Stochastic Simulation for Systems Biology

Stephen Gilmore

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

Galileo workshop, Pisa, October 2007

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.
- 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;
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
- 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

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

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

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: 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: 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 $t + \mathrm{d}t$.

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

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

Stochastic: State change probabilities

$\Pr(\text{no state change over } \mathrm{d}t)$

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

$\Pr(\text{state change to } \mathbf{X} \text{ over } \mathrm{d}t)$

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

Stochastic: Partial derivatives

We are considering the behaviour of the system in the limit as $\mathrm{d}t$ tends to zero. This leads us to consider *partial derivatives*, which are defined thus:

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

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).

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

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.

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.
- 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: 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.

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.

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*.

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.

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.

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.

Binomial leap methods (2004)

Independently Tian and Burrage, and Chatterjee and Vlachos, proposed replacing the use of the Poisson distribution with the binomial distribution.

Unlike Poisson random variables whose range of sample values is from zero to infinity, binomial random variables have a finite range of sample values.

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.

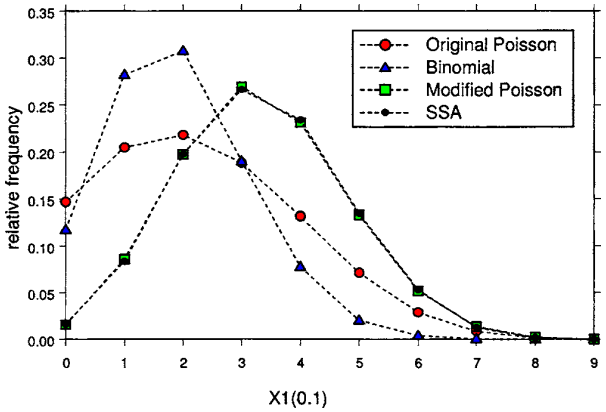
Gillespie's Modified Poisson tau-leap methods (2006)

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.

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.

Comparing Poisson and Binomial leap results



Computation time

ϵ	Original Poisson Time (s)	Original Poisson Leaps	Binomial Time (s)	Binomial Leaps	Mod. Poisson Time (s)	Mod. Poisson Leaps
0.03	57	5.15×10^5	89	7.75×10^5	72	6.31×10^5
0.05	36	3.20×10^5	85	7.73×10^5	47	4.13×10^5

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	

Example: A monostable system

Molecular noise acting upon dynamical structures can generate disorderly behaviour in homeostatic systems.

Consider a simple example where protein molecules X and Y are synthesized from the reservoirs A and B at an equal rate k. X and Y are assumed to associate irreversibly with association rate constant k_a in the formation of a heterodimer C. Molecules of X and Y can also decay with first-order rate constant α_1 and α_2 respectively.

Example: A monostable system (deterministically)

$$A \xrightarrow{k} X \xrightarrow{\alpha_1} \phi; \qquad B \xrightarrow{k} Y \xrightarrow{\alpha_2} \phi; \qquad X + Y \xrightarrow{k_a} C$$

$$\begin{aligned} \frac{d\phi_1}{dt} &= k - \alpha_1\phi_1 - k_a\phi_1\phi_2 \\ \frac{d\phi_2}{dt} &= k - \alpha_2\phi_2 - k_a\phi_1\phi_2 \end{aligned}$$

$$\begin{aligned} k &= 10 \\ \alpha_1 &= 10^{-6} \\ \alpha_2 &= 10^{-5} \\ k_a &= 10^{-5} \end{aligned} \qquad \phi_1^{ss} \simeq \sqrt{\frac{k}{k_a}} = 1000 \qquad \phi_2^{ss} \simeq 100$$

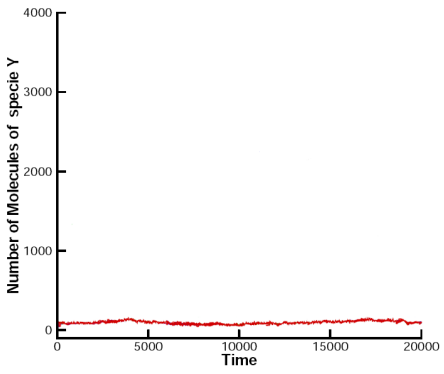
Example: A monostable system (deterministically)

$$A \xrightarrow{k} X \xrightarrow{\alpha_1} \phi; \qquad B \xrightarrow{k} Y \xrightarrow{\alpha_2} \phi; \qquad X + Y \xrightarrow{k_a} C$$

$$\begin{aligned} \frac{d\phi_1}{dt} &= k - \alpha_1\phi_1 - k_a\phi_1\phi_2 \\ \frac{d\phi_2}{dt} &= k - \alpha_2\phi_2 - k_a\phi_1\phi_2 \end{aligned}$$

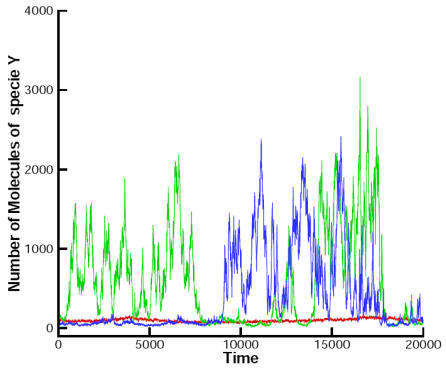
$$\begin{aligned} k &= 10^3 \\ \alpha_1 &= 10^{-4} \\ \alpha_2 &= 10^{-3} \\ k_a &= 10^{-3} \end{aligned} \qquad \phi_1^{ss} \simeq \sqrt{\frac{k}{k_a}} = 1000 \qquad \phi_2^{ss} \simeq 100$$

Example: A monostable system (stochastically)



$$\begin{aligned} k &= 10 \\ \alpha_1 &= 10^{-6} \\ \alpha_2 &= 10^{-5} \\ k_a &= 10^{-5} \end{aligned}$$

Example: A monostable system (stochastically)



$$\begin{aligned} k &= 10^3 \\ \alpha_1 &= 10^{-4} \\ \alpha_2 &= 10^{-3} \\ k_a &= 10^{-3} \end{aligned}$$

Example: A monostable system (Conclusion)

- For the second set of parameters there is a noticeable discrepancy between the behaviour of the mean and that of the deterministic trajectory.
- Stochastic excursions indicate a severe effect of noise on the system.
- Such an effect indicates that a deterministic approach to the analysis of such a system can be misleading and calls for a thorough stochastic treatment.

Circadian clock

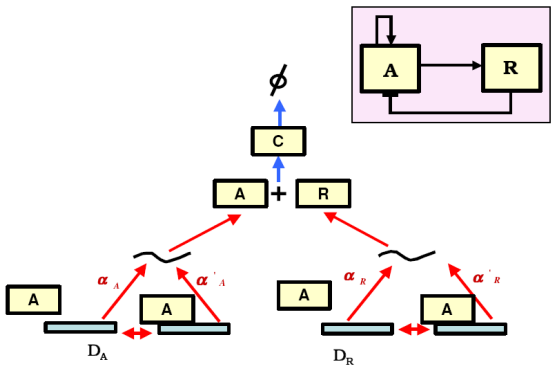
To adapt to natural periodicity, such as the alternation of day and night, most living organisms have developed the capability of generating oscillating expressions of proteins in their cells with a period close to 24 hours (circadian rhythm).

The Vilar-Kueh-Barkai-Leibler (VKBL in short) description of the circadian oscillator incorporates an abstraction of a minimal set of essential, experimentally determined mechanisms for the circadian system.

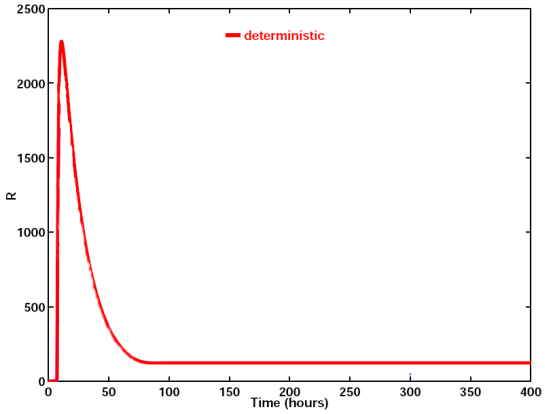
Circadian clock

- The VKBL model involves two genes, an *activator* A and a *repressor* R, which are transcribed into *mRNA* and subsequently translated into proteins.
- The activator A binds to the A and R promoters and increases their expression rate.
- Therefore, A implements a *positive loop* acting on its own transcription.
- At the same time, R sequesters A to form a complex C, therefore inhibiting it from binding to the gene promoter and acting as a *negative feedback loop*.

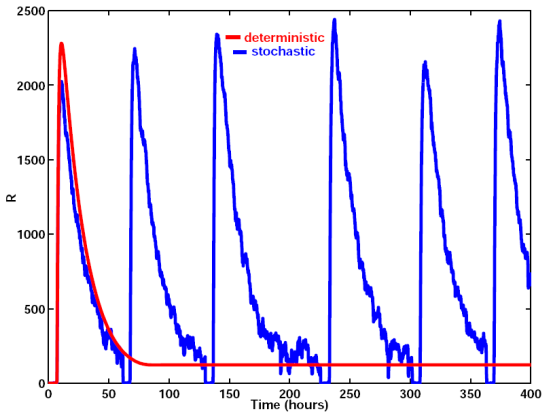
Circadian clock (cartoon)



Circadian clock (deterministically . . .)



Circadian clock (. . . and stochastically)



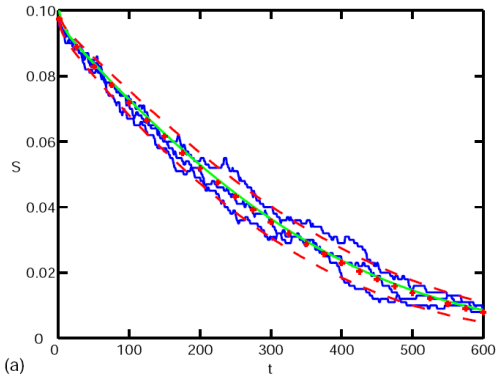
Circadian clock (Conclusions)

- For some parameter values a differential equation model exhibits autonomous oscillations.
- These oscillations disappear from the deterministic model as the degradation rate of the repressor δ_R is decreased.
- The system of ODEs undergoes a bifurcation at this point and the unique deterministic equilibrium of the system becomes stable.
- However, if the effects of molecular noise are incorporated the oscillations in the stochastic system pertain.
- This phenomenon is a manifestation of *coherence resonance*, and illustrates the crucial interplay between noise and dynamics.

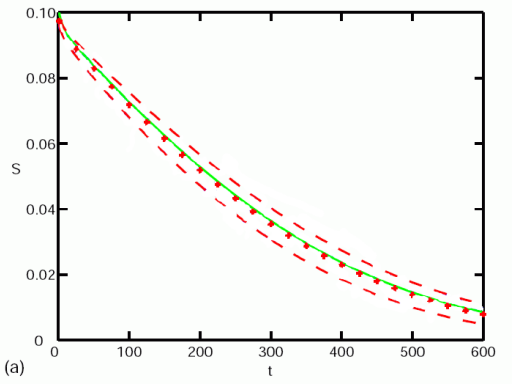
Comparing stochastic simulation and ODEs

It is relatively straightforward to contrast the results of the two methods. We compare the results of 2000 runs of the stochastic algorithm simulating a system with initial molecular populations $S_0 = 100, E_0 = 10, C_0 = 0, P_0 = 0$ and a volume of 1000 units.

Results for $S_0 = 100, E_0 = 10, C_0 = 0, P_0 = 0$ (vol 1000)



Results for $S_0 = 100, E_0 = 10, C_0 = 0, P_0 = 0$ (vol 1000)



Comparing stochastic simulation and ODEs

It is clear that there is a close correspondence between the predictions of the deterministic approach and the stochastic approach, with the deterministic curve falling well within one standard deviation (S.D.) of the stochastic mean.

This is a very close match, especially considering our stochastic simulation is modelling a system containing just 110 molecules—well within what we might consider to be the *microscopic* domain.

The variance of the stochastic approach

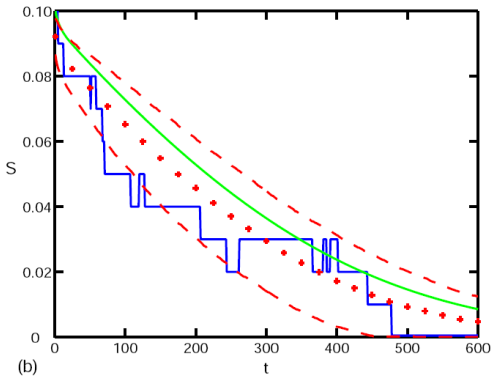
However, it is worth bearing in mind that an actual *in vivo* biochemical reaction would follow just one of the many random curves that average together producing the closely fitting mean. This curve may deviate significantly from that of the deterministic approach, and thus call into question its validity.

Hence, it is perhaps most important to consider the *variance* of the stochastic approach—with a larger variance indicating a greater deviation from the mean and hence from the deterministic curve.

Comparing results at lower population sizes

Consider exactly the same simulation setup, except this time we are modelling a system consisting of just 11 molecules within a volume of 100 units [thus the molecular *concentrations* are equal to those earlier].

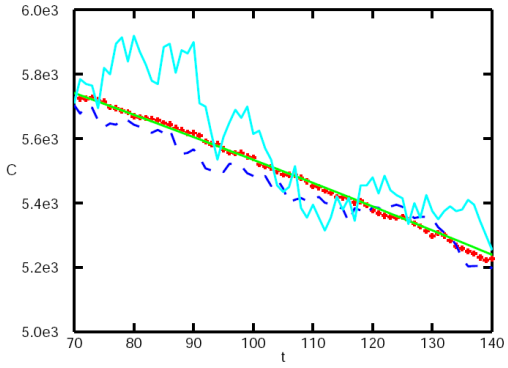
Results for $S_0 = 10, E_0 = 1, C_0 = 0, P_0 = 0$ (vol 100)



Compatibility of the two approaches

On average, the stochastic approach tends to the same solution as the deterministic approach as the number of molecules in the system increases, and we hence move from the microscopic to the macroscopic domain.

Mean results for 11, 110 and 1100 molecules



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.

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

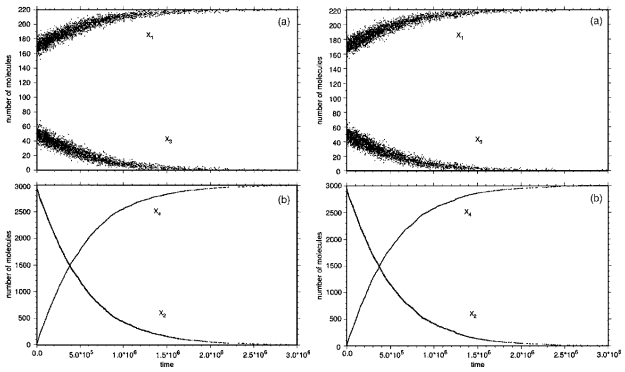
$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \overset{k_2}{\rightarrow} E + P$$

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



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

Daniel T. Gillespie.
Stochastic Simulation of Chemical Kinetics,
Annual Review of Physical Chemistry 58:35-55, 2007.
doi:10.1146/annurev.physchem.58.032806.104637

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

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

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

Hong Li, Yang Cao, Linda R. Petzold, Daniel T. Gillespie
Algorithms and software for stochastic simulation of biochemical reacting systems.
To appear in *Biotechnology Progress*, 2007.
http://www.engineering.ucsb.edu/∼cse/StochKit/