SAND2007-7865P

Uncertainty Quantification in Stochastic Reaction Networks

KHACHIK SARGSYAN, BERT DEBUSSCHERE, HABIB NAJM, YOUSSEF MARZOUK

Sandia National Laboratories, Livermore, CA

Stochastic Reaction Networks

- Jump Markov Process
- Governed by Chemical Master Equation
- Reduces to deterministic Rate Equations only with large number of molecules
- Numerically, Gillespie's Stochastic Simulation Algorithm (SSA) [1]

Motivation

Reaction networks involving small number of molecules necessitate the use of *stochastic* modeling instead of the *deterministic* one. E.g.

- Immune system signaling reactions
- Microbial reactions
- Surface catalytic reactions

Goals

Develop tools for dynamical, sensitivity and predictability analyses of stochastic reaction networks that account for

- Inherent variability
- Model / parameter uncertainty
- Limited data

Methods

PC expansion of a stochastic process: Quadrature-Based Spectral Methods

Polynomial Chaos (PC) expansion of a stochastic process $X(t,\omega)$

$$X(t,\omega) = \sum_{k=0}^{\infty} X_k(t)\Psi_k(\xi)$$

with standard polynomials $\Psi_k(\xi)$ of standard random variables ξ.

$\xi \sim N(0,1) \Longrightarrow \Psi_k$ Hermite polynomials

$\xi \sim U[-1,1] \Longrightarrow \Psi_k$ Legendre polynomials

PC coefficients recovered by quadrature integration

$$X_k(t) = \frac{\langle X(t,\omega)\Psi_k(\xi)\rangle}{\langle \Psi_k^2(\xi)\rangle}$$

- Generalizes to multiple dimensions
- ξ is viewed as a random input parameter
- Bypass SSA in generating the stochastic process
- Need to put ω and ξ on the same stochastic space

Inferring response curve: Inference-Based Spectral Methods

Inferring PC coefficients of the response curve of an observable $u = \langle f(X(T,\omega))\rangle$ versus a parameter λ

$$u(\lambda) = \sum_{k=0}^{\infty} u_k \Psi_k(\lambda)$$

- Using Monte Carlo Markov Chain (MCMC) to infer u_k from limited set of data $u(\lambda_i)$
- Obtaining the probability distribution of u_k's, i.e. response curve $u(\lambda)$ is found with quantified uncertainty

Reduced order modeling: Karhunen-Loève Decomposition

A stochastic process $X(t,\omega)$ with covariance function $C(t_1,t_2)$ of eigen-set $\{\lambda_n, f_n(t)\}$ admits the KL decomposition[2, 5]

$$X(t,\omega) = \bar{X}(t) + \sum_{n=1}^{\infty} \xi_n(\omega)\sqrt{\lambda_n} f_n(t)$$

with uncorrelated (but not independent!), zero-mean random variables ξ_n.

KL coefficients recovered by integration

$$\xi_n(\omega) = \frac{1}{\sqrt{\lambda_n}} \int_0^{\infty} (X(t,\omega) - \bar{X}(t)) f_n(t) dt.$$

- Truncation to finite sum reduces the order
- Re-generate the stochastic process by sampling ξ_n
- ξ_n can be sampled using inverse Rosenblatt transformation [6]

Generation of random variables: Inverse Rosenblatt Transformation

Rosenblatt transformation [6] is a map of any set of random variables $\xi_1,\dots,\xi_n$ (with known or statistically estimated joint CDF $F(\cdot)$) to a set of independent uniform random variables $\eta_1,\dots\eta_n$.

Inverse Rosenblatt transformation $\vec{\xi} = R^{-1}(\vec{\eta})$ allows to re-generate KL coefficients as functions of uniform random variables in order to integrate them with polynomials in PC expansion.

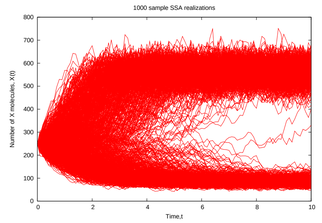
$$\begin{aligned}
\eta_1 &= F_1(\xi_1)\\
\eta_2 &= F_{2|1}(\xi_2|\xi_1)\\
\eta_3 &= F_{3|2,1}(\xi_3|\xi_2,\xi_1)\\
&\vdots\\
\eta_n &= F_{n|n-1,\dots,1}(\xi_n|\xi_{n-1},\dots,\xi_1),
\end{aligned}$$

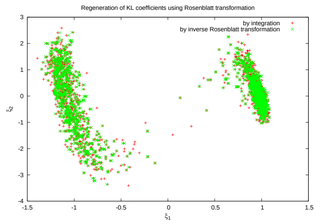
Model Problem.1

Karhunen-Loève representation of Schlögl Model

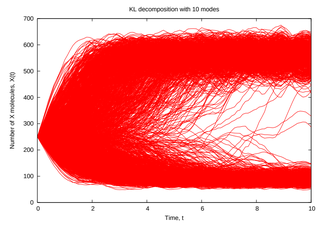
$$\begin{aligned}
A + 2X &\underset{a_2}{\overset{a_1}{\rightleftharpoons}} 3X,\\
B &\underset{a_4}{\overset{a_3}{\rightleftharpoons}} X.
\end{aligned}$$

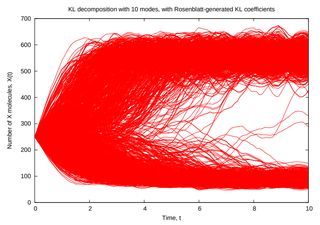
- Interested in the dynamics of number of molecules X
- Models bistable behavior
- Rare switches between quasistable states, unlike the deterministic analog





The first KL coefficient ξ_1 is a random variable with bimodal PDF, each maximum corresponding to one of the branches of the time series. In order to map ξ_n's to a known, standard stochastic space, we use the inverse Rosenblatt transformation. In other words, $(\xi_1,\xi_2,\dots) = \vec{\xi} = R^{-1}(\vec{\eta})$, where $\vec{\eta}$ is a vector of i.i.d. uniform random variables.





The first figure shows 10-term finite KL decompostion of the stochastic process $X(t)$. The second figure is the same, only with KL coefficients ξ_i's generated using inverse Rosenblatt transformation, i.e. using uniform random variables.

References

[1] D.T. Gillespie. Exact Stochastic Simulation of Coupled Chemical Reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[2] K. Karhunen. Zur spektraltheorie stochastischer prozesse. *Ann. Acad. Sci. Fennicae*, 37, 1946.

[3] D. Kim, B.J. Debusschere, and H.N. Najm. Spectral Methods for Parametric Sensitivity in Stochastic Dynamical Systems. *Biophys. J.*, 92(2):379 – 393, Jan 2007.

[4] O.P. Le Maître, H.N. Najm, R.G. Ghanem, and O.M. Knio. Multi-resolution analysis of Wiener-type uncertainty propagation schemes. *Journal of Computational Physics*, 197:502–531, 2004.

[5] M.M. Loève. *Probability Theory*. Van Nostrand, Princeton, NJ, 1955.

[6] M. Rosenblatt. Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23(3):470 – 472, 1952.

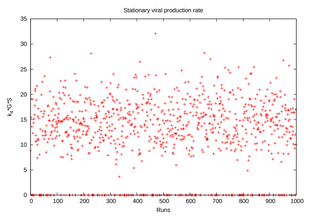
Model Problem.2

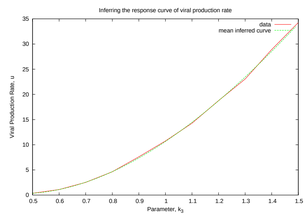
Parametric uncertainty in Viral Infection Model

$$\begin{aligned}
G &\xrightarrow{k_1} T,\\
T &\xrightarrow{k_2} \emptyset,\\
T &\xrightarrow{k_3} G + T,\\
G + S &\xrightarrow{k_4} V,\\
T &\xrightarrow{k_5} S + T,\\
S &\xrightarrow{k_6} \emptyset.
\end{aligned}$$

- G-genomic nucleid acid, T-template nucleid acid, S-structural protein, V-viral progeny
- Initial setting: $G(0) = S(0) = V(0) = 0$, $T(0) = 1$ (Muntipilicy of Infection $= 1$)
- Main observable is the expectation of the stationary viral production rate: $u = \langle k_4 G(\infty) S(\infty)\rangle$ over several realizations

Sensitivity analysis has been performed in [3]. Currently we are extending it to a predictability analysis, with large uncertainties in parameters k_i. With inference-based approach, the PC coefficients of the viral producion rate $u = \langle k_4 GS\rangle$ have been inferred together with the uncertainty due to limited data and intrinsic stochasticity.





The first figure shows the stationary viral production rate over a span of 1000 realizations. Since Multiplicity of Infection $= 1$, several runs show vanishing viral production rate.
The second figure illustrates the MCMC inferred virus production rate response versus a range of different values of k_3 - the most important parameter in terms of sensitivity analysis[3]. Although only the inferred curve is plotted, MCMC allows to obtain the full probability distribution functions of the coefficients of inferred curve. In other words, the uncertainty produced by intrinsic stochasticity and limited data has been taken into account and quantified.

Current Work

Adaptive Multiwavelet Analysis

Global Spectral Methods fail in some circumstances:

- *Multimodal* PDF of the stochastic process at some finite time, i.e. *bistable* stochastic process
- Parameter range includes a *bifurcation*, i.e. *discontinuous* response curve
- Strong non-linearities are present

To tackle both cases, the Polynomial Chaos expansion is generalized to Multiwavelet Expansion [4] of polynomial basis functions on a cascade of shrinking domains. Current work is focused on implementing adaptive schemes of domain decomposition in order to infer a discontinuous response curve, or to represent a multistable stochastic process.

Sandia National Laboratories

National Nuclear Security Administration

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.