Accelerated Maximum Likelihood Estimation for Stochastic Biochemical Systems

Bernie J. Daigle, Jr.[1,2], Min K. Roh[1], Matthew Wheeler[3], Linda R. Petzold[1], and Jarad Niemi[3]

***Short Abstract* — Parameter estimation is a central task for the mechanistic modeling of biochemical systems. Given the probabilistic nature of stochastic systems, a natural approach involves calculating maximum likelihood parameter estimates (MLEs). We have developed a novel variant of the Monte Carlo expectation-maximization algorithm that efficiently computes MLEs along with a multivariate parameter uncertainty estimate. Our method overcomes shortcomings of existing MLE methods, and we demonstrate its favorable performance by applying it to several stochastic models.**

***Keywords* — Parameter estimation, stochastic simulation, maximum likelihood, expectation-maximization, rare event simulation.**

I. Background

A prerequisite for the mechanistic simulation of a biochemical system is detailed knowledge of its kinetic parameters. Despite recent advances in experimental methodologies, the estimation of (usually unknown) parameter values from observed data is considered the bottleneck for obtaining accurate simulation results [1]. Although many methods exist for parameter estimation in deterministic biochemical systems, similar methods for stochastic systems are less well developed [2].

Given the probabilistic nature of stochastic biochemical models, a natural approach is to choose parameter values that maximize the probability of the observed data with respect to the unknown parameters. This approach gives rise to maximum likelihood parameter estimates (MLEs) which, in the case of fully observed data, can be calculated analytically. Since realistic biochemical systems are discretely and partially observed, computational MLE methods are necessary. Existing approaches typically iterate between two steps: (1) estimating parameter likelihoods using Monte Carlo sampling and (2) maximizing those estimates with respect to the unknown parameters using an optimization algorithm. Execution of (1) requires the generation of multiple system trajectories that are consistent with experimental data, which for models with unknown parameters can be an extremely rare occurrence. This presents a computational challenge recently addressed by either requiring parameter bounds [1,3] or generating system trajectories in a heuristic manner [2,4]. However, parameter bounds are not available for all systems, and it is not clear whether heuristically generated trajectories can be used to accurately and efficiently parameterize all systems.

II. Results

We have developed an accelerated method for calculating MLEs along with uncertainty estimates that combines advances in rare event simulation [5,6] with a computationally efficient version of the Monte Carlo expectation-maximization (MCEM) algorithm [7]. Our method does not require prior bounds on parameters, and it generates probabilistically coherent system trajectories using the stochastic simulation algorithm [8]. We applied the method to several stochastic models of increasing complexity, progressing from an analytically tractable toy model to more realistic (and computationally demanding) biologically-inspired systems. Our results demonstrate that the method substantially accelerates MLE computation on all tested models when compared to a stand-alone version of MCEM. Additionally, our method accurately identified parameter values for systems in which a recently proposed computationally efficient method [2] was ineffective.

III. Conclusion

This work provides a novel, accelerated version of a likelihood-based parameter estimation method that can be readily applied to stochastic biochemical systems. In addition, our results suggest opportunities for added efficiency improvements that will further enhance our ability to mechanistically simulate biological processes.

References

[1] Poovathingal SK, Gunawan R (2010) Global parameter estimation methods for stochastic biochemical systems. *BMC Bioinformatics* **11**, 414.

[2] Wang Y, et al. (2010) Parameter inference for discretely observed stochastic kinetic models using stochastic gradient descent. *BMC Systems Biology* **4**, 99.

[3] Tian T, et al. (2007) Simulated maximum likelihood method for estimating kinetic rates in gene expression. *Bioinformatics* **23**, 84-91.

[4] Horvath A, Manini D (2008) Parameter estimation of kinetic rates in stochastic reaction networks by the EM method. *International Conference on BioMedical Engineering and Informatics* **1**, 713-717.

[5] Gillespie DT, Roh MK, Petzold LR (2009) Refining the weighted stochastic simulation algorithm. *J Chem Phys* **130**, 174103.

[6] Daigle Jr. BJ, et al. (2011) Automated estimation of rare event probabilities in biochemical systems. *J Chem Phys* **134**, 044110.

[7] Caffo BS, Jank W, Jones GL (2005) Ascent-based Monte Carlo expectation-maximization. *Journal of the Royal Statistical Society Series B* **67**, 235-251.

[8] Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* **81**, 2340-2361.

Acknowledgements: This work was funded by grants from NSF IGERT, the US Army Institute for Collaborative Biotechnologies, Army Research Office, NIBIB, and DOE.

[1]Department of Computer Science, University of California Santa Barbara.

[2]E-mail: bdaigle@gmail.com

[3]Department of Statistics and Applied Probability, University of California Santa Barbara.