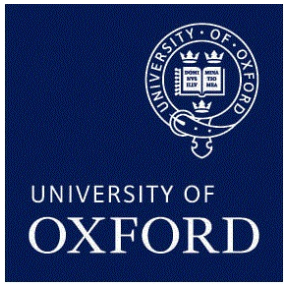
Stochastic Analysis of Chemical Reaction Networks using Linear Noise Approximation

Luca Cardelli[1,2], Marta Kwiatkowska[1], **Luca Laurent**[1]

[1] Department of Computer Science, University of Oxford

[2] Microsoft Research



Microsoft®
Research

Summary

- Motivation
- Background
- Linear Noise Approximation (LNA)
- Stochastic Evolution Logic (SEL)
- Experimental Results

Motivation

- Biochemical systems are generally analysed considering deterministic models. However, deterministic models are accurate only when the molecular population is large

- When the interacting entities are in low numbers there is a need of considering a stochastic model

- Existing methods for analysis of discrete state space stochastic processes are not scalable and highly dependent on the initial number of molecules

- **Question:** Can we derive a formal method to analyse the stochastic semantics of biochemical systems that is scalable and independent of the initial number of molecules?

Chemical Reaction Networks (CRNs)

- **A CRN $\mathcal{C}=(\Lambda,R)$ is a pair of sets**

- Λ is a finite set of **species** $\{\lambda_1, \lambda_2, ..., \lambda_{|\Lambda|}\}$

- R is a finite set of **reactions** $\{\tau_1, \tau_2, ..., \tau_{|R|}\}$
  - *e.g.* $\tau_i : \lambda_1 + \lambda_2 \xrightarrow{k} \lambda_3$
  - k is the **rate constant**

➢ A **configuration** or **state** of the system, $x \in \mathbb{N}^{|\Lambda|}$, is given by the number of molecules of each species in that configuration

Deterministic Semantics of CRNs

- Set of autonomous polynomial ODEs:

$$d\Phi/dt = \mathrm{F}(\mathrm{t})\Phi(t)$$

$$\Phi(0) = x\downarrow 0\ /N$$

- $N = Volume \cdot N\downarrow A$ is the **Volumetric factor** of the system
- $\Phi(t) \in R\uparrow |\Lambda|$ represents the species concentration at time t
- $F(t)$ is determined by **mass action kinetics**
- Number of differential equations equals number of species
- Valid only for high number of molecules
- Does not take into account the stochastic nature of molecular interactions

Stochastic Semantics of CRNs

- It is a **continous time Markov process** $(X\uparrow N\ (t), t \geq 0)$ with discrete state space S and infinitesimal generator matrix Q determined by the reactions

- The transient evoloution of $X\uparrow N$ is described by the **Chemical Master Equation (CME)**
  - Assuming $P(x,t) = Prob\{X\uparrow N\ (t) = x \mid X\uparrow N\ (0) = x \downarrow 0\}$ then the CME can be written as

$$dP(x,t)/dt = P(x,t)Q$$

State Space Explosion Problem

- One differential equation in the CME for any reachable state
- S highly dependent on the initial number of molecules
- Set of reachable states can be huge or even infinite

➢ Not possible to solve the CME for large molecular populations and/or large CRNs

- **Solution: USE LINEAR NOISE APPROXIMATION !**

Linear Noise Approximation (LNA)

- Technique pioneered by Van Kampen in his CME espansion

$$X\uparrow N\,(t) \approx Y\uparrow N\,(t) = N\Phi(t) + \sqrt{N}\,Z(t)$$

- $\Phi(t)$ solution of the deterministic semantics
- $Z(t)$ is a **Gaussian Process** independent of N
  - $E[Z(t)] = 0 \;\; for \;\; t \geq 0$
  - $dC[Z(t)]/dt = J\downarrow F\,(\Phi(t))C[Z(t)] + C[Z(t)]J\downarrow F\uparrow T\,(\Phi(t)) + G(\Phi(t))$
    - $J\downarrow F\,(\Phi(\mathrm{t}))$ Jacobian of $F(\Phi(t))$
    - $G = 1/N \sum \tau \in R\uparrow \blacksquare v\downarrow \tau \; v\downarrow \tau \uparrow T \; \alpha\downarrow \tau$

Linear Noise Approximation (LNA)

- For any CRN, assuming mass action kinetics, the LNA is always accurate at least for a limited time (it is enough to increase M)
- Independence of the initial number of molecules
  - The number of differential equations depends only on the number of species
- Number of differential equations quadratic in the number of species
- Still good approximation for a large class of CRNs even for quite small molecular populations
  - Not able to handle multinomial distributions

LNA Also Known as Gaussian Approximation

- $Z(t)$ is a Gaussian process

- $Y^N(t) = N\Phi(t) + \sqrt{N}\,Z(t)$ is a Gaussian process

- $B \in \mathbb{R}^{|\Lambda|}$, the linear combination of species $B^T Y^N(t)$ is still Gaussian
  - $E[B^T Y^N(t)] = B^T E[Y^N(t)] = N(B^T \Phi(t))$
  - $C[B^T Y^N(t)] = BC[Y^N(t)]B^T = BC[Z(t)]B^T$

- Probability is calculated by solving Gaussian integrals

Stochastic Evolution Logic (SEL)

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

- Q={ $supV$, $infV$,$supE$,$infE$}
- I set of closed disjoint intervals
- $B \in Z\downarrow \geq 0 \uparrow |\Lambda|$

- $P\downarrow \sim p\ [B,I\ ]\downarrow [t\downarrow 1\ ,t\downarrow 2\ ]$: probabilistic operator
- $supV|infV\downarrow \sim v\ [B]\downarrow [t\downarrow 1\ ,t\downarrow 2\ ]$: supremum/infimum of variance operators
- $supE|infE\downarrow \sim v\ [B]\downarrow [t\downarrow 1\ ,t\downarrow 2\ ]$: supremum/infimum of expected value operators

Semantics (SEL)

$$Pr^{X^N}_{B,I}(t) = Prob\{\omega \in Path(X^N, x_0) \,|\, \omega, t \models (B, I)\}$$

$$X^N, x_0 \models P_{\sim p}[B, I]_{[t_1,t_1]} \quad \leftrightarrow \quad Pr^{X^N}_{B,I}(t_1) \sim p$$

$$X^N, x_0 \models P_{\sim p}[B, I]_{[t_1,t_2]} \quad \leftrightarrow \quad \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} Pr^{X^N}_{B,I}(t)\,\mathrm{d}t \sim p$$

$$X^N, x_0 \models supV_{\sim v}[B]_{[t_1,t_2]} \quad \leftrightarrow \quad sup(C[B(X^N)], [t_1, t_2]) \sim v$$

$$X^N, x_0 \models infV_{\sim v}[B]_{[t_1,t_2]} \quad \leftrightarrow \quad inf(C[B(X^N)], [t_1, t_2]) \sim v$$

$$X^N, x_0 \models supE_{\sim v}[B]_{[t_1,t_2]} \quad \leftrightarrow \quad sup(E[B(X^N)], [t_1, t_2]) \sim v$$

$$X^N, x_0 \models infE_{\sim v}[B]_{[t_1,t_2]} \quad \leftrightarrow \quad inf(E[B(X^N)], [t_1, t_2]) \sim v$$

$$X^N, x_0 \models \eta_1 \wedge \eta_2 \quad \leftrightarrow \quad X^N, x_0 \models \eta_1 \wedge X^N, x_0 \models \eta_2$$

$$X^N, x_0 \models \eta_1 \vee \eta_2 \quad \leftrightarrow \quad X^N, x_0 \models \eta_1 \vee X^N, x_0 \models \eta_2$$

Approximate Model Checking

- We use the LNA for a numerical approximate model checking algorithm of SEL
  - $X\uparrow N$ is approximated by the Gaussian Process $Y\uparrow N$
  - The probability that $B\uparrow T\ Y\uparrow N$ is within the interval $[l,r]$ at time t is:

    $\int_{l\uparrow r} g_{x E[B\uparrow T\ Y\uparrow N\ (t)], C[B\uparrow T\ Y\uparrow N\ (t)]}\ dt$

    where $g(x|E,C)$ is the **Gaussian distribution** with expected value E and variance C.
  - $E[B\uparrow T\ Y\uparrow N\ (t)]$ and $C[B\uparrow T\ Y\uparrow N\ (t)]$ are obtained by solving the LNA for the given initial condition

Phosphorelay Network

Species = $\{L1, L1p, L2, L2p, L3, L3p, B\}$

$\tau\downarrow 1 : L1 + B \rightarrow \downarrow\uparrow k\downarrow 1 \;\; B + L1p$
$\tau\downarrow 2 : L1p + L2 \rightarrow \uparrow k\downarrow 2 \;\; L1 + L2p$
$\tau\downarrow 3 : L2p + L3 \rightarrow \uparrow k\downarrow 2 \;\; L2 + L3p$
$\tau\downarrow 4 : L3p \rightarrow \uparrow k\downarrow 3 \;\; L3$

Initial Condition:

$x\downarrow 0\ (L1) = x\downarrow 0\ (L2) = x\downarrow 0\ (L3) = Init;$
$x\downarrow 0\ (B) = 3 \cdot Init;$

where Init is a variable with values in N

Property to check:

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

Comparison with standard Uniformization

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

FGF Pathways

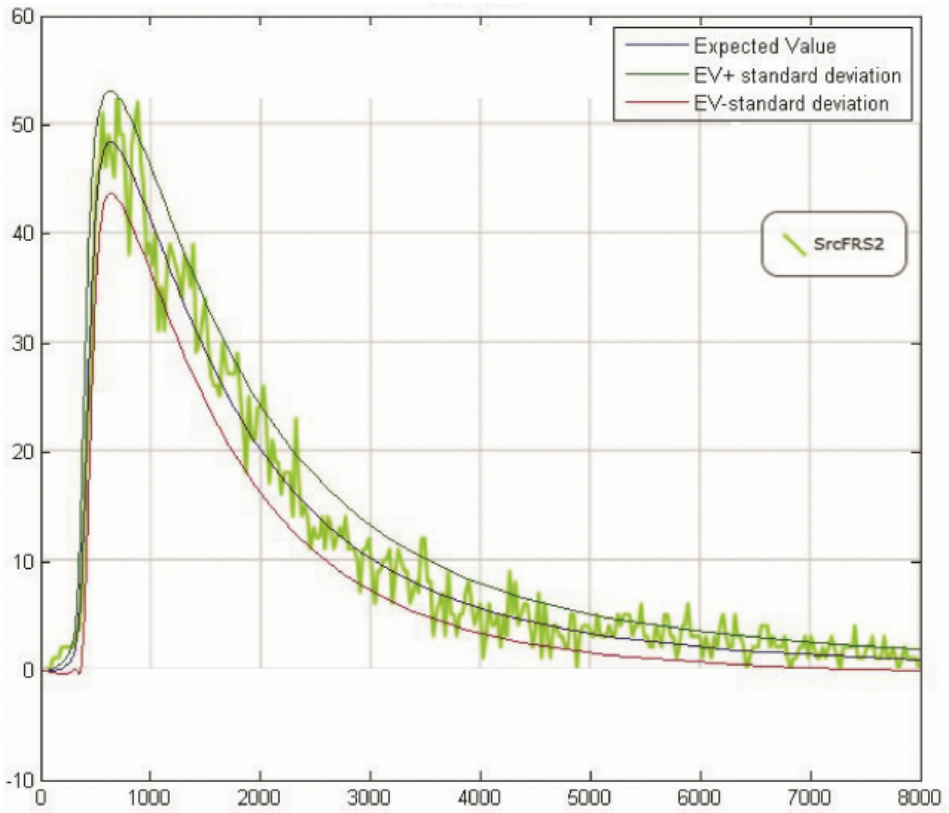
- CRN composed of more than 50 reactions and species!
  - Initial condition such that all species with non zero concentration have 105 molecules
  - Exploration of state space infeasible
  - Simulations are time consuming for such a biochemical system

Comparison of SEL and single stochastic simulation

$supE\downarrow=?\ [\#src{:}FRS2]\downarrow[T,T]$

$supV\downarrow=?\ [\#Src{:}FRS2]\downarrow[T,T]$

$For\ \ T\in[0,8000]$



Conclusion

- We have presented **SEL** with an approximate model checking algorithm based on the LNA
- Our method can be useful for a **fast stochastic characterization** of biochemical systems or for stochastic analysis of systems too large to be checked with standard techniques.
- Increasing the number of molecules the LNA is **always** a valid model assuming mass action kinetics, but can be accurate even far from the thermodynamic limit for a large class of CRNs

Some References

- 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.
- Van Kampen, Nicolaas Godfried. ***Stochastic processes in physics and chemistry***. Vol. 1. Elsevier, 1992. e
- Gillespie, Daniel T. **Deterministic limit of stochastic chemical kinetics.** *The Journal of Physical Chemistry B* 113.6 (2009): 1640-1644.
- Luca Cardelli, Marta Kwiatkowska, Luca Laurenti. **Stochastic Analysis of Chemical Reaction Networks Using Linear Noise Approximation.** *arXiv preprint arXiv:1506.07861* (2015).

THANK YOU!