A Modified Stochastic Simulation Algorithm for Time-Dependent Intensity Rates

Conference Paper · May 2013

DOI: 10.1109/CSCS.2013.101

2 authors:

Raluca Roxana Purnichescu Purtan
Polytechnic University of Bucharest
30 PUBLICATIONS **269** CITATIONS

SEE PROFILE

Andreea Udrea
Polytechnic University of Bucharest
61 PUBLICATIONS **361** CITATIONS

SEE PROFILE

2013 19th International Conference on Control Systems and Computer Science

A Modified Stochastic Simulation Algorithm for Time-Dependent Intensity Rates

Raluca Roxana Purnichescu Purtan
Faculty of Applied Science, Department of Mathematical Methods and Models
University "Politehnica" of Bucharest
Bucharest, Romania
raluca.purtan@gmail.com

Andreea Udrea
Faculty of Automatic Control and Computers, Department of Automatic Control and Systems Engineering
University "Politehnica" of Bucharest
Bucharest, Romania
udrea.andreea@yahoo.com

Abstract **— There are two main approaches in the mathematical modeling of coupled systems of (bio)chemical reactions: continuous, represented by differential equations whose variables are concentrations or discrete, represented by stochastic processes whose variables are numbers of molecules. The latter approach is used mostly for biochemical systems with a low to moderate number of molecules of certain species and this kind of systems are typically modeled as continuous time – discrete state Markov Process. There are exact stochastic algorithms to simulate state trajectories of discrete, stochastic systems and these algorithms are based on methods that are rigorously equivalent to the Master Equation approach. Two of the most widely used methods for simulating the stochastic dynamics of a chemical system are the exact stochastic simulation algorithm (SSA, known also as Gillespie algorithm) and its approximate variant, the tau-leaping algorithm.**

This paper describes a modified version of SSA - First reaction method - by letting the intensity rates of the reactions to be functions of time. The importance of this adaptation is obvious when considering some classes of biological models (for example, the one involving circadian rhythm). The underlying assumptions are that the system is well stirred such that at any moment, each reactions occur with equal probability at any position, that each reaction, once occurred, completes instantaneously (there are no reactions with delay involved) and that the system is non stiff (there are no different time scales of the reactions involved).

Keywords **— chemical kinetics, stochastic simulation algorithm, time-dependent intensity rates**

I. INTRODUCTION

Stochastic chemical kinetics describes the time evolution of a well-stirred chemically reacting system in a way that takes into account the fact that molecules dynamics come in whole numbers and exhibit some degree of randomness in their dynamical behavior. In the last 30 years, more and more researchers are using this approach to chemical kinetics in the analysis of cellular systems in biology, where small molecular populations of only a few reactant species can lead to deviations from the predictions of the deterministic differential equations of classical chemical kinetics.

By representing the chemical reacting system as a jump Markov process, the state of the system is modeled in terms of molecule numbers (non-negative integers) and as a chemical reaction fire, discrete state transitions occur, in continuous time. Many of the existing numerical methods developed for modeling well-mixed chemical kinetic systems are based on the work of Gillespie [5][6], who developed the stochastic simulation algorithm (SSA) in two variants, the Direct and First Reaction Methods. More recently, Gibson and Bruck [4] created the Next Reaction Method and Cao et all [3] demonstrated that for certain classes of systems the Direct method can be faster than the Next Reaction method. The stochastic simulation algorithm produces an exact solution to the dynamical chemical system by simulating each molecular reaction event. For systems with fast reversible or numerous reactions, the exact simulation algorithm becomes computationally costly.

The basic idea of the SSA is that, at each point in time, a waiting time to the next reaction and the most likely reaction to occur must be sampled from a joint probability density function. If the rate constants and/or the number of molecules in the system are large, then the time step of the algorithm can be very small and a very large number of steps must be employed for the completion of the simulation. Because of this, Gillespie [7] introduced the Poisson τ-leap method (an Euler-type method) in which more than one reaction are allowed to fire in a given time interval τ with a frequency extracted from a Poisson distribution. Since then, many extensions of this idea have been developed, most of them regarding the improvement of the computational efficiency and finding efficient methods for simulating stiff systems. In the last 10 years, especially in the biochemical field, the variability in external factors (temperature fluctuation, change in volume, circadian rhythms) has been taken more and more into account, and the need for mathematical models which can include them increased. There have been limited studies concerning this subject [8], [1], [2]. A very interesting extension of the SSA can be found in the recent paper [9] where the rate constant is replaced with a random variable, i.e. possess a stationary distribution with a known density.

In this paper we extend the SSA (First Reaction method) by letting the reaction intensity be a general function of time. In particular, we formulate an optimization problem and discuss it for a class of non-negative and continuous functions from the point of view of numerical implementation and time costs. We also come with new stopping conditions for the proposed algorithm in terms of reaction chain verisimilitude. Until now some conditions for non-negative populations of

978-0-7695-4980-4/13 $26.00 © 2013 IEEE
DOI 10.1109/CSCS.2013.101

molecules were discussed and implemented for the approximate binomial τ-leap method [10].

II. THEORETICAL AND COMPUTATIONAL FRAMEWORK

The state of a chemical system in the stochastic framework is defined by the number of each species involved and changes discretely with every reaction that is executed. This simplifying approach is made under the assumption that nonreactive collisions occur far more often than reactive collisions and, hence, the fast dynamics of motion can be neglected.

A typical process can be defined as a well-stirred system of molecules of N chemical species $\{S_1, S_2, ..., S_N\}$, which interact through M *elemental*[1] reaction channels $\{R_1, R_2, ..., R_M\}$. Our goal is to estimate the state vector $X(t) = \left(X_1(t), X_2(t), ..., X_N(t)\right)$, given that the system starts in state $X(t_0) = x_0$ at some initial time t_0. Assuming that the system obeys the mass-action law, each reaction channel R_μ can be characterized mathematically by two quantities [5]:

(P1) The *state-change vector ("jump" vector)* $v_\mu = (v_{\mu 1}, v_{\mu 2}, ..., v_{\mu N})$, where $v_{\mu i}$ is the stoichiometric vector for the specific reaction R_μ, defined to be the change in the $\{S_i\}$ vector of molecular populations induced by a single occurrence of the particular R_μ reaction event; thus, an R_μ reaction induces the state change $x \to x + v_\mu$.

(P2) The *intensity* (*propensity) rate* a_μ defined by: $a_\mu(x)dt$ the probability, given $X(t) = x$, that *one particular* R_μ reaction will occur inside Ω in the next infinitesimal time interval $[t, t + dt)$. The propensity rates are *transition probabilities per unit time* for all M reactions.

Definition (P2) can be viewed as the *fundamental premise* of stochastic chemical kinetics, because everything else in the theory follows from it ([5], [6]). This requires finding the specific conditions under which functions $a_\mu(x)$ having the property (P2) exist, and then determining the forms of those functions. This effort typically focuses on identifying for each R_μ a *reaction probability rate constant* c_μ[2]. Then, using the addition law of probability theory, the probability (P2) can be computed as the sum of $c_\mu dt$ *over all distinct combinations* of R_μ reactant molecules in the current state $x = X(t) = \left(X_1(t), X_2(t), ..., X_N(t)\right)$.

If we denote the number of all distinct combinations of R_μ reactant molecules in the system at any time t, when there are exactly X_i of the S_i molecules $(i = 1, 2, ..., N)$ by $h_\mu(X_1(t), X_2(t), ..., X_N(t))$, we can write the propensity functions for each reaction channels R_μ as $a_\mu(x)dt = c_\mu h_\mu(x)dt$.

The vectors $h_\mu(x)$ are products of combinatorial functions, corresponding to each particular chemical reaction structure. Each vector is easily obtained from the reaction channel by counting the numbers of each molecule species that are consumed and produced in the reaction.

The form of c_μ is *highly reaction specific* – experimentally deduced and usually depends on the *system volume* and *temperature*.

If c_μ are *constants*, the propensity rates are *time independent* the fluctuations in the system are due to intrinsic noise.

If c_μ is a time-dependent function, it corresponds to *time dependent propensity rates* and the fluctuations in the system arises from variability in external factors (extrinsic noise).

Example: Consider a system with 4 molecular populations $(N = 4)$ which interact via 3 reaction channels:

$$S_1 + S_2 \rightarrow 2S_1 + S_3$$

$$S_2 + 2S_3 \rightarrow S_4$$

$$S_4 \rightarrow S_2$$

We denote by $x = (x_1, x_2, x_3, x_4)$ the state of the system at a time moment t.

---

[1] An elemental reaction is one that occurs essentially instantaneously. In practice, there are only two types of elemental reaction: *unimolecular*, in which a single molecule changes form; and *bimolecular*, in which two molecules collide and a chemical change occurs as a result. All other types of reaction (trimolecular, reversible, etc.) are made up of a series of two or more elemental reactions.

[2] $c_\mu dt$ = probability that *a particular combination* (randomly chosen) of R_μ reactant molecules will react accordingly in the next infinitesimal time interval $[t, t + dt)$

TABLE I System characteristics

Type	Reaction description	$h_\mu(x)$ and $v_\mu(x)$
R_1	$S_1 + S_2 \rightarrow 2S_1 + S_3$	$h_1(x) = x_1 x_2$ $v_1 = (1,-1,1,0)$
R_2	$S_2 + 2S_3 \rightarrow S_4$	$h_2(x) = \frac{1}{2} x_2 x_3 (x_3 - 1)$ $v_2 = (0,-1,-2,1)$
R_3	$S_4 \rightarrow S_2$	$h_3(x) = x_4$ $v_3 = (0,1,0,-1)$

After any of the reactions fires, the state of the system will change accordingly: $x \rightarrow x + v_\mu$.

A. The Chemical Master Equation

It has been proven ([5], [6]) that the behavior of the N species population vector $X(t)$ is a jump type Markov process on the non-negative N-dimensional integer lattice, i.e. the transition probability depends only on the current state and not on the previous states. The probability that a certain reaction R_μ will take place in the next infinitesimal time interval $[t + dt)$ is given by:

$$P\{N(t+dt) - N(t) = 1\} = a_\mu(X(t))dt + o(dt)^3 \qquad (1)$$

Consequently, the probability that no reaction will occur in $[t + dt)$ is:

$$P\{N(t+dt) - N(t) = 0\} = 1 - \sum_{\mu=1}^{M} a_\mu(X(t))dt + o(dt) \qquad (2)$$

and the probability that more than one reaction occur in $[t + dt)$ is:

$$P\{N(t+dt) - N(t) > 1\} = o(t) \qquad (3)$$

The corresponding probabilities for these three events that can take place in the interval $[t + dt)$ are rigorously deduced by D.T.Gillespie in [5], and we denoted by $N(t)$ the number of events (reactions R_μ) that occurred by the time t.

The Chemical Master Equation (CME) is just another name for the Kolmogorov forward equation of Markov Processes. It contains one probability variable for each possible state of the system and represents a set of linear differential equations with constant coefficients for the conditional probability:

$$P(X_t, t \mid X_s, s) = P\{X(t) = X_t, \text{ given that } X(s) = X_s, \text{ for } s \le t\}$$

$$\frac{\partial P(X_t, t \mid X_s, s)}{\partial t} = \sum_{\mu=1}^{M} [P(X_t - v_\mu, t \mid X_s, s) \times c_\mu h_\mu(X_t - v_\mu) - P(X_t, t \mid X_s, s) \times c_\mu h_\mu(X_t)] \qquad (4)$$

or, after replacing $a_\mu(x)$ with $c_\mu h_\mu(x)$:

$$\frac{\partial P(X_t, t \mid X_s, s)}{\partial t} = \sum_{\mu=1}^{M} [P(X_t - v_\mu, t \mid X_s, s) \times a_\mu(X_t - v_\mu) - P(X_t, t \mid X_s, s) \times a_\mu(X_t)] \qquad (5)$$

In principle, the CME completely determines the function $P(X_t, t \mid X_s, s)$, subject to some initial conditions. The CME equation is an *exact* consequence from the reaction channels characterized by intensity rates and state-change vectors. If the equation can be solved (i.e. for a system with very few states), then the process $X(t)$ is fully tractable. However, an exact solution of CME can rarely be obtained and the difficulty comes from the *high dimension*, which equals the total number of possible states of the system under study.

B. The Stochastic Simulation Algorithm

The intractability of the CME lies also in the fact that the deterministic approach tries to solve simultaneously for the probability of all possible trajectories. An alternate approach for describing the behavior of the system is to generate a numerical realization of $X(t)$, i.e., a single trajectory of $X(t)$ versus t. This is *not* the same as solving the CME numerically [5]. The SSA proposed by Gillespie in [5] with both of its variants (The Direct Method and The First Reaction Method) is an *exact* algorithm because the generated trajectories are chosen according to *the correct probability distributions*, i.e. the probability of generating a given trajectory with the simulation algorithm is exactly the probability that would come out of the solution of the CME. Due to the fact that SSA is generating trajectories with the correct probability, any parameters of interest can be estimated by generating many trajectories.

If we consider a system of N molecular species (states) interacting via M chemical reactions (transitions), at each time instant the system is in exactly one state and there are at most M possible transitions from that state to another. The main two questions are „which reaction occurs next ?" and „when does it occur ?". In answer, Gillespie proposed two exact simulation methods.

Further, we will focus on the First Reaction Method, assuming, for now, that all reaction probability rates c_μ are time independent.

Taking into account that relations (1), (2), (3) hold for each reaction channel, the reactions are independent Poisson processes with intensity (propensity) rate $a_\mu(x)$. Therefore the distribution of the first event time (firing of the each reaction) is exponentially distributed with a parameter of $a_\mu(x)$. Considering the whole system, the reaction that will fire first

[3] $o(dt)$ denotes terms that are negligible for small dt

(and will be the only reaction firing in the time interval $[t, t+dt)$, according to (1), (2) and (3)) is, naturally, the one with the smallest time of first event.

In order to find the time of the first event for each reaction, is a common practice to generate a sequence of M random variables $(r_1, r_2, ..., r_\mu, ..., r_M)$, uniform distributed in $(0,1)$ and apply the Probability Integral Transform to obtain a corresponding sequence of M random variables exponentially distributed : $a_\mu(x) \cdot e^{-a_\mu(x)\cdot\tau_\mu}$. The time of the first firing for the μ-th reaction will be $\tau_\mu = \frac{1}{a_\mu(x)} \ln\left(\frac{1}{r_\mu}\right)$ and the reaction that will fire is the reaction corresponding to $t_{\min} = \min_{\mu=1...M}\{\tau_\mu\}$.

This is the main core of the First Reaction Method. Updating the state vector with the corresponding state change vector and the incrementing the time by $t_{\min}$, the algorithm is running until a stop condition is enabled.

The First Reaction Method is the algorithm implementing the above strategies. In [5], Gillespie proved that is an exact simulation algorithm.

III. MODIFIED FIRST REACTION METHOD ALGORITHM

Consider a system in which the probability rates c_μ are time dependent, i.e. $c_\mu(t)$. For example, any modification in the temperature or volume of the reacting system could lead to changes in the specific probabilities rates $c_\mu(t)$. In biochemistry is a well-known fact that external factors as circadian rhythms can play a critical role in absorption and metabolize of different substances. Some of these external influences have been monitored by repeatedly and carefully conducted experiments, so there is a large amount of data we can rely on for constructing some pertinent classes of *specific* time dependent functions.

Using time dependent probability rates $c_\mu(t)$ (at least for some of the reactions of the system), we have to change accordingly the intensity rates: $a_\mu(X(t),t)dt = c_\mu(t)h_\mu(X(t))dt$. Every reaction is a non-homogeneous Poisson process. The number of events for each reaction R_μ in any interval $(t_1, t_2]$ has a Poisson distribution with parameter $\int_{t_1}^{t_2} a_\mu(X(t),t)dt$.

In other words:

$$P\{X(t_1) - X(t_2) = k\} = \frac{\left[\int_{t_1}^{t_2} a_\mu(X(t),t)dt\right]^k \cdot e^{-\int_{t_1}^{t_2} a_\mu(X(t),t)dt}}{k!}$$

It can be shown by straightforward calculations that relations (1), (2) and (3) hold for time dependent intensity rates.

As in the homogeneous case, we are interested to find the time of the first event for each reaction. We shall generate a sequence of M random variables $(r_1, r_2, ..., r_\mu, ..., r_M)$, uniform distributed in $(0,1)$ and the time of the first firing for the μ-th reaction will be the positive solution τ_μ (with a minimal value, if there are multiple solutions) of the integral equation:

$$\int_t^{t+\tau_\mu} a_\mu(X(t),u)du = \ln\left(\frac{1}{r_\mu}\right) \qquad (6)$$

and the reaction that will fire is the reaction corresponding to $t_{\min} = \min_{\mu=1...M}\{\tau_\mu\}$. Finally, updating the state vector with the corresponding state change vector and the incrementing the time by $t_{\min}$, the algorithm is running until a stop condition is enabled.

The classical stopping condition is: the number of desired steps or the sum of the populations dimensions.

In addition we consider that the following criterion must be incorporated in the stop condition: strict negative values for each one of the populations.

Note that $X(t)$ is constant in the above integrals (i.e. the h_μ vectors are constants) because no reaction take place in the time interval $[t, t+\tau_\mu)$. In case of a time independent intensity (propensity) rate $a_\mu(x)$, equation (6) becomes $\tau_\mu = \frac{1}{a_\mu(x)} \ln\left(\frac{1}{r_\mu}\right)$ (the same as for the homogeneous case).

Another important aspect of the above approach is the specific class of $c_\mu(t)$ functions. In our future work we will study in depth this aspect.

Modified Algorithm for First Reaction Method

1) Initialization:
- Set the initial numbers of molecules for each species (in the state vector $X(t)$);
- Set the specific probability rate constants $c_\mu(t)$ for all elemental reactions;
- Set the functions h_μ for all reactions;
- Set the state-change vectors v_μ for all reactions;
- Set $t = 0$.

2) Calculate the propensity functions $a_\mu(X(t),t)$ for each reaction .

3) Generate the sequence of random numbers $\{r_1, r_2, ..., r_\mu, ...r_M\}$ from a unit-interval uniform distribution.

4) For every reaction, find the positive solution τ_μ (with a minimal value, if there are multiple solutions) of the integral equation $\int_t^{t+\tau_\mu} a_\mu(X(t),u)du = \ln\left(\frac{1}{r_\mu}\right)$;

5) Find $t_{\min} = \min_{\mu=1\ldots M}\{\tau_\mu\}$ and store the index of the reaction with $\tau_\mu = t_{\min}$;

6) Update the state vector to reflect that reaction R_μ fired: $X(t) = X(t) + v_\mu$; Set $t \leftarrow t + t_{\min}$;

7) Go to step 2 or stop.

As it was already pointed out, SSA is time consuming especially if the number of reactions and populations is large. If we consider the numerical implementation of step (4) we are in fact confronted with an optimization problem: searching for the minimal solution of the equation over the time interval $(t, t_{\max})$,where t is the present time and $t_{\max}$ is the simulation duration in order to determine the value for τ_μ.

There are some general methods used for determining the most suitable value for $t_{\max}$ function of the number of populations and reactions involved [5].

Numerically computing the integral at step (4) over the time interval $(t, t_{\max})$ and looking for the minimal solution of the equation it would certainly raise a computation time problem.

We consider that, in spite of this aspect, there are certain classes of functions that are extremely common in bio-chemistry for which the numerical algorithm can be successfully implemented in terms of time costs.

One of these classes is represented by periodic functions (modeling, for example, the circadian rhythms, that have sinusoidal properties). For this specific class of functions the search for a minimal solution takes place only on the interval $\left(t, \left(\left[\frac{t}{T}\right]+1\right)\cdot T\right)$ where T denotes the period of the considered function.

Also, keeping in mind that: $a_\mu(X(t),t)dt = c_\mu(t)h_\mu(X(t))dt$

where, in fact, $h_\mu(X(t))$ has a constant value during the numerical implementation of this step, one only has to compute the intersection between the sinusoid and a straight line, parallel to the time axes denoting the constant:

$$\frac{1}{h_\mu(X(t))}\ln\left(\frac{1}{r_\mu}\right).$$

Other classes of functions can be found for which the algorithm does not dramatically increase in terms of duration.

In spite of the fact the time increases for the simulation duration, the proposed algorithm solves the general problem of time varying intensity rates in a general manner. The numerical implementation can be adapted for each class of functions characterizing the intensity rates and optimal solution that decrease the simulation duration can be found.

IV. Conclusions and Further Work

In this paper we propose a modified version of SSA - First reaction method considering the intensity rates of the reactions to be functions of time. This modification gives an added value to the original algorithm in the context of many bio-chemical processes with variable comportments. The mathematical aspects and a series of numerical implementations problems and solutions are discussed. An example of class of functions for which the algorithm can be applied without the otherwise inherent drawbacks in terms of is given.

Improvements and further work:

- simplifying the expression of the equality in the 4th step of the algorithm in order to permit a more efficient numerical implementation

- defining and analyzing other classes of functions inspired directly from bio-chemistry.

Acknowledgment

For the first author, this research is supported in part by POSDRU/107/1.5/S/76813.

References

[1] D.F. Anderson, "A modified next reaction method for simulating chemical systems with time dependent propensities and delays", Journal of Chemical Physics, vol.127, 2007, 214107

[2] A.P. Arkin, J. Ross and H.H. McAdams, "Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells", Genetics, vol.149, no.4,1998, pp.1633-1648

[3] Y. Cao, H. Li and L. Petzold, "Efficient Formulation of the Stochastic Simulation Algorithm for Chemically Reacting Systems", Journal of Chemical Physics, vol.121, no.9, 2004, pp.4059-4067

[4] M.A. Gibson and J. Bruck, "Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels", Journal of Physical Chemistry, no.104, 2000, pp.1876-1889

[5] D.T. Gillespie, "Exact stochastic simulation of coupled chemical reactions", Journal of Physical Chemistry, no.81, (1977)pp.2340-2361

[6] D.T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions", J. Comput. Phys., no.22, 1976, pp.403-434

[7] D.T. Gillespie, "Approximate accelerated stochastic simulation of chemically reacting systems", Journal of Chemical Physics, vol.115, no.4, 2001, pp.1716-1733

[8] P. Lecca, "A time-dependent extension of Gillespie algorithm for biochemical stochastic π-calculus", 2006, Proceedings of the 2006 ACM symposium on Applied Computing 23-27 April 2006, Dijon, pp.137-144

[9] C.Y Lee, "Stochastic simulation of biochemical systems with randomly fluctuating rate constants", 2012, arXiv:1202.1266[q-bio.QM]

[10] T. Tian, K. Burrage, "Binomial leap methods for simulating stochastic chemical kinetics", Journal of Chemical Physics, vol.121, no.21, 2004, pp.10356-10364