Stochastic Analysis of Chemical Reaction Networks Using Linear Noise Approximation

Luca Cardelli, Marta Kwiatkowska, and Luca Laurenti
Department of Computer Science
UNIVERSITY OF OXFORD

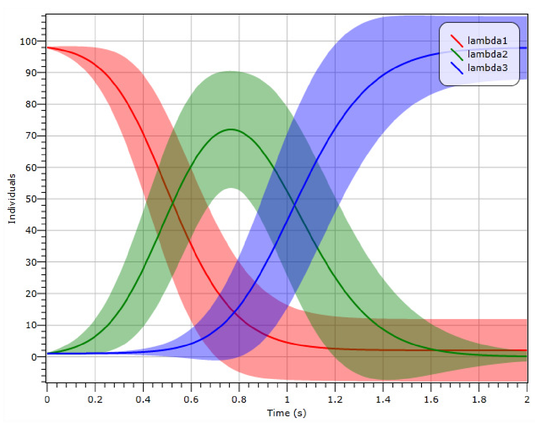
Introduction

A Chemical Reaction Network (CRN) $C = (\Lambda, R)$, where Λ is a set of species which react according to the reactions in R, describes a reactive system. Its analysis is generally performed either assuming a deterministic semantics, and so solving a set of autonomous ordinary differential equations, or by considering a stochastic semantics, generally a Markov process. The deterministic approach is valid only for high number of molecules, while the stochastic semantics is a valid model even for small molecular counts. Nevertheless, transient analysis of a Markov process is rarely feasible for real systems, because of the curse of dimensionality issue.

Question: Can we develop methods to analyze the stochastic semantics of a CRN, while still maintaining the scalability of the deterministic approach?

Motivation

Consider the CRN $C = (\{\lambda_1, \lambda_2, \lambda_3\}, R)$, where $R = \{(\lambda_1 + \lambda_2 \to^{10} \lambda_2 + \lambda_2), (\lambda_2 + \lambda_3 \to^{10} \lambda_3 + \lambda_3)\}$. The following figure plots the time evolution of expected value and variance of the number of molecules of each species for a particular initial condition:



We are interested in checking properties about expected value, variance and probability of the linear combination of the species of C over time, while maintaining the scalability of the deterministic approach. For instance: Is the probability that $\#\lambda_1 - (\#\lambda_2 + \#\lambda_3) > 0$ during the first two seconds greater than 0.8? Or: Is the variance of $\#\lambda_1$ greater than the variance of $\#\lambda_2$ during the first 2 seconds?

Stochastic Evolution Logic (SEL)

The syntax of SEL is given by

$$\eta := P_{\sim p}[B, I]_{[t_1,t_2]} \quad | \quad Q_{\sim v}[B]_{[t_1,t_2]} \quad | \quad \eta_1 \wedge \eta_2 \quad | \quad \eta_1 \vee \eta_2$$

where $Q = \{supV, infV, supE, infE\}$, $\sim = \{<, >\}$, $p \in [0, 1]$ and $v \in \mathbb{R}$. $I = \{[l_i, u_i] \,|\, l_i, u_i \in \mathbb{R} \cup [+\infty, -\infty]\}$ and $B \in \mathbb{Z}^{|\Lambda|}$.

- $P_{\sim p}[B, I]_{[t_1,t_2]}$: is the probability that within $[t_1, t_2]$ the random variable representing the linear combination of the species defined in B has a value inside I greater (smaller) than p?
- $Q_{\sim v}[B]_{[t_1,t_2]}$: is the average value of infimum (supremum) of expected value (variance) of the random variable representing the linear combination of the species defined in B within $[t_1, t_2]$ greater (smaller) than v?

Linear Noise Approximation (LNA)

The stochastic semantics of a CRN is given by a time-homogeneous Continuous time Markov Chain (CTMC) $(X(t), t \in \mathbb{R}_{\geq 0})$ with state space $S = \mathbb{N}^{|\Lambda|}$. The LNA approximate $X(t)$ with the stochastic process $Y(t) = N\Phi(t) + \frac{Z(t)}{\sqrt{N}}$, where $\Phi(t)$ is the solution of the deterministic rate-equations and $Z(t)$ is a Gaussian noise term, independent of N. $Y(t)$ is a Gaussian Process, therefore given $B \in \mathbb{Z}^{|\Lambda|}$, for any time t, $BY(t)$ is a normal random variable with

- $E[BY(t)] = N\Phi(t)$
- $C[BY(t)] = BNC[Z(t)]B^T$

The LNA is guaranteed to be exact for any mass action kinetics system, if sufficiently increasing the molecular population, at least for a limited time, but still gives a good approximation for a large class of biochemical systems, even for quite small counts of molecules. [1,3]

Case Study 1: Phosphorelay Network

The CRN is

$$L1 + B \to B + L1p$$
$$L1p + L2 \to L2p + L1$$
$$L2p + L3 \to L3p + L2$$
$$L3p \to L3$$

We consider the following SEL property:

$$P_{>0.7}[(\#L1p - \#L3p), [0, +\infty]]_{[0,100]} \wedge$$
$$P_{>0.98}[(\#L3p - \#L1p), [0, +\infty]]_{[300,600]}$$

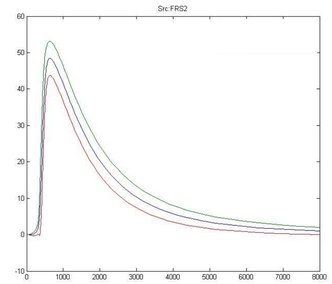
Case Study 1 (Continued)

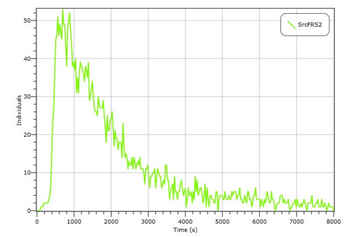
We compare the result of LNA-based model checking of SEL with the corresponding property verified using standard uniformization, for different initial numbers of molecules of each species (Init).

Init	Time (LNA)	Time (Unif)	MaxErr	AvgErr
20	0.22 sec	2 min	0.0675	0.0519
32	0.23 sec	5 min	0.059	0.02
64	0.26 sec	$>$ 2 hr	0.0448	0.0027
100	0.3 sec	$>$ 2 hr	0.03	0.0011

Case Study 2: FGF Pathway

We consider the CRN described in [2]. It is composed by more than 50 reactions and species, with initial counts of molecules for each non-zero concentration species equal to 105. We use SEL with LNA-based model checking (left figure) to calculate expected value and variance of $\#Src{:}FRS2$ over time and compare the result with a single stochastic simulation of the same system (right figure)





Result confirms the utility of SEL and LNA-based model checking for systems that cannot be analyzed by statistical model checking (time consuming) or by exploration of the state space (state-space explosion problem).

References

[1] Wallace, E. W. J., et al. *Linear noise approximation is valid over limited times for any chemical system that is sufficiently large.* IET systems biology 6.4 (2012): 102-115.
[2] Heath, John, et al. *Probabilistic model checking of complex biological pathways.* Theoretical Computer Science 391.3 (2008): 239-257.
[3] Cardelli L., Kwiatkowska, M., Laurenti L., *Stochastic Analysis of Chemical Reaction Networks Using Linear Noise Approximation.* Computational Methods for System Biology (CMSB) 2015 (to appear).

The authors are supported by a Royal Society Professorship and ERC AdG VERIWARE.

{luca.laurenti,marta.kwiatkowska}@cs.ox.ac.uk, luca@microsoft.com