Fast Simulation of Robust Stochastic Chemical Reaction Networks

David Soloveichik
Erik Winfree group

Outline

• Review models of chemical kinetics: mass-action and stochastic

• Formulate the asymptotic computational complexity of simulation algorithms as a function of maximum molecular count and time

• Toward a rigorous theory of how approximation decreases computation time

• Which systems are robust to this kind of approximation?

• New approximate simulation algorithm (Bounded tau-leaping) with a provable upper-bound on number of leaps

• Lower-bounding the computational complexity of predicting the outcome of simulation
Background:
two widely used models of chemical kinetics

**mass-action**

- works well for macroscopic reaction solutions (e.g., beaker)
- describes variation of concentrations over time
- continuous, deterministic
- simulated by ODE solvers

**stochastic**

- small volume chemistry (e.g., living cell)
- keeps track of the exact number of molecules and each reaction event
- discrete, stochastic
- simulated by Gillespie’s SSA
Stochastic Model is Essential for Understanding Cellular Function

Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage λ-Infected Escherichia coli Cells

Adam Arkin,*,1 John Ross1 and Harley H. McAdams*
*Department of Developmental Biology and Department of Chemistry, Stanford University, Stanford, California 94305
Manuscript received March 5, 1998. Accepted for publication April 30, 1998

Stochastic Gene Expression in a Single Cell

Michael B. Elowitz,1,2* Arnold J. Levine,1 Eric D. Siggia,2 Peter S. Swain2

Clonal populations of cells exhibit substantial phenotypic variation. Such heterogeneity can be essential for many biological processes and is conjectured to arise from still poorly understood stochastic mechanisms. We use the two chemical genetic tools, temperature-sensitive growth and antibiotic resistance, to generate the two reporter strains. We then use the amplitudes and phases of the reporter gene expression to study how the stochastic and deterministic levels of gene expression are related. Our results have implications for the origin of cancer and for the design and study of synthetic genetic circuits.

Stochastic mechanisms in gene expression
(prokaryotic genetics/transcriptional regulation/simulation of genetic regulation/stochastic behavior)

Harley H. McAdams*† and Adam Arkin‡
*†724 Esplanade Way, Stanford, CA 94305; and ‡Department of Developmental Biology, Beckman Center, Stanford University School of Medicine, Stanford, CA 94305
Communicated by Stanley N. Cohen, September 14, 1996

An excitable gene regulatory circuit induces transient cellular differentiation

Gürol M. Süel1, Jordi Garcia-Ojalvo2, Louisa M. Liberman1 & Michael B. Elowitz1

Certain types of cellular differentiation are probabilistic and transient. In such systems individual cells can switch to an alternative state and, after some time, switch back again. In Bacillus subtilis competence is an example of such a transiently

traditional methods that average over large populations fail. To overcome these limitations, we built strains in which activities of different genes are controlled in a cooperative manner. This allows us to study how genetic networks are responsible for desynchronizing cell populations and how they can be used to control the differentiation process.
Review of Stochastic Chemical Kinetics

Assumptions:
- well-mixedness
- bouncing ball interactions
- instantaneous reactions

The model is a continuous-time Poisson process:

\[ A + B \xrightarrow{k_1} 2B \]
\[ C + B \xrightarrow{k_2} D \]
The model is a continuous-time Poisson process:

<table>
<thead>
<tr>
<th>rxn type</th>
<th>propensity $a_j$ (prob of rxn per time instant $dt$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$k \cdot #A$</td>
</tr>
<tr>
<td>$A + B$</td>
<td>$k \cdot #A \cdot #B/V$</td>
</tr>
<tr>
<td>$A + A$</td>
<td>$k \cdot #A(#A - 1)/V$</td>
</tr>
</tbody>
</table>

time until next reaction is exponential random variable with rate $\sum a_j$

probability that next reaction is $j^*$ is $a_j^*/\sum a_j$

Naive simulation algorithm: SSA
Simulation of stochastic kinetics by SSA is often too slow

“[A]ny procedure that simulates every reaction event one at a time, no matter how efficiently it does that, will simply be too slow for many practical applications.”

–Gillespie
We use computational complexity theory to evaluate speed of simulation algorithms.

How does computation time vary asymptotically with the following parameters?

- maximum molecular count $m$,
- duration of simulated time $t$.

Reactions and rate constants are fixed.

For which chemical systems?

- "worst case" ones.

Chemical kinetics can be very complex: (e.g. Turing-universal).
Comparing computation time of mass-action vs stochastic simulation algorithms

ODE solvers:
- if concentration is bounded, volume does not matter
- size of the system does not matter!

Gillespie’s algorithm (SSA):
- computation time increases with molecular counts
- size of the system does matter!
- becomes impractical for many systems (even on cellular scale)
Can approximation in stochastic simulation provably decrease computation time?

Desired properties of approximation:

- Can be exploited for speedup
- Arbitrary degrees of approximation
- Systems and behaviors of interest should be robust to it
- Simple to formalize
Propensities $a_j$ governing the approximate process deviate slightly (measured as percent change) from correct propensities $a_j$ calculated according to SSA.

Deviations are time varying, history dependent.

$$A + B \xrightarrow{k_1} 2B$$

$$C + B \xrightarrow{k_2} D$$

It is tempting to believe that many natural systems are robust to this kind of perturbation because they continue to work well in not completely well-mixed solutions, different temperatures and salt conditions, etc.

Formal definition of approximation

The perturbed process is governed by perturbed propensities rather than correct SSA propensities:

\[ \{\xi_j(t)\} \text{ are random variables indexed by reaction channel } j \text{ and time } t \]

\[ a_j'(\vec{x}, t) = \xi_j(t) a_j(\vec{x}) \]

The perturbed process is called a \( \rho \)-perturbation (\( 0 < \rho < 1 \)) if:

\[ \forall j, t, (1 - \rho) \leq \xi_j(t) \leq (1 + \rho) \]
Robustness to approximation

SSA process is said to be \((\rho, \delta)\)-robust with respect to outcome \(\Gamma\) at time \(t\) if for any \(\rho\)-perturbation the probability of being in \(\Gamma\) at time \(t\) is within \(\delta\) of the SSA process.

In general seems hard to prove robustness of particular systems.
Robustness to approximation

For certain simple systems, can prove degrees of robustness.

Example:

We say an SSA process is *monotonic* with respect to outcome $\Gamma$ if every species is a reactant in at most one reaction, and there is a set $\{n_j\}$ such that $\Gamma$ occurs as soon as every reaction $j$ has fired at least $n_j$ times.

Then $\Pr[\text{in } \Gamma \text{ at time } t]$ for any $\rho$-perturbation is bounded by $\Pr[\text{in } \Gamma \text{ at time } t]$ for $\rho$-perturbations with constant $\xi = 1 - \rho$ and $\xi = 1 + \rho$. 
Figure 1: Examples of SCRNs exhibiting contrasting degrees of robustness. The SSA process $C$ and outcome $\Gamma$ are defined for the two systems by: (a) Rate constants: $k_1 = 1$, $k_2 = 0.001$; start state: $x^0 = (300, 0, 300, 0)$; outcome $\Gamma$: $x_4 \geq 150$. (b) Rate constants: $k_1 = 0.01$, $k_2 = 0.01$; start state: $x^0 = (300, 10, 10)$; outcome $\Gamma$: $x_2 \geq 160$. Plots show $F^{\Gamma}(\cdot, t)$ for an SSA process or $\rho$-perturbation estimated from $10^3$ SSA runs. (Dashed line with circles) Original SSA process $C$. (Dashed lines without circles) The two extremal $\rho$-perturbations: $\tilde{C}^+\rho$ with constant $\xi_j(t) = 1 + \rho$, and $\tilde{C}^-\rho$ with constant $\xi_j(t) = 1 - \rho$. For SCRN (b) we also plot $F^{\Gamma}(\cdot, t)$ for a $\rho$-perturbation with constant $\xi_1(t) = 1 + \rho$, $\xi_2(t) = 1 - \rho$ (triangles), or constant $\xi_1(t) = 1 - \rho$, $\xi_2(t) = 1 + \rho$ (diamonds). Perturbation parameter $\rho = 0.1$ throughout.
Intuition: approximating reaction propensities can speed up simulation

\[ A + B \xrightarrow{k_1} 2B \]
\[ C + B \xrightarrow{k_2} D \]

SSA

assume propensities remain constant for duration of leap
⇒ number of rxn occurrences can be drawn from some distribution

Tau-leaping

leap by some time \( \tau \)

Gillespie 2001, 2006
Cao 2005, etc
New “Bounded tau-leaping” algorithm

1. Compute bounds $b_j$ on occurrences of each reaction.
2. Change state, recompute propensities $a_j$.
3. “Roll dice” to determine when reactions would violate bounds if they continued at current rates. First violating reaction determines leap length $\tau_j$.
4. Determine how many times $n_j$ each reaction occurs in time $\tau$.
   - First violating reaction: already determined.
   - Other reactions: “roll dice” $n_j \sim \text{Binomial}(b_j-1, \tau/\tau_j)$.

- satisfies our definition of approximation.
- cleaner implementation than Gillespie’s tau-leaping, naturally avoids pitfalls like negative molecular counts.
- provable upper-bound on the number of leaps.
Upper Bound on Number of Leaps of Bounded Tau-Leaping: Intuition

How many times can a species violate?

- If decrease only, maximum violations: \( O(1) t \left( \frac{m}{V} + O(1) \right) \)
- Can’t be too many increases: after \( \log m \) of them we exceed \( m \)

Key idea: for rxn to decrease A, \( \#A \) must appear as a reactant and therefore as factor in propensity.
**Upper Bound on Number of Leaps**

SSA

\[ O(1) t m \left( \frac{m}{V} + O(1) \right) \]

concentration: usually bounded

t: duration of simulated time
m: maximum number of molecules
V: volume

with high probability

potentially huge!

Bounded tau-leaping algorithm (and maybe other forms of tau-leaping)

\[ O(1) (\log m)^{O(1)} t \left( \frac{m}{V} + O(1) \right) \]

with high probability

If bounded concentration, molecular count almost doesn’t matter. Approaches speed of ODE solvers
Toward capturing the computational complexity of the prediction problem for robust processes

Recall, the number of leaps Bounded tau-leaping makes takes:

\[ O(1) \left( \log m \right)^{O(1)} t \left( \frac{m}{V} + O(1) \right) \]

A probabilistic time hierarchy conjecture (Computational Complexity):

For any “reasonable” bounds on computation time \( t(n) \) and space \( s(n) \), there are boolean functions that can be computed with bounded error within these time and space bounds but not in time \( O(t(n)^\alpha) \), for any \( \alpha < 1 \), even allowing larger space.

\[ \Rightarrow \text{Implies computation time} \]

\[ O(1) \left( \log m \right)^\beta t^\eta \left( \frac{m}{V} + O(1) \right)^\gamma \]

for \( \eta < 1 \) or \( \gamma < 1 \) is impossible

Based on: Angluin 2006, Soloveichik et al 2008
Contributions:

- initiated study of asymptotic, worst-case computational complexity wrt molecular count of simulating stochastic chemical kinetics
- new simulation algorithm (Bounded tau-leaping) with a provable upper-bound on number of leaps
- defined class of robust chemical processes and tried to capture the computational complexity of predicting their behavior

Open questions:

- better ways to prove that a system is robust
- relate to other notions of robustness
- capturing the computational complexity of non-robust systems
Acknowledgments

Erik Winfree
Matthew Cook
Winfree group members
NSF