

High-accuracy R-leaping: Implementing and exploring a potentially exact method for accelerated stochastic simulation

David Orendorff^{1,*}, Philippe Chatelain², Eric Mjolsness¹

1. Department of Computer Science, UC Irvine, Irvine, USA
2. Eidgenossische Technische Hochschule, Zurich, Switzerland

*E-mail: david.orendorff@uci.edu

Abstract

The stochastic simulation algorithm (SSA) [1] has become an integral part of Systems Biology. However, simulating large number of molecules with SSA is inefficient. Recently faster leaping simulation algorithms have been proposed, but they may permit small simulation errors. Present work implements and explores a new algorithm that accelerates SSA with a reaction-leap method and appears to be theoretically exact. Test results support an SSA-equivalent simulator with an absolute maximum Z-score below 3.0 on tests from the CaliBayes suite [2]. A roughly 3-fold speedup over SSA is demonstrated on a simple network. This work supports the prospect of simulating larger biochemical systems with an exact, accelerated stochastic method.

Introduction

The stochastic simulation algorithm (SSA) [1] evolves a well stirred chemically reacting mixture in time by independently sampling (next reaction, next time step) pairs. For large number of molecules in solution, the SSA may become prohibitively slow. Several *leaping* methods have been proposed for speeding up SSA [3, 4] but in doing so may permit small errors.

We explore a reaction leaping method (*ER-leaping*) [5] that is theoretically exact even for a small number of molecules for each species. As with *R-leaping* [4], a number L of events are sampled from a multinomial distribution and their time span is sampled from a Gamma distribution. However, we use rejection sampling to correct the erroneous distribution.

An empirical test of exactness is presented by comparison to about half of the numerical tests with analytically known mean and standard deviation solutions in the CaliBayes [2] test suite. Additionally, a demonstration of speedup is shown with a simple model network.

Implementation

We implemented a C++ version of *ER-leap*. It uses efficient random generators and is code optimized. The libSBML [6] library is used for very basic v2.1 SBML [7] support.

Numerical Validation of Exactness

ER-leap equivalence to SSA is demonstrated empirically. Mass-action stochastic kinetics are assumed. Consider the birth-death process $\{X \rightarrow 2X, X \rightarrow \emptyset\}$. Species X is measured in molecules and is initially 100. The rate parameter of birth is $\lambda = 1$. The rate of death is $\mu = 1.1$.

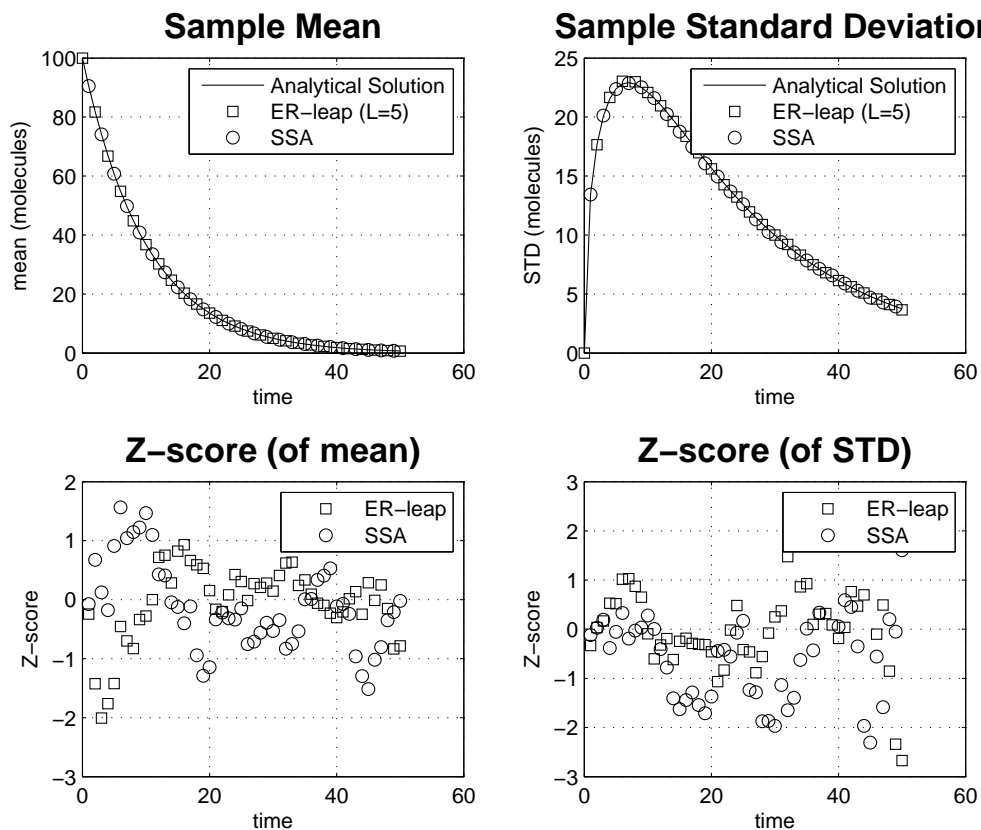


Figure 1: ER-leap with $L=5$ and SSA compared against the analytical mean and standard deviation for birth-death process. The Z-scores will be normally distributed under the null hypothesis of simulator correctness. Z values from $(-3,3)$ are considered reasonable. Simulation time is 50 seconds. Z-score calculated at one second intervals. Results from 20,000 runs.

The analytical solution for the mean and variance of a linear birth-death process is known. Assuming a normal distribution, we can standardize and calculate the Z-score (Figure 1).

All three models in the CaliBayes [2] test suite were run: birth-death, immigration-death and dimerisation. Similar Z-scores are achieved although not presented here.

Optimization

ER-leap can accelerate SSA by a factor of at most L . However, the rejection sampling scheme modifies this efficiency when the acceptance rate drops, as must happen for sufficiently large L . Therefore if L is too high, performance will degrade rapidly. So there is an optimal value of L which can shift with simulated time, as shown numerically in Figure 2.

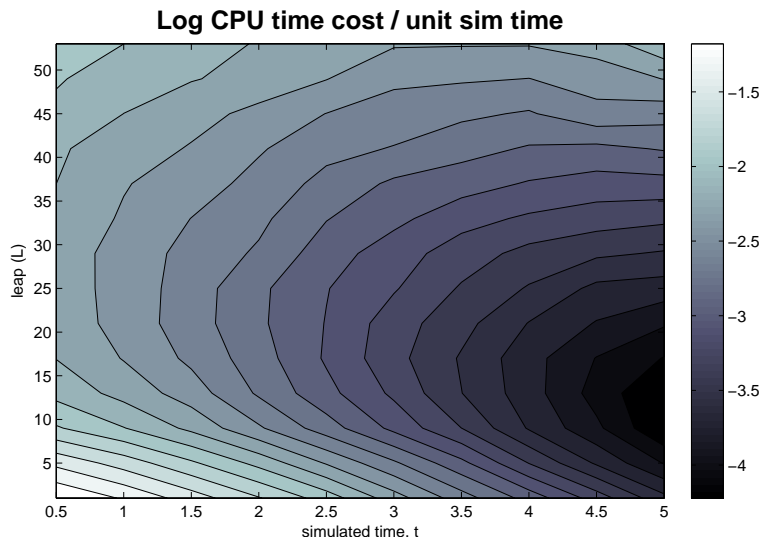


Figure 2: Contour plot of Log CPU time per unit simulation time vs. simulation time and leap. Basic cascading network $\{S1 \rightarrow S2, S2 \rightarrow S3, S3 \rightarrow S4\}$. Initial values: $S1 = 41500, S2 = 39565, S3 = 34450$ and $S4 = 0$. All rates 1.0. Averaged over 20 runs.

Conclusion

The Z-scores from Figure 1 suggest that *ER-leap* has no more bias in mean and variance than SSA. Figure 2 shows the optimality of large leap number, L , and a roughly 3-fold algorithmic speedup over SSA in this case.

References

- [1] D. T. Gillespie, J. Phys. Chem. 81, 2340 1977.
- [2] <http://www.calibayes.ncl.ac.uk/Resources/dsmts/>
- [3] D. T. Gillespie, J. Chem. Phys. 115, 1716 2001.
- [4] A. Auger, J. Chem. Phys. 125, 084103 2006.
- [5] E. Mjolsness, D. Orendorff, P. Chatelain, P. Koumoutsakos. Manuscript in preparation.
- [6] libSBML <http://www.sbml.org/software/libsbml/>
- [7] Hucka M, Bolouri H, Finney A, Sauro H, Doyle J, H K, Arkin A, Bornstein B, Bray D, Cornish-Bowden A, Cuellar A, Dronov S, Ginkel M, Gor V, Goryanin I, Hedley W, Hodgman T, Hunter P, Juty N, Kasberger J, Kremling A, Kummer U, Le Novre N, Loew L, Lucio D, Mendes P, Mjolsness E, Nakayama Y, Nelson M, Nielsen P, Sakurada T, Schaff J, Shapiro B, Shimizu T, Spence H, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J: The Systems Biology Markup Language (SBML): A Medium for Representation and Exchange of Biochemical Network Models. Bioinformatics 2003, 19:524-531.