An Introduction to Stochastic Simulation

Stephen Gilmore

Laboratory for Foundations of Computer Science
School of Informatics
University of Edinburgh

PASTA workshop, London, 29th June 2006

Background

- The modelling of chemical reactions using deterministic rate laws has proven extremely successful in both chemistry and biochemistry for many years.

Background

- The modelling of chemical reactions using deterministic rate laws has proven extremely successful in both chemistry and biochemistry for many years.
- This deterministic approach has at its core the law of mass action, an empirical law giving a simple relation between reaction rates and molecular component concentrations.

Background

- The modelling of chemical reactions using deterministic rate laws has proven extremely successful in both chemistry and biochemistry for many years.
- This deterministic approach has at its core the law of mass action, an empirical law giving a simple relation between reaction rates and molecular component concentrations.
- Given knowledge of initial molecular concentrations, the law of mass action provides a complete picture of the component concentrations at all future time points.

Background: Law of Mass Action

- The law of mass action considers chemical reactions to be macroscopic under convective or diffusive stirring, continuous and deterministic.

Background: Law of Mass Action

- The law of mass action considers chemical reactions to be macroscopic under convective or diffusive stirring, continuous and deterministic.
- These are evidently simplifications, as it is well understood that chemical reactions involve discrete, random collisions between individual molecules.

Background: Law of Mass Action

- The law of mass action considers chemical reactions to be macroscopic under convective or diffusive stirring, continuous and deterministic.
- These are evidently simplifications, as it is well understood that chemical reactions involve discrete, random collisions between individual molecules.
- As we consider smaller and smaller systems, the validity of a continuous approach becomes ever more tenuous.

Background: Law of Mass Action

- The law of mass action considers chemical reactions to be macroscopic under convective or diffusive stirring, continuous and deterministic.
- These are evidently simplifications, as it is well understood that chemical reactions involve discrete, random collisions between individual molecules.
- As we consider smaller and smaller systems, the validity of a continuous approach becomes ever more tenuous.
- As such, the adequacy of the law of mass action has been questioned for describing intracellular reactions.

Background: Application of Stochastic Models

Arguments for the application of stochastic models for chemical reactions come from at least three directions, since the models:

1. take into consideration the discrete character of the quantity of components and the inherently random character of the phenomena;

Background: Application of Stochastic Models

Arguments for the application of stochastic models for chemical reactions come from at least three directions, since the models:

1. take into consideration the discrete character of the quantity of components and the inherently random character of the phenomena;
2. are in accordance with the theories of thermodynamics and stochastic processes; and

Background: Application of Stochastic Models

Arguments for the application of stochastic models for chemical reactions come from at least three directions, since the models:

1. take into consideration the discrete character of the quantity of components and the inherently random character of the phenomena;
2. are in accordance with the theories of thermodynamics and stochastic processes; and
3. are appropriate to describe "small systems" and instability phenomena.

Background: Simulation

- Stochastic simulation methods

Background: Simulation

- Stochastic simulation methods
- Nothing new?

Background: Simulation

- Stochastic simulation methods
- Nothing new?
- Not just discrete-event simulation

Background: Simulation

- Stochastic simulation methods
- Nothing new?
- Not just discrete-event simulation
- Specialist method well-suited to large-scale systems

Acknowledgements

H. Bolouri, J.T. Bradley, J. Bruck, K. Burrage, M. Calder, Y. Cao, K.-H. Cho, A.J. Duguid, C. van Gend, M.A. Gibson, D.T. Gillespie, J. Hillston, M. Khammash, W. Kolch, U. Kummer, D. Orrell, L. Petzold, S. Ramsey, H.E. Samad, S. Schnell, N.T. Thomas, T.E. Turner, M. Ullah, O. Wolkenhauer

Outline

Outline

1. The deterministic and stochastic approaches
2. Stochastic simulation algorithms
3. Comparing stochastic simulation and ODEs
4. Modelling challenges

Deterministic: The law of mass action

The fundamental empirical law governing reaction rates in biochemistry is the law of mass action.

Deterministic: The law of mass action

The fundamental empirical law governing reaction rates in biochemistry is the law of mass action.

This states that for a reaction in a homogeneous, free medium, the reaction rate will be proportional to the concentrations of the individual reactants involved.

Deterministic: Michaelis-Menten kinetics

Consider the simple Michaelis-Menten reaction

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_2} E + P$$

Deterministic: Michaelis-Menten kinetics

Consider the simple Michaelis-Menten reaction

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_2} E + P$$

For example, we have

$$\frac{\mathrm{d}C}{\mathrm{d}t} = k_1 SE - (k_{-1} + k_2)C$$

Deterministic: Michaelis-Menten kinetics

Consider the simple Michaelis-Menten reaction

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_2} E + P$$

For example, we have

$$\frac{\mathrm{d}C}{\mathrm{d}t} = k_1 SE - (k_{-1} + k_2)C$$

Hence, we can express any chemical system as a collection of coupled non-linear first order differential equations.

Stochastic: Random processes

- Whereas the deterministic approach outlined above is essentially an empirical law, derived from *in vitro* experiments, the stochastic approach is far more physically rigorous.
- Fundamental to the principle of stochastic modelling is the idea that molecular reactions are essentially random processes; it is impossible to say with complete certainty the time at which the next reaction within a volume will occur.

Stochastic: Random processes

- Whereas the deterministic approach outlined above is essentially an empirical law, derived from *in vitro* experiments, the stochastic approach is far more physically rigorous.
- Fundamental to the principle of stochastic modelling is the idea that molecular reactions are essentially random processes; it is impossible to say with complete certainty the time at which the next reaction within a volume will occur.

Stochastic: Random processes

- Whereas the deterministic approach outlined above is essentially an empirical law, derived from *in vitro* experiments, the stochastic approach is far more physically rigorous.
- Fundamental to the principle of stochastic modelling is the idea that molecular reactions are essentially random processes; it is impossible to say with complete certainty the time at which the next reaction within a volume will occur.

Stochastic: Predictability of macroscopic states

- In macroscopic systems, with a large number of interacting molecules, the randomness of this behaviour averages out so that the overall macroscopic state of the system becomes highly predictable.
- It is this property of large scale random systems that enables a deterministic approach to be adopted; however, the validity of this assumption becomes strained in *in vivo* conditions as we examine small-scale cellular reaction environments with limited reactant populations.

Stochastic: Predictability of macroscopic states

- In macroscopic systems, with a large number of interacting molecules, the randomness of this behaviour averages out so that the overall macroscopic state of the system becomes highly predictable.
- It is this property of large scale random systems that enables a deterministic approach to be adopted; however, the validity of this assumption becomes strained in *in vivo* conditions as we examine small-scale cellular reaction environments with limited reactant populations.

Stochastic: Predictability of macroscopic states

- In macroscopic systems, with a large number of interacting molecules, the randomness of this behaviour averages out so that the overall macroscopic state of the system becomes highly predictable.
- It is this property of large scale random systems that enables a deterministic approach to be adopted; however, the validity of this assumption becomes strained in *in vivo* conditions as we examine small-scale cellular reaction environments with limited reactant populations.

Stochastic: Propensity function

As explicitly derived by Gillespie, the stochastic model uses basic Newtonian physics and thermodynamics to arrive at a form often termed the propensity function that gives the probability a_μ of reaction μ occurring in time interval $(t, t + \mathrm{d}t)$.

$$a_\mu \mathrm{d}t = h_\mu c_\mu \mathrm{d}t$$

where the M reaction mechanisms are given an arbitrary index μ $(1 \leq \mu \leq M)$, h_μ denotes the number of possible combinations of reactant molecules involved in reaction μ, and c_μ is a stochastic rate constant.

Stochastic: Fundamental hypothesis

The rate constant c_μ is dependent on the radii of the molecules involved in the reaction, and their average relative velocities – a property that is itself a direct function of the temperature of the system and the individual molecular masses.

Stochastic: Fundamental hypothesis

The rate constant c_μ is dependent on the radii of the molecules involved in the reaction, and their average relative velocities – a property that is itself a direct function of the temperature of the system and the individual molecular masses.

These quantities are basic chemical properties which for most systems are either well known or easily measurable. Thus, for a given chemical system, the propensity functions, a_μ can be easily determined.

Stochastic: Grand probability function

The stochastic formulation proceeds by considering the grand probability function $\Pr(\mathbf{X}; t) \equiv$ probability that there will be present in the volume V at time t, X_i of species S_i, where $\mathbf{X} \equiv (X_1, X_2, \ldots X_N)$ is a vector of molecular species populations.

Stochastic: Grand probability function

The stochastic formulation proceeds by considering the grand probability function $\Pr(\mathbf{X}; t) \equiv$ probability that there will be present in the volume V at time t, X_i of species S_i, where $\mathbf{X} \equiv (X_1, X_2, \ldots X_N)$ is a vector of molecular species populations.

Evidently, knowledge of this function provides a complete understanding of the probability distribution of all possible states at all times.

Stochastic: Infinitesimal time interval

By considering a discrete infinitesimal time interval $(t, t + \mathrm{d}t)$ in which either 0 or 1 reactions occur we see that there exist only $M + 1$ distinct configurations at time t that can lead to the state $\mathbf{X}$ at time $t + \mathrm{d}t$.

Stochastic: Infinitesimal time interval

By considering a discrete infinitesimal time interval $(t, t + \mathrm{d}t)$ in which either 0 or 1 reactions occur we see that there exist only $M + 1$ distinct configurations at time t that can lead to the state $\mathbf{X}$ at time $t + \mathrm{d}t$.

$$\begin{aligned}&\Pr(\mathbf{X}; t + \mathrm{d}t)\\&\qquad = \Pr(\mathbf{X}; t)\,\Pr(\text{no state change over } \mathrm{d}t)\end{aligned}$$

Stochastic: Infinitesimal time interval

By considering a discrete infinitesimal time interval $(t, t + \mathrm{d}t)$ in which either 0 or 1 reactions occur we see that there exist only $M + 1$ distinct configurations at time t that can lead to the state $\mathbf{X}$ at time $t + \mathrm{d}t$.

$$
\begin{aligned}
&\Pr(\mathbf{X}; t + \mathrm{d}t) \\
&\qquad = \Pr(\mathbf{X}; t)\,\Pr(\text{no state change over } \mathrm{d}t) \\
&\qquad + \textstyle\sum_{\mu=1}^{M} \Pr(\mathbf{X} - \mathbf{v}_\mu; t)\,\Pr(\text{state change to } \mathbf{X} \text{ over } \mathrm{d}t)
\end{aligned}
$$

Stochastic: Infinitesimal time interval

By considering a discrete infinitesimal time interval $(t, t + \mathrm{d}t)$ in which either 0 or 1 reactions occur we see that there exist only $M + 1$ distinct configurations at time t that can lead to the state $\mathbf{X}$ at time $t + \mathrm{d}t$.

$$\begin{aligned} &\Pr(\mathbf{X}; t + \mathrm{d}t) \\ &\qquad = \Pr(\mathbf{X}; t) \Pr(\text{no state change over } \mathrm{d}t) \\ &\qquad + \textstyle\sum_{\mu=1}^{M} \Pr(\mathbf{X} - \mathbf{v}_\mu; t) \Pr(\text{state change to } \mathbf{X} \text{ over } \mathrm{d}t) \end{aligned}$$

where $\mathbf{v}_\mu$ is a stoichiometric vector defining the result of reaction μ on state vector $\mathbf{X}$, i.e. $\mathbf{X} \rightarrow \mathbf{X} + \mathbf{v}_\mu$ after an occurrence of reaction μ.

Stochastic: State change probabilities

Pr(no state change over dt)

$$1 - \sum_{\mu=1}^{M} a_\mu(\mathbf{X})\mathrm{d}t$$

Stochastic: State change probabilities

Pr(no state change over dt)

$$1 - \sum_{\mu=1}^{M} a_\mu(\mathbf{X})\mathrm{d}t$$

Pr(state change to **X** over dt)

$$\sum_{\mu=1}^{M} \Pr(\mathbf{X} - \mathbf{v}_\mu; t) a_\mu(\mathbf{X} - \mathbf{v}_\mu)\mathrm{d}t$$

Stochastic: Partial derivatives

$$\frac{\partial \Pr(\mathbf{X}; t)}{\partial t} = \lim_{\mathrm{d}t \to 0} \frac{\Pr(\mathbf{X}; t + \mathrm{d}t) - \Pr(\mathbf{X}; t)}{\mathrm{d}t}$$

Stochastic: Chemical Master Equation

Applying this, and re-arranging the former, leads us to an important *partial differential equation* (PDE) known as the Chemical Master Equation (CME).

$$\frac{\partial \Pr(\mathbf{X}; t)}{\partial t} = \sum_{\mu=1}^{M} a_\mu(\mathbf{X} - \mathbf{v}_\mu) \Pr(\mathbf{X} - \mathbf{v}_\mu; t) - a_\mu(\mathbf{X}) \Pr(\mathbf{X}; t)$$

Stochastic: Chemical Master Equation

Applying this, and re-arranging the former, leads us to an important *partial differential equation* (PDE) known as the Chemical Master Equation (CME).

$$\frac{\partial \Pr(\mathbf{X}; t)}{\partial t} = \sum_{\mu=1}^{M} a_\mu(\mathbf{X} - \mathbf{v}_\mu) \Pr(\mathbf{X} - \mathbf{v}_\mu; t) - a_\mu(\mathbf{X}) \Pr(\mathbf{X}; t)$$

Stochastic: Chemical Master Equation

Applying this, and re-arranging the former, leads us to an important *partial differential equation* (PDE) known as the Chemical Master Equation (CME).

$$\frac{\partial \Pr(\mathbf{X}; t)}{\partial t} = \sum_{\mu=1}^{M} a_\mu(\mathbf{X} - \mathbf{v}_\mu) \Pr(\mathbf{X} - \mathbf{v}_\mu; t) - a_\mu(\mathbf{X}) \Pr(\mathbf{X}; t)$$

Stochastic: Chemical Master Equation

Applying this, and re-arranging the former, leads us to an important *partial differential equation* (PDE) known as the Chemical Master Equation (CME).

$$\frac{\partial \Pr(\mathbf{X};t)}{\partial t} = \sum_{\mu=1}^{M} a_\mu(\mathbf{X}-\mathbf{v}_\mu)\Pr(\mathbf{X}-\mathbf{v}_\mu;t) - a_\mu(\mathbf{X})\Pr(\mathbf{X};t)$$

The problem with the Chemical Master Equation

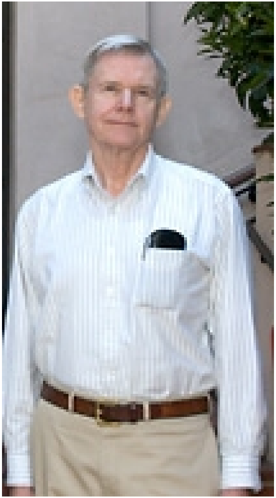
- The CME is really a set of nearly as many coupled ordinary differential equations as there are combinations of molecules that can exist in the system!
- The CME can be solved analytically for only a very few very simple systems, and numerical solutions are usually prohibitively difficult.

D. Gillespie and L. Petzold.
chapter *Numerical Simulation for Biochemical Kinetics*, in *System Modelling in Cellular Biology*, editors Z. Szallasi, J. Stelling and V. Periwal.
MIT Press, 2006.

Outline

1. The deterministic and stochastic approaches
2. Stochastic simulation algorithms
3. Comparing stochastic simulation and ODEs
4. Modelling challenges

Breakthrough: Gillespie's Stochastic simulation algorithms



Biography: Daniel T. Gillespie

1960	BA from Rice University
1968	PhD from Johns Hopkins University
1968–1971	Postdoc at the University of Maryland's Institute for Molecular Physics.
1971–2001	Research Physicist in the Earth & Planetary Sciences Division of the Naval Air Warfare Center in China Lake, California.
2001	Retirement from Civil Service. Begins consultancy for California Institute of Technology and the Molecular Sciences Institute, working mostly with Linda Petzold and her group at the University of California at Santa Barbara.

Books by Daniel T. Gillespie

- A Quantum Mechanics Primer (1970)
- Markov Processes: An Introduction for Physical Scientists (1992)
- Biography of radio comedy writer Tom Koch (2004)

Stochastic simulation algorithms

Gillespie's Stochastic Simulation Algorithm (SSA) is essentially an exact procedure for numerically simulating the time evolution of a well-stirred chemically reacting system by taking proper account of the randomness inherent in such a system.

Stochastic simulation algorithms

Gillespie's Stochastic Simulation Algorithm (SSA) is essentially an exact procedure for numerically simulating the time evolution of a well-stirred chemically reacting system by taking proper account of the randomness inherent in such a system.

It is rigorously based on the same microphysical premise that underlies the chemical master equation and gives a more realistic representation of a system's evolution than the deterministic reaction rate equation (RRE) represented mathematically by ODEs.

Stochastic simulation algorithms

Gillespie's Stochastic Simulation Algorithm (SSA) is essentially an exact procedure for numerically simulating the time evolution of a well-stirred chemically reacting system by taking proper account of the randomness inherent in such a system.

It is rigorously based on the same microphysical premise that underlies the chemical master equation and gives a more realistic representation of a system's evolution than the deterministic reaction rate equation (RRE) represented mathematically by ODEs.

As with the chemical master equation, the SSA converges, in the limit of large numbers of reactants, to the same solution as the law of mass action.

Gillespie's exact SSA (1977)

- The algorithm takes time steps of variable length, based on the rate constants and population size of each chemical species.

Gillespie's exact SSA (1977)

- The algorithm takes time steps of variable length, based on the rate constants and population size of each chemical species.
- The probability of one reaction occurring relative to another is dictated by their relative propensity functions.

Gillespie's exact SSA (1977)

- The algorithm takes time steps of variable length, based on the rate constants and population size of each chemical species.
- The probability of one reaction occurring relative to another is dictated by their relative propensity functions.
- According to the correct probability distribution derived from the statistical thermodynamics theory, a random variable is then used to choose which reaction will occur, and another random variable determines how long the step will last.

Gillespie's exact SSA (1977)

- The algorithm takes time steps of variable length, based on the rate constants and population size of each chemical species.
- The probability of one reaction occurring relative to another is dictated by their relative propensity functions.
- According to the correct probability distribution derived from the statistical thermodynamics theory, a random variable is then used to choose which reaction will occur, and another random variable determines how long the step will last.
- The chemical populations are altered according to the stoichiometry of the reaction and the process is repeated.

Stochastic simulation: Job done!

Stochastic simulation: realisations and ensembles

The SSA computes one realisation of a dynamic trajectory of a chemically reacting system. Often an ensemble of trajectories is computed, to obtain an estimate of the probability density function of the system.

The dynamic evolution of the probability density function is given by the Chemical Master Equation.

Gillespie's SSA is a Monte Carlo Markov Chain simulation

The SSA is a Monte Carlo type method. With the SSA one may approximate any variable of interest by generating many trajectories and observing the statistics of the values of the variable. Since many trajectories are needed to obtain a reasonable approximation, the efficiency of the SSA is of critical importance.

Computational cost of Gillespie's exact algorithm

The cost of this detailed stochastic simulation algorithm is the likely large amounts of computing time.

The key issue is that the time step for the next reaction can be very small indeed if we are to guarantee that only one reaction can take place in a given time interval.

Computational cost of Gillespie's exact algorithm

The cost of this detailed stochastic simulation algorithm is the likely large amounts of computing time.

The key issue is that the time step for the next reaction can be very small indeed if we are to guarantee that only one reaction can take place in a given time interval.

Increasing the molecular population or number of reaction mechanisms necessarily requires a corresponding decrease in the time interval. The SSA can be very computationally inefficient especially when there are large numbers of molecules or the propensity functions are large.

Gibson and Bruck (2000)

Gibson and Bruck refined the first reaction SSA of Gillespie by reducing the number of random variables that need to be simulated.

This can be effective for systems in which some reactions occur much more frequently than others.

M.A. Gibson and J. Bruck.
Efficient exact stochastic simulation of chemical systems with many species and many channels.
J. Comp. Phys., 104:1876–1889, 2000.

Variants of SSA

Gillespie developed two different but equivalent formulations of the SSA: the Direct Method (DM) and the First Reaction Method (FRM). A third formulation of the SSA is the Next Reaction Method (NRM) of Gibson and Bruck. The NRM can be viewed as an extension of the FRM, but it is much more efficient than the latter.

Variants of SSA

Gillespie developed two different but equivalent formulations of the SSA: the Direct Method (DM) and the First Reaction Method (FRM). A third formulation of the SSA is the Next Reaction Method (NRM) of Gibson and Bruck. The NRM can be viewed as an extension of the FRM, but it is much more efficient than the latter.

It was widely believed that Gibson and Bruck's method (the Next Reaction Method) was more efficient than Gillespie's Direct Method (DM). This conclusion is based on a count of arithmetic operations.

Gibson and Bruck challenged (2004)

It was established by Cao, Li and Petzold (2004) that Gibson and Bruck's analysis misses the dominant cost of the NRM, which is maintaining the priority queue data structure of the tentative reaction times and that good implementations of DM such as the Optimised Direct Method (ODM) have lower asymptotic complexity than Gibson and Bruck's method.

Y. Cao, H. Li, and L. Petzold.
Efficient formulation of the stochastic simulation algorithm for chemically reacting systems.
J. Chem. Phys, 121(9):4059–4067, 2004.

Enhanced stochastic simulation techniques

If the system under study possesses a macroscopically infinitesimal timescale so that during any dt all of the reaction channels can fire many times, yet none of the propensity functions change appreciably, then the discrete Markov process as described by the SSA can be approximated by a continuous Markov process.

Enhanced stochastic simulation techniques

If the system under study possesses a macroscopically infinitesimal timescale so that during any dt all of the reaction channels can fire many times, yet none of the propensity functions change appreciably, then the discrete Markov process as described by the SSA can be approximated by a continuous Markov process.

Enhanced stochastic simulation techniques

If the system under study possesses a macroscopically infinitesimal timescale so that during any dt all of the reaction channels can fire many times, yet none of the propensity functions change appreciably, then the discrete Markov process as described by the SSA can be approximated by a continuous Markov process.

Enhanced stochastic simulation techniques

If the system under study possesses a macroscopically infinitesimal timescale so that during any dt all of the reaction channels can fire many times, yet none of the propensity functions change appreciably, then the discrete Markov process as described by the SSA can be approximated by a continuous Markov process.

Enhanced stochastic simulation techniques

If the system under study possesses a macroscopically infinitesimal timescale so that during any dt all of the reaction channels can fire many times, yet none of the propensity functions change appreciably, then the discrete Markov process as described by the SSA can be approximated by a continuous Markov process.

This Markov process is described by the Chemical Langevin Equation (CLE), which is a stochastic ordinary differential equation (SDE).

Stochastic Differential Equations

A stochastic differential equation (SDE)

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t$$

is interpreted as a stochastic integral equation

$$X_t = X_{t_0} + \int_{t_0}^{t} a(s, X_s)ds + \int_{t_0}^{t} b(s, X_s)dW_s$$

where the first integral is a Lebesgue (or Riemann) integral for each sample path and the second integral is usually an Ito integral.

Chemical Langevin Equation

The Langevin equation

$$dX_t = -aX_t dt + dW_t$$

is a linear SDE with additive noise. The solution for $t_0 = 0$ is

$$X_t = X_0 e^{-at} + e^{-at} \int_0^t e^{as} dW_s$$

Gillespie's tau-leap method (2001)

Gillespie proposed two new methods, namely the τ-leap method and the midpoint τ-leap method in order to improve the efficiency of the SSA while maintaining acceptable losses in accuracy.

Daniel T. Gillespie.
Approximate accelerated stochastic simulation of chemically reacting systems.
J. Comp. Phys., 115(4):1716–1733, 2001.

Gillespie's tau-leap method (2001)

Gillespie proposed two new methods, namely the τ-leap method and the midpoint τ-leap method in order to improve the efficiency of the SSA while maintaining acceptable losses in accuracy.

Daniel T. Gillespie.
Approximate accelerated stochastic simulation of chemically reacting systems.
J. Comp. Phys., 115(4):1716–1733, 2001.

The key idea here is to take a larger time step and allow for more reactions to take place in that step, but under the proviso that the propensity functions do not change too much in that interval. By means of a Poisson approximation, the tau-leaping method can "leap over" many reactions.

Gillespie’s tau-leap method (significance)

For many problems, the tau-leaping method can approximate the stochastic behaviour of the system very well.

Gillespie's tau-leap method (significance)

For many problems, the tau-leaping method can approximate the stochastic behaviour of the system very well.

The tau-leaping method connects the SSA in the discrete stochastic regime to the explicit Euler method for the chemical Langevin equation in the continuous stochastic regime and the RRE in the continuous deterministic regime.

Gillespie's tau-leap method (significance)

For many problems, the tau-leaping method can approximate the stochastic behaviour of the system very well.

The tau-leaping method connects the SSA in the discrete stochastic regime to the explicit Euler method for the chemical Langevin equation in the continuous stochastic regime and the RRE in the continuous deterministic regime.

Gillespie's tau-leap method (significance)

For many problems, the tau-leaping method can approximate the stochastic behaviour of the system very well.

The tau-leaping method connects the SSA in the discrete stochastic regime to the explicit Euler method for the chemical Langevin equation in the continuous stochastic regime and the RRE in the continuous deterministic regime.

Gillespie's tau-leap method (significance)

For many problems, the tau-leaping method can approximate the stochastic behaviour of the system very well.

The tau-leaping method connects the SSA in the discrete stochastic regime to the explicit Euler method for the chemical Langevin equation in the continuous stochastic regime and the RRE in the continuous deterministic regime.

Gillespie's tau-leap method (drawback)

The use of approximation in Poisson methods leads to the possibility of negative molecular numbers being predicted — something with no physical explanation.

Gillespie's Modified Poisson tau-leap methods (2005)

Gillespie's modified Poisson tau-leaping method introduces a second control parameter whose value dials the procedure from the original Poisson tau-leaping method at one extreme to the exact SSA at the other.

Gillespie's Modified Poisson tau-leap methods (2005)

Gillespie's modified Poisson tau-leaping method introduces a second control parameter whose value dials the procedure from the original Poisson tau-leaping method at one extreme to the exact SSA at the other.

Any reaction channel with a positive propensity function which is within n_c firings of exhausting its reactants is termed a *critical* reaction.

Y. Cao, D. Gillespie, and L. Petzold.
Avoiding negative populations in explicit tau leaping.
J. Chem. Phys, 123(054104), 2005.

Gillespie's Modified Poisson tau-leap methods (2006)

The modified algorithm chooses τ in such a way that no more than *one* firing of *all* the critical reactions can occur during the leap. The probability of producing a negative population is reduced to nearly zero.

Gillespie's Modified Poisson tau-leap methods (2006)

The modified algorithm chooses τ in such a way that no more than *one* firing of *all* the critical reactions can occur during the leap. The probability of producing a negative population is reduced to nearly zero.

If a negative population *does* occur the leap can simply be rejected and repeated with τ reduced by half, or the entire simulation can be abandoned and repeated for larger n_c.

Y. Cao, D. Gillespie, and L. Petzold.
Efficient stepsize selection for the tau-leaping method.
J. Chem. Phys, 2006.
To appear.

Family of stochastic simulation algorithms

FASTEST, BEST

Discrete, exact	**Continuous, approximate**
	Modified Poisson τ leap (2005)
	τ leap (2001)
Logarithmic Direct Method (2006)	
Sorting Direct Method (2005)	
Optimised Direct Method (2004)	
Next Reaction Method (2000)	
Direct Method (1977)	
First Reaction Method (1977)	

SLOWEST, WORST

Outline

Comparing stochastic simulation and ODEs

We know that stochastic simulation can allow us to observe phenomena which ODEs cannot.

Comparing stochastic simulation and ODEs

We know that stochastic simulation can allow us to observe phenomena which ODEs cannot.

Are there places where they agree?

A simple example: processors and resources

$$\begin{aligned}
Proc_0 &\stackrel{\mathit{def}}{=} (task1, \top).Proc_1 \\
Proc_1 &\stackrel{\mathit{def}}{=} (task2, r_2).Proc_0 \\
Res_0 &\stackrel{\mathit{def}}{=} (task1, r_1).Res_1 \\
Res_1 &\stackrel{\mathit{def}}{=} (reset, s).Res_0
\end{aligned}$$

$$Proc_0[P] \underset{\{task1\}}{\bowtie} Res_0[R]$$

A simple example: processors and resources

$$
\begin{aligned}
Proc_0 &\overset{def}{=} (task1, \top).Proc_1 \\
Proc_1 &\overset{def}{=} (task2, r_2).Proc_0 \\
Res_0 &\overset{def}{=} (task1, r_1).Res_1 \\
Res_1 &\overset{def}{=} (reset, s).Res_0
\end{aligned}
$$

$$Proc_0[P] \underset{\{task1\}}{\bowtie} Res_0[R]$$

CTMC interpretation

Processors (P)	Resources (R)	States (2^{P+R})
1	1	4
2	1	8
2	2	16
3	2	32
3	3	64
4	3	128
4	4	256
5	4	512
5	5	1024
6	5	2048
6	6	4096
7	6	8192
7	7	16384
8	7	32768
8	8	65536
9	8	131072
9	9	262144
10	9	524288
10	10	1048576

A simple example: processors and resources

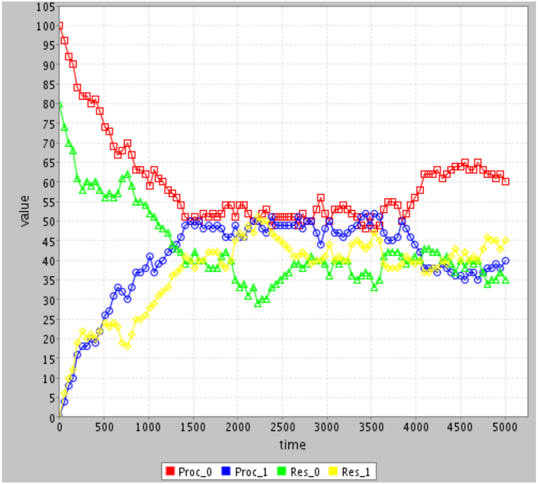
$$
\begin{aligned}
Proc_0 &\overset{def}{=} (task1, \top).Proc_1 \\
Proc_1 &\overset{def}{=} (task2, r_2).Proc_0 \\
Res_0 &\overset{def}{=} (task1, r_1).Res_1 \\
Res_1 &\overset{def}{=} (reset, s).Res_0
\end{aligned}
$$

$$Proc_0[P] \underset{\{task1\}}{\bowtie} Res_0[R]$$

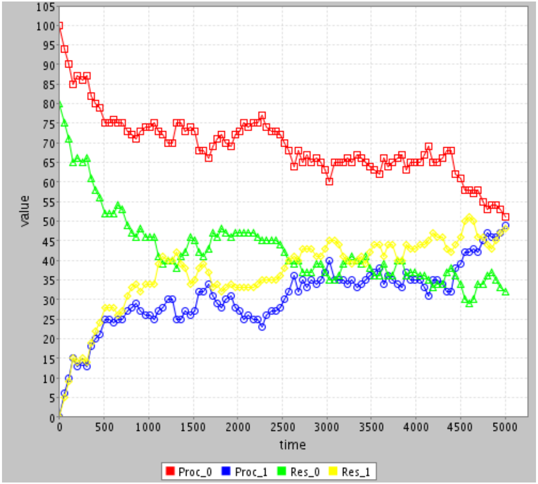
ODE interpretation

$$
\begin{aligned}
\frac{\mathrm{d}Proc_0}{\mathrm{d}t} &= -r_1 \min(Proc_0, Res_0) \\
&\quad + r_2\, Proc_1 \\
\frac{\mathrm{d}Proc_1}{\mathrm{d}t} &= r_1 \min(Proc_0, Res_0) \\
&\quad - r_2\, Proc_1 \\
\frac{\mathrm{d}Res_0}{\mathrm{d}t} &= -r_1 \min(Proc_0, Res_0) \\
&\quad + s\, Res_1 \\
\frac{\mathrm{d}Res_1}{\mathrm{d}t} &= r_1 \min(Proc_0, Res_0) \\
&\quad - s\, Res_1
\end{aligned}
$$

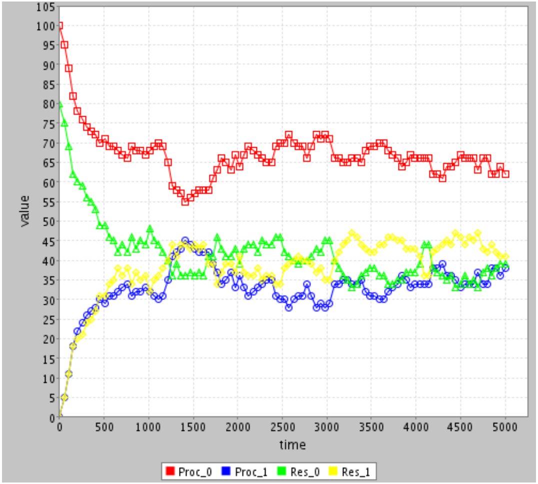
Processors and resources (SSA run A)



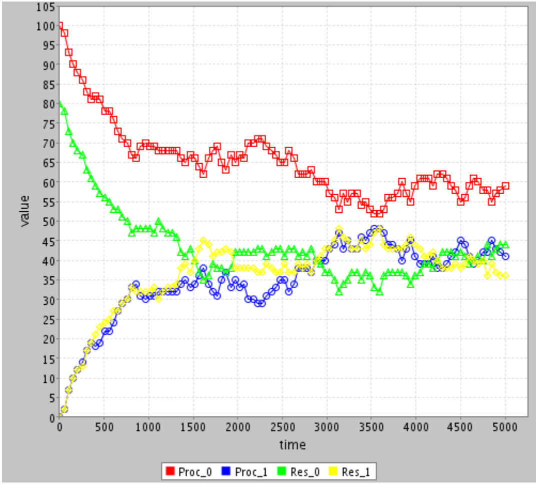
Processors and resources (SSA run B)



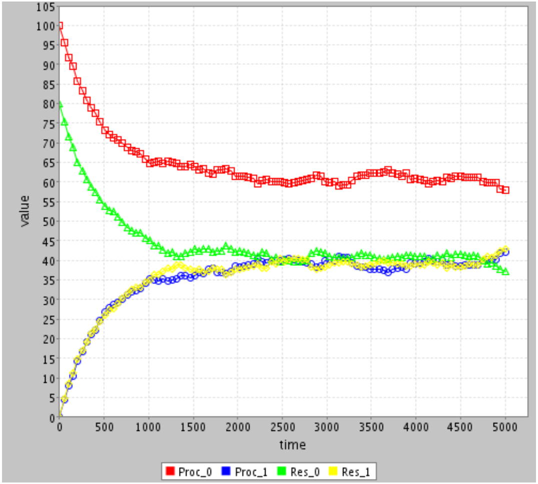
Processors and resources (SSA run C)



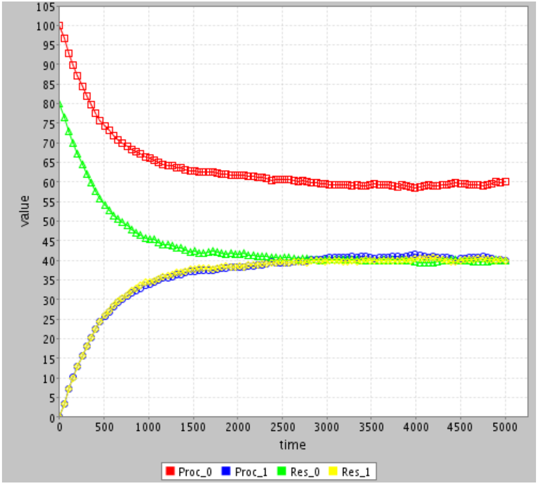
Processors and resources (SSA run D)



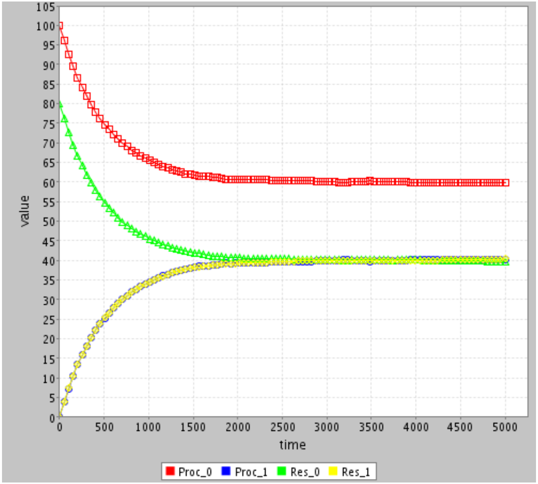
Processors and resources (average of 10 runs)



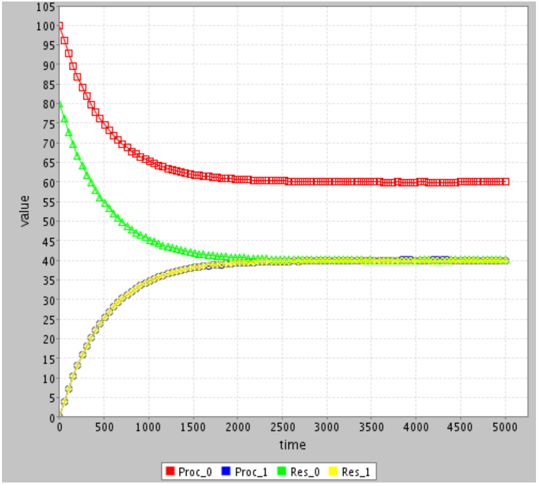
Processors and resources (average of 100 runs)



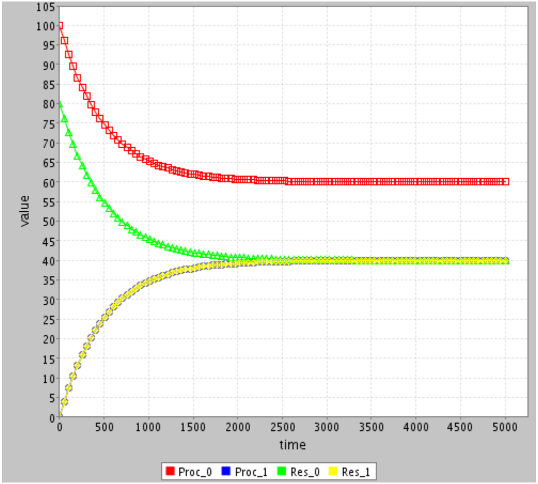
Processors and resources (average of 1000 runs)



Processors and resources (average of 10000 runs)



Processors and resources (ODE solution)



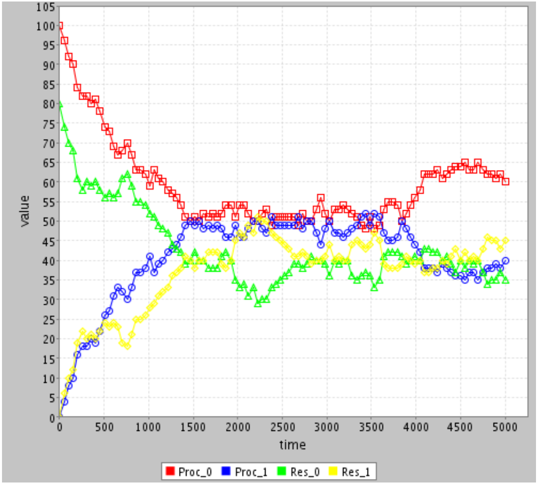
From realisations to ensembles

As we repeatedly sample from the underlying random number distributions the average of the samples tends to the mean.

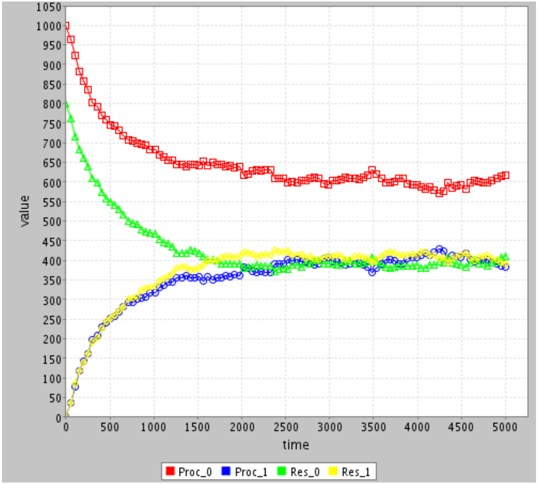
Processors and resources: scaling out

What effect does increasing the number of copies have?

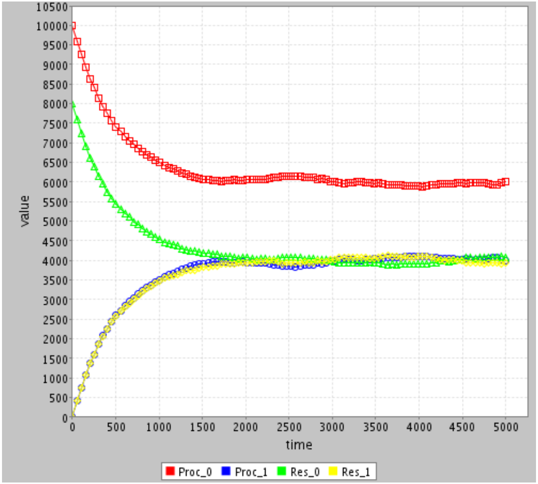
Processors and resources (single SSA run, 100/80)



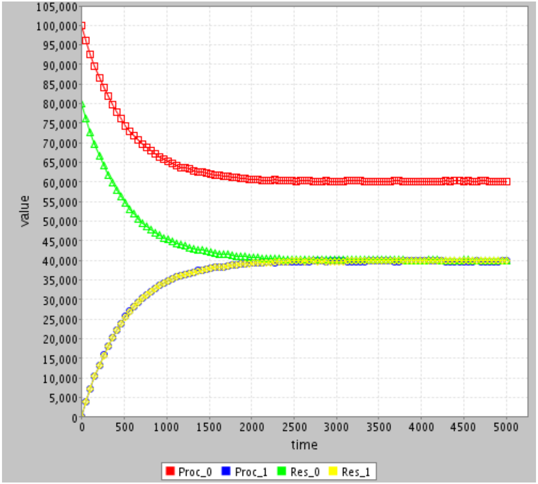
Processors and resources (single SSA run, 1,000/800)



Processors and resources (single SSA run, 10,000/8,000)



Processors and resources (single SSA run, 100,000/80,000)



From the microscopic to the macroscopic domain

Each specific run is individually in closer and closer agreement with the deterministic approach as the number of molecules in the system increases.

This is a direct effect of the inherent averaging of macroscopic properties of a system of many particles.

Conclusions from the comparison

1. These results provide clear verification of the compatibility of the deterministic and stochastic approaches.
2. They also illustrate the validity of the deterministic approach in systems containing as few as 100 copies of components.

Outline

Modelling challenges: stiffness

A problem for modelling temporal evolution is *stiffness*. Some reactions are much faster than others and quickly reach a stable state. The dynamics of the system is driven by the slow reactions.

Modelling challenges: stiffness

A problem for modelling temporal evolution is *stiffness*. Some reactions are much faster than others and quickly reach a stable state. The dynamics of the system is driven by the slow reactions.

Most chemical systems, whether considered at a scale appropriate to stochastic or to deterministic simulation, involve several widely varying time scales, so such systems are nearly always stiff.

Modelling challenges: stiffness

A problem for modelling temporal evolution is *stiffness*. Some reactions are much faster than others and quickly reach a stable state. The dynamics of the system is driven by the slow reactions.

Most chemical systems, whether considered at a scale appropriate to stochastic or to deterministic simulation, involve several widely varying time scales, so such systems are nearly always stiff.

Modelling challenges: multiscale populations

The *multiscale population* problem arises when some species are present in relatively small quantities and should be modelled by a discrete stochastic process, whereas other species are present in larger quantities and are more efficiently modelled by a deterministic ordinary differential equation (or at some scale in between). SSA treats all of the species as discrete stochastic processes.

Gillespie's multiscale SSA methods (2005)

SSA is used for slow reactions or species with small populations. The multiscale SSA method generalizes this idea to the case in which species with small population are involved in fast reactions.

Gillespie's slow-scale SSA methods (2005)

The setting for Gillespie's slow-scale SSA method is

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_2} E + P$$

where

$$k_{-1} \gg k_2$$

Gillespie's slow-scale SSA methods (2005)

The setting for Gillespie's slow-scale SSA method is

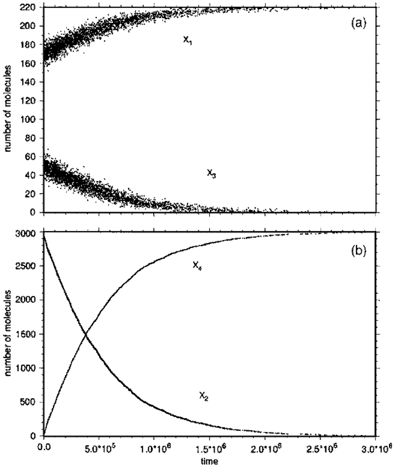
$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_2} E + P$$

where

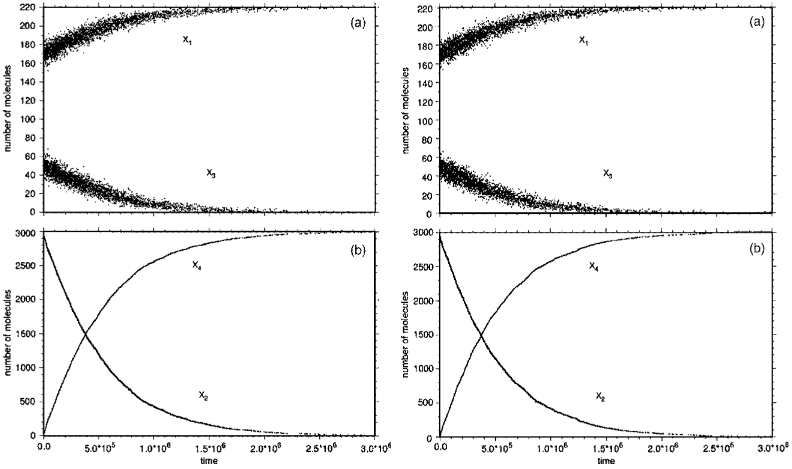
$$k_{-1} \gg k_2$$

Slow-scale SSA explicitly simulates only the relatively rare conversion reactions, skipping over occurrences of the other two less interesting but much more frequent reactions.

Comparing SSA and Slow-Scale SSA results



Comparing SSA and Slow-Scale SSA results



Conclusions

- Stochastic simulation is a well-founded method for simulating *in vivo* reactions.
- Gillespie's SSA can be more accurate than ODEs at low molecular numbers; compatible with them at large molecular numbers.
- Recent explosion of interest in the subject with many new variants of the SSA algorithm.

Excellent introductory papers

T.E. Turner, S. Schnell, and K. Burrage.
Stochastic approaches for modelling in vivo reactions.
Computational Biology and Chemistry, 28:165–178, 2004.

D. Gillespie and L. Petzold.
System Modelling in Cellular Biology, chapter Numerical Simulation for Biochemical Kinetics,.
MIT Press, 2006.

Stochastic simulation software

S. Ramsey, D. Orrell, and H. Bolouri.
Dizzy: stochastic simulation of large-scale genetic regulatory networks.
J. Bioinf. Comp. Biol., 3(2):415–436, 2005.
http://magnet.systemsbiology.net/software/Dizzy