Studying intracellular processes using a Markov chain truncation method

Slaven Peleš[1]

***Abstract*— Complexity and stochastic nature of intracellular processes pose serious challenges for quantitative analysis and require developing efficient computational methods. Thus far, Gillespie's kinetic Monte Carlo algorithm has been a gold standard for simulating events inside a cell. We demonstrate that a different approach based on Markov chain truncation provides dramatic improvements over kinetic Monte Carlo methods, both in terms of speed and accuracy. Here we introduce a method for analysis of processes that reach their asymptotic state quickly with respect to the observation time. This method is complementary to earlier proposed Finite State Projection, which is accurate only over a finite period of time.**

***Index Terms*— master equation, stochastic dynamics, simulations, gene networks, signal transduction.**

I. Master equation

INTRACELLULAR processes are often described by a set of chemical reactions between proteins, RNA and DNA molecules. Because copy numbers of relevant chemical species are typically low, macroscopic chemical kinetics does not give accurate quantitative prediction for such a system and stochastic effects must be taken into consideration. Intracellular processes can be viewed as jump Markov processes and modeled using master equation

$$\dot{p}_i(t) = \sum_{j \neq i} [w_{ij} p_j(t) - w_{ji} p_i(t)] \qquad (1)$$

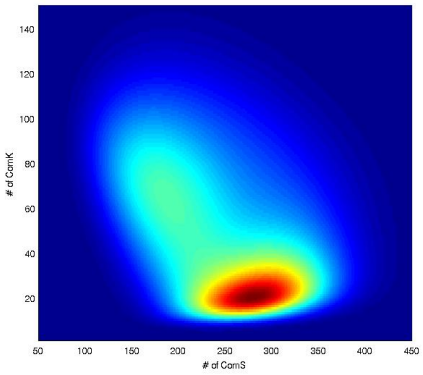
Master equation describes time evolution of probabilities p_i of finding system in a given state i. Here w_{ij} is a mesoscopic transition rate from the state j to the state i. While this is an elegant way to describe dynamics of a stochastic system, solving master equation, even numerically, is a difficult problem. One way to do so is to carry out kinetic Monte Carlo simulations. However, even for relatively simple systems computational cost may be prohibitive.

II. State space truncation

A more efficient approach can be formulated if we use a little bit of intuition about the problem at hand. Cell has limited size and resources and, hence, it is reasonable to assume that all processes will take place within a narrow subset of the state space. We can then approximate probabilities of all unlikely states (say all $p_{i>n}$) with some mean probability field p_∞. If we are interested only in a steady state probability distribution, we can calculate p^s_∞ from the condition $\sum_{j=1}^n w_{\infty j}\, p^s_j = p^s_\infty \sum_{j=1}^n w_{j\infty}$ and substitute it back in the truncated system. In this way we obtain a closed system of equations from which we can compute approximate steady state distribution $\tilde{p}^s_i$ using only simple linear algebraic operations. The approximation error has strict upper and lower bounds. As an illustration we apply truncation method and Gillespie's algorithm to a toy heat shock model [1] and compare computational cost and accuracy of the two approaches.

Method	Execution time	1-norm error
State space truncation	30s	$< 5 \times 10^{-5}$
Gillespie algorithm (1×10^5 samples)	49s	0.024
Gillespie algorithm (3×10^5 samples)	145s	0.017

There is a large number of problems where it is important to find steady state probability distribution. For example, an intricate interplay between ComK and ComS proteins in competence model for *Bacillus Subtilis* [2] is summarized by a probability distribution in the figure below. From there one can easily find what fraction of cells enter the competent state.



III. Conclusion

We demonstrate that Markov chain truncation approach may be a viable alternative to kinetic Monte Carlo simulations. They generally have better convergence properties, allow for more accurate error control and do not require random number generators. Furthermore, since the system is reduced to a finite size, a time scale separation can be readily applied [3]. Here we described a method for calculating steady state probability distribution. In case where transient processes are of interest one should consider Finite State Projection method [4].

References

[1] H. E. Samad and M. Khammash, *Biophys. J.*, vol. 90, pp. 3749–3761, 2006.
[2] G. S. at al., *Nature*, vol. 440, p. 545, 2006.
[3] S. P. at al., *J. Chem. Phys.*, vol. 125, no. 20, p. 204104, 2006.
[4] B. M. at al., *J. Chem. Phys.*, vol. 124, p. 044104, January 2006.

[1]Institute for Collaborative Biotechnologies, University of California, Santa Barbara, CA 93106.