UNBIASED τ-LEAP METHODS FOR STOCHASTIC SIMULATION OF BIOCHEMICAL SYSTEMS

Zhouyi Xu and Xiaodong Cai

ECE Dept., University of Miami, Coral Gables, FL 33124
E-mail: zxu@umsis.miami.edu, x.cai@miami.edu

ABSTRACT

Stochastic simulation of biological systems has received much attention recently. A very promising stochastic simulation method is the τ-leap method, which can significantly accelerate simulation with controllable accuracy. However, all current τ-leap methods produce biased results, which can cause large simulation errors. In this paper, we analyze the expected number of reactions occurring during each leap. Relying on the analytical results, we develop an unbiased Poisson τ-leap method and an unbiased binomial τ-leap method. Simulations demonstrate that our new unbiased τ-leap method can significantly improve simulation accuracy without sacrificing simulation speed.

Index Terms— Biological system modeling, Stochastic simulation, Parameter estimation, Cell signalling pathway

1. INTRODUCTION

Biochemical systems in living cells are inherently stochastic due to small number of reactant molecules involved. To investigate the stochastic dynamics of such systems, stochastic approaches rather than the traditional deterministic chemical kinetics are needed. The exact stochastic simulation algorithm(SSA), developed by Gillespie [1], can simulate every reacting event occurring in a well-stirred chemical reaction system, thereby revealing the stochastic nature of the system. However, Gillespie's exact SSA requires large computation and is impractical for simulating many real systems.

To accelerate simulation speed, Gillespie developed an approximate simulation method termed Poisson τ-leap method [2, 3]. In the Poisson τ-leap method, a number of reactions are allowed to occur in a relative larger time interval of duration τ. The number of times, K_m, that the mth reaction channel fires in a time interval is approximated by a Poisson random variable. The value of τ is carefully chosen to tradeoff accuracy for simulation speed. As the realization of a Poisson random variable can be any nonnegative integer, there always is a certain probability that some reaction channels fire so many times during a leap that more molecules of some reactants will be consumed than are actually available. When this occurs, the numbers of molecules of those reactants become negative. To deal with the problem of negative number of molecules, the binomial τ-leap method, where the Poisson random variables used in the Poisson τ-leap method are substituted by binomial random variables with the same mean, was proposed in [4, 5]. More recently, a multinomial τ-leap method was developed in [6] to improve the simulation efficiency of the binomial τ-leap method. Since more than one reactions occur during a leap, there always are certain changes in propensity functions. To reduce the simulation errors caused by such changes in propensity functions, Gillespie proposed a midpoint τ-leap method, where the mean of the Poisson random variables is calculated from the propensity functions at an estimated midpoint during a leap [2]. The midpoint τ-leap method was also applied to the binomial τ-leap method [4]. For a review of stochastic simulation methods, we refer the readers to [7].

Due to the changes in propensity functions during a leap, the mean of K_m used in all the τ-leap methods mentioned earlier is not equal to the true mean, as we demonstrated in [8]. Hence, all current τ-leap methods produce bias in simulation results, which can cause large simulation errors. In this paper, we first analyze the true mean of K_m based on the chemical master equation (CME). After getting the true mean, we develop an unbiased Poisson τ-leap method and an unbiased binomial τ-leap method, which can significantly improve simulation accuracy.

2. SYSTEM DESCRIPTION

We are concerned here a well-stirred mixture of $N \geq 1$ molecular species $\{S_1, \cdots, S_N\}$ that chemically interact through $M \geq 1$ reaction channels $\{R_1, \cdots, R_M\}$. We describe the dynamic state of this chemical system by the state vector $\mathbf{X}(t) = [X_1(t), \cdots, X_N(t)]^T$, where $X_n(t)$, $n = 1, \cdots, N$, is the number of S_n molecules at time t, and $[\cdot]^T$ denotes the transpose of the vector in the bracket. Following Gillespie [2, 3], we define the dynamics of reaction R_m by a state-change vector $\boldsymbol{\nu}_m = [\nu_{1m}, \cdots, \nu_{Nm}]^T$, where ν_{nm} gives the changes in the S_n molecular population produced by one R_m reaction, and a propensity function $a_m(\mathbf{x})$ together with the

fundamental premise of stochastic chemical kinetics:

$$a_m(\mathbf{x})dt \triangleq \text{the probability, given } \mathbf{X}(t)=\mathbf{x}, \text{ that one reaction} R_m \text{ will occur in the next infinitesimal time interval } [t, t+dt). \quad (1)$$

Define the probability rate constant c_m. Let $h_m(\mathbf{x})$ be the number of distinct combinations of R_m reactant molecules in the system at time t, then the propensity function is given by $a_m(\mathbf{x}) = c_m h_m(\mathbf{x})$ [1]. Gillespie showed that for the bimolecular reaction R_m of the form $S_1 + S_2 \rightarrow$ product(s), we have $a_m(\mathbf{x}) = c_m x_1 x_2$. For the bimolecular reaction R_m of the form $S_1 + S_1 \rightarrow$product(s), we have $a_m(\mathbf{x}) = c_m x_1(x_1-1)/2$. For a monomolecular reaction R_m: $S_1 \rightarrow$ product(s), we have $a_m(\mathbf{x}) = c_m x_1$.

Based on the fundamental premise (1), Gillespie developed an exact SSA to simulate the occurrence of every reaction when the time evolves [1]. However, the exact SSA requires huge computation, when the system population and/or the number of reaction channels are relatively large. To reduce computation burden, Gillespie further developed the τ-leap method that will be described in the following.

2.1. The Poisson τ-leap method

The τ-leap method attempts to accelerate stochastic simulation by allowing each reaction channel to fire more than one times during a time interval of duration τ [2, 3]. The deterministic value τ is also referred to as the step size of a leap and is selected to satisfy the following leap condition [2, 3]:

C1 *The change in the state during* $[t, t+\tau]$ *is so slight that no propensity function will suffer an appreciable change in its value, i.e.,* $a_m(\mathbf{X}(t')) \approx a_m(\mathbf{x})$, $\forall t' \in [t,\ t+\tau]$, $\forall m \in [1, M]$.

Letting $\Delta a_m(\tau;\mathbf{x}) \triangleq a_m\big(\mathbf{X}(t+\tau)\big) - a_m(\mathbf{x})$, Gillespie imposed the following constraint to satisfy the leap condition C1 [2]:

$$|\Delta a_m(\tau;\mathbf{x})| \le \epsilon a_0(\mathbf{x}),\ \forall m = 1, \cdots, M, \quad (2)$$

where ϵ is a prespecified error control parameter satisfying $0 < \epsilon \ll 1$. Gillespie derived a formula for selecting τ to satisfy (2), which is given by Eq. (6) of [3].

Let K_m, for any $\tau > 0$, be the number of R_m reactions that occur in the time interval $[t, t+\tau]$. If the leap condition C1 is satisfied, Gillespie showed that K_m, $m = 1, \cdots, M$, are well approximated by independent Poisson random variables with mean $a_m(\mathbf{x})\tau$ [2, 3]. Therefore, Gillespie's Poisson τ-leap SSA executes the following steps during each leap: calculates τ from Eq. (6) of [3], then generates a realization of $K_m, m = 1, \cdots, M$, according to the Poisson distribution, and updates the state after a leap as follows:

$$\mathbf{X}(t+\tau) = \mathbf{X}(t) + \boldsymbol{\nu}\mathbf{K}, \quad (3)$$

where $\boldsymbol{\nu} = [\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_M]$ and $\mathbf{K} = [K_1, \cdots, K_M]^T$. We summarize the Poisson τ-leap algorithm as follows [3]:

Algorithm 1 (Poisson τ-Leap)

1. *Initialization (set* $\mathbf{X}(0)$ *and* $t \leftarrow 0$*).*
2. *Calculate* $a_m(\mathbf{x})$, $m = 1, \cdots, M$.
3. *Calculate* τ *from Eq. (6) of [3].*
4. *Generate* K_m, $m = 1, \cdots, M$, *according to the Poisson distribution with mean* $a_m(\mathbf{x})\tau$.
5. *Set* $t \leftarrow t+\tau$, *and update the state vector using* (3).
6. *Go to step 2 until reaching the end time* t_{end}.

3. UNBIASED τ-LEAP METHODS

Using three elementary reactions, we demonstrated that K_m, $m = 1, \cdots, M$ generated in all current τ-leap methods, including the (midpoint) Poisson [3], binomial [4, 5] and multinomial [6] τ-leap methods, are biased [8]. In order to improve simulation accuracy, we need to develop an unbiased τ-leap method. Towards this end, we first need to find the mean of $K_1, \cdots, K_M$ in each leap, and then generate $K_1, \cdots, K_M$ from their probability distributions with the true mean. The mean of $K_1, \cdots, K_M$ can be derived from the CME of the probability mass function (PMF) of $\mathbf{K}$, $P(\mathbf{K};\tau)$, which is given by [9]:

$$\frac{\partial P(\mathbf{K};\tau)}{\partial \tau} = \sum_{m=1}^{M} a_m(\mathbf{K}-\mathbf{e}_m)P(\mathbf{K}-\mathbf{e}_m;\tau) - a_m(\mathbf{K})P(\mathbf{K};\tau), \quad (4)$$

with initial condition P(**0**;0)=1, where $\mathbf{e}_m$ is the mth column of the $M \times M$ identity matrix, and

$$a_m(\mathbf{K}) \triangleq c_m h_m(\mathbf{X}(t) + \boldsymbol{\nu}\mathbf{K}). \quad (5)$$

If we define $\boldsymbol{\mu}(\tau) = E[\mathbf{K}]$, then we can obtain the following ordinary differential equation (ODE) from the CME (4) [8]:

$$\frac{d\boldsymbol{\mu}(\tau)}{d\tau} = E[\mathbf{a}], \quad (6)$$

where $\mathbf{a} \triangleq [a_1(\mathbf{K}), \cdots, a_M(\mathbf{K})]^T$. If $a_m(\mathbf{K}), m = 1, \cdots, M$ are linear functions of $\mathbf{K}$, which is true if all reactions are zeroth or first order reactions, then $E[\mathbf{a}]$ can be written as a linear function of $\boldsymbol{\mu}(\tau)$. In this case, we obtain a first order linear ODE for $\boldsymbol{\mu}(\tau)$, which is ready to be solved analytically or using an efficient numerical method. However, it is often that $a_m(\mathbf{K})$, $m = 1, \cdots, M$ are nonlinear functions of $\mathbf{K}$, due to the reactions with an order higher than 1 involved. In this case, $E[\mathbf{a}]$ involves not only $\boldsymbol{\mu}(\tau)$ but also the second and possibly higher order moments of $\mathbf{K}$, and thus, it is difficult to obtain $\boldsymbol{\mu}(\tau)$ by solving (6). To overcome this problem, we approximate $\mathbf{a}$ by its first order Taylor expansion and then

show in [8] that (6) can be approximated by the following first order linear ODE:

$$\frac{d\boldsymbol{\mu}(\tau)}{d\tau} = \mathbf{F}\boldsymbol{\mu}(\tau) + \mathbf{a}_0 \tag{7}$$

where $\mathbf{a}_0 = [a_1(0), \cdots, a_M(0)]^T$ containing the propensity functions at time t, and $\mathbf{F}$ is an $M \times M$ matrix whose entry on the mth row and m'th column is given by [8]:

$$[\mathbf{F}]_{mm'} = \left[\frac{\partial a_m(\mathbf{x})}{\partial \mathbf{x}}\right]^T \boldsymbol{\nu}_{m'},\ m, m' = 1, \cdots, M. \tag{8}$$

The initial condition of the ODE (7) is $\boldsymbol{\mu}(0) = \mathbf{0}$. It is easy to solve the ODE (7) analytically or using an efficient numerical methods to get $\boldsymbol{\mu}(\tau)$.

3.1. The unbiased Poisson τ-leap method

After we obtain $\boldsymbol{\mu}(\tau)$, we can generate $K_1, \cdots, K_M$ using their distributions with a mean equal to $\boldsymbol{\mu}(\tau)$. Applying this idea to the Poisson τ-leap method, we keep steps 1, 2, 3, 5 and 6 in Algorithm 1 unchanged, but modify step 4 as follows: find $\boldsymbol{\mu}(\tau)$ by solving ODE (7) and then generate K_m, $1 \cdots, K_M$, according to the Poisson distribution with mean $\boldsymbol{\mu}(\tau)$. We refer to our new τ-leap method as unbiased Poisson τ-leap method. Strictly speaking, our unbiased Poisson τ-leap method does not completely eliminate the bias, because the ODE (7), that is used to obtain $\boldsymbol{\mu}(\tau)$, is an approximation of (6). However, the leap step size τ is typically small, and thus, the approximation error introduced by (7) is also typically very small. Moreover, very accurate and efficient numerical methods for solving (7) are available [10].

The variance of a Poisson random variable is equal to the mean. Although the unbiased Poisson τ-leap method can remove the bias, it may not be able to remove the errors in variance, if the variance of K_m is significantly different from the mean, which is possibly the case when changes in propensity functions are relatively large. In [8], we derived the ODE for the variance of K_m, $m = 1, \cdots, M$, and then developed an unbiased Poisson/Gaussian τ-leap method to remove the bias and correct errors in variance. Due to space limitation, we will not present this method in this paper.

3.2. The unbiased binomial τ-leap method

Applying the same idea of the unbiased Poisson τ-leap method, we can also remove the bias in the binomial τ-leap method. Specifically, we can get the mean of K_m, $m = 1, \cdots, M$, $\mu_m(\tau)$, from (7), and then generate K_m, $m = 1, \cdots, M$, from a binomial random variable $\mathcal{B}(k_{m,\max}, p_m)$, where $k_{m,\max}$ is obtained in the same way as in [5] or [4] and $p_m = \mu_m(\tau)/k_{m,\max}$. Since the binomial τ-leap method in [4] cannot handle the case where more than two reaction channels share certain reactants, we now modify the binomial τ-leap method of Chatterjee *et al.* in [5] to obtain the unbiased binomial τ-leap method. More specifically, the unbiased binomial τ-leap algorithm keeps steps 1, 2, 3, 5 and 6 in Algorithm 1, but changes step 4 as follows: find the mean of K_m, $m = 1, \cdots, M$, $\mu_m(\tau)$, from (7), set $\tilde{x}_n = X_n(t)$ and for $m = 1$ to M reaction channels, do the following:

(a) Find $k_{m,\max} = \min_{\nu_{im}<0,\ i\in[1,N]} \lfloor \widetilde{x}_i/|\nu_{im}| \rfloor$, where $\lfloor x \rfloor$ denotes the largest integer less than x.

(b) Calculate $p = \mu_m(\tau)/k_{m,\max}$ and generate K_m from the binomial distribution with parameter $k_{m,\max}$ and p.

(c) Set $\tilde{x}_n = \tilde{x}_n + \nu_{nm}K_m$ for $n = 1, \cdots, N$, if $\nu_{nm} < 0$.

Table 1. The average population of four species in the EGFR signaling pathway

Species	Exact SSA	Unbiased binomial	Binomial
Grb	26007	26006	25986
Sh-G	61516	61517	61537
Shc	5810.8	5810.7	5818.2
Ra	12595	12595	12601

4. SIMULATIONS

Each cell in a multicellular organism has been programmed during development to respond to a specific set of extracellular signals. Such extracellular signals are transduced into the cell through cell signaling pathways. Signalling pathways through the receptor tyrosine kinase (RTK) family of receptors regulates a wide range of biological phenomena, including cell proliferation and differentiation. The epidermal growth factor receptor (EGFR) is an important member of the RTK family. A number of computational models have been employed to investigate the dynamical behavior of the EGFR pathway.

Here we simulate the EGFR signaling pathway based on a computational model described in [6, 11]. This model consists of 23 molecular species and 47 reaction channels, which are listed in Table I of [6]. In our simulations, we used all rate constants and the initial condition listed in the table I of [6], except that the initial concentration of the epidermal growth factor (EGF) was chose to be 1 nM. From the initial concentration of EGF, the initial population of EGF can be found as 1.152×10^6. The initial populations of other species are the same as those in [6]. We run simulations 10^4 times, and each time starts at $t = 0$ and ends at $t = 8$ using exact SSA, the binomial τ-leap method in [5], the midpoint binomial τ-leap method and our unbiased binomial τ-leap method.

Table 1 lists the mean of the number of molecules for four species at $t = 8$. It is seen that our unbiased binomial τ-leap method produces almost the same mean as the exact SSA, while the binomial τ-leap method produces considerable bias. For other species that are not listed in Table 1, all three leap methods yield almost the same mean as the exact SSA. Figure 1 depicts the PDF of the number of Grb

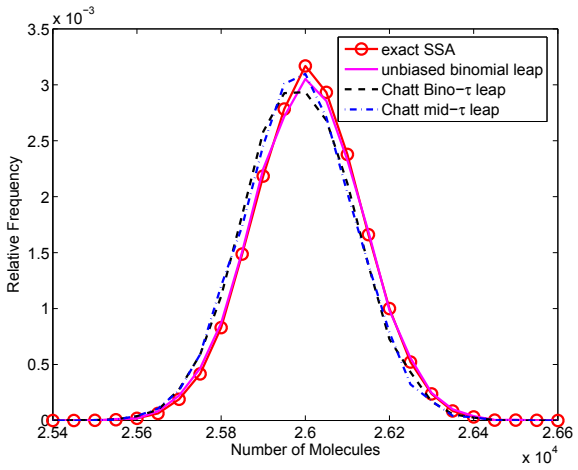

Fig. 1. The estimated PDF of Grb at $t = 8$

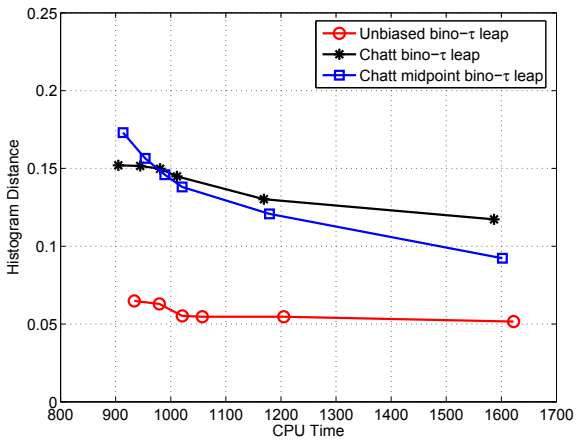

Fig. 2. Histogram distance of Grb at $t = 8$ versus CPU time

molecules at $t = 8$ estimated from the results of 10^4 simulation runs. It is observed that the PDF obtained from our unbiased binomial τ-leap method matches that obtained from the exact SSA, while the PDFs obtained from the (midpoint) binomial τ-leap method exhibits bias. The histogram distances between the results of the exact SSA and those of a leap method was proposed in [12] to measure the simulation accuracy: a small distance implies high accuracy. Figure 2 depicts the histogram distance of *Grb* versus CPU time. It is seen that our unbiased binomial τ-leap method yields much smaller histogram distance than the (midpoint) binomial τ-leap method, while requiring almost the same CPU time for a given ϵ. For other species, our unbiased binomial τ-leap method offers smaller or almost the same histogram distance as the (midpoint) binomial τ-leap method.

5. CONCLUSION

We have developed an unbiased Poisson τ-leap method and an unbiased binomial τ-leap method for stochastic simulation of biological systems. Since bias is absent in our new unbiased τ-leap methods, our leap methods can significantly improve simulation accuracy, compared to existing τ-leap methods. Simulation results have corroborated the superiority of our unbiased τ-leap methods.

6. REFERENCES

[1] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reaction," *J. Phys. Chem.*, vol. 81, pp. 2340–2361, 1977.

[2] ——, "Approximate accelerated stochastic simulation of chemically reacting systems," *J. Chem. Phys.*, vol. 115, pp. 1716–1733, 2001.

[3] D. T. Gillespie and L. R. Petzold, "Improved leap-size selection for accelerated stochastic simulation," *J. Chem. Phys.*, vol. 119, no. 6, pp. 8229–8234, 2003.

[4] T. Tian and K. Burrage, "Binomial leap methods for simulating stochastic chemical kinetics," *J. Chem. Phys.*, vol. 121, pp. 10 356–10 364, 2004.

[5] A. Chatterjee, D. G. Vlachos, and M. A. Katsoulakis, "Binomial distribution based τ-leap accelerated stochastic simulation," *J. Chem. Phys.*, vol. 122, art. no. 024112, 2005.

[6] M. F. Pettigrew and H. Resat, "Multinomial tau-leaping method for stochastic kinetic simulations," *J. Chem. Phys.*, vol. 126, no. 8, Feb. 2007.

[7] X. Cai and X. Wang, "Stochastic modeling and simulation of gene networks," *IEEE Signal Processing Mag.*, vol. 24, no. 1, pp. 27–36, Jan. 2007.

[8] Z. Xu and X. Cai, "Unbiased τ-leap methods for stochastic simulation of chemically reacting systems," *J. Chem. Phys.*, submitted, Oct. 2007.

[9] J. Goutsias, "Quasiequlibrium approximation of fast reaction kinetics in stochastic biochemical systems," *J. Chem. Phys.*, vol. 122, art. no. 184102, 2005.

[10] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*. Cambidge Univ. Press, 1995.

[11] B. N. Kholodenko, O. V. Demin, G. Moehren, and J. B. Hoek, "Quantification of short term signaling by the epidermal growth factor receptor," *J. Biol. Chem.*, vol. 274, pp. 30 169–30 181, Oct. 1999.

[12] Y. Cao, D. T. Gillespie, and L. R. Petzold, "Efficient stepsize selection for the tau-leap simulation method," *J. Chem. Phys.*, vol. 124, no. 4, art. no. 044109, 2006.