Numerical approximation methods for stochastically modeled biochemical reaction networks

David F. Anderson*

*anderson@math.wisc.edu

Department of Mathematics
University of Wisconsin - Madison

Illinois Math Bio Seminar

February 24th, 2011
Outline

1. Discuss Poisson process and time changes.


3. Numerical methods and error approximation.

4. Mathematical question for today: how do we effectively quantify how well different methods approximate the continuous time Markov chain model.

5. Discuss/use multi-scale nature of biochemical reaction networks.
The models I am going to discuss are a subset of the class of models termed continuous time Markov chains.

The simplest continuous time Markov chain is the Poisson process.
The Poisson process

A Poisson process, \( Y(\cdot) \), is a model for a series of random observations occurring in time.

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

In the figure above, \( Y(t) = 6 \).

Intuition: The unit rate Poisson process is simply the number of points hit when we run along the time frame at rate one.
The Poisson process

A Poisson process, $Y(\cdot)$, is a model for a series of random observations occurring in time.

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

(b) Now, put points down on a line with spacing equal to the $\xi_i$.

Let $Y(t)$ denote the number of points hit by time $t$.

In the figure above, $Y(t) = 6$.

Intuition: The unit rate Poisson process is simply the number of points hit when we run along the time frame at rate one.
The Poisson process

A Poisson process, $Y(\cdot)$, is a model for a series of random observations occurring in time.

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

(b) Now, put points down on a line with spacing equal to the $\xi_i$.

\[ \begin{array}{cccccccc}
  x & x & & x & x & x & x & x \\
\xi_1 & \xi_2 & \xi_3 & \cdots & t \\
\end{array} \]

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

Intuition: The unit rate Poisson process is simply the number of points hit when we run along the time frame at rate one.
The Poisson process

Let

- $Y$ be a unit rate Poisson process.

- Define $Y_\lambda(t) \equiv Y(\lambda t)$,

Then $Y_\lambda$ is a Poisson process with parameter $\lambda$.

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

Thus, we have “changed time” to convert a unit-rate Poisson process to one which has rate $\lambda$. 
The Poisson process

Let

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

Then \( Y_\lambda \) is a Poisson process with parameter \( \lambda \).

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

Thus, we have “changed time” to convert a unit-rate Poisson process to one which has rate \( \lambda \).
The Poisson process

There is no reason $\lambda$ needs to be constant in time, in which case

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

is an inhomogeneous Poisson process.
The Poisson process

There is no reason $\lambda$ needs to be constant in time, in which case

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

is an inhomogeneous Poisson process.

Key property: $Y(T + \Delta) - Y(T) = \text{Poisson}(\Delta)$.
The Poisson process

There is no reason $\lambda$ needs to be constant in time, in which case

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

is an inhomogeneous Poisson process.

Key property: $Y(T + \Delta) - Y(T) = \text{Poisson}(\Delta)$.

Thus

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

There is no reason $\lambda$ needs to be constant in time, in which case

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

is an inhomogeneous Poisson process.

Key property: $Y(T + \Delta) - Y(T) = \text{Poisson}(\Delta)$.

Thus

$$P\{Y_\lambda(t + \Delta t) - Y_\lambda(t) > 0 | F_t\} = 1 - \exp\left\{- \int_t^{t+\Delta t} \lambda(t)dt\right\} \approx \lambda(t)\Delta t.$$ 

Points:

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

2. Will use more complicated time changes of unit-rate processes to build models of interest.
Outline

1. Discuss Poisson process and time changes.


3. Numerical methods and error approximation.

4. Mathematical question for today: how do we effectively quantify how well different methods approximate the continuous time Markov chain model.

5. Discuss/use multi-scale nature of biochemical reaction networks.
Consider the simple system

\[ A + B \rightarrow C \]

where one molecule each of \( A \) and \( B \) is being converted to one of \( C \).
Models of interest

Consider the simple system

\[ A + B \rightarrow C \]

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

Intuition for standard stochastic model:

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

where

- \( \mathcal{F}_t \) represents the information about the system available at time \( t \) and
- \( \kappa \) is a positive constant, the reaction rate constant.
Models of interest

\[ A + B \rightarrow C \]

Simple book-keeping: if \( X(t) = (X_A(t), X_B(t), X_C(t)) \) gives the state at time \( t \),

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

where

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

\[ A + B \to C \]

Simple book-keeping: if \( X(t) = (X_A(t), X_B(t), X_C(t)) \) gives the state at time \( t \),

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

where

\[ R(t) \] is the \# of times the reaction has occurred by time \( t \) and

\[ X(0) \] is the initial condition.

\( R \) is a counting process:

\[ R(0) = 0 \] and

\[ R \] is constant except for jumps of plus one.

Goal: represent \( R(t) \) in terms of Poisson process.
Models of interest

Recall that for $A + B \rightarrow C$ our intuition was

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

This suggests we can model $R(t) = Y(\int_{0}^{t} \kappa X_A(s) X_B(s) ds)$, where $Y$ is a unit-rate Poisson process. Hence

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

This equation uniquely determines $X$ for all $t > 0$. 
Models of interest

Recall that for $A + B \rightarrow C$ our intuition was

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

and that for an inhomogeneous Poisson process with rate $\lambda(t)$ we have

$$P\{Y_\lambda(t + \Delta t) - Y_\lambda(t) > 0 | \mathcal{F}_t\} = 1 - \exp \left\{ - \int_0^{\Delta t} \lambda(t) dt \right\} \approx \lambda(t) \Delta t.$$
Models of interest

Recall that for \( A + B \rightarrow C \) our intuition was

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

and that for an inhomogeneous Poisson process with rate \( \lambda(t) \) we have

\[
P\{Y_\lambda(t + \Delta t) - Y_\lambda(t) > 0 | \mathcal{F}_t\} = 1 - \exp \left\{- \int_0^{\Delta t} \lambda(t) dt \right\} \approx \lambda(t) \Delta t.
\]

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 $A + B \rightarrow C$ our intuition was

$$P\{\text{reaction occurs in } (t, t + \Delta t)|F_t\} \approx \kappa X_A(t)X_B(t)\Delta t,$$

and that for an inhomogeneous Poisson process with rate $\lambda(t)$ we have

$$P\{Y_\lambda(t + \Delta t) - Y_\lambda(t) > 0|F_t\} = 1 - \exp\left\{-\int_0^{\Delta t} \lambda(t)dt\right\} \approx \lambda(t)\Delta t.$$

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 $t > 0$. 
General stochastic models of biochemical reactions

- We consider a network of reactions involving \(d\) chemical species, \(S_1, \ldots, S_d\):

\[
\sum_{i=1}^{d} \nu_{ik} S_i \rightarrow \sum_{i=1}^{d} \nu'_{ik} S_i
\]

- \(\nu_k \in \mathbb{Z}_{\geq 0}^d\): number of molecules of each chemical species consumed in the \(k\)th reaction.

- \(\nu'_k \in \mathbb{Z}_{\geq 0}^d\): number of molecules of each chemical species created in the \(k\)th reaction.

Denote reaction vector as \(\zeta_k = \nu'_k - \nu_k\), so that if \(k\)th reaction occurs at time \(t\), the new state becomes

\[
X(t) = X(t-) + \nu'_k - \nu_k = X(t-) + \zeta_k,
\]
General stochastic models of biochemical reactions

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

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

- $\nu_k \in \mathbb{Z}_{\geq 0}^d$: number of molecules of each chemical species consumed in the $k$th reaction.

- $\nu'_k \in \mathbb{Z}_{\geq 0}^d$: number of molecules of each chemical species created in the $k$th reaction.

Denote reaction vector as $\zeta_k = \nu'_k - \nu_k$, so that if $k$th reaction occurs at time $t$, the new state becomes

$$
X(t) = X(t-) + \nu'_k - \nu_k = X(t-) + \zeta_k,
$$

- The intensity (or propensity) of $k$th reaction is $\lambda_k : \mathbb{Z}_{\geq 0}^d \rightarrow \mathbb{R}$.

- By analogy with before

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

$Y_k$ are independent, unit-rate Poisson processes.
Mass-action kinetics

The standard intensity function chosen is mass-action kinetics:

\[
\lambda_k(x) = \kappa_k \left( \prod_i \nu_i ! \right) \left( \frac{x}{\nu_k} \right) = \kappa_k \prod_i \frac{x_i !}{(x_i - \nu_{ik}) !}.
\]

1. Rate is proportional to the number of distinct subsets of the molecules present that can form inputs for the reaction.
2. This assumes vessel is “well-stirred”.

Example: If \( S_1 \rightarrow \text{anything} \), then \( \lambda_k(x) = \kappa_k x_1 \).

Example: If \( S_1 + S_2 \rightarrow \text{anything} \), then \( \lambda_k(x) = \kappa_k x_1 x_2 \).

Example: If \( 2S_2 \rightarrow \text{anything} \), then \( \lambda_k(x) = \kappa_k x_2(x_2 - 1) \).
Some examples

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

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Intensity</th>
<th>Reaction</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\emptyset \rightarrow A_8$</td>
<td>$4.00 \times 10^0$</td>
<td>$A_6 + A_8 \rightarrow A_9$</td>
<td>$3.62 \times 10^{-4} X_{A_6} X_{A_8}$</td>
</tr>
<tr>
<td>$A_2 \rightarrow A_3$</td>
<td>$7.00 \times 10^{-1} X_{A_2}$</td>
<td>$A_8 \rightarrow \emptyset$</td>
<td>$9.99 \times 10^{-5} X_{A_8}$</td>
</tr>
<tr>
<td>$A_3 \rightarrow A_2$</td>
<td>$1.30 \times 10^{-1} X_{A_3}$</td>
<td>$A_9 \rightarrow A_6 + A_8$</td>
<td>$4.40 \times 10^{-5} X_{A_9}$</td>
</tr>
<tr>
<td>$\emptyset \xrightarrow{A_1} A_2$</td>
<td>$7.00 \times 10^{-3} X_{A_1}$</td>
<td>$\emptyset \rightarrow A_1$</td>
<td>$1.40 \times 10^{-5}$</td>
</tr>
<tr>
<td>stuff + $A_3 \rightarrow A_5 + A_2$</td>
<td>$6.30 \times 10^{-3} X_{A_3}$</td>
<td>$A_1 \rightarrow \emptyset$</td>
<td>$1.40 \times 10^{-6} X_{A_1}$</td>
</tr>
<tr>
<td>stuff + $A_3 \rightarrow A_4 + A_2$</td>
<td>$4.88 \times 10^{-3} X_{A_3}$</td>
<td>$A_7 \xrightarrow{A_4} A_6$</td>
<td>$1.42 \times 10^{-6} X_{A_4} X_{A_7}$</td>
</tr>
<tr>
<td>stuff + $A_3 \rightarrow A_6 + A_2$</td>
<td>$4.88 \times 10^{-3} X_{A_3}$</td>
<td>$A_5 \rightarrow \emptyset$</td>
<td>$1.80 \times 10^{-8} X_{A_5}$</td>
</tr>
<tr>
<td>$A_7 \rightarrow A_2 + A_6$</td>
<td>$4.40 \times 10^{-4} X_{A_7}$</td>
<td>$A_6 \rightarrow \emptyset$</td>
<td>$6.40 \times 10^{-10} X_{A_6}$</td>
</tr>
<tr>
<td>$A_2 + A_6 \rightarrow A_7$</td>
<td>$3.62 \times 10^{-4} X_{A_2} X_{A_6}$</td>
<td>$A_4 \rightarrow \emptyset$</td>
<td>$7.40 \times 10^{-11} X_{A_4}$</td>
</tr>
</tbody>
</table>

---

\(^1\) Hye Won Kang, presentation at SPA in 2007.
Some examples

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

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Intensity</th>
<th>Reaction</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\emptyset \rightarrow A_8$</td>
<td>$4.00 \times 10^0$</td>
<td>$A_6 + A_8 \rightarrow A_9$</td>
<td>$3.62 \times 10^{-4}X_{A_6}X_{A_8}$</td>
</tr>
<tr>
<td>$A_2 \rightarrow A_3$</td>
<td>$7.00 \times 10^{-1}X_{A_2}$</td>
<td>$A_8 \rightarrow \emptyset$</td>
<td>$9.99 \times 10^{-5}X_{A_8}$</td>
</tr>
<tr>
<td>$A_3 \rightarrow A_2$</td>
<td>$1.30 \times 10^{-1}X_{A_3}$</td>
<td>$A_9 \rightarrow A_6 + A_8$</td>
<td>$4.40 \times 10^{-5}X_{A_9}$</td>
</tr>
<tr>
<td>$\emptyset \rightarrow A_2$</td>
<td>$7.00 \times 10^{-3}X_{A_1}$</td>
<td>$\emptyset \rightarrow A_1$</td>
<td>$1.40 \times 10^{-5}$</td>
</tr>
<tr>
<td>stuff $+ A_3 \rightarrow A_5 + A_2$</td>
<td>$6.30 \times 10^{-3}X_{A_3}$</td>
<td>$A_1 \rightarrow \emptyset$</td>
<td>$1.40 \times 10^{-6}X_{A_1}$</td>
</tr>
<tr>
<td>stuff $+ A_3 \rightarrow A_4 + A_2$</td>
<td>$4.88 \times 10^{-3}X_{A_3}$</td>
<td>$A_7 \rightarrow A_6$</td>
<td>$1.42 \times 10^{-6}X_{A_4}X_{A_7}$</td>
</tr>
<tr>
<td>stuff $+ A_3 \rightarrow A_6 + A_2$</td>
<td>$4.88 \times 10^{-3}X_{A_3}$</td>
<td>$A_5 \rightarrow \emptyset$</td>
<td>$1.80 \times 10^{-8}X_{A_5}$</td>
</tr>
<tr>
<td>$A_7 \rightarrow A_2 + A_6$</td>
<td>$4.40 \times 10^{-4}X_{A_7}$</td>
<td>$A_6 \rightarrow \emptyset$</td>
<td>$6.40 \times 10^{-10}X_{A_6}$</td>
</tr>
<tr>
<td>$A_2 + A_6 \rightarrow A_7$</td>
<td>$3.62 \times 10^{-4}X_{A_2}X_{A_6}$</td>
<td>$A_4 \rightarrow \emptyset$</td>
<td>$7.40 \times 10^{-11}X_{A_4}$</td>
</tr>
</tbody>
</table>

Lotka-Volterra predator-prey model: think of $A$ as a prey and $B$ as a predator.

\[
A \xrightarrow{\kappa_1} 2A, \quad A + B \xrightarrow{\kappa_2} 2B, \quad B \xrightarrow{\kappa_3} \emptyset,
\]

\footnote{Hye Won Kang, presentation at SPA in 2007.}
Markov chain models

Path-wise representation

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

Kolmogorov’s forward equation (”Chemical Master Equation”) describes the evolution of the distribution of the state of the system

\[ \frac{d}{dt} P_x(t) = \sum_k \lambda_k(x - \zeta_k) P_{x - \zeta_k}(t) - \sum_k \lambda_k(x) P_x(t), \]

where \( P_x(t) \) is probability \( X(t) = x. \)

Model I’ve described is equivalent to this “chemical master equation model”.
Population Example: Lotka-Volterra predator-prey model

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

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

with $A(0) = B(0) = 1000$ and $\kappa_1 = 2$, $\kappa_2 = .002$, $\kappa_3 = 2$. 
Population Example: Lotka-Volterra predator-prey model

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

\[ A \overset{\kappa_1}{\rightarrow} 2A, \quad A + B \overset{\kappa_2}{\rightarrow} 2B, \quad B \overset{\kappa_3}{\rightarrow} \emptyset, \]

with $A(0) = B(0) = 1000$ and $\kappa_1 = 2$, $\kappa_2 = .002$, $\kappa_3 = 2$.

Deterministic model. Let $x(t) = [A(t), B(t)]^T$.

\[
x(t) = x(0) + \kappa_1 \int_0^t x_1(s)ds \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \kappa_2 \int_0^t x_1(s)x_2(s)ds \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \kappa_3 \int_0^t x_2(s)ds \begin{bmatrix} 0 \\ -1 \end{bmatrix}
\]
Population Example: Lotka-Volterra predator-prey model

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

\[
A \overset{\kappa_1}{\rightarrow} 2A,
A + B \overset{\kappa_2}{\rightarrow} 2B,
B \overset{\kappa_3}{\rightarrow} 0,
\]

with $A(0) = B(0) = 1000$ and $\kappa_1 = 2$, $\kappa_2 = .002$, $\kappa_3 = 2$.

Deterministic model. Let $x(t) = [A(t), B(t)]^T$.

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

Stochastic model. Let $X(t) = [A(t), B(t)]^T$.

\[
X(t) = X(0) + Y_1 \left( \kappa_1 \int_0^t X_1(s)ds \right) \begin{bmatrix} 1 \\ 0 \end{bmatrix} + Y_2 \left( \kappa_2 \int_0^t X_1(s)X_2(s)ds \right) \begin{bmatrix} -1 \\ 1 \end{bmatrix} + Y_3 \left( \kappa_3 \int_0^t X_2(s)ds \right) \begin{bmatrix} 0 \\ -1 \end{bmatrix}
\]
Lotka-Volterra

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

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

with $A(0) = B(0) = 1000$ and $\kappa_1 = 2$, $\kappa_2 = .002$, $\kappa_3 = 2$. 
Pathwise Representations – Random time changes

A representation for path-wise solutions of our model is given by random time-changes of Poisson processes

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

where the \( Y_k \) are independent, unit-rate Poisson processes.

Random time changes have interesting history:

(Wolfgang Doeblin)
Outline

1. Discuss Poisson process and time changes.


3. Numerical methods and error approximation.

4. Mathematical question for today: how do we effectively quantify how well different methods approximate the continuous time Markov chain model.

5. Discuss/use multi-scale nature of biochemical reaction networks.
Numerical methods

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

**Good news.** There are a number of numerical methods that produce statistically exact sample paths:

1. Gillespie’s algorithm.
2. The first reaction method.
3. The next reaction method.
Numerical methods

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

**Good news.** There are a number of numerical methods that produce statistically exact sample paths:

1. Gillespie’s algorithm.
2. The first reaction method.
3. The next reaction method.

For each step of these methods one must find:

(i) the amount of time that passes until the next reaction takes place:

\[ \Delta_n \sim \exp \left( \sum_k \lambda_k(X(t)) \right) \]

(the minimum of exponential RVs)

(ii) which reaction takes place at that time.
**Numerical methods**

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

**Good news.** There are a number of numerical methods that produce statistically exact sample paths:

1. Gillespie’s algorithm.
2. The first reaction method.
3. The next reaction method.

For each step of these methods one must find:

(i) the amount of time that passes until the next reaction takes place:

\[ \Delta_n \sim \exp \left( \sum_k \lambda_k(X(t)) \right) \]

(the minimum of exponential RVs)

(ii) which reaction takes place at that time.

**Bad news.** If \( \sum_k \lambda_k(X(t)) \gg 1 \), then \( \Delta_n \approx \frac{1}{\sum_k \lambda_k(X(t))} \ll 1 \)

* time to produce a single path over an interval [0, T] can be prohibitive.
Explicit “$\tau$-leaping” was developed by Dan Gillespie in an effort to overcome the problem that $\Delta n$ may be prohibitively small.

**Tau-leaping** is essentially an Euler approximation of $\int_0^t \lambda_k(X(s))ds$:

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

$$\approx Z(0) + \sum_k Y_k \left( \lambda_k(Z(0)) \, \tau \right) \zeta_k$$

$$\overset{d}{=} Z(0) + \sum_k \text{Poisson} \left( \lambda_k(Z(0)) \, \tau \right) \zeta_k.$$

---

One “intuitive” representation for $Z(t)$ generated by $\tau$-leaping is

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

where

$$\eta(s) = t_n, \quad \text{if} \quad t_n \leq s < t_{n+1} = t_n + \tau$$

is a step function giving left endpoints of time discretization.
Another algorithm: A midpoint method

For a time discretization $0 = t_0 < t_1 < \cdots < t_N = T$, with $\tau = t_n - t_{n-1}$, let

$$\rho(z) = z + \frac{1}{2} \tau \sum_k \lambda_k(z) \zeta_k,$$

be a “deterministic” midpoint approximation

and let $Z(t)$ solve:

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

where $\eta(s) = t_n$, if $t_n \leq s < t_{n+1} = t_n + \tau$. 
A new algorithm: Weak trapezoidal method

Fixing a $\theta \in (0, 1)$, we define

$$\xi_1 = \frac{1}{2} \frac{1}{\theta(1 - \theta)}$$
and

$$\xi_2 \overset{\text{def}}{=} \frac{1}{2} \left( \frac{(1 - \theta)^2 + \theta^2}{\theta(1 - \theta)} \right).$$

Repeat the following steps until $t_{n+1} > T$, in which we first compute a $\theta$-midpoint $y^*$, and then the new value $Z(t_{n+1})$:

1. Set $y^* = Z(t_n) + \sum_k y_{k,n,1}(\lambda_k(Z(t_n))\theta h)\zeta_k$.

2. Set $Z(t_{n+1}) = y^* + \sum_k y_{k,n,2}\left([\xi_1 \lambda_k(y^*) - \xi_2 \lambda_k(t_n)]^+(1 - \theta)h\right)\zeta_k$.

3. Set $t_{n+1} = t_n + h$.

Notes:

1. If $\theta = 1/2$, then $\xi_1 = 2$ and $\xi_2 = 1$.
2. First step is just Euler's method on partial interval.
Weak trapezoidal method: intuition

Alternative representation for

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

is

\[ X(t) = X(0) + \sum_k \zeta_k \int_0^t \int_0^\infty 1_{[0,\lambda_k(X(s))]}(u) \mu_k(du \times ds), \]

where \( \mu_k \) are independent Poisson random measures with Lebesgue mean measure (area):
Alternative representation for

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

is

\[ X(t) = X(0) + \sum_k \zeta_k \int_0^t \int_0^\infty 1_{[0, \lambda_k(X(s))]}(u) \mu_k(du \times ds), \]

where \( \mu_k \) are independent Poisson random measures with Lebesgue mean measure (area):

1. For disjoint \( A, B \subset \mathbb{R}^2 \),

2. \( \mu_k(A) \) and \( \mu_k(B) \) are Poisson with parameters \( \text{Area}(A) \) and \( \text{Area}(B) \), respectively.
Why this works: $\theta = 1/2$
Why this works: $\theta = 1/2$
Why this works: $\theta = 1/2$
Error analysis

- Let $X(t)$ denote some process and $Z_h(t)$ be an approximation.
- Multiple ways to understand how good an approximation $Z_h$ is to $X$. 

1. Strong error approximation. Bound:
   
   $$\sup_{t \leq T} (E_x^0 | X(t) - Z_h(t)|^{p/2}) \leq Ch^q (E_x^0 \sup_{t \leq T} |X(t) - Z_h(t)|^{p/2})$$

   Common choices are $p = 1, 2$.

2. Weak error approximation. For large class of real valued test functions $f$,
   
   $$\sup_{t \leq T} |E_x^0 f(X(t)) - E_x^0 f(Z_h(t))| \leq Ch^q.$$
Let $X(t)$ denote some process and $Z_h(t)$ be an approximation.

Multiple ways to understand how good an approximation $Z_h$ is to $X$.

1. Strong error approximation. Bound:

$$\sup_{t \leq T} \left( \mathbb{E}_{x_0} |X(t) - Z_h(t)|^p \right)^{1/p} \leq C h^q$$

$$\left( \mathbb{E}_{x_0} \sup_{t \leq T} |X(t) - Z_h(t)|^p \right)^{1/p} \leq C h^q$$

Common choices are $p = 1, 2$. 
Error analysis

- Let $X(t)$ denote some process and $Z_h(t)$ be an approximation.

- Multiple ways to understand how good an approximation $Z_h$ is to $X$.

1. **Strong error approximation.** Bound:

   $$\sup_{t \leq T} \left( \mathbb{E}_{x_0} |X(t) - Z_h(t)|^p \right)^{1/p} \leq Ch^q$$

   $$\left( \mathbb{E}_{x_0} \sup_{t \leq T} |X(t) - Z_h(t)|^p \right)^{1/p} \leq Ch^q$$

   Common choices are $p = 1, 2$.

2. **Weak error approximation.** For large class of real valued test functions $f$, bound

   $$\sup_{t \leq T} \left| \mathbb{E}_{x_0} f(X(t)) - \mathbb{E}_{x_0} f(Z_h(t)) \right| \leq Ch^q.$$
Error analysis

Under the scaling $h \to 0$:

1. Li$^3$ and also Rathinam, Petzold, Cao, and Gillespie$^4$ showed, among other things (implicit methods, etc.), that explicit Euler tau-leaping has a strong error (in the $L^2$ norm) of order $1/2$ and a weak error of order one:

$$\sup_{n \leq N} \sqrt{\mathbb{E}_{X_0} |Z(t_n) - X(t_n)|^2} \leq Ch^{1/2}$$

$$|\mathbb{E}_{X_0} f(Z(T)) - \mathbb{E}_{X_0} f(X(T))| \leq Ch,$$

where $0 = t_0 < t_1 < \cdots < t_N = T$ is a partition of $[0, T]$.

2. The midpoint method has the same order accurate as explicit Euler tau-leaping as $h \to 0$.

---

Error analysis

Under the scaling $h \to 0$:

1. Li\textsuperscript{3} and also Rathinam, Petzold, Cao, and Gillespie\textsuperscript{4} showed, among other things (implicit methods, etc.), that explicit Euler tau-leaping has a strong error (in the $L^2$ norm) of order $1/2$ and a weak error of order one:

$$\sup_{n \leq N} \sqrt{\mathbb{E}_{x_0} |Z(t_n) - X(t_n)|^2} \leq Ch^{1/2}$$

$$|\mathbb{E}_{x_0} f(Z(T)) - \mathbb{E}_{x_0} f(X(T))| \leq Ch,$$

where $0 = t_0 < t_1 < \cdots < t_N = T$ is a partition of $[0, T]$.

2. The midpoint method has the same order accurate as explicit Euler tau-leaping as $h \to 0$.

3. Many examples show midpoint method more accurate than Euler: led to work in “classical scaling.” \textsuperscript{5}


\textsuperscript{5}A., Ganguly, Kurtz, Ann. Applied Prob., 2011
Example

Again think of \( A \) as a prey and \( B \) as a predator.

\[
A \xrightarrow{\kappa_1} 2A, \quad A + B \xrightarrow{\kappa_2} 2B, \quad B \xrightarrow{\kappa_3} \emptyset,
\]

with \( A(0) = B(0) = 1000 \) and \( \kappa_1 = 2, \kappa_2 = .002, \kappa_3 = 2. \)

Letting \( \tau = 1/20 \) and simulating 30,000 sample paths with each method yields the following approximate distributions for \( B(10) \):
Recall, tau-leaping methods are used when \( h \gg \mathbb{E}\Delta_n \), for otherwise an exact method would be performed. Therefore, we should require that

\[
h \gg \frac{1}{\sum_k \lambda_k(X(t))} = \mathbb{E}\Delta_n \quad \text{while} \quad \sum_k \lambda_k(X(t)) \gg 1.
\]

So \( h \to 0 \) does not tell the whole story \cdots what to do?
Recall, tau-leaping methods are used when $h \gg \mathbb{E}\Delta_n$, for otherwise an exact method would be performed. Therefore, we should require that

$$
\frac{1}{\sum_k \lambda_k(X(t))} = \mathbb{E}\Delta_n \quad \text{while} \quad \sum_k \lambda_k(X(t)) \gg 1.
$$

So $h \to 0$ does not tell the whole story ⋯ what to do?

Explicitly take the multi-scale nature of the systems into account.
Let $N \gg 1$. Assume that we are given a model of the form

$$X(t) = X(0) + \sum_k Y_k \left( \int_0^t \lambda'_k(X(s)) ds \right) \zeta_k,$$

where the $\lambda'_k$ are of the form

$$\lambda'_k(x) = \kappa'_k \prod_i \nu_{ik}! \prod_i \binom{x_i}{\nu_{ik}}.$$
Let \( N \gg 1 \). Assume that we are given a model of the form

\[
X(t) = X(0) + \sum_k Y_k \left( \int_0^t \lambda'_k(X(s)) ds \right) \zeta_k,
\]

where the \( \lambda'_k \) are of the form

\[
\lambda'_k(x) = \kappa'_k \prod_i \nu_{ik}! \prod_i \binom{x_i}{\nu_{ik}}.
\]

For each species \( i \), define the normalized abundance

\[
X^N_i(t) = \frac{X_i(t)}{N^{\alpha_i}},
\]

where \( \alpha_i \geq 0 \) should be selected so that \( X^N_i = O(1) \).

Here \( X^N_i \) may be the species number (\( \alpha_i = 0 \)) or the species concentration or something else.
Rate constants, $\kappa'_k$, may also vary over several orders of magnitude.

We write $\kappa'_k = \kappa_k N^\beta_k$ where the $\beta_k$ are selected so that $\kappa_k = O(1)$.

Thus:

1. For a binary reaction $S_i + S_j \rightarrow \ast$ we have $\kappa'_k x_i x_j = N^\beta_k + \alpha_i + \alpha_j \kappa_k z_i z_j$, and we can write $\beta_k + \alpha_i + \alpha_j = \beta_k + \nu_k \cdot \alpha$.

2. We also have for $S_i \rightarrow \ast$ and $2S_i \rightarrow \ast$, respectively, $\kappa'_k x_i = N^\beta_k + \nu_k \cdot \alpha z_i$, $\kappa'_k x_i (x_i - 1) = N^\beta_k + \nu_k \cdot \alpha z_i (z_i - N - \alpha_i)$, with similar expressions for intensities involving higher order reactions.
Rate constants, $\kappa'_k$, may also vary over several orders of magnitude. We write

$$\kappa'_k = \kappa_k N^{\beta_k}$$

where the $\beta_k$ are selected so that $\kappa_k = O(1)$. 

1. For a binary reaction $S_i + S_j \rightarrow \ast$, we can write $\beta_k + \alpha_i + \alpha_j = \beta_k + \nu_k \cdot \alpha$.

2. We also have for $S_i \rightarrow \ast$ and $2S_i \rightarrow \ast$, respectively,

$$\kappa'_k = \kappa_k N^{\beta_k}$$

$$\kappa'_k (x_i - 1) = \kappa_k N^{\beta_k} (z_i - N - \alpha_i),$$

with similar expressions for intensities involving higher order reactions.
Rate constants, $\kappa_k'$, may also vary over several orders of magnitude. We write

$$\kappa_k' = \kappa_k N^{\beta_k}$$

where the $\beta_k$ are selected so that $\kappa_k = O(1)$.

Thus:

1. For a binary reaction $S_i + S_j \rightarrow{*}$

$$\kappa_k' x_i x_j = N^{\beta_k + \alpha_i + \alpha_j} \kappa_k z_i z_j,$$

and we can write

$$\beta_k + \alpha_i + \alpha_j = \beta_k + \nu_k \cdot \alpha.$$
Multiscale nature of biochemical reaction networks

Rate constants, $\kappa'_k$, may also vary over several orders of magnitude. We write

$$\kappa'_k = \kappa_k N^{\beta_k}$$

where the $\beta_k$ are selected so that $\kappa_k = O(1)$.

Thus:

1. For a binary reaction $S_i + S_j \rightarrow \ast$

$$\kappa'_k x_i x_j = N^{\beta_k + \alpha_i + \alpha_j} \kappa_k z_i z_j,$$

and we can write

$$\beta_k + \alpha_i + \alpha_j = \beta_k + \nu_k \cdot \alpha.$$

2. We also have for $S_i \rightarrow \ast$ and $2S_i \rightarrow \ast$, respectively,

$$\kappa'_k x_i = N^{\beta_k + \nu_k \cdot \alpha} z_i, \quad \kappa'_k x_i (x_i - 1) = N^{\beta_k + \nu_k \cdot \alpha} z_i (z_i - N^{-\alpha_i}),$$

with similar expressions for intensities involving higher order reactions.
Mulit-scale nature of biochemical reaction networks

Our model has become

\[ X_i^N(t) = X_i^N(0) + \sum_k N^{-\alpha_i} Y_k \left( N^{\beta_k + \nu_k \cdot \alpha} \int_0^t \lambda_k(X^N(s)) ds \right) \zeta_k, \quad i \in \{1, \ldots, d\}. \]
Mulit-scale nature of biochemical reaction networks

Our model has become

\[ X^N_i(t) = X^N_i(0) + \sum_k N^{-\alpha_i} Y_k \left( N^{\beta_k + \nu_k \cdot \alpha} \int_0^t \lambda_k(X^N(s)) ds \right) \zeta_k, \quad i \in \{1, \ldots, d\}. \]

Remark:

If

\[ \beta_k + \nu_k \cdot \alpha = \alpha_i = 1 \]

for all \( i, k \), then model is

\[ X^N(t) = X^N(0) + \sum_k \frac{1}{N} Y_k \left( N \int_0^t \lambda_k(X^N(s)) ds \right) \zeta_k. \]

which is typically called the classical scaling.

- In this case it is natural to consider \( X^N \) as a vector whose \( i \)th component gives the concentration, in moles per unit volume, of the \( i \)th species.
Mulit-scale nature of biochemical reaction networks

Our model has become

$$X_i^N(t) = X_i^N(0) + \sum_k N^{-\alpha_i} Y_k \left( N^{\beta_k + \nu_k \cdot \alpha} \int_0^t \lambda_k(X^N(s)) ds \right) \zeta_k, \quad i \in \{1, \ldots, d\}.$$ 

Remark:

If

$$\beta_k + \nu_k \cdot \alpha = \alpha_i = 1$$

for all $i, k$, then model is

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

which is typically called the classical scaling.

- In this case it is natural to consider $X^N$ as a vector whose $i$th component gives the concentration, in moles per unit volume, of the $i$th species.
- For large $N$, can use Langevin/diffusion approximation.
Mulit-scale nature of biochemical reaction networks

Our model has become

\[ X_i^N(t) = X_i^N(0) + \sum_k N^{-\alpha_i} Y_k \left( N^{\beta_k + \nu_k \cdot \alpha} \int_0^t \lambda_k(X^N(s)) ds \right) \zeta_k, \quad i \in \{1, \ldots, d\}. \]

Remark:

If

\[ \beta_k + \nu_k \cdot \alpha = \alpha_i = 1 \]

for all \( i, k \), then model is

\[ X^N(t) = X^N(0) + \sum_k \frac{1}{N} Y_k \left( N \int_0^t \lambda_k(X^N(s)) ds \right) \zeta_k. \]

which is typically called the classical scaling.

- In this case it is natural to consider \( X^N \) as a vector whose \( i \)th component gives the concentration, in moles per unit volume, of the \( i \)th species.
- For large \( N \), can use Langevin/diffusion approximation.
- As \( N \to \infty \), converge to ODE model with mass action kinetics.
Mulit-scale nature of biochemical reaction networks

Our model has become

\[ X_i^N(t) = X_i^N(0) + \sum_k N^{-\alpha_i} Y_k \left( N^{\beta_k + \nu_k \cdot \alpha} \int_0^t \lambda_k(X^N(s)) ds \right) \zeta_k, \quad i \in \{1, \ldots, d\}. \]

Remark:

If \( \beta_k + \nu_k \cdot \alpha = \alpha_i = 1 \)

for all \( i, k \), then model is

\[ X^N(t) = X^N(0) + \sum_k \frac{1}{N} Y_k \left( N \int_0^t \lambda_k(X^N(s)) ds \right) \zeta_k. \]

which is typically called the classical scaling.

- In this case it is natural to consider \( X^N \) as a vector whose \( i \)th component gives the concentration, in moles per unit volume, of the \( i \)th species.
- For large \( N \), can use Langevin/diffusion approximation.
- As \( N \to \infty \), converge to ODE model with mass action kinetics.
- It was specifically this scaling that was used in the analyses of Euler and midpoint \( \tau \)-leaping found in A., Ganguly, Kurtz, Ann. Appl. Prob, 2011.
A case for simulation

Our model has become

\[ X_i^N(t) = X_i^N(0) + \sum_k N^{-\alpha_i} Y_k \left( N^{\beta_k + \nu_k \cdot \alpha} \int_0^t \lambda_k(X^N(s)) ds \right) \zeta_k, \quad i \in \{1, \ldots, d\}. \]

In this general setting, methods of model simplification/reduction:

1. Langevin/Diffusion approximations, or

2. linear noise/Van Kampen approximations, or

3. Averaging ideas to reduce system, or

4. ODE/Law of large numbers types of approximations

are exceedingly difficult.

Very active line of current research: Kurtz, Popovic, Kang, etc.
Weak error analysis on normalized processes

\[ X_i^N(t) = X_i^N(0) + \sum_k N^{-\alpha_i} Y_k \left( N^{\beta_k + \nu_k \cdot \alpha_i} \int_0^t \lambda_k(X^N(s))ds \right) \zeta_k, \quad i \in \{1, \ldots, d\}. \]

- Denote exact process as \( X^N \) and approximate by \( Z^N \).

- For \( f \in C_0(\mathbb{R}^d, \mathbb{R}) \), we define the operators \( \mathcal{P}_t \) and \( P_t \) by

\[
\begin{align*}
\mathcal{P}_t f(x) & \overset{\text{def}}{=} \mathbb{E}_x f(X^N(t)) \\
P_t f(x) & \overset{\text{def}}{=} \mathbb{E}_x f(Z^N(t)).
\end{align*}
\]
Weak error analysis on normalized processes

\[ X_i^N(t) = X_i^N(0) + \sum_k N^{-\alpha_i} Y_k \left( N^{\beta_k + \nu_k \cdot \alpha} \int_0^t \lambda_k(X^N(s)) \, ds \right) \zeta_k, \quad i \in \{1, \ldots, d\}. \]

- Denote exact process as \( X^N \) and approximate by \( Z^N \).
- For \( f \in C_0(\mathbb{R}^d, \mathbb{R}) \), we define the operators \( \mathcal{P}_t \) and \( P_t \) by
  \[ \mathcal{P}_t f(x) \overset{\text{def}}{=} \mathbb{E}_x f(X^N(t)) \]
  \[ P_t f(x) \overset{\text{def}}{=} \mathbb{E}_x f(Z^N(t)). \]

Want to understand

\[ \mathcal{P}_t f(x) - P_t f(x) = \mathbb{E}_x f(X^N(t)) - \mathbb{E}_x f(Z^N(t)) \]
Weak error analysis on normalized processes

\[ X_i^N(t) = X_i^N(0) + \sum_k N^{-\alpha_i} Y_k \left( N^{\beta_k + \nu_k \cdot \alpha} \int_0^t \lambda_k(X^N(s)) \, ds \right) \zeta_k, \quad i \in \{1, \ldots, d\}. \]

Define

\[ \gamma \overset{\text{def}}{=} \max \left\{ \beta_k + \nu_k \cdot \alpha - \alpha_i : i,k : \zeta_{ik} \neq 0 \right\}. \tag{1} \]

Then, \( N^\gamma \) should be interpreted as the “time-scale” of the system.
Weak error analysis on normalized processes

\[ X_i^N(t) = X_i^N(0) + \sum_k N^{-\alpha_i} Y_k \left( N^{\beta_k + \nu_k \cdot \alpha} \int_0^t \lambda_k(X^N(s))\,ds \right) \zeta_k, \quad i \in \{1, \ldots, d\}. \]

Define

\[ \gamma \overset{\text{def}}{=} \max_{i,k : \zeta_{ik} \neq 0} \{ \beta_k + \nu_k \cdot \alpha - \alpha_i \}. \] (1)

Then, \( N^\gamma \) should be interpreted as the “time-scale” of the system.

**Example.** If system is

\[ S \overset{100}{\rightarrow} \emptyset, \quad X(0) = 10,000. \]
Weak error analysis on normalized processes

\[ X_i^N(t) = X_i^N(0) + \sum_k N^{-\alpha_i} Y_k \left( N^{\beta_k + \nu_k \cdot \alpha} \int_0^t \lambda_k(X^N(s))ds \right) \zeta_k, \quad i \in \{1, \ldots, d\}. \]

Define

\[ \gamma \overset{\text{def}}{=} \max_{\{i,k : \zeta_{ik} \neq 0\}} \{ \beta_k + \nu_k \cdot \alpha - \alpha_i \}. \] (1)

Then, \( N^\gamma \) should be interpreted as the “time-scale” of the system.

**Example.** If system is

\[ S^{100} \rightarrow \emptyset, \quad X(0) = 10,000. \]

Then, \( N = 10,000, \kappa = 100 = N^{1/2} \). Normalized system is

\[ X^N(t) = 1 - Y \left( N^{1/2} N \int_0^t X_N(s)ds \right) \frac{1}{N}. \]

Here,

\[ \gamma = 1/2, \quad N^\gamma = 100. \]
Weak error analysis on normalized processes

Define $\zeta_k^N$ component-wise

$$\zeta_k^N \overset{\text{def}}{=} N^{-\alpha_i} \zeta_i.$$

Model is

$$X^N(t) = X^N(0) + \sum_k Y_k \left( N^{\gamma_k + \nu_k} \int_0^t \lambda_k(X^N(s)) ds \right) \zeta_k^N.$$
Weak error analysis on normalized processes

Define $\zeta_N$ component-wise

$$\zeta_N^{ik} \overset{\text{def}}{=} N^{-\alpha_i} \zeta_{ik}.$$  

Model is

$$X^N(t) = X^N(0) + \sum_k Y_k \left(N^\gamma N^{\beta_k + \nu_k \cdot \alpha - \gamma} \int_0^t \lambda_k(X^N(s))ds\right) \zeta_N^k.$$  

Define a “directional derivative” in direction $\zeta_N^k$:

$$\nabla_k^N f(x) \overset{\text{def}}{=} N^{\beta_k + \nu_k \cdot \alpha - \gamma} (f(x + \zeta_N^k) - f(x)).$$
Weak error analysis on normalized processes

Define $\zeta^N_k$ component-wise

$$\zeta^N_{ik} \overset{\text{def}}{=} N^{-\alpha_i} \zeta_{ik}. $$

Model is

$$X^N(t) = X^N(0) + \sum_k Y_k \left( N^{-\gamma} N^{\beta_k + \nu_k \cdot \alpha - \gamma} \int_0^t \lambda_k(X^N(s)) ds \right) \zeta^N_k. $$

Define a “directional derivative” in direction $\zeta^N_k$:

$$\nabla^N_k f(x) \overset{\text{def}}{=} N^{\beta_k + \nu_k \cdot \alpha - \gamma} (f(x + \zeta^N_k) - f(x)). $$

Now define a “$j$-norm”:

$$\|f\|_{j}^{\nabla^N} = \sup \left\{ \left\| \left( \prod_{i=1}^{p} \nabla^N_{\ell_i} \right) f \right\|_{\infty}, 1 \leq \ell_i \leq m, p \leq j \right\}$$
Weak error analysis on normalized processes

Define $\zeta_k^N$ component-wise

$$\zeta_{ik}^N \overset{\text{def}}{=} N^{-\alpha_i} \zeta_{ik}.$$  

Model is

$$X^N(t) = X^N(0) + \sum_k Y_k \left( N^{\gamma} N^{\beta_k + \nu_k \cdot \alpha - \gamma} \int_0^t \lambda_k(X^N(s)) ds \right) \zeta_k^N.$$  

Define a “directional derivative” in direction $\zeta_k^N$:

$$\nabla_k^N f(x) \overset{\text{def}}{=} N^{\beta_k + \nu_k \cdot \alpha - \gamma} (f(x + \zeta_k^N) - f(x)).$$  

Now define a “$j$-norm”:

$$\|f\|_{\nabla^N}^j = \sup \left\{ \left\| \left( \prod_{i=1}^p \nabla_{\ell_i}^N \right) f \right\|_{\infty} , 1 \leq \ell_i \leq m, p \leq j \right\}$$  

Finally, define, for any operator $Q : C_0(\mathbb{R}^d, \mathbb{R}) \rightarrow C_0(\mathbb{R}^d, \mathbb{R})$,

$$\|Q\|_{\nabla^N}^j = \sup_{f \in C_0^j} \frac{\|Qf\|_{\nabla^N}^j}{\|f\|_{\nabla^N}^j}.$$
Weak error analysis on normalized processes

We want

\[ |\mathbb{E}_x f(X^N(t)) - \mathbb{E}_x f(Z^N(t))| \leq \|P_t f - P_t f\|_{\nabla N} \leq \|P_t f - P_t f\|_{\nabla N} \rightarrow 0 \|f\|_{\nabla N} \]

Want: For \( t \leq T \) \( \|P_t f - P_t f\|_{\nabla N} \rightarrow 0 = O(h^q) \) for the different methods.
We want

\[ |\mathbb{E}_x f(X^N(t)) - \mathbb{E}_x f(Z^N(t))| \leq \|P_t f - P_t f\|_\infty \]
Weak error analysis on normalized processes

We want

\[ |\mathbb{E}_x f(X^N(t)) - \mathbb{E}_x f(Z^N(t))| \leq \|P_t f - P_t f\|_\infty = \|(P_t - P_t)f\|_{\nabla^N} \]
We want

\[ |\mathbb{E}_x f(X^N(t)) - \mathbb{E}_x f(Z^N(t))| \leq \|P_t f - P_t f\|_\infty = \|\mathcal{P}_t - \mathcal{P}_t\|^\nabla_N \leq \|\mathcal{P}_t - \mathcal{P}_t\|^\nabla_{N,0} \leq \|\mathcal{P}_t - \mathcal{P}_t\|_{m \to 0} \|f\|^{\nabla_N}. \]

**Want:** For \( t \leq T \)

\[ \|\mathcal{P}_t - \mathcal{P}_t\|^\nabla_{N,0} = O(h^q) \]

for the different methods.
Weak error analysis

\[ P_t f(x) \overset{\text{def}}{=} \mathbb{E}_x f(X^N(t)), \quad P_t f(x) \overset{\text{def}}{=} \mathbb{E}_x f(Z^N(t)). \]

Theorem (A., Koyama – 2011) For any \( n, m \geq 0, \) and \( h > 0 \) (think: \( T = nh \))

\[
\| P_h^n - P_{nh} \|_{\nabla^N m \to 0} = O(n \| P_h - P_h \|_{\nabla^N m \to 0} \max_{\ell \in \{1, \ldots, n\}} \{ \| P_{\ell h} \|_{\nabla^N m \to m} \})
\]

Points:

1. Last part is independent of the method.
   Just depends on behavior of exact process.

2. Rest says:

   one step (local error) is \( O(h^{q+1}) \) \( \implies \) global error is \( O(h^q) \).

\[ \implies \text{Can focus on local, one step error.} \]
Proof

Note that

\[ \| P_h f \|_0 = \sup_x \mathbb{E}_x f(Z^N(h)) \leq \| f \|_0. \]

Thus, \( P_h \) is a contraction, i.e. \( \| P_h \|_{0 \to 0} \leq 1 \).
Weak error analysis

Proof
Note that
\[ \| P_h f \|_0 = \sup_x |\mathbb{E}_x f(Z^N(h))| \leq \|f\|_0. \]

Thus, \( P_h \) is a contraction, i.e. \( \| P_h \|_{0 \to 0} \leq 1 \). With this in mind

\[
\|(P^n_h - \mathcal{P}_{nh}) f\|_0 = \left\| \sum_{j=1}^{n} (P_h^j \mathcal{P}_{h(n-j)} - P_h^{j-1} \mathcal{P}_{h(n-j+1)}) f \right\|_0
\leq \sum_{j=1}^{n} \| P_h^{j-1} (P_h - \mathcal{P}_h) \mathcal{P}_{h(n-j)} f \|_0 \]

(2)

\[
\leq \sum_{j=1}^{n} \| P_h^{j-1} \|_{0 \to 0} \| P_h - \mathcal{P}_h \|_{m \to 0} \| \mathcal{P}_{h(n-j)} \|_{m \to m} \| f \|_{m} \]

Since \( P_h \) is a contraction, i.e. \( \| P_h \|_{0 \to 0} \leq 1 \), the result is shown.
One step error analyses

**Theorem** (A., Koyama – 2011) For forward Euler, if $h < N^{-\gamma}$

$$\|P_h - \mathcal{P}_h\|_{2 \rightarrow 0}^{\mathcal{V}^N} \in O(N^{2\gamma}h^2).$$

**Corollary** If $Z^N$ is generated via forward Euler, then

$$\sup_{t \leq T} |\mathbb{E}_x f(X^N(t)) - \mathbb{E}_x f(Z^N(t))| = O(N^{2\gamma}h).$$
One step error analyses

Define

\[ \eta_k \overset{\text{def}}{=} \min\{\alpha_i : \zeta_{ik}^N \neq 0\} . \]

Example: if \( \zeta_k^N = \begin{bmatrix} N^{-2} & 0 & -N^{-1} \\ 0 & -N^{-1} & N^{-1/2} \end{bmatrix} \), then \( \eta_k = 1/2 \). Thus, \( |\zeta_k| = O(N^{-\eta_k}) \).
One step error analyses

Define

$$\eta_k \overset{\text{def}}{=} \min\{\alpha_i : \zeta_{ik} \neq 0\}.$$  

Example: if $$\zeta_k^N = \begin{bmatrix} N^{-2} \\ 0 \\ -N^{-1} \\ N^{-1/2} \end{bmatrix}$$, then $$\eta_k = 1/2$$. Thus, $$|\zeta_k| = O(N^{-\eta_k})$$.

Theorem (A., Koyama – 2011) For midpoint method, if $$h < N^{-\gamma}$$

$$\|P_h - P_h^{\nabla N}\|_{3 \to 0} = O(N^{3\gamma} h^3 + N^{2\gamma - \min\{\eta_k\}} h^2).$$

Corollary If $$Z^N$$ is generated via midpoint method, then

$$\sup_{t \leq T} \left| \mathbb{E}_x f(X^N(t)) - \mathbb{E}_x f(Z^N(t)) \right| = O(N^{3\gamma} h^2 + N^{2\gamma - \min\{\eta_k\}} h).$$
One step error analyses

Define

\[ \eta_k \overset{\text{def}}{=} \min \{ \alpha_i : \zeta_{ik}^N \neq 0 \} . \]

Example: if \( \zeta_k^N = \begin{bmatrix} N^{-2} \\ 0 \\ -N^{-1} \\ N^{-1/2} \end{bmatrix} \), then \( \eta_k = 1/2 \). Thus, \( |\zeta_k| = O(N^{-\eta_k}) \).

Theorem (A., Koyama – 2011) For midpoint method, if \( h < N^{-\gamma} \)

\[ \| P_h - \mathcal{P}_h \|_{3 \to 0} = O(N^{3\gamma} h^3 + N^{2\gamma - \min\{\eta_k\}} h^2) . \]

Corollary If \( Z^N \) is generated via midpoint method, then

\[ \sup_{t \leq T} \left| \mathbb{E}_x f(X^N(t)) - \mathbb{E}_x f(Z^N(t)) \right| = O(N^{3\gamma} h^2 + N^{2\gamma - \min\{\eta_k\}} h) . \]

Notes

1. In classical scaling, \( \eta_k \equiv 1 \), and \( \gamma = 0 \). Further, \( h \gg 1/N \). So, error is \( O(h^2) \).
2. In other scalings, can by \( O(h) \).
3. \( O(h^2) \) if \( h \gg N^{1-\gamma - \min\{\eta_k\}} \).
One step error analyses

**Theorem** (A., Koyama – 2011) For weak trapezoidal, if $h < N^{-\gamma}$

$$\| P_h - P_h \|_{3 \to 0}^{\nabla^N} \in O(N^{3\gamma} h^3).$$

**Corollary** If $Z^N$ is generated via weak trapezoidal, then

$$\sup_{t \leq T} |\mathbb{E}_x f(X^N(t)) - \mathbb{E}_x f(Z^N(t))| = O(N^{3\gamma} h^2).$$
Flavor of proof: Euler’s method

Generator for any Markov process $X$ is defined to be operator $A$ satisfying

$$Af(x) = \lim_{h \to 0} \frac{\mathbb{E}_x f(X(h)) - f(x)}{h}.$$
Flavor of proof: Euler’s method

Generator for any Markov process $X$ is defined to be operator $\mathcal{A}$ satisfying

$$\mathcal{A}f(x) = \lim_{h \to 0} \frac{\mathbb{E}_x f(X(h)) - f(x)}{h}.$$ 

Our model is

$$X^N(t) = X^N(0) + \sum_k Y_k \left( N^\gamma N^{\beta_k + \nu_k \cdot \alpha - \gamma} \int_0^t \lambda_k(X^N(s)) \, ds \right) \zeta_k^N.$$ 

Generator for our process is operator $\mathcal{A}^N$

$$\mathcal{A}^N f(x) = N^\gamma \sum_k N^{\beta_k + \nu_k \cdot \alpha - \gamma} \lambda_k(x)(f(x + \zeta_k^N) - f(x)).$$
Flavor of proof: Euler’s method

Generator for any Markov process $X$ is defined to be operator $A$ satisfying

$$Af(x) = \lim_{h \to 0} \frac{\mathbb{E}_x f(X(h)) - f(x)}{h}.$$ 

Our model is

$$X^N(t) = X^N(0) + \sum_k Y_k \left( N^\gamma N^{\beta_k + \nu_k \cdot \alpha - \gamma} \int_0^t \lambda_k(X^N(s)) ds \right) \zeta_k^N.$$ 

Generator for our process is operator $A^N$

$$A^N f(x) = N^\gamma \sum_k N^{\beta_k + \nu_k \cdot \alpha - \gamma} \lambda_k(x) (f(x + \zeta_k^N) - f(x)).$$

Recalling that we defined

$$\nabla_k^N f(x) \overset{\text{def}}{=} N^{\beta_k + \nu_k \cdot \alpha - \gamma} (f(x + \zeta_k^N) - f(x)).$$

it is

$$A^N f(x) = N^\gamma \sum_k \lambda_k(x) \nabla_k^N f(x) = N^\gamma (\lambda \cdot \nabla^N) f(x)$$
Flavor of proof: Euler’s method

Generators useful

$$\mathbb{E}_x f(X^N(t)) = f(x) + \mathbb{E}_x \int_0^t A^N f(X^N(s)) ds$$

Thus, they are right analytical tool for us to understand $\mathbb{E}_x f(X^N(t))$. 
Flavor of proof: Euler’s method

Euler’s method solves

$$Z^N(h) = x_0 + \sum_k Y_k \left( N^\gamma N^{\beta_k + \nu_k \cdot \alpha - \gamma} \int_0^h \lambda_k(x_0) \, ds \right) \zeta_k^N.$$ 

Generator of process is

$$\mathcal{B}^N f(x) \overset{\text{def}}{=} N^\gamma \sum_k N^{\beta_k + \nu_k \cdot \alpha - \gamma} \lambda_k(x_0) (f(x + \zeta_k^N) - f(x)) = N^\gamma (\lambda(x_0) \cdot \nabla^N) f(x)$$

$$\mathbb{E}_{x_0} f(Z^N(h)) = f(x_0) + \mathbb{E}_{x_0} \int_0^h \mathcal{B}^N f(Z^N(s)) \, ds.$$
Flavor of proof: Euler’s method

\[ \mathbb{E}_{x_0} f(X^N(h)) = f(x_0) + \mathbb{E}_{x_0} \int_0^h A^N f(X^N(s)) ds \]

\[ = f(x_0) + hA^N f(x_0) + \mathbb{E}_{x_0} \int_0^h \int_0^s (A^N)^2 f(X^N(r)) dr \]

\[ = f(x_0) + hA^N f(x_0) + \frac{h^2}{2} (A^N)^2 f(x_0) + O(N^{3\gamma} \| f \|_{\nabla^N} h^3) \]
Flavor of proof: Euler’s method

\[ \mathbb{E}_{x_0} f(X^N(h)) = f(x_0) + \mathbb{E}_{x_0} \int_0^h A^N f(X^N(s)) \, ds \]

\[ = f(x_0) + hA^N f(x_0) + \mathbb{E}_{x_0} \int_0^h \int_0^s (A^N)^2 f(X^N(r)) \, dr \]

\[ = f(x_0) + hA^N f(x_0) + \frac{h^2}{2} (A^N)^2 f(x_0) + O(N^{3\gamma} \|f\|_3 \nabla^N h^3) \]

and

\[ \mathbb{E}_{x_0} f(Z^N(h)) = f(x_0) + hB^N f(x_0) + \frac{h^2}{2} (B^N)^2 f(x_0) + O(N^{3\gamma} \|f\|_3 \nabla^N h^3), \]

Now do lots of computations and compare. \(\Box\)
Flavor of proof: other methods

1. **Midpoint method simply has different generator** $B^N$. Repeat basic analysis.
Flavor of proof: other methods

1. **Midpoint method simply has different generator $B^N$.** Repeat basic analysis.

2. Weak trapezoidal is harder because generator changes half-way through time step. Instead use that

$$\mathbb{E}_{x_0} f(Z(h)) = \mathbb{E}_{x_0} [\mathbb{E}_{x_0} [f(Z(h)) | y^*]].$$

*i.* Work on inner expectation with one operator, and then

*ii.* work on outer expectation with different operator.
Example

Consider

\[ A \equiv B \equiv C, \]

with each \( \kappa_k \equiv 1 \) and \( X(0) = [1000, 1000, 4000] \). Log-Log plot of \( |\mathbb{E}Z_C(1) - \mathbb{E}X_C(1)| \):

Slopes:

\[ \text{Euler} = .9957, \quad \text{Midpoint} = 2.7, \quad \text{Weak Trap} = 2.32. \]
Consider

\[ A \leftrightarrow B \leftrightarrow C, \]

with each \( \kappa_k \equiv 1 \) and \( X(0) = [1000, 1000, 4000] \). Log-Log plot of second moments:
Consider

\[ A \leftrightarrow B \leftrightarrow C, \]

with each \( \kappa_k \equiv 1 \) and \( X(0) = [1000, 1000, 4000] \). Log-Log plot of

\[ | \text{Var}(Z_{C(1)}) - \text{Var}(X_{C(1)}) | : \]

Slopes:

Euler = .9831, \hspace{1em} Midpoint = 1.0525, \hspace{1em} Weak Trap = 2.2291.
Careful!

For Midpoint method in classical scaling:

\[ E(\frac{Z}{N})^2 - E(\frac{X}{N})^2 \approx c_1 h^2 + c_2 \frac{1}{N} h \]

\[ E(\frac{Z}{N}) - E(\frac{X}{N}) \approx c_3 h^2 + c_4 \frac{1}{N} h \]

Thus,

\[ \text{Var}(X) - \text{Var}(Z) \approx (c_1 h^2 N^2 - c_2 h N) - 2(c_3 h^2 N + c_4 h) E[Z], \]

which could be (and was on previous example) \( O(hN) \).
Take home messages

1. Simulation of continuous time Markov chains is easy…..unless it’s not.

2. Calculating error estimates must be done with care.

3. Incorporating the natural scales of the process plays an important role.
Take home messages

1. Simulation of continuous time Markov chains is easy.....unless it’s not.

2. Calculating error estimates must be done with care.

3. Incorporating the natural scales of the process plays an important role.

Thank you!