Stochastic models of biochemical systems

David F. Anderson*

*anderson@math.wisc.edu

Department of Mathematics

University of Wisconsin - Madison

University of Amsterdam

November 14th, 2012

Stochastic models of biochemical systems

Goal:

- give broad introduction to stochastic models of biochemical systems,
- with minimal technical details.

Outline

1. Construct useful representation for most common continuous time Markov chain model for population processes.
2. Discuss some computational methods – sensitivity analysis.
3. Discuss various approximate models for these CTMCs.

Example: ODE Lotka-Volterra predator-prey model

Think of A as a prey and B as a predator.

$$A \xrightarrow{\kappa_1} 2A, \qquad A + B \xrightarrow{\kappa_2} 2B, \qquad B \xrightarrow{\kappa_3} \emptyset,$$

with $\kappa_1 = 2$, $\kappa_2 = .002$, $\kappa_3 = 2$.

Example: ODE Lotka-Volterra predator-prey model

Think of A as a prey and B as a predator.

$$A \xrightarrow{\kappa_1} 2A, \qquad A + B \xrightarrow{\kappa_2} 2B, \qquad B \xrightarrow{\kappa_3} \emptyset,$$

with $\kappa_1 = 2$, $\kappa_2 = .002$, $\kappa_3 = 2$.

Deterministic model. Let $x(t) = [\#\text{ prey at } t, \#\text{ predator at } t]^T$

$$\dot{x(t)} = \kappa_1 x_1(t) \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \kappa_2 x_1(t) x_2(t) \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \kappa_3 x_2(t) \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

or

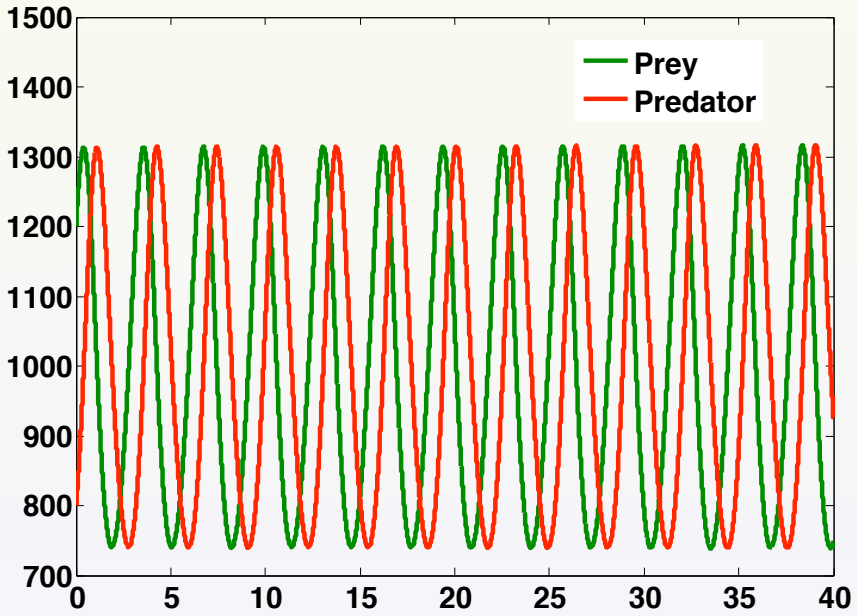
$$x(t) = x(0) + \kappa_1 \int_0^t x_1(s)ds \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \kappa_2 \int_0^t x_1(s)x_2(s)ds \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \kappa_3 \int_0^t x_2(s)ds \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Lotka-Volterra

Think of A as a prey and B as a predator.

$$A \xrightarrow{\kappa_1} 2A, \qquad A + B \xrightarrow{\kappa_2} 2B, \qquad B \xrightarrow{\kappa_3} \emptyset,$$

with $\kappa_1 = 2$, $\kappa_2 = .002$, $\kappa_3 = 2$.



Biological example: transcription-translation

Gene transcription & translation:

$$G \xrightarrow{\kappa_1} G + M \quad \textit{transcription}$$

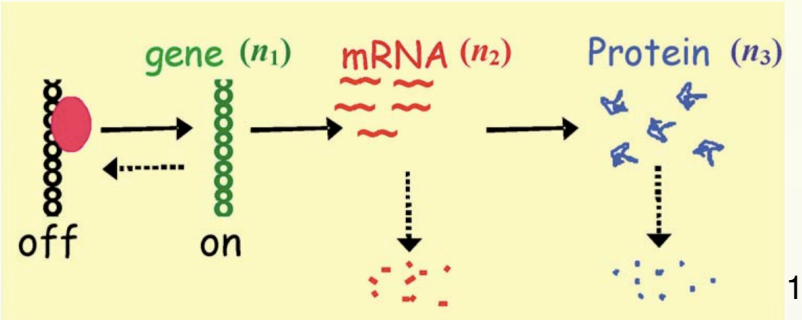
$$M \xrightarrow{\kappa_2} M + P \quad \textit{translation}$$

$$M \xrightarrow{\kappa_3} \emptyset \quad \textit{degradation}$$

$$P \xrightarrow{\kappa_4} \emptyset \quad \textit{degradation}$$

$$G + P \underset{\kappa_{-5}}{\overset{\kappa_5}{\rightleftarrows}} B \quad \text{Binding/unbinding of Gene}$$

Cartoon representation:



[1]

$$X_1 \underset{N^\alpha q_2}{\overset{N^\alpha q_1}{\rightleftarrows}} X_2,$$

$$X_1 \xrightarrow{N\cdot\lambda_1} M, \qquad X_2 \xrightarrow{N\cdot\lambda_2} M,$$

$$M \xrightarrow{\mu} \emptyset.$$

[1] J. Paulsson, Physics of Life Reviews, **2**, 2005 157 – 175.

Another example: Viral infection

Let

1. $T =$ viral template.
2. $G =$ viral genome.
3. $S =$ viral structure.
4. $V =$ virus.

Reactions:

R1) $T +$ "stuff" $\stackrel{\kappa_1}{\to} T + G$	$\kappa_1 = 1$
R2) $G \stackrel{\kappa_2}{\to} T$	$\kappa_2 = 0.025$
R3) $T +$ "stuff" $\stackrel{\kappa_3}{\to} T + S$	$\kappa_3 = 1000$
R4) $T \stackrel{\kappa_4}{\to} \emptyset$	$\kappa_4 = 0.25$
R5) $S \stackrel{\kappa_5}{\to} \emptyset$	$\kappa_5 = 2$
R6) $G + S \stackrel{\kappa_6}{\to} V$	$\kappa_6 = 7.5 \times 10^{-6}$

- R. Srivastava, L. You, J. Summers, and J. Yin, J. Theoret. Biol., 2002.
- E. Haseltine and J. Rawlings, J. Chem. Phys, 2002.
- K. Ball, T. Kurtz, L. Popovic, and G. Rempala, Annals of Applied Probability, 2006.
- W. E, D. Liu, and E. Vanden-Eijden, J. Comput. Phys, 2006.

Some examples

E. coli Heat Shock Response Model. 9 species, 18 reactions.

Reaction	Intensity	Reaction	Intensity
$\emptyset \to A_8$	4.00×10^{0}	$A_6 + A_8 \to A_9$	$3.62 \times 10^{-4} X_{A_6} X_{A_8}$
$A_2 \to A_3$	$7.00 \times 10^{-1} X_{A_2}$	$A_8 \to \emptyset$	$9.99 \times 10^{-5} X_{A_8}$
$A_3 \to A_2$	$1.30 \times 10^{-1} X_{A_3}$	$A_9 \to A_6 + A_8$	$4.40 \times 10^{-5} X_{A_9}$
$\emptyset \xrightarrow{A_1} A_2$	$7.00 \times 10^{-3} X_{A_1}$	$\emptyset \to A_1$	1.40×10^{-5}
$\text{stuff} + A_3 \to A_5 + A_2$	$6.30 \times 10^{-3} X_{A_3}$	$A_1 \to \emptyset$	$1.40 \times 10^{-6} X_{A_1}$
$\text{stuff} + A_3 \to A_4 + A_2$	$4.88 \times 10^{-3} X_{A_3}$	$A_7 \xrightarrow{A_4} A_6$	$1.42 \times 10^{-6} X_{A_4} X_{A_7}$
$\text{stuff} + A_3 \to A_6 + A_2$	$4.88 \times 10^{-3} X_{A_3}$	$A_5 \to \emptyset$	$1.80 \times 10^{-8} X_{A_5}$
$A_7 \to A_2 + A_6$	$4.40 \times 10^{-4} X_{A_7}$	$A_6 \to \emptyset$	$6.40 \times 10^{-10} X_{A_6}$
$A_2 + A_6 \to A_7$	$3.62 \times 10^{-4} X_{A_2} X_{A_6}$	$A_4 \to \emptyset$	$7.40 \times 10^{-11} X_{A_4}$

[2]

[2] Hye Won Kang, presentation at SPA in 2007.

Modeling

1. These models (and much more complicated ones) have historically been predominantly modeled using ODEs.

2. However:

   2.1 there are often low numbers of molecules, which makes timing of reactions more random (less averaging),

   2.2 when a reaction occurs, the system jumps to new state by non-trivial amount: $1 \gg 0$.

Modeling

1. These models (and much more complicated ones) have historically been predominantly modeled using ODEs.

2. However:

   2.1 there are often low numbers of molecules, which makes timing of reactions more random (less averaging),

   2.2 when a reaction occurs, the system jumps to new state by non-trivial amount: $1 \gg 0$.

3. Researchers (mostly) lived with these shortcomings until the late 1990s and early 2000s when it was shown ODE models can not capture important qualitative behavior of certain models:

   - λ-phage lysis-lysogeny decision mechanism (Arkin-McAdams 1998).
   - Green fluorescent protein.

ODEs were often the wrong modeling choice.

Specifying infinitesimal behavior

Q: What is a better modeling choice? Should be

1. discrete space, since counting molecules, and
2. stochastic dynamics.

Let's return to development of ODEs.

An ordinary differential equation is specified by describing how a function should vary over a small period of time

$$X(t + \Delta t) - X(t) \approx F(X(t))\Delta t$$

Specifying infinitesimal behavior

Q: What is a better modeling choice? Should be

1. discrete space, since counting molecules, and
2. stochastic dynamics.

Let's return to development of ODEs.

An ordinary differential equation is specified by describing how a function should vary over a small period of time

$$X(t + \Delta t) - X(t) \approx F(X(t))\Delta t$$

A more precise description (consider a telescoping sum)

$$X(t) = X(0) + \int_0^t F(X(s))ds$$

Infinitesimal behavior for jump processes

We are interested in functions that are piecewise constant and random.

Changes, when they occur, won't be small. If "reaction k" occurs at time t,

$$X(t) - X(t-) = \zeta_k \in \mathbb{Z}^d$$

Infinitesimal behavior for jump processes

We are interested in functions that are piecewise constant and random.

Changes, when they occur, won't be small. If "reaction k" occurs at time t,

$$X(t) - X(t-) = \zeta_k \in \mathbb{Z}^d$$

What is small? The probability of seeing a jump of a particular size.

$$P\{X(t + \Delta t) - X(t) = \zeta_k \mid \mathcal{F}_t\} \approx \lambda_{\zeta_k}(t)\Delta t$$

Question: Can we specify the λ_{ζ_k} in some way that determines X?

- For the ODE, F depended on X.
- Maybe λ_{ζ_k} should depend on X?

Simple model

For example, consider the simple system

$$A + B \to C$$

where one molecule each of A and B is being converted to one of C.

Simple model

For example, consider the simple system

$$A + B \to C$$

where one molecule each of A and B is being converted to one of C.

Intuition for standard stochastic model:

$$P\{\text{reaction occurs in } (t, t+\Delta t] \big| \mathcal{F}_t\} \approx \kappa X_A(t) X_B(t) \Delta t$$

where

- κ is a positive constant, the reaction rate constant.
- $\mathcal{F}_t$ is all the information pertaining to the process up through time t.

Can we specify a reasonable model satisfying this assumption?

Background information: The Poisson process

Will view a Poisson process, $Y(\cdot)$, through the lens of an underlying point process.

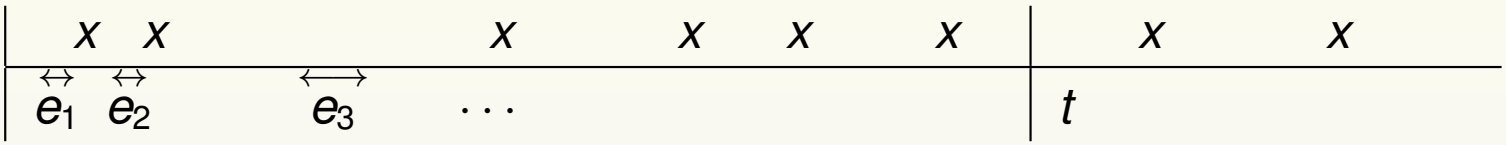
(a) Let $\{e_i\}$ be i.i.d. exponential random variables with parameter one.

Background information: The Poisson process

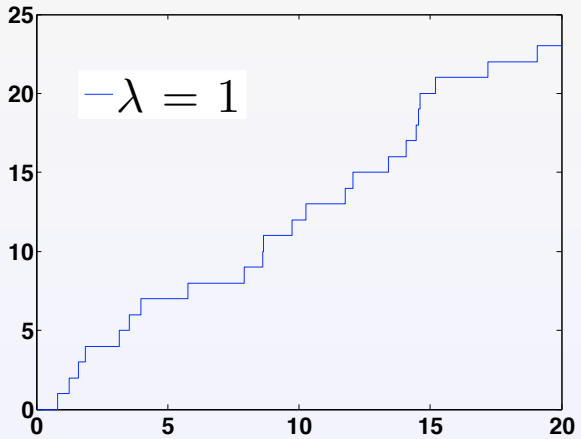
Will view a Poisson process, $Y(\cdot)$, through the lens of an underlying point process.

(a) Let $\{e_i\}$ be i.i.d. exponential random variables with parameter one.

(b) Now, put points down on a line with spacing equal to the e_i:



- Let $Y_1(t)$ denote the number of points hit by time t.
- In the figure above, $Y_1(t) = 6$.

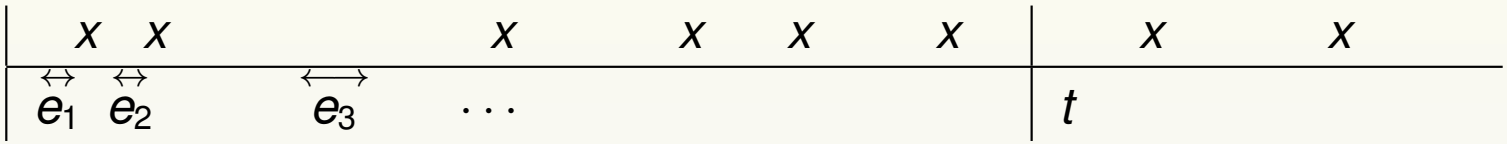


The Poisson process

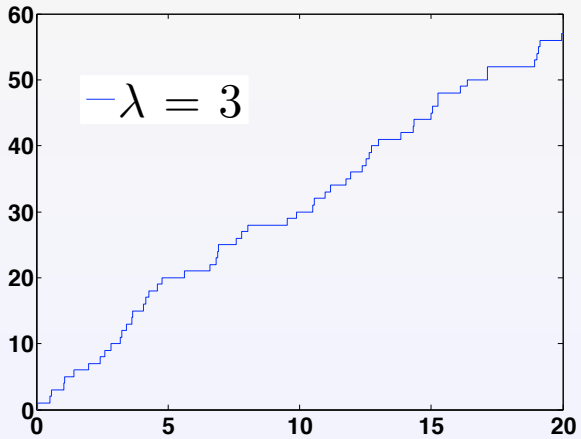
Let

- Y_1 be a unit rate Poisson process.
- Define $Y_\lambda(t) \equiv Y_1(\lambda t)$,

Then Y_λ is a Poisson process with parameter λ.



Intuition: The Poisson process with rate λ is simply the number of points hit (of the unit-rate point process) when we run along the time frame at rate λ.



The Poisson process

There is no reason λ needs to be constant in time, in which case

$$Y_\lambda(t) \equiv Y\left(\int_0^t \lambda(s)ds\right)$$

is a non-homogeneous Poisson process with propensity/intensity $\lambda(t) \geq 0$.

Thus

$$P\{Y_\lambda(t+\Delta t) - Y_\lambda(t) > 0 | \mathcal{F}_t\} = 1 - \exp\left\{-\int_t^{t+\Delta t} \lambda(s)ds\right\} \approx \lambda(t)\Delta t.$$

The Poisson process

There is no reason λ needs to be constant in time, in which case

$$Y_\lambda(t) \equiv Y\left(\int_0^t \lambda(s)ds\right)$$

is a non-homogeneous Poisson process with propensity/intensity $\lambda(t) \geq 0$.

Thus

$$P\{Y_\lambda(t+\Delta t) - Y_\lambda(t) > 0 | \mathcal{F}_t\} = 1 - \exp\left\{-\int_t^{t+\Delta t} \lambda(s)ds\right\} \approx \lambda(t)\Delta t.$$

Points:

1. We have "changed time" to convert a unit-rate Poisson process to one which has rate or intensity or propensity $\lambda(t)$.

2. Will use similar time changes of unit-rate processes to build models of interest.

Return to models of interest

Consider the simple system

$$A + B \to C$$

where one molecule each of A and B is being converted to one of C.

Intuition for standard stochastic model:

$$P\{\text{reaction occurs in } (t, t+\Delta t] \,\big|\, \mathcal{F}_t\} \approx \kappa X_A(t) X_B(t) \Delta t$$

where

- κ is a positive constant, the reaction rate constant.
- $\mathcal{F}_t$ is all the information pertaining to the process up through time t.

Models of interest

$$A + B \to C$$

Simple book-keeping says: if

$$X(t) = \begin{bmatrix} X_A(t) \\ X_B(t) \\ X_C(t) \end{bmatrix}$$

gives the state at time t, then

$$X(t) = X(0) + R(t) \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix},$$

where

- $R(t)$ is the # of times the reaction has occurred by time t and
- $X(0)$ is the initial condition.

Goal: represent $R(t)$ in terms of Poisson process.

Models of interest

Recall that for $A + B \to C$ our intuition was to specify infinitesimal behavior

$$P\{\text{reaction occurs in } (t, t + \Delta t] \big| \mathcal{F}_t\} \approx \kappa X_A(t) X_B(t) \Delta t,$$

Models of interest

Recall that for $A + B \to C$ our intuition was to specify infinitesimal behavior

$$P\{\text{reaction occurs in } (t, t+\Delta t] \big| \mathcal{F}_t\} \approx \kappa X_A(t) X_B(t) \Delta t,$$

and that for a counting process with specified intensity $\lambda(t)$ we have

$$P\{ Y_\lambda(t+\Delta t) - Y_\lambda(t) = 1 | \mathcal{F}_t\} \approx \lambda(t) \Delta t.$$

Models of interest

Recall that for $A + B \to C$ our intuition was to specify infinitesimal behavior

$$P\{\text{reaction occurs in } (t, t+\Delta t] \big| \mathcal{F}_t\} \approx \kappa X_A(t) X_B(t) \Delta t,$$

and that for a counting process with specified intensity $\lambda(t)$ we have

$$P\{Y_\lambda(t+\Delta t) - Y_\lambda(t) = 1 | \mathcal{F}_t\} \approx \lambda(t) \Delta t.$$

This suggests we can model

$$R(t) = Y\left(\int_0^t \kappa X_A(s) X_B(s) ds\right)$$

where Y is a unit-rate Poisson process.

Models of interest

Recall that for $A + B \to C$ our intuition was to specify infinitesimal behavior

$$P\{\text{reaction occurs in } (t, t+\Delta t] \big| \mathcal{F}_t\} \approx \kappa X_A(t) X_B(t) \Delta t,$$

and that for a counting process with specified intensity $\lambda(t)$ we have

$$P\{Y_\lambda(t+\Delta t) - Y_\lambda(t) = 1 | \mathcal{F}_t\} \approx \lambda(t) \Delta t.$$

This suggests we can model

$$R(t) = Y\left(\int_0^t \kappa X_A(s) X_B(s) ds\right)$$

where Y is a unit-rate Poisson process.
Hence

$$\begin{pmatrix} X_A(t) \\ X_B(t) \\ X_C(t) \end{pmatrix} \equiv X(t) = X(0) + \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} Y\left(\int_0^t \kappa X_A(s) X_B(s) ds\right).$$

This equation uniquely determines X for all $t \geq 0$.

Build up model: Random time change representation of Kurtz

- Now consider a network of reactions involving d chemical species, $S_1, \ldots, S_d$:

$$\sum_{i=1}^{d} \nu_{ik} S_i \longrightarrow \sum_{i=1}^{d} \nu'_{ik} S_i$$

Denote reaction vector as

$$\zeta_k = \nu'_k - \nu_k,$$

so that if reaction k occurs at time t

$$X(t) = X(t-) + \zeta_k.$$

Build up model: Random time change representation of Kurtz

- Now consider a network of reactions involving d chemical species, $S_1, \ldots, S_d$:

$$\sum_{i=1}^{d} \nu_{ik} S_i \longrightarrow \sum_{i=1}^{d} \nu'_{ik} S_i$$

Denote reaction vector as

$$\zeta_k = \nu'_k - \nu_k,$$

so that if reaction k occurs at time t

$$X(t) = X(t-) + \zeta_k.$$

- The intensity (or propensity) of kth reaction is $\lambda_k : \mathbb{Z}^d_{\geq 0} \to \mathbb{R}$.
- By analogy with before:

$$X(t) = X(0) + \sum_k R_k(t)\zeta_k,$$

with

$$X(t) = X(0) + \sum_k Y_k \left(\int_0^t \lambda_k(X(s))ds \right) \zeta_k,$$

Y_k are independent, unit-rate Poisson processes.

Mass-action kinetics

The standard intensity function chosen is mass-action kinetics:

$$\lambda_k(x) = \kappa_k(\prod_i \nu_{ik}!)\binom{x}{\nu_k} = \kappa_k \prod_i \frac{x_i!}{(x_i - \nu_{ik})!}.$$

Example: If $S_1 \to$ anything, then $\lambda_k(x) = \kappa_k x_1$.

Example: If $S_1 + S_2 \to$ anything, then $\lambda_k(x) = \kappa_k x_1 x_2$.

Example: If $2S_2 \to$ anything, then $\lambda_k(x) = \kappa_k x_2(x_2 - 1)$.

Other ways to understand model

The infinitesimal generator of a Markov process determines the process:

$$\mathcal{A}f(x) \stackrel{def}{=} \lim_{h \to 0} \frac{1}{h} \left[\mathbb{E}_x f(X(h)) - f(x) \right]$$

Other ways to understand model

The infinitesimal generator of a Markov process determines the process:

$$\begin{aligned} \mathcal{A}f(x) &\stackrel{def}{=} \lim_{h\to 0} \frac{1}{h}\left[\mathbb{E}_x f(X(h)) - f(x)\right] \\ &= \lim_{h\to 0} \frac{1}{h}\left[\sum_k (f(x+\zeta_k) - f(x))P(R_k(h)=1)\right] + O(h) \end{aligned}$$

Other ways to understand model

The infinitesimal generator of a Markov process determines the process:

$$\begin{aligned}
\mathcal{A}f(x) &\stackrel{def}{=} \lim_{h\to 0} \frac{1}{h}\left[\mathbb{E}_x f(X(h)) - f(x)\right] \\
&= \lim_{h\to 0} \frac{1}{h}\left[\sum_k (f(x+\zeta_k) - f(x)) P(R_k(h) = 1)\right] + O(h) \\
&= \lim_{h\to 0} \frac{1}{h}\left[\sum_k (f(x+\zeta_k) - f(x))\lambda_k(x)h\right] + O(h)
\end{aligned}$$

Other ways to understand model

The infinitesimal generator of a Markov process determines the process:

$$\begin{aligned}
\mathcal{A}f(x) &\stackrel{def}{=} \lim_{h\to 0}\frac{1}{h}\left[\mathbb{E}_x f(X(h)) - f(x)\right] \\
&= \lim_{h\to 0}\frac{1}{h}\left[\sum_k (f(x+\zeta_k) - f(x))P(R_k(h)=1)\right] + O(h) \\
&= \lim_{h\to 0}\frac{1}{h}\left[\sum_k (f(x+\zeta_k) - f(x))\lambda_k(x)h\right] + O(h) \\
&= \sum_k \lambda_k(x)(f(x+\zeta_k) - f(x)).
\end{aligned}$$

Other ways to understand model

And we have Dynkin's formula (See Ethier and Kurtz, 1986, Ch. 1)

$$\mathbb{E}f(X(t)) - f(X_0) = \mathbb{E}\int_0^t \mathcal{A}f(X(s))ds,$$

Other ways to understand model

And we have Dynkin's formula (See Ethier and Kurtz, 1986, Ch. 1)

$$\mathbb{E}f(X(t)) - f(X_0) = \mathbb{E}\int_0^t \mathcal{A}f(X(s))ds,$$

Letting $f(y) = 1_x(y)$, above so that

$$\mathbb{E}[f(X(t))] = P\{X(t) = x\} = p_x(t),$$

gives Kolmogorov forward equation (chemical master equation)

$$p_t'(x) = \sum_k \lambda(x - \zeta_k)p_t(x - \zeta_k) - p_t(x)\sum_k \lambda_k(x)$$

Equivalence of formulations

We now have three ways of making the infinitesimal specification

$$P\{X(t+\Delta t) - X(t) = \xi_k | \mathcal{F}_t^X\} \approx \lambda_k(X(t))\Delta t$$

precise:

1. The stochastic equation:

$$X(t) = X(0) + \sum_k Y_k \left(\int_0^t \lambda_k(X(s))ds \right) \zeta_k$$

2. The process is Markov with infinitesimal generator

$$(\mathcal{A}f)(x) = \sum_k \lambda_k(x)(f(x+\zeta_k) - f(x))$$

3. The master (forward) equation for the probability distributions:

$$p_x'(t) = \sum_k \lambda_k(x-\zeta_k)p_t(x-\zeta_k) - p_t(x)\sum_k \lambda_k(x)$$

Fortunately, if the solution of the stochastic equation doesn't blow up, the three are equivalent.

This model is an example of a continuous time Markov chain.

Example: ODE Lotka-Volterra predator-prey model

Think of A as a prey and B as a predator.

$$A \xrightarrow{\kappa_1} 2A, \qquad A + B \xrightarrow{\kappa_2} 2B, \qquad B \xrightarrow{\kappa_3} \emptyset,$$

with $\kappa_1 = 2$, $\kappa_2 = .002$, $\kappa_3 = 2$.

Deterministic model. Let $x(t) = [\#\text{prey}, \#\text{predators}]^T$

$$x(t) = x(0) + \kappa_1 \int_0^t x_1(s)ds \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \kappa_2 \int_0^t x_1(s)x_2(s)ds \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \kappa_3 \int_0^t x_2(s)ds \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Example: ODE Lotka-Volterra predator-prey model

Think of A as a prey and B as a predator.

$$A \xrightarrow{\kappa_1} 2A, \qquad A + B \xrightarrow{\kappa_2} 2B, \qquad B \xrightarrow{\kappa_3} \emptyset,$$

with $\kappa_1 = 2$, $\kappa_2 = .002$, $\kappa_3 = 2$.

Deterministic model. Let $x(t) = [\#\text{prey}, \#\text{predators}]^T$

$$x(t) = x(0) + \kappa_1 \int_0^t x_1(s)ds \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \kappa_2 \int_0^t x_1(s)x_2(s)ds \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \kappa_3 \int_0^t x_2(s)ds \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Stochastic model. Let $X(t) = [\#\text{prey}, \#\text{predators}]^T$

$$X(t) = X(0) + Y_1\left(\kappa_1 \int_0^t X_1(s)ds\right) \begin{bmatrix} 1 \\ 0 \end{bmatrix} + Y_2\left(\kappa_2 \int_0^t X_1(s)X_2(s)ds\right) \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$+ Y_3\left(\kappa_3 \int_0^t X_2(s)ds\right) \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Another example: Viral infection

Let

1. $T =$ viral template.
2. $G =$ viral genome.
3. $S =$ viral structure.
4. $V =$ virus.

Reactions:

R1) $T +$ "stuff" $\stackrel{\kappa_1}{\to} T + G$	$\kappa_1 = 1$
R2) $G \stackrel{\kappa_2}{\to} T$	$\kappa_2 = 0.025$
R3) $T +$ "stuff" $\stackrel{\kappa_3}{\to} T + S$	$\kappa_3 = 1000$
R4) $T \stackrel{\kappa_4}{\to} \emptyset$	$\kappa_4 = 0.25$
R5) $S \stackrel{\kappa_5}{\to} \emptyset$	$\kappa_5 = 2$
R6) $G + S \stackrel{\kappa_6}{\to} V$	$\kappa_6 = 7.5 \times 10^{-6}$

- R. Srivastava, L. You, J. Summers, and J. Yin, J. Theoret. Biol., 2002.
- E. Haseltine and J. Rawlings, J. Chem. Phys, 2002.
- K. Ball, T. Kurtz, L. Popovic, and G. Rempala, Annals of Applied Probability, 2006.
- W. E, D. Liu, and E. Vanden-Eijden, J. Comput. Phys, 2006.

Another example: Viral infection

Stochastic equations for $X = (X_G, X_S, X_T, X_V)$ are

$$X_1(t) = X_1(0) + Y_1\left(\int_0^t X_3(s)ds\right) - Y_2\left(0.025\int_0^t X_1(s)ds\right) - Y_6\left(7.5\times 10^{-6}\int_0^t X_1(s)X_2(s)ds\right)$$

$$X_2(t) = X_2(0) + Y_3\left(1000\int_0^t X_3(s)ds\right) - Y_5\left(2\int_0^t X_2(s)ds\right) - Y_6\left(7.5\times 10^{-6}\int_0^t X_1(s)X_2(s)ds\right)$$

$$X_3(t) = X_3(0) + Y_2\left(0.025\int_0^t X_1(s)ds\right) - Y_4\left(0.25\int_0^t X_3(s)ds\right)$$

$$X_4(t) = X_4(0) + Y_6\left(7.5\times 10^{-6}\int_0^t X_1(s)X_2(s)ds\right).$$

Computational methods

These are continuous time Markov chains!

Simulation/computation should be easy.

The most common simulation methods include

1. Gillespie's Algorithm – Answer where and when independently.
2. The next reaction method of Gibson and Bruck.
3. Each is an example of discrete event simulation.

Numerical methods

Each exact method produces sample paths that can approximate values such as (which I will talk about tomorrow at CWI)

$$\mathbb{E}f(X(t)) \approx \frac{1}{n}\sum_{i=1}^{n} f(X_{[i]}(t))$$

For example,

1. Means – expected virus yield.
2. Variances.
3. Probabilities.

or sensitivities

$$\frac{d}{d\theta}\mathbb{E}f(\theta, X^{\theta}(t)).$$

Numerical methods

Each exact method produces sample paths that can approximate values such as (which I will talk about tomorrow at CWI)

$$\mathbb{E}f(X(t)) \approx \frac{1}{n}\sum_{i=1}^{n} f(X_{[i]}(t))$$

For example,

1. Means – expected virus yield.
2. Variances.
3. Probabilities.

or sensitivities

$$\frac{d}{d\theta}\mathbb{E}f(\theta, X^{\theta}(t)).$$

Problem: solving using these algorithms can be computationally expensive:

1. Each path may require significant number of computational steps.
2. $\rightarrow$ May require significant number of paths.

Solution: Need to use novel stochastic representations to get good methods.

Specific computational problem: Gradient estimation/sensitivity analysis

We have

$$X^\theta(t) = X^\theta(0) + \sum_k Y_k \left(\int_0^t \lambda_k(\theta, X^\theta(s)) ds \right) \zeta_k,$$

with $\theta \in \mathbb{R}^s$, and we define

$$J(\theta) = \mathbb{E} f(\theta, X^\theta(t)].$$

We know how to estimate $J(\theta)$ using Monte Carlo.

Specific computational problem: Gradient estimation/sensitivity analysis

We have

$$X^\theta(t) = X^\theta(0) + \sum_k Y_k \left(\int_0^t \lambda_k(\theta, X^\theta(s)) ds \right) \zeta_k,$$

with $\theta \in \mathbb{R}^s$, and we define

$$J(\theta) = \mathbb{E} f(\theta, X^\theta(t)].$$

We know how to estimate $J(\theta)$ using Monte Carlo.

However, what if we want

$$J'(\theta) = \frac{d}{d\theta} \mathbb{E} f(\theta, X^\theta(t)).$$

Thus, we want to know how sensitive our statistic is to perturbations in θ.
Tells us, for example:

1. Robustness of system to perturbations in parameters.
2. Which parameters we need to estimate well from data, etc.

Specific computational problem: Gradient estimation/sensitivity analysis

We have

$$X^\theta(t) = X^\theta(0) + \sum_k Y_k \left(\int_0^t \lambda_k(\theta, X^\theta(s)) ds \right) \zeta_k,$$

with $\theta \in \mathbb{R}^s$, and we define

$$J(\theta) = \mathbb{E} f(\theta, X^\theta(t)].$$

We know how to estimate $J(\theta)$ using Monte Carlo.

However, what if we want

$$J'(\theta) = \frac{d}{d\theta} \mathbb{E} f(\theta, X^\theta(t)).$$

Thus, we want to know how sensitive our statistic is to perturbations in θ.
Tells us, for example:

1. Robustness of system to perturbations in parameters.
2. Which parameters we need to estimate well from data, etc.

There are multiple methods. We will consider:

- Finite differences.

Finite differencing

This method is pretty straightforward and is therefore used most.

Finite differencing

This method is pretty straightforward and is therefore used most.

Simply note that

$$J'(\theta) = \frac{J(\theta+\epsilon) - J(\theta)}{\epsilon} + O(\epsilon) = \mathbb{E}\left[\frac{f(\theta+\epsilon, X^{\theta+\epsilon}(t)) - f(\theta, X^{\theta}(t))}{\epsilon}\right] + O(\epsilon).$$

Centered differencing reduces bias to $O(\epsilon^2)$.

Finite differencing

This method is pretty straightforward and is therefore used most.

Simply note that

$$J'(\theta) = \frac{J(\theta+\epsilon) - J(\theta)}{\epsilon} + O(\epsilon) = \mathbb{E}\left[\frac{f(\theta+\epsilon, X^{\theta+\epsilon}(t)) - f(\theta, X^{\theta}(t))}{\epsilon}\right] + O(\epsilon).$$

Centered differencing reduces bias to $O(\epsilon^2)$.

The usual finite difference estimator is

$$D_N(\epsilon) = \frac{1}{N}\sum_{i=1}^{N} \frac{f(\theta+\epsilon, X_{[i]}^{\theta+\epsilon}(t)) - f(\theta, X_{[i]}^{\theta}(t))}{\epsilon}$$

Letting $\delta > 0$ be some desired accuracy (for confidence interval), we need N so that

$$\sqrt{Var(D_N(\epsilon))} \leq \delta.$$

Finite differencing

Want

$$\sqrt{Var(D_N(\epsilon))} \leq \delta.$$

with

$$D_N(\epsilon) = \frac{1}{N}\sum_{i=1}^{N} \frac{f(\theta, X_{[i]}^{\theta+\epsilon}(t)) - f(\theta, X_{[i]}^{\theta}(t))}{\epsilon}$$

If paths generated independently, then

$$\begin{aligned} \text{Var}(D_N(\epsilon)) &= N^{-1}\epsilon^{-2}\text{Var}(f(\theta, X_{[i]}^{\theta+\epsilon}(t)) - f(\theta, X_{[i]}^{\theta}(t))) \\ &= O(N^{-1}\epsilon^{-2}), \end{aligned}$$

implying

$$\frac{1}{\sqrt{N}}\frac{1}{\epsilon} = O(\delta) \implies N = O(\epsilon^{-2}\delta^{-2})$$

Terrible. Worse than expectations.

How about common random numbers for variance reduction?

Common random numbers

It's exactly what it sounds like. Reuse the random numbers used in the generation of

$$X_{[i]}^{\theta+\epsilon}(t) \quad \text{and} \quad X_{[i]}^{\theta}(t).$$

Why?

Common random numbers

It's exactly what it sounds like. Reuse the random numbers used in the generation of

$$X_{[i]}^{\theta+\epsilon}(t) \quad \text{and} \quad X_{[i]}^{\theta}(t).$$

Why? Because:

$$\begin{aligned}\text{Var}(f(\theta, X_{[i]}^{\theta+\epsilon}(t)) - f(\theta, X_{[i]}^{\theta+\epsilon}(t))) = \text{Var}(f(\theta, X_{[i]}^{\theta+\epsilon}(t))) + \text{Var}(f(\theta, X_{[i]}^{\theta}(t))) \\ - 2\text{Cov}(f(\theta, X_{[i]}^{\theta+\epsilon}(t)), f(\theta, X_{[i]}^{\theta}(t))).\end{aligned}$$

So, if we can "couple" the random variables, we can get a variance reduction! Sometimes substantial.

Common random numbers

- In the context of Gillespie's algorithm, we simply reuse all the same random numbers (uniforms).
- This can be achieved simply by setting the "seed" of the random number generator before generating $X^{\theta+\epsilon}$ and X^{θ}.

Common random numbers

CRN + Gillespie is good idea.

1. Costs little in terms of implementation.
2. Variance reduction and gains in efficiency can be **huge**.

Thus, it is probably the most common method used today.

But:

- Over time, the processes decouple, often completely.

Can we do better?

Coupling

Using common random numbers in previous fashion is a way of “coupling” the two processes together.

Coupling

Using common random numbers in previous fashion is a way of “coupling” the two processes together.

Is there a natural way to couple processes using random time change? Can we couple the Poisson processes?

Coupling

Using common random numbers in previous fashion is a way of “coupling” the two processes together.

Is there a natural way to couple processes using random time change? Can we couple the Poisson processes?

Answer: yes. Multiple ways. I will show one which works very well.

How do we generate processes simultaneously

Suppose I want to generate:

- A Poisson process with intensity 13.1.
- A Poisson process with intensity 13.

How do we generate processes simultaneously

Suppose I want to generate:

- A Poisson process with intensity 13.1.
- A Poisson process with intensity 13.

- We could let Y_1 and Y_2 be independent, unit-rate Poisson processes, and set

$$Z_{13.1}(t) = Y_1(13.1t),$$
$$Z_{13}(t) = Y_2(13t),$$

Using this representation, these processes are independent and, hence, not coupled.

How do we generate processes simultaneously

Suppose I want to generate:

- A Poisson process with intensity 13.1.
- A Poisson process with intensity 13.

- We could let Y_1 and Y_2 be independent, unit-rate Poisson processes, and set

$$Z_{13.1}(t) = Y_1(13.1t),$$
$$Z_{13}(t) = Y_2(13t),$$

Using this representation, these processes are independent and, hence, not coupled.

The variance of difference is large:

$$\begin{aligned}\mathrm{Var}(Z_{13.1}(t) - Z_{13}(t)) &= \mathrm{Var}(Y_1(13.1t)) + \mathrm{Var}(Y_2(13t)) \\ &= 26.1t.\end{aligned}$$

How do we generate processes simultaneously

Suppose I want to generate:

- A Poisson process with intensity 13.1.
- A Poisson process with intensity 13.

How do we generate processes simultaneously

Suppose I want to generate:

- A Poisson process with intensity 13.1.
- A Poisson process with intensity 13.

- We could let Y_1 and Y_2 be independent unit-rate Poisson processes, and set

$$Z_{13.1}(t) = Y_1(13t) + Y_2(0.1t)$$
$$Z_{13}(t) = Y_1(13t),$$

The variance of difference is much smaller:

$$\mathrm{Var}(Z_{13.1}(t) - Z_{13}(t)) = \mathrm{Var}\left(Y_2(0.1t)\right) = 0.1t.$$

Using a fact: sum of homogeneous Poisson process is again a Poisson process.

How do we generate processes simultaneously

More generally, suppose we want

1. non-homogeneous Poisson process with intensity $f(t)$ and
2. non-homogeneous Poisson process with intensity $g(t)$.

How do we generate processes simultaneously

More generally, suppose we want

1. non-homogeneous Poisson process with intensity $f(t)$ and
2. non-homogeneous Poisson process with intensity $g(t)$.

We can can let Y_1, Y_2, and Y_3 be independent, unit-rate Poisson processes and define

$$Z_f(t) = Y_1\left(\int_0^t f(s) \wedge g(s) ds\right) + Y_2\left(\int_0^t f(s) - (f(s) \wedge g(s))\, ds\right),$$

$$Z_g(t) = Y_1\left(\int_0^t f(s) \wedge g(s) ds\right) + Y_3\left(\int_0^t g(s) - (f(s) \wedge g(s))\, ds\right),$$

How do we generate processes simultaneously

More generally, suppose we want

1. non-homogeneous Poisson process with intensity $f(t)$ and
2. non-homogeneous Poisson process with intensity $g(t)$.

We can can let Y_1, Y_2, and Y_3 be independent, unit-rate Poisson processes and define

$$Z_f(t) = Y_1\left(\int_0^t f(s) \wedge g(s) ds\right) + Y_2\left(\int_0^t f(s) - (f(s) \wedge g(s))\, ds\right),$$
$$Z_g(t) = Y_1\left(\int_0^t f(s) \wedge g(s) ds\right) + Y_3\left(\int_0^t g(s) - (f(s) \wedge g(s))\, ds\right),$$

where we are using that, for example,

$$Y_1\left(\int_0^t f(s) \wedge g(s) ds\right) + Y_2\left(\int_0^t f(s) - (f(s) \wedge g(s))\, ds\right) = Y\left(\int_0^t f(s) ds\right),$$

where Y is a unit rate Poisson process.

Parameter sensitivities.

Couple the processes.

$$X^{\theta+\epsilon}(t) = X^{\theta+\epsilon}(0) + \sum_k Y_{k,1}\left(\int_0^t \lambda_k^{\theta+\epsilon}(X^{\theta+\epsilon}(s)) \wedge \lambda_k^{\theta}(X^{\theta}(s))ds\right)\zeta_k$$
$$+ \sum_k Y_{k,2}\left(\int_0^t \lambda_k^{\theta+\epsilon}(X^{\theta+\epsilon}(s)) - \lambda_k^{\theta+\epsilon}(X^{\theta+\epsilon}(s)) \wedge \lambda_k^{\theta}(X^{\theta}(s))ds\right)\zeta_k$$

$$X^{\theta}(t) = X^{\theta}(0) + \sum_k Y_{k,1}\left(\int_0^t \lambda_k^{\theta+\epsilon}(X^{\theta+\epsilon}(s)) \wedge \lambda_k^{\theta}(X^{\theta}(s))ds\right)\zeta_k$$
$$+ \sum_k Y_{k,3}\left(\int_0^t \lambda_k^{\theta}(X^{\theta}(s)) - \lambda_k^{\theta+\epsilon}(X^{\theta+\epsilon}(s)) \wedge \lambda_k^{\theta}(X^{\theta}(s))ds\right)\zeta_k,$$

Parameter sensitivities.

Theorem

[3] *Suppose* $(X^{\theta+\epsilon}, X^{\theta})$ *satisfy coupling. Then, for any* $T > 0$ *there is a* $C_{T,f} > 0$ *for which*

$$\mathbb{E} \sup_{t \leq T} \left(f(\theta + \epsilon, X^{\theta+\epsilon}(t)) - f(\theta, X^{\theta}(t)) \right)^2 \leq C_{T,f} \epsilon.$$

[3] David F. Anderson, *An Efficient Finite Difference Method for Parameter Sensitivities of Continuous Time Markov Chains*, SIAM: Journal on Numerical Analysis, Vol. 50, No. 5, 2012.

Parameter sensitivities.

Theorem

[3] *Suppose* $(X^{\theta+\epsilon}, X^{\theta})$ *satisfy coupling. Then, for any* $T > 0$ *there is a* $C_{T,f} > 0$ *for which*

$$\mathbb{E} \sup_{t \leq T} \left(f(\theta + \epsilon, X^{\theta+\epsilon}(t)) - f(\theta, X^{\theta}(t)) \right)^2 \leq C_{T,f} \epsilon.$$

This lowers variance of estimator from

$$O(N^{-1} \epsilon^{-2}),$$

to

$$O(N^{-1} \epsilon^{-1}).$$

Lowered by order of magnitude (in ϵ).

Point: a deeper mathematical understanding led to better computational method.

---

[3] David F. Anderson, *An Efficient Finite Difference Method for Parameter Sensitivities of Continuous Time Markov Chains*, SIAM: Journal on Numerical Analysis, Vol. 50, No. 5, 2012.

Analysis

Theorem

Suppose $(X^{\theta+\epsilon}, X^{\theta})$ *satisfy coupling. Then, for any* $T > 0$ *there is a* $C_{T,f} > 0$ *for which*

$$\mathbb{E} \sup_{t \leq T} \left(f(\theta + \epsilon, X^{\theta+\epsilon}(t)) - f(\theta, X^{\theta}(t)) \right)^2 \leq C_{T,f} \epsilon.$$

Proof:

Analysis

Theorem

Suppose $(X^{\theta+\epsilon}, X^{\theta})$ *satisfy coupling. Then, for any* $T > 0$ *there is a* $C_{T,f} > 0$ *for which*

$$\mathbb{E} \sup_{t \leq T} \left(f(\theta + \epsilon, X^{\theta+\epsilon}(t)) - f(\theta, X^{\theta}(t)) \right)^2 \leq C_{T,f} \epsilon.$$

Proof:

Key observation of proof:

$$X^{\theta+\epsilon}(t) - X^{\theta}(t) = M^{\theta,\epsilon}(t) + \int_0^t F^{\theta+\epsilon}(X^{\theta+\epsilon}(s)) - F^{\theta}(X^{\theta}(s)) ds,$$

where "most" of the jumps have vanished.

Now work on Martingale and absolutely continuous part.

Example: gene transcription and translation

$$G \xrightarrow{2} G + M,$$

$$M \xrightarrow{10} M + P,$$

$$M \xrightarrow{k} \emptyset,$$

$$P \xrightarrow{1} \emptyset.$$

Want

$$\frac{\partial}{\partial \theta} \mathbb{E}\left[X^{\theta}_{\text{protein}}(30)\right], \quad \theta \approx 1/4.$$

Example: gene transcription and translation

$$G \xrightarrow{2} G + M,$$

$$M \xrightarrow{10} M + P,$$

$$M \xrightarrow{k} \emptyset,$$

$$P \xrightarrow{1} \emptyset.$$

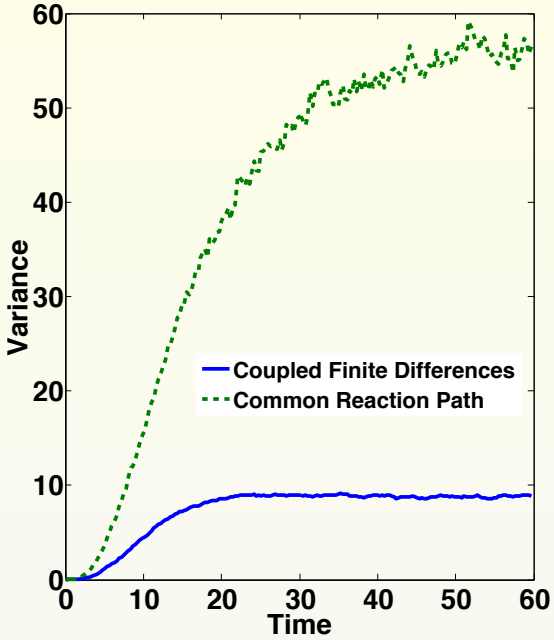
Want

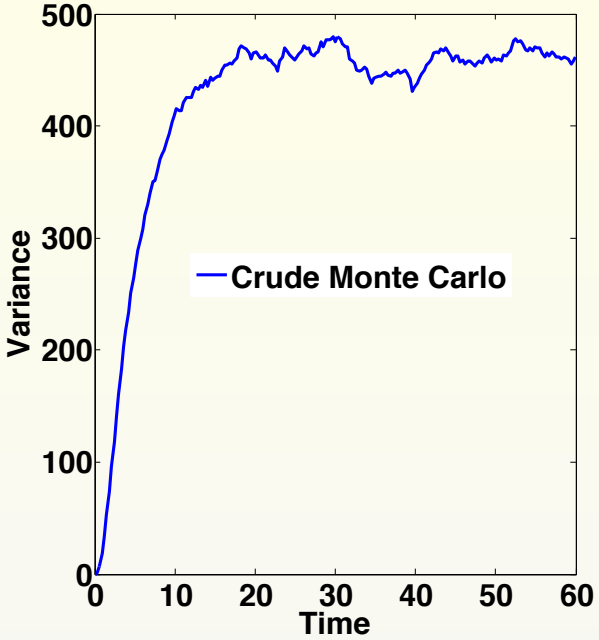
$$\frac{\partial}{\partial \theta} \mathbb{E}\left[X^{\theta}_{\text{protein}}(30)\right], \quad \theta \approx 1/4.$$

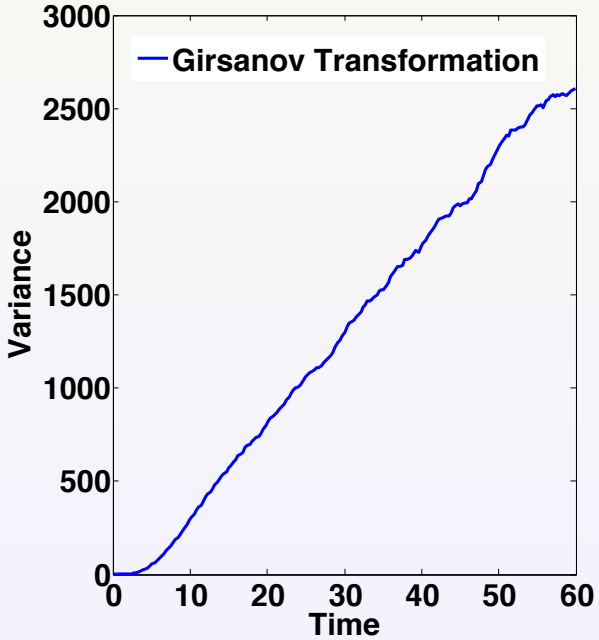
Method	R	95% CI	# updates	CPU Time
Likelihood	689,600	-312.1 $\pm$ 6.0	2.9×10^9	3,506.6 S
CMC	246,000	-319.3 $\pm$ 6.0	2.1×10^9	2,364.8 S
CRP/CRN	25,980	-316.7 $\pm$ 6.0	2.2×10^8	270.9 S
CFD	4,580	-319.9 $\pm$ 6.0	2.0×10^7	29.2 S

Table: Each finite difference method used $\epsilon = 1/40$. The exact value is $J(1/4) = -318.073$.

Comparison from 5,000 samples each with $\epsilon = 1/40$







Example: genetic toggle switch

$$\emptyset \underset{\lambda_2(X)}{\overset{\lambda_1(X)}{\rightleftarrows}} X_1, \quad \emptyset \underset{\lambda_4(X)}{\overset{\lambda_3(X)}{\rightleftarrows}} X_2, \tag{1}$$

with intensity functions

$$\lambda_1(X(t)) = \frac{\alpha_1}{1 + X_2(t)^\beta}, \quad \lambda_2(X(t)) = X_1(t)$$

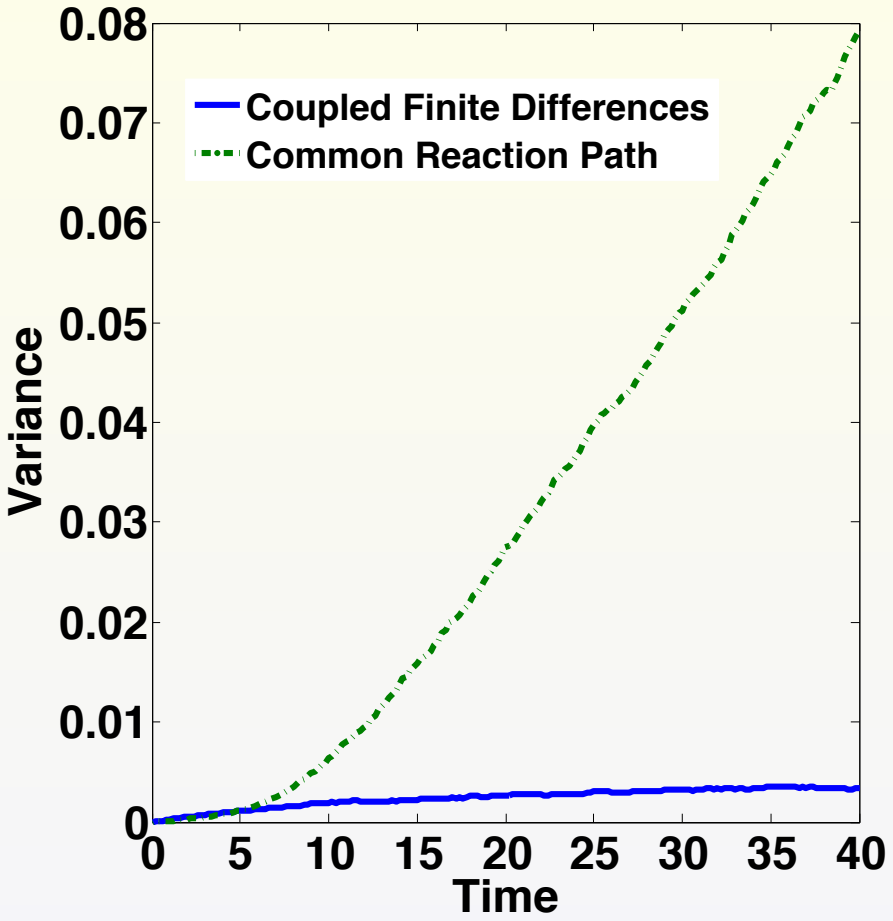
$$\lambda_3(X(t)) = \frac{\alpha_2}{1 + X_1(t)^\gamma}. \quad \lambda_4(X(t)) = X_2(t),$$

and parameter choice

$$\alpha_1 = 50, \quad \alpha_2 = 16, \quad \beta = 2.5, \quad \gamma = 1.$$

- Begin the process with initial condition $[0, 0]$ and
- consider the sensitivity of X_1 as a function of α_1.

Example: genetic toggle switch



(a) Variance to time $T = 40$

Figure: Time plot of the variance of the Coupled Finite Difference estimator versus the Common Reaction Path estimator for the model (1). Each plot was generated using 10,000 sample paths. A perturbation of $\epsilon = 1/10$ was used.

Are these representations only good for simulation? – LLN and ODEs. Tom Kurtz $\sim$ 1970's

Suppose $X_N(t) = O(N)$. Denote concentrations via

$$\overline{X}_N(t) = N^{-1} X_N(t) = O(1).$$

Are these representations only good for simulation? – LLN and ODEs. Tom Kurtz $\sim$ 1970's

Suppose $X_N(t) = O(N)$. Denote concentrations via

$$\overline{X}_N(t) = N^{-1} X_N(t) = O(1).$$

Under mild assumptions, have

$$\lambda_k(X_N(t)) = \lambda_k(N \cdot X_N(t)/N) = N \cdot \overline{\lambda_k}(\overline{X}_N(t)).$$

Are these representations only good for simulation? – LLN and ODEs. Tom Kurtz $\sim$ 1970's

Suppose $X_N(t) = O(N)$. Denote concentrations via

$$\overline{X}_N(t) = N^{-1} X_N(t) = O(1).$$

Under mild assumptions, have

$$\lambda_k(X_N(t)) = \lambda_k(N \cdot X_N(t)/N) = N \cdot \overline{\lambda_k}(\overline{X}_N(t)).$$

$$X_N(t) = X_N(0) + \sum_k Y_k \left(\int_0^t \lambda_k(X_N(s)) ds \right) \xi_k$$

becomes

$$\overline{X}_N(t) = \overline{X}_N(0) + \sum_k N^{-1} Y_k \left(N \int_0^t \overline{\lambda_k}(\overline{X_N}(s)) ds \right) \xi_k$$

Are these representations only good for simulation? – LLN and ODEs. Tom Kurtz $\sim$ 1970's

Suppose $X_N(t) = O(N)$. Denote concentrations via

$$\overline{X}_N(t) = N^{-1} X_N(t) = O(1).$$

Under mild assumptions, have

$$\lambda_k(X_N(t)) = \lambda_k(N \cdot X_N(t)/N) = N \cdot \overline{\lambda_k}(\overline{X}_N(t)).$$

$$X_N(t) = X_N(0) + \sum_k Y_k \left(\int_0^t \lambda_k(X_N(s)) ds \right) \xi_k$$

becomes

$$\overline{X}_N(t) = \overline{X}_N(0) + \sum_k N^{-1} Y_k \left(N \int_0^t \overline{\lambda_k}(\overline{X_N}(s)) ds \right) \xi_k$$

use that

$$\lim_{N \to \infty} \sup_{\{u \leq U\}} \left| N^{-1} Y(Nu) - u \right| = 0,$$

Are these representations only good for simulation? – LLN and ODEs. Tom Kurtz $\sim$ 1970's

Suppose $X_N(t) = O(N)$. Denote concentrations via

$$\overline{X}_N(t) = N^{-1} X_N(t) = O(1).$$

Under mild assumptions, have

$$\lambda_k(X_N(t)) = \lambda_k(N \cdot X_N(t)/N) = N \cdot \overline{\lambda_k}(\overline{X}_N(t)).$$

$$X_N(t) = X_N(0) + \sum_k Y_k \left(\int_0^t \lambda_k(X_N(s))ds \right) \xi_k$$

becomes

$$\overline{X}_N(t) = \overline{X}_N(0) + \sum_k N^{-1} Y_k \left(N \int_0^t \overline{\lambda_k}(\overline{X_N}(s))ds \right) \xi_k$$

use that

$$\lim_{N\to\infty} \sup_{\{u \le U\}} \left| N^{-1} Y(Nu) - u \right| = 0,$$

find $\overline{X}_N(t)$ converges to solution of classical ODE

$$x(t) = x(0) + \sum_k \int_0^t \overline{\lambda}_k(x(s))ds \cdot \xi_k \equiv x(0) + \int_0^t F(X(s))ds.$$

Diffusions? Argument due to Tom Kurtz

Suppose $X_N(t) = O(N)$. Denote concentrations via

$$\overline{X}_N(t) = N^{-1} X_N(t) = O(1).$$

Diffusions? Argument due to Tom Kurtz

Suppose $X_N(t) = O(N)$. Denote concentrations via

$$\overline{X}_N(t) = N^{-1} X_N(t) = O(1).$$

$$X_N(t) = X_N(0) + \sum_k Y_k \left(\int_0^t \lambda_k(X_N(s)) ds \right) \xi_k$$

becomes

$$\overline{X}_N(t) = \overline{X}_N(0) + \sum_k N^{-1} Y_k \left(N \int_0^t \overline{\lambda_k}(\overline{X_N}(s)) ds \right) \xi_k$$

Diffusions? Argument due to Tom Kurtz

Suppose $X_N(t) = O(N)$. Denote concentrations via

$$\overline{X}_N(t) = N^{-1} X_N(t) = O(1).$$

$$X_N(t) = X_N(0) + \sum_k Y_k \left(\int_0^t \lambda_k(X_N(s)) ds \right) \xi_k$$

becomes

$$\overline{X}_N(t) = \overline{X}_N(0) + \sum_k N^{-1} Y_k \left(N \int_0^t \overline{\lambda_k}(\overline{X_N}(s)) ds \right) \xi_k$$

use that

$$\frac{1}{\sqrt{N}} (Y_k(Nu) - Nu) \approx W_k(u)$$

Diffusions? Argument due to Tom Kurtz

Suppose $X_N(t) = O(N)$. Denote concentrations via

$$\overline{X}_N(t) = N^{-1} X_N(t) = O(1).$$

$$X_N(t) = X_N(0) + \sum_k Y_k \left(\int_0^t \lambda_k(X_N(s)) ds \right) \xi_k$$

becomes

$$\overline{X}_N(t) = \overline{X}_N(0) + \sum_k N^{-1} Y_k \left(N \int_0^t \overline{\lambda_k}(\overline{X_N}(s)) ds \right) \xi_k$$

use that

$$\frac{1}{\sqrt{N}} (Y_k(Nu) - Nu) \approx W_k(u)$$

find $\overline{X}_N(t)$ well approximated by chemical Langevin process

$$X(t) = X(0) + \sum_k \zeta_k \int_0^t \lambda_k(X(s)) ds + \frac{1}{\sqrt{N}} \sum_k \int_0^t \sqrt{\lambda_k(X(s))} dW_k(s).$$

or

$$dX(t) = F(X(t))dt + N^{-1/2} \sum_k \zeta_k \sqrt{\lambda_k(X(t))} dW_k(t).$$

Central limit theorem - Kurtz/ Van Kampen

Suppose $X_N(t) = O(N)$. Denote concentrations via

$$\overline{X}_N(t) = N^{-1} X_N(t) = O(1), \quad x(t) = \text{ODE solution.}$$

Let

$$U_N(t) = \sqrt{N}\left(\overline{X}_N(t) - x(t)\right) = \frac{X_n(t) - Nx(t)}{\sqrt{N}}.$$

Central limit theorem - Kurtz/ Van Kampen

Suppose $X_N(t) = O(N)$. Denote concentrations via

$$\overline{X}_N(t) = N^{-1} X_N(t) = O(1), \quad x(t) = \text{ODE solution.}$$

Let

$$U_N(t) = \sqrt{N}\left(\overline{X}_N(t) - x(t)\right) = \frac{X_n(t) - Nx(t)}{\sqrt{N}}.$$

Then,

$$\begin{aligned} U_N(t) &= \sqrt{N}\left(N^{-1}\sum_k \zeta_k \tilde{Y}_k\left(N\int_0^t \overline{\lambda}_k(\overline{X}_n(s))ds\right)\right) \\ &\quad + \sqrt{N}\int_0^t (F(\overline{X}_n(s)) - F(x(s)))ds \\ &\approx \frac{1}{\sqrt{N}}\sum_k \zeta_k \tilde{Y}_k\left(N\int_0^t \overline{\lambda}_k(\overline{X}_n(s))ds\right) + \int_0^t DF(x(s))U_n(s)ds. \end{aligned}$$

Central limit theorem - Kurtz/ Van Kampen

Suppose $X_N(t) = O(N)$. Denote concentrations via

$$\overline{X}_N(t) = N^{-1} X_N(t) = O(1), \quad x(t) = \text{ODE solution.}$$

Let

$$U_N(t) = \sqrt{N}\left(\overline{X}_N(t) - x(t)\right) = \frac{X_n(t) - Nx(t)}{\sqrt{N}}.$$

Then,

$$\begin{aligned} U_N(t) &= \sqrt{N}\left(N^{-1}\sum_k \zeta_k \tilde{Y}_k\left(N\int_0^t \overline{\lambda}_k(\overline{X}_n(s))ds\right)\right) \\ &\quad + \sqrt{N}\int_0^t (F(\overline{X}_n(s)) - F(x(s)))ds \\ &\approx \frac{1}{\sqrt{N}}\sum_k \zeta_k \tilde{Y}_k\left(N\int_0^t \overline{\lambda}_k(\overline{X}_n(s))ds\right) + \int_0^t DF(x(s))U_n(s)ds. \end{aligned}$$

use martingale central limit theorem to show that

$$\frac{1}{\sqrt{N}}\tilde{Y}_k(N\cdot) \Rightarrow W_k(\cdot),$$

get $U_n \Rightarrow U$,

$$U(t) = \sum_k \zeta_k W_k\left(\int_0^t \overline{\lambda}_k(x(s))ds\right) + \int_0^t DF(x(s))U(s)ds$$

Thanks!

References:

1. David F. Anderson, *An Efficient Finite Difference Method for Parameter Sensitivities of Continuous Time Markov Chains*, SIAM: Journal on Numerical Analysis, Vol. 50, No. 5, 2012.

2. David F. Anderson and Thomas G. Kurtz, *Continuous Time Markov Chain Models for Chemical Reaction Networks*, in Design and Analysis of biomolecular circuits, Springer, 2011, Eds. Heinz Koeppl et al.

Funding: NSF-DMS-1009275.