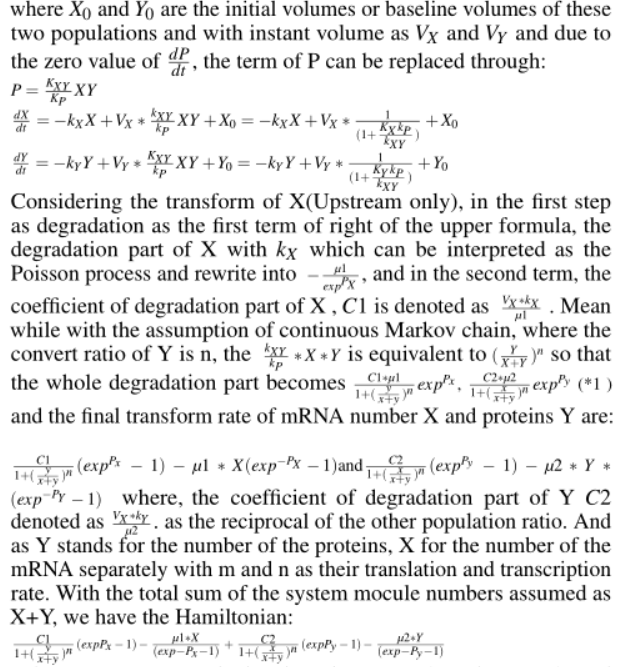


2.1 switching model with least action

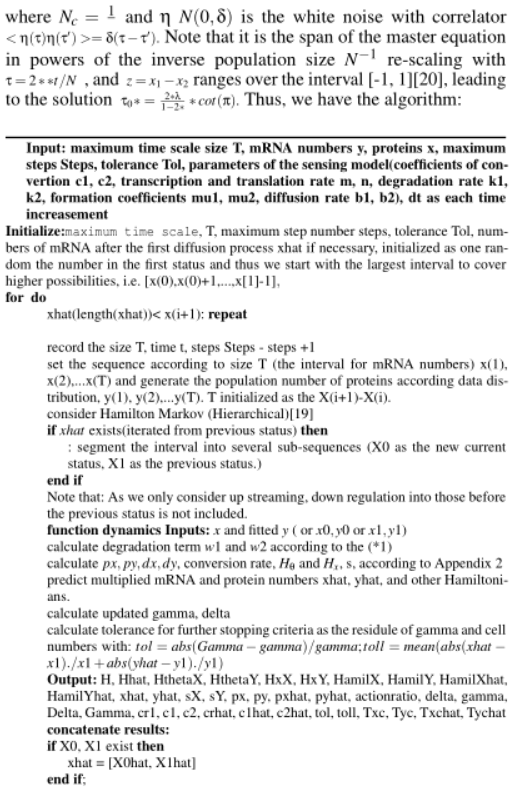


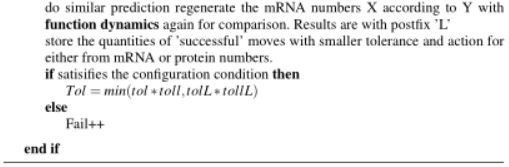
2.2 diffusion Approximation with numerical methods on the convert ratio referencing bacteria sensoring and MTS on difference mapping

As to study the switching model interpreting the process 









2.3 Stochastic Model on the Uncertainty of Reasoning Switching Model

Stochastic model checking is a method for calculating the likelihood of the occurrence of certain events during the execution of a system while linear-time properties specify the traces to be exhibited through a transition system, including the atomic propositions AP, as the subset of (2AP)w, satisfaction relation and etc. Although the definition of such language with LT property is composed by infinite word, the transition systems fulfilled the LT property is of finite states. When the state is visited transitional, each time a state is visited, we adversarial pick a transition distribution that respects the interval constraints, and take a probabilistic step according to the chosen distribution.

2.3.1 Probabilistic Uncertainty Conditional Model

Considering the formal semantic system, the normal logic formula are utilized to construct the belief graph,with the agents defined as the molecules in the transfer process (mRNA and proteins in this paper.) As the sensing model is highly sensitive to external factors, it is suitable for the agents to be modeled as epistemic states which is based on the prior of the given[22] transfer instead of the deterministic probability. That is reasoning the actions in the presence of sensing dynamically with regards to epistemic agents [23] rather than the whole convert system.

And, with each action states S(a, p) defined, the belief graph is supported by reasoning bidirectionally about both the transcription and translation to ensure trust and safety in the interaction (as shown in Appendix A).

2.3.2 Model Check for Stochastic Models Combining Continuous Time Markov Chain with Embedding in Reward Computation

The logic applied on a probabilistic notion regards to the belief graph is based on the trust which is reflected by the reliability and predictability. Specifically, the language of the stochastic models used for computing CTMC is the Continuous Stochastic Logic(CSL) developed and extended by some research[24.]

A CTMC is a tuple C = (S,s,R,L) where S is the finite set of states, s is the initial state； R is S\*S->R>0  is the transition rate matrix; L: S->2AP is a labelling function which assigns to each state s € S is the set L(s) of atomic propositions valid in the state. Instead of the case of DTMCs, a fixed set of atomic propositions AP is applied, the transition rate matrix R assigns rates to each pair of states in the CTMC, used as parameters of the exponential distribution. A transition can only occur between states s and s’ if R(s,s)>0, representing the probability of this transition being triggered within t time-units equals 1-e-R(s,s’)t. Time spent in state s, before such transition occurs, is exponentially distributed with rate E(s), where:

E(s) = sum(R(s,s’))

where E(s) is known as the exit rate of state s.

The embedded DTMC of a CTMC, is the probability of each state s’ transitioned from the precious s, independent of the time, defined as:

Emb(C) = (S,s,Pemb(C),L) where for s,s’ € S:

,Pemb(C)(s,s’) = { R(s,s’) /E(s), if E(s) ≠0

1. if E(s) =0 and s =s’
2. otherwis

where the behavior of the CTMC in the alternative way remains in a state s delayed and exponentially distributed with rate E(s) and transit with Pemb(C)(s,s’).

The infinitesimal generator matrix for the CTMC C=(S,s,R,L) is the matrix Q: S\*S->R defined as:

Q(s,s’) = R(s,s’), if s is not s’

- ∑s’’ ≠s, R(s,s’) otherwise

The CTMC stores the transition from s to s’ in ratio format instead of the possibility in DTMC[25].

However, the probability measures Prs on ∑PathC(s) as the unique measure such that Prs(C(s)) = 1 and for any cylinder C(s,I,..,In-1,sn,I’,s’), Prs(C(s, I,…,In-1,Sn, I’,s’)) equals:

Prs(C(s,I,…,In-1,sn))= Prs(C(s,I,…,In-1,sn))\*P1emb(C1)(sn,s’)(e-E(sn)\*inf I’ -e-E(sn)\*sup I’)

In our case, such model check as with PCTL, we can easily derive the path formulae for the states between S0 and S7 separately with 6 time intervals I= [t0,ti] :

P~p[◊Iφ] = P~p[true UI φ]，

P~p[□Iφ] = P~p[exist UI φ]，

φ = ‘transit

Stands for the probability that a transition occurs in time interval I=[t0, ti]，

And thus, For determing the least solution,ProbC,(s,φi,U[0,t],ѱ)

=∫∑Pemb(C)(s,s’)\*E(s)\*e-E(s)\*x\*ProbC(s’, φi,U[0,t], ѱ)

=ProbC(φ,U[t,œ])

=Prob{ProbC(s, φ,U[0,t’-t], ѱ), if s|=φ

0 otherwise

And define the rewards function a CTMC D=(S,s,R,L), the semantics is defined as:

S |= R~r[I= t]  ExpC(s,XI=t)~r

2.3.3 Model Check for stochastic models combining Discrete Time Markov Chain with Switching Diffusion in Reachability Computation

In the second application of model check, the continuous dynamics described by switching diffusions is studied with reachability and dually safety properties on DTMC[21] Compared with the MC on continuous time domain, DTMC is defined with a fixed, finite set of atomic propositions used to label states. The DTMC D is a tuple similar as CTMC (S,s,P,L), where S is a finite set of states; s is the initial states; P: S\*S->[0,1] is the transition probability matrix where ∑s P(s,s’) = 1 for all s€S where L(s) of atomic propositions are valid.

max{dq(t1,t2)}<=Kd\*|t2-t1|, where dq(t1,t2) pseudally defined as dq(t1,t2) = sqrt(E[(Xq(t2)- Xq (t1))2]), and K>=12 is the Dudley metric universal constant.

Let h defined larger than 0 be a sampling time and the mean E and the covariance C to simulate a nomal distribution N(x| E,C). Then, the discreete kernel is T((A,qj),(x,qi)) =

∫AN(x|eF(qi)\*hx,gamma(I,h)) dx\*e-Ᾱh if qi = qj

∫A(∫N(x|Eqi,x(s),C qi,x(s)\*) dx\*lambda\*Ᾱij/Ᾱi\*Ᾱi\*h\*e(-Ᾱh) if qi = qj,

Where gamma(i,t) = ∫A(eF(qi)\*(t-m))\*G(qi)\*G(qi)T(eF(qi)\*(t-m))Tdm,

ῼ Ᾱi,Ᾱi,t(s) = (Ᾱi-Ᾱi)\*e(Ᾱj\*s-Ᾱi\*t-Ai\*s)/( eᾹj\*s- eᾹj\*s),

Eqi,x(s) = eF(qi)seF(q)(h-s)x,

Cqi,x(s) = eF(qi)s gamma(i,t) eF(q)(h-s)x+ gamma(j,h-t),

0=<eps<=1-e- Ᾱi\*h- Ᾱi\*h\*e-- Ᾱi\*h

With the events on t€I An ={X(t) €S| P~p[1, ] = P~p[true UI φ]}，

Bn={X(t) €S| P~p[◊Iφ] = P~p [exist UI φ]，

Psafe(X,S,I) = limP(An^Bc), Preach(X,Sc,I) =1- limP(An^Bc)

Tdx(z1, z2) = T(z1,z2), if z1,z2€Sdx

1-∑zj€SdxrzT(z), if z1€Sdx,z2/, z2€S

1, if z1,z2€ φ,

0, if z1€φ, z2€Sdx

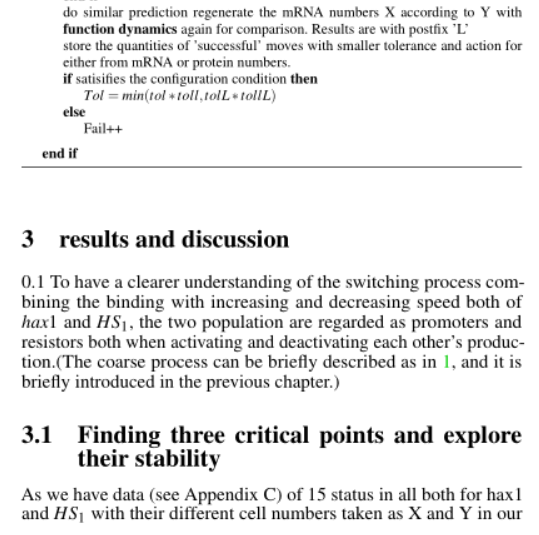
Continuous kernel proof see Appendix B.

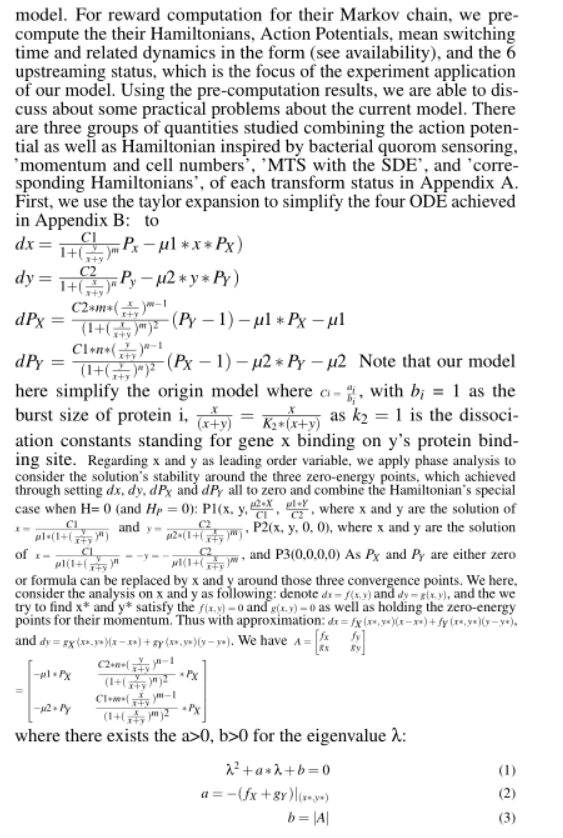
To compute the reachability/safety properties, we introduce the scheme based on Discrete Time Markov Chain(DTMC) which discretize the state space to approximate the Discrete Time Markov Process(DTMP) results from the original switching diffusion process H, a tuple H=(Q, K, F,G,W,^), where Q = {q1,…,q|Q|} is the set of discrete modes instead of the matrix in CTMC and Y=(X, α) its solution. For any q €Q, call Xq the solution of the SDE:

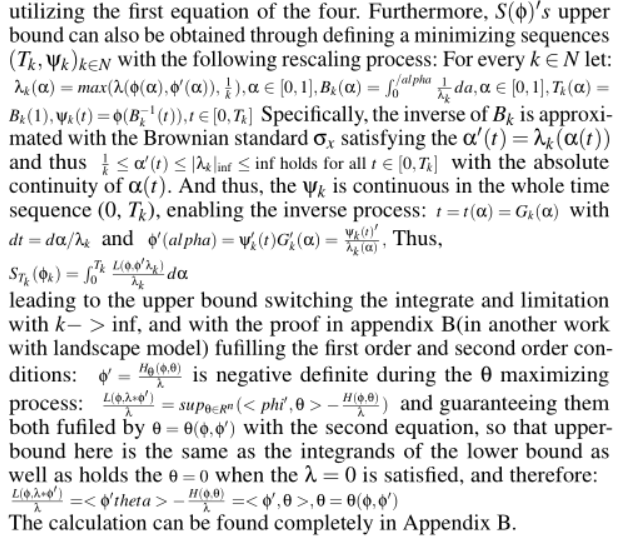
dXq(t) = F(q)\*Xq(t)dt+G(q)\*dW(t) (\*)

In this section we assure that Xq is a ui-dimensional, zero mean Gaussig,an process (GP). Xq is almost surely bounded within the interval I by Assumption.

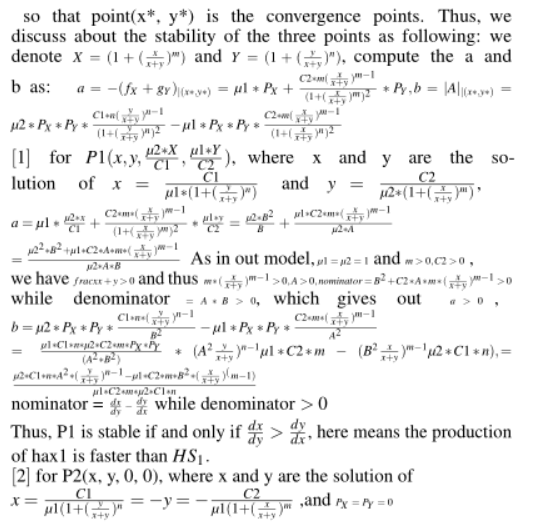
Set h = min{2(-n)/(2\*sqrt(2)\*K2\*Kd),2 -n} and €n = 2(-n/2), where n€N, and Kd is a constant such that for any t1, t2€I





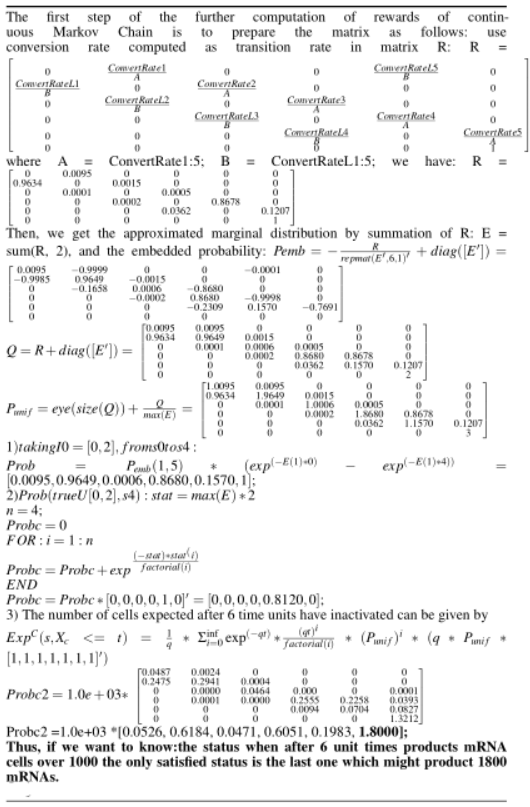






Thus, for P2, a = b = 0. It’s unstable.

[3] for P3(0, 0, 0, 0), same as P2, a = b = 0 and it’s unstable.



**3.3 Conclusion**

chain random walk with its faster convergence. As in 2(g) and 2(h), the convergence(variation to mean) of the markov chain hamilton is in blue line and the red line for clinical data and simulation on more possible transition status, giving different convergence but similar phase interval(according to 2(i)), interestingly. The last status transition converge the worst followed by the fist transition. And the result simulated with more markov chain status converges better than the clinical results. And according to the convert rate, the mRNA to Protein transfer ratio should be the highest when starting, and goes especially lower in the last two status which is in assistance to the protein binding as we cut off the process around the convergence point where the two population has reached metastability F M [29].According to the simulation result, the protein has gone through the switching process changing from normals to persisters and back to normals(bursts in optimal time in 2(h) might also due to the switch.).

Mean while, as the second population providing food(protein) to the other’s binding site and either activiate or deactivate it, it works as the extrinsic noice induced the excitability or exhibition of the other gene. Here, as we choose hax1 and HS 1 , they work as promoters for

each others. One noticable computation is the reward computaiton based on stochastic model selection which is useful in predict the possible status of the cell numbers easily with precomputation. And we can consider correct the transition matrix with simualted clinical tested results to improve the prediction as well. On the other hand, the most important calculation action potential is easier to be achieved through Hamilton as we proved with geometric minimum action and stochastic approximation. Ohter methods can cover Hamilton Jacobian matrix, WKB and etc.[28]. As we also improve the algorithm with adding hierarchical markov in calculating number of cells in different status only record successful move according to the tolerance based on action potential and residual of prediction numbers both, the convergence of the algorithm is guaranteed. And

further research can be conducted on the whole process from bifurcation to catastrophe and extinction as well. Problem with multi population is also possible. As hax1 is observed to have function in signaling and regulating of genes especially in learning systems and motor related brain function, this switching model study related to its binding might help to predict the cell numbers and production or degradation rate especially later with further study into both with promoters and persisters as to test different drug and their efficiency

on the aging process related disease.

In the computation of reachability, the approximation with DTMC[30] mainly compute the kernels of Brownian with shift, finally discretize the original switching diffusion process. As the DTMC gives out the kernel with probability instead of the ratio, it is then convenient to be written into transition matrix Pdx which is discretized from Sdx on finite space state and gives out the reachability with error I/h\*(Kdx+exp(-2^n-2^n/2+1)). As the proof in Appendix, the error bounds with Lipschitz constants converged with prominent K= mh1+Lh2. The computed result N\*K\*dxis here 0.453 with N = 6, m = L = 2,dx = 0.002 and h1 = 0.001, and h2 = ceil(h2\*N)=1.71. The result is

Since the final result of the continuous process is not of probability range thus we normalize it with P = ratio/sum(ratio) and the DTMC approximation shown in the figure is the approachability(1-safety.) The result of the ttest tested continuous embedded matrix and h =0,p = 0.0515,ci = -0.7212 , 0.0029,stats = struct with stat: -2.2103( df: 10), do not reject the hypothesis that the two process. As the final safety consistently for two methods gives highest concentration for the last state showing the example computation’s direction from off to on, although there is the slight difference that the forth in the continuous process is relative lower comparing to its other five states as well as the one in the discrete process states. The DTMC gives strictly increasing concentration from off to on during the 6 states.

As the switching diffusion is a commonly used model in genetic field, not only useful in the transmission of different molecules but also can be derived into analytical models giving straight transfer information abou tsome process with either concentration[31] change or energy change.

Acknowledgments

Although I have graduated with excellence graduates awards,I am still working in the university currently cooperated with university hospital. Thanks to all the supervisors and colleagues I have worked with.

Availability

Please find supplementary data, code and experimental results in this link:

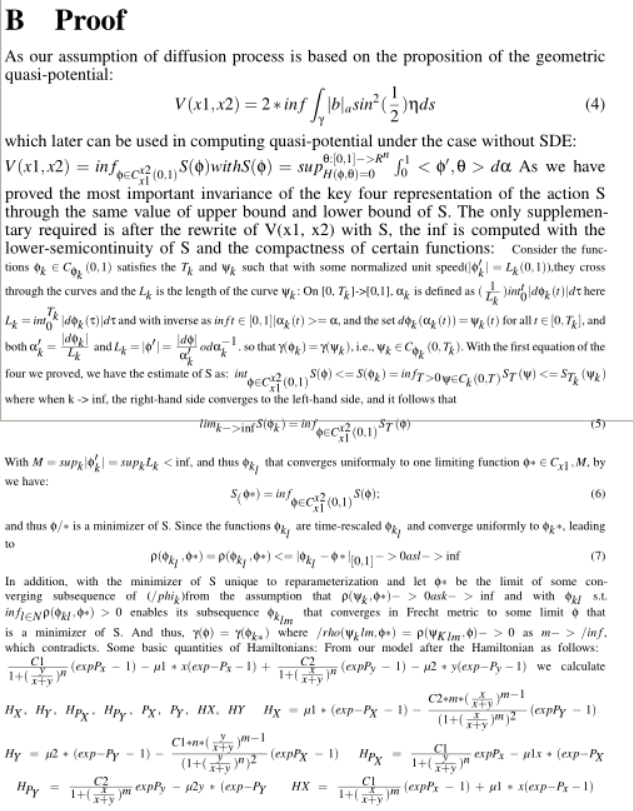
<https://github.com/dashboard>

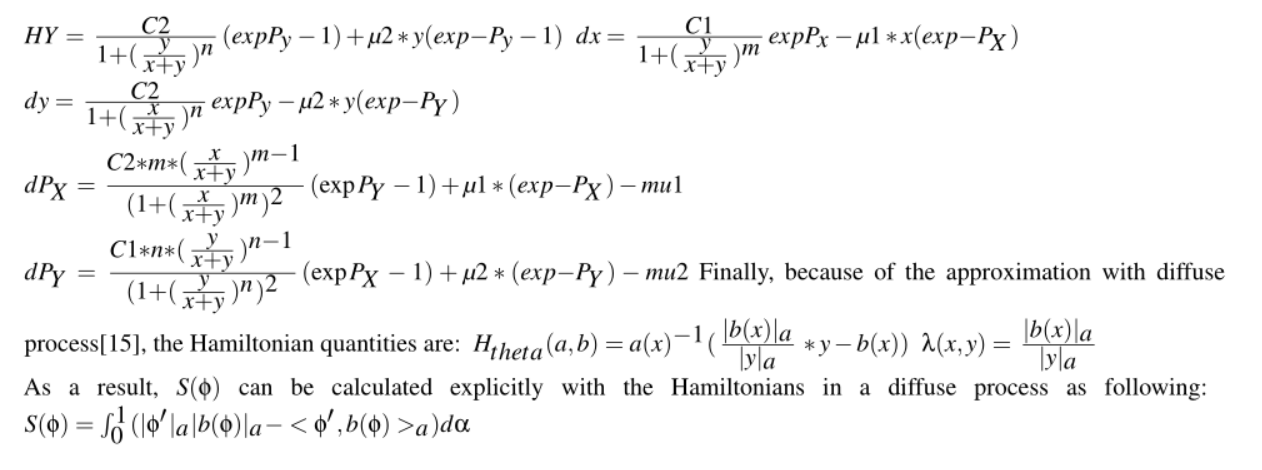
**Figure**





|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |







1. linear regression method on data, (Y = B\*X) :

-0.00114 -6.69764 1.11999 4345.69955 3.25582 NA 0.33371

And the analysis summary :

#Coefficients: (1 not defined because of .66singularities)

# Estimate Std. Error t value Pr(>|t|)

#(Intercept) -0.00114 0.060137 88 0.001425

#as.matrix(X)V2 1.11999 0.003344 66.79 0.0001425 \*\*

#as.matrix(X)V4 3.25582 0.000215 271.46 0.0001425 \*\*

#as.matrix(X)V5 -0.44712 0.000600 1.02 0.0001425 \*\*

#as.matrix(X)V6 0.33371 0.287314 2e10\*\*2 0.0001425 \*\*

#Residual standard error: 0.231 on 456 degrees of freedom

#Multiple R-squared: 1, Adjusted R-squared: 0.9946

F-statistic: 1.21e+02 on 2 and 456 DF, pvalue 254

The estimated effect of V2,V4, V5, V6 on the convert is 1.11999 3.25582 , -0.44712, 0.3371.

This means that for every 1% increase in V2 on X, there is a correlated 1.11% decrease in the incidence of Y. Similar to V4, V5, V6

The standard errors for these regression coefficients are very small, and the t-statistics are very large (66.79,271.46 , 1.02 and 200 , respectively). The p-values reflect these small errors and large t-statistics. And for both parameters, there is almost zero probability that this effect is due to chance.(\*\* gives the variance in 0.001)

**C.code**

MaxStep = 50;

W = eye(size(z));

X = z;

v = var(X)+0.000000001;

for steps = 1:MaxStep

temp = ones(size(v))./v;

Vtemp = inv(X'\*inv(W)\*X+temp);

V = Vtemp\*eye(size(Vtemp))\*Vtemp';

L = chol(abs(V),'lower');

S = V\*X';

B = S\*inv(W)\*z;

% observations for normal with W

H = X.\*S;

CW = H./(W-H);

CW(isnan(CW))=0;

m = X.\*B;

m = m-CW.\*(z-m);

qtemp = CW.\*(CW+1);

qtemp(isnan(qtemp)) = 0;

% draw Z from truncated normal

q = sqrt(qtemp\*eye(size(qtemp))\*qtemp');

Z = X;

R = X;

%Z(:,i) = mvnpdf(X,mean(m,2)',q);

temp = mean(m,2)';

for i = 1:size(X,1)

Z(:,i) = normpdf(X(:,i),temp(i),q(i,i));

end

%update B

B = B +((Z-X)./W).\*S;

B(isinf(B)) = 0;

%update beta

beta = B + L\*T;

%observations for logistics

m = beta.\*X;

for i = 1:size(X,1)

temp = makedist('Logistic','mu',mean(m(:,i)),'sigma',abs(std(m(:,i))));

Z(:,i) = pdf(temp, Z(:,i));

R(:,i) = Z(:,i)- m(:,i);

end

%sampling lambda

Y = normpdf(Z,0,1);

Y = Y.^2;

Y = 1+(Y-sqrt(Y.\*(4\*R+Y)))./(2\*R);

lambda = Z;

for i = 1:6

Ztemp = R(:,i).\*Y(:,i);

Ztemp2 = R(:,i)./Y(:,i);

lambda(:,i) = Ztemp;

lambda(Z(:,i)>(ones(size(Y(:,i)))./(1+Y(:,i)))) =

if mean(lambda(:,i)) > 4/3

Z(:,i) = rightmost(Z(:,i),lambda(:,i));

else

Z(:,i) = leftmost(Z(:,i),lambda(:,i));

end

end

end

function Z = rightmost(U,lambda)

X = exp(-0.5\*lambda);

Z = X;

% squeezing

for t = 1:length(X)-1

Z(t) = Z(t) - (t+1)^2\*X(t)^(t+1)^2-1;

t = t+1;

Z(t) = Z(t) + (t+1)^2\*X(t)^(t+1)^2-1;

end

Z(Z<U)= 0;

End

function Z = leftmost(U,Lambda)

H = 0.5\*log(2)+2.5\*log(pi)-2.5\*log(Lambda)-repmat(pi^2,6,1)./(2\*Lambda)+0.5\*Lambda;

lU = log(U);

X = exp(-pi^2/(2\*Lambda));

Z = X;

K = Lambda/pi^2;

% squeezing

for t = 1:length(X)-1

Z(t) = Z(t) - K(t)^(t^2-1);

t = t+1;

Z(t) = Z(t) + K(t)^(t^2-1);

end

Z((reshape(H,6,1)+reshape(log(Z),6,1))<reshape(lU,6,1))= 0;

End