



Stochastic Simulation Methods for Biochemical Reaction Systems

Wanda Strychalski¹, David Adalsteinsson¹, and Timothy C. Elston²

¹University of North Carolina at Chapel Hill, Department of Mathematics, ²University of North Carolina at Chapel Hill, Department of Pharmacology

Abstract

Transcription from DNA to RNA in protein production occurs in the cell and involves a small number of mRNA (messenger RNA) molecules and two copies of each gene. Because of the small number of molecules, stochastic effects must be considered. Several numerical methods simulate time series for this type of reaction network. Gillespie methods generate exact trajectories from the solution of the master equation. These methods can be computationally expensive when there are many reactions that occur often. Several hybrid methods have been developed to approximate the trajectories of “fast” species in the reaction system. The time series of these “fast” species can be simulated using a Langevin equation, and the “slow” species can be simulated using a Gillespie method.

Background

A system of chemical reactions consists of species reacting with each other (for example $A \rightarrow B$ at a rate of k_f). We assume mass action kinetics. When a reaction occurs, a species is either increased or decreased by a finite number. Therefore we can treat the process as a jump Markov process whose state space is the integers. The master equation governs the conditional probability density for the process. The master equation is derived from the Chapman-Kolmogorov equation. The general form of the master equation for a system of chemical reactions with N species and M reactions is

$$\frac{dp_{\mathbf{n}}}{dt} = - \sum_{i=1}^M (F_i + B_i) p_{\mathbf{n}} + \sum_{i=1}^M F_i p_{\mathbf{n}-\delta_i} + \sum_{i=1}^M B_i p_{\mathbf{n}+\delta_i}$$

The vector of N species is \mathbf{n} , F_i and B_i are the forward and backward reaction rates, and the vectors δ_i are the stoichiometric constants for the i^{th} reaction. There is no general method to solve the master equation for time-dependent processes, and solving it analytically is not tractable because it requires knowing possibly an infinite number of states.

Gillespie’s Methods

To address this problem, Gillespie derived the next reaction probability, the probability at time t that the next reaction in the time interval $(t+\tau, t+\tau+d\tau)$ will be an R_μ reaction. The next reaction probability density is:

$$p(\tau, \mu) = a_\mu \exp \left[- \sum_{i=1}^M a_i \tau \right]$$

By using conditional probabilities, he separated the joint probabilities into two separate ones and use different ways of generating random numbers to derive the Direct Method and the Next Reaction Method.

Hybrid Methods

If the number of each species is large (much greater than one), Gillespie’s methods become slow. The time in between reactions becomes very small. In this case we can approximate the master equation by a stochastic differential equation, the chemical Langevin equation,

$$\frac{dN_j}{dt} = A_j(\mathbf{N}) + \sum_{k=1}^M \Delta_{k,j} \sqrt{F_k(\mathbf{N}) + B_k(\mathbf{N})} w_k(t)$$

where $\Delta_{k,j}$ is the k_j^{th} entry in the matrix with rows δ_k , the $w_k(t)$ are independent Gaussian white noise processes, and

$$A_j(\mathbf{N}) = \sum_{i=1}^M \Delta_{j,i} (F_i(\mathbf{N}) - B_i(\mathbf{N})).$$

If some species in the biochemical reaction system have a large number of molecules and others have a small number, one approach is to use a hybrid method. In [1], the slow reaction rates are summed and multiplied by the time step in the SDE, Δt . Call this number p_t . If p_t is less than some small number ϵ , then a uniform random number R is generated. If $R < p_t$, a transition occurred and R/p_t determines which slow reaction occurred. If $p_t > \epsilon$, the slow reactions are updated with a Gillespie method and the SDEs are updated with the time step Δt . The hybrid method in [4] dynamically partitions the species in the system as fast or slow. The jump equations of the system give the time that a slow reaction occurs. When $R_j(t)=0$ a slow reaction has occurred.

$$\int_{t_0}^{t_0+t} a_j(t') dt' + \log(r_j) = R_j(t)$$

By comparing the hybrid methods we find which methods are more accurate and faster for a given reaction system. We also explore which hybrid method is best for systems with spatial and stochastic effects.

Numerical Results

In our simulations, we use the Self-Promoter system described in [2]. In this system the gene product acts as an activating regulatory protein for its gene.

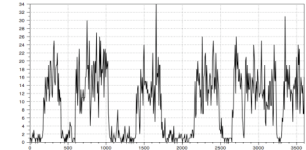
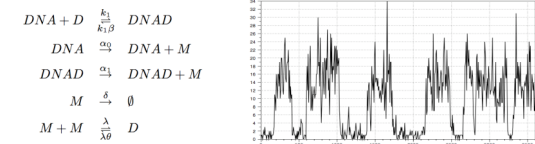
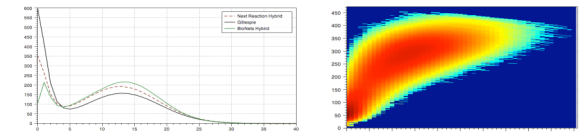
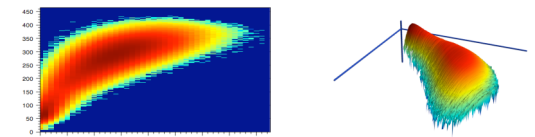


Figure 1: The amount of protein dimer in Self-Promoter system (Gillespie method)



Figures 2 and 3: A histogram comparison of the three methods for the protein dimer and a 2D histogram of D and M from Gillespie’s method



Figures 3 and 4: A 2D histogram of D and M from BioNetS method and a 3D view of the 2D histogram from the Next Reaction Hybrid

References:

- [1] D. Adalsteinsson, T. Elston, D. McMillan *Biochemical Network Stochastic Simulator (BioNetS): Software for stochastic modeling of biochemical networks*, BMC Bioinformatics **15** (2004).
- [2] T. Elston, and T. Kepler *Stochasticity in Transcriptional Regulation: Origins, Consequences, and Mathematical Representations*, Biophysical Journal, **81** (2001), 3116-3136.
- [3] D.T. Gillespie *A General Method for Numerically Solving the Stochastic Time Evolution of Coupled Chemical Reactions*, J. Comp. Phys., **22** (1976), 403-434.
- [4] Y. Kaznessis, and H. Salis *Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions*, J. Chem. Phys., **122** (2004).