

Dynamics of Biological Systems

Part II - Stochastic simulation

Introduction

- The modelling of chemical reactions using deterministic rate laws (such as the law of mass action) has proven extremely successful in both chemistry and biochemistry for many years
- Rate laws consider chemical reactions to be macroscopic, 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 (such as intracellular reactions) the validity of a continuous approach becomes even more tenuous

Introduction

The fundamental principle behind 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

In 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

Gillespie's Stochastic Approach

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

Given a set of reactions $\mathcal{R} = \{R_1, \dots, R_n\}$, the SSA:

- assumes a stochastic reaction constant c_μ for each chemical reaction $R_\mu \in \mathcal{R}$
- $c_\mu dt$ is the probability that a particular combination of reactant molecules of R_μ react in an infinitesimal time interval dt

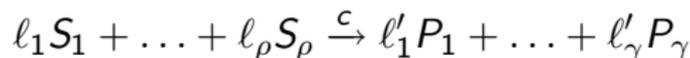
Gillespie's Stochastic Approach

The constant c_μ is used to compute the **propensity** (or stochastic rate) of R_μ to occur in the whole chemical solution, denoted a_μ , as follows:

$$a_\mu = h_\mu c_\mu$$

where h_μ is the number of distinct molecular reactant combinations.

Let R_μ be



In accordance with standard combinatorics, the number of distinct reactant combinations of R_μ in a solution with X_i molecules of S_i , with $1 \leq i \leq \rho$, is given by

$$h_\mu = \prod_{i=1}^{\rho} \binom{X_i}{\ell_i}$$

Gillespie's Stochastic Approach

Example:

solution with X_1 molecules S_1 and X_2 molecules S_2

reaction $R_1 : S_1 + S_2 \rightarrow 2S_1$

- $h_1 = \binom{X_1}{1} \binom{X_2}{1} = X_1 X_2$
- $a_1 = X_1 X_2 c_1$

reaction $R_2 : 2S_1 \rightarrow S_1 + S_2$

- $h_2 = \binom{X_1}{2} = \frac{X_1(X_1-1)}{2}$
- $a_2 = \frac{X_1(X_1-1)}{2} c_2$

Note that propensity a_μ is similar, for suitable kinetic constants, to the mass action rates:

- For R_1 with $k_1 = c_1$, the law of mass action gives $k_1[S_1][S_2] \approx a_1$
- For R_2 with $k_2 = c_2/2$, the law of mass action gives $k_2[S_1]^2 \approx a_2$

Gillespie's Stochastic Approach

Propensity a_μ is used in Gillespie's approach as the parameter of an exponential probability distribution modelling the time between subsequent occurrences of reactions R_μ .

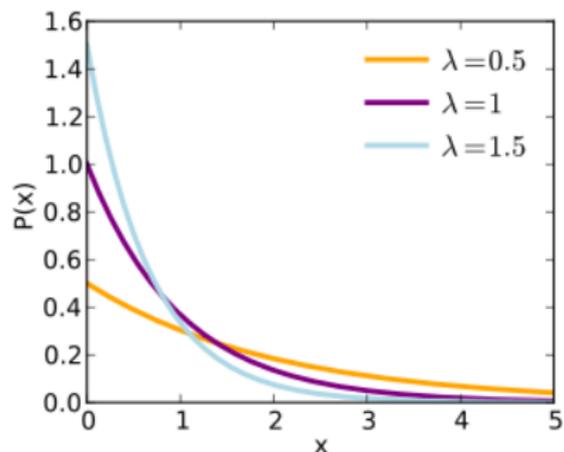
Exponential distribution is a continuous probability distribution (on $[0, \infty]$) describing the timing between events in a Poisson process, namely a process in which events occur continuously and independently at a constant average rate (taken as parameter).

The probability density function f and the cumulative distribution function F of an exponential distribution with parameter λ are as follows:

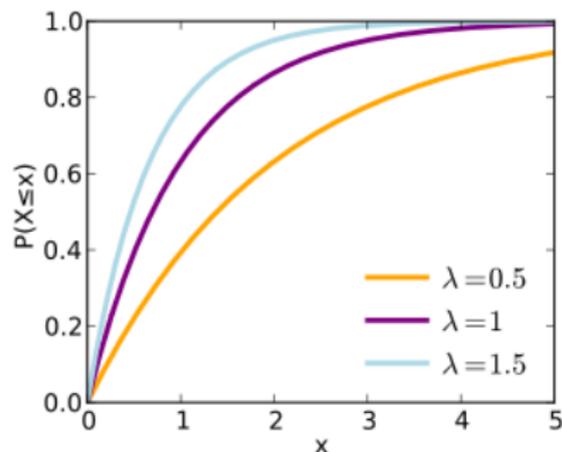
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Gillespie's Stochastic Approach

Probability density function



Cumulative distribution function



The mean of an exponentially distributed variable with parameter λ is $\frac{1}{\lambda}$.

Gillespie's Stochastic Approach

Two important properties of the exponential distribution hold:

- The exponential distribution is **memoryless**:
 $P(X > t + s \mid X > s) = P(X > t)$. This allows a simulation algorithm in which the exponential distribution is used to forget about the history of the simulation
- Let X_1, \dots, X_n be independent exponentially distributed random variables with parameters $\lambda_1, \dots, \lambda_n$. Then $X = \min(X_1, \dots, X_n)$ is also exponentially distributed with parameter $\lambda = \lambda_1 + \dots + \lambda_n$. This allows a simulation algorithm to use a unique exponential distribution for the whole set of reactions to be simulated

Gillespie's Stochastic Simulation Algorithm (SSA)

Given:

- a set of molecular species $\{S_1, \dots, S_n\}$
- initial numbers of molecules of each species $\{X_1, \dots, X_n\}$ with $X_i \in \mathbb{N}$
- a set of chemical reactions $\{R_1, \dots, R_M\}$

Gillespie's algorithm computes a **possible evolution** of the system

Gillespie's algorithm

The state of the simulation:

- is a vector representing the multiset of molecules in the chemical solution (initially $[X_1, \dots, X_n]$)
- a real variable t representing the simulation time (initially $t = 0$)

The algorithm iterates the following steps until t reaches a final value t_{stop} .

- 1 The time $t + \tau$ at which the next reaction will occur is randomly chosen with τ exponentially distributed with parameter $\sum_{\nu=1}^M a_{\nu}$;
- 2 The reaction R_{μ} that has to occur at time $t + \tau$ is randomly chosen with probability $a_{\mu} dt$.

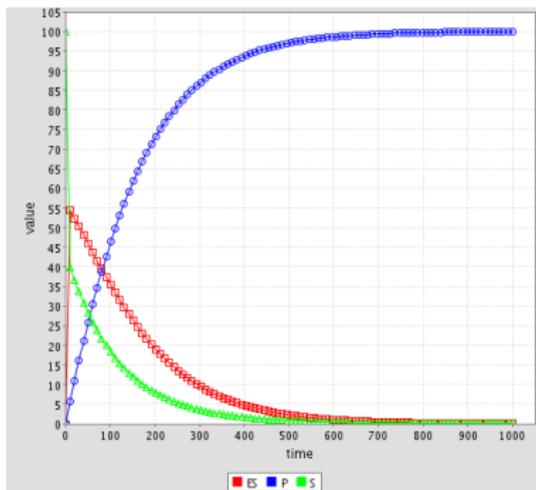
At each step t is incremented by τ and the multiset representing the chemical solution is updated.

ODEs vs SSA

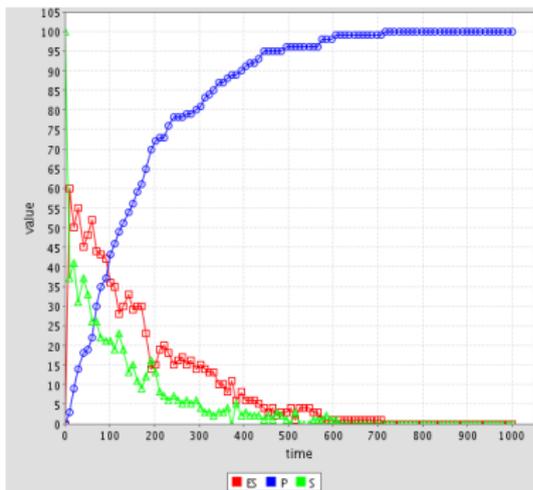
Let us compare the deterministic and stochastic approach with some examples of (bio)chemical reactions:

First example: Enzymatic activity: $E + S \xrightleftharpoons[10.0]{0.3} ES \xrightarrow{0.01} E + P$

Starting with: 100E and 100S.



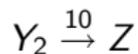
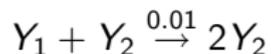
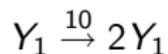
ODEs



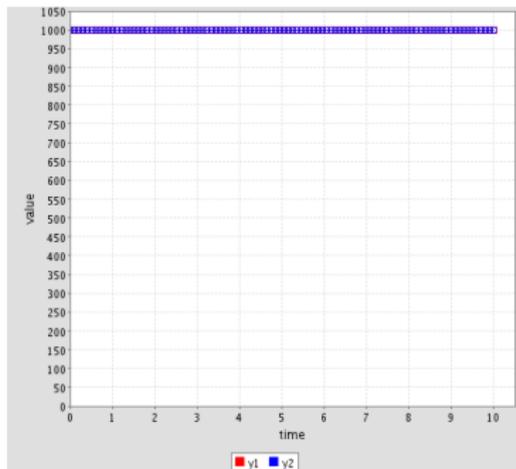
SSA

ODEs vs SSA

Second example: Lotka reactions:



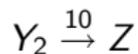
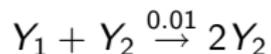
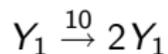
Starting with $1000Y_1$ and $1000Y_2$



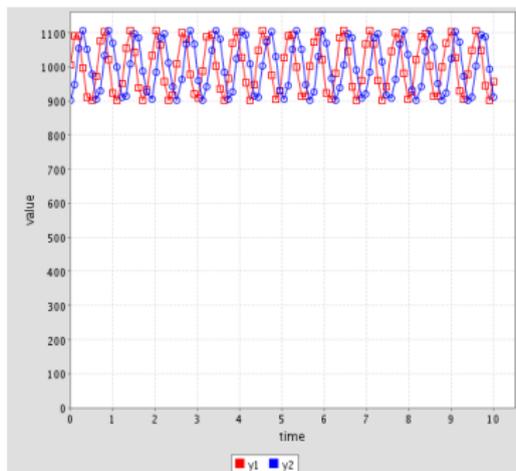
ODEs

ODEs vs SSA

Second example: Lotka reactions:



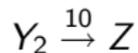
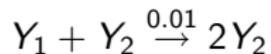
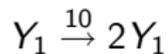
Starting with $1000Y_1$ and $900Y_2$ (slight perturbation).



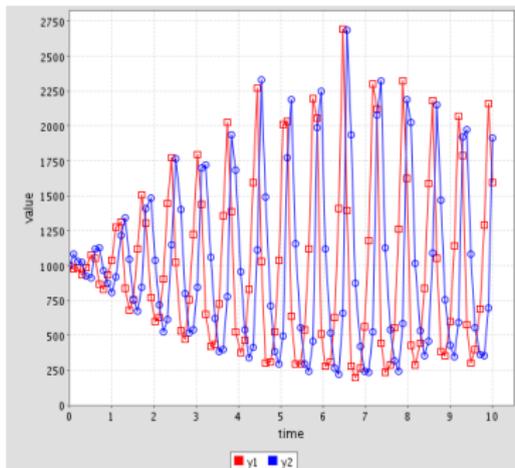
ODEs

ODEs vs SSA

Second example: Lotka reactions:



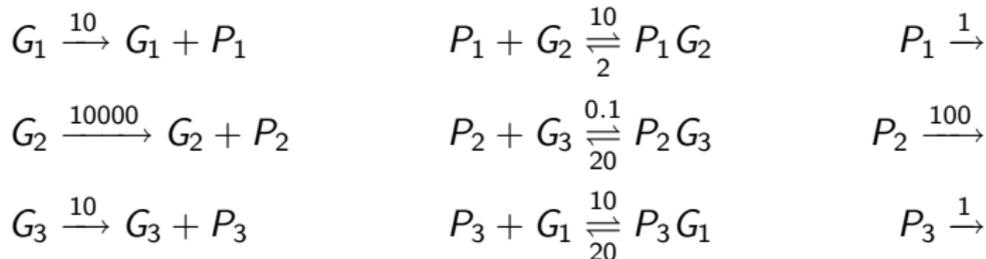
Starting with $1000Y_1$ and $1000Y_2$



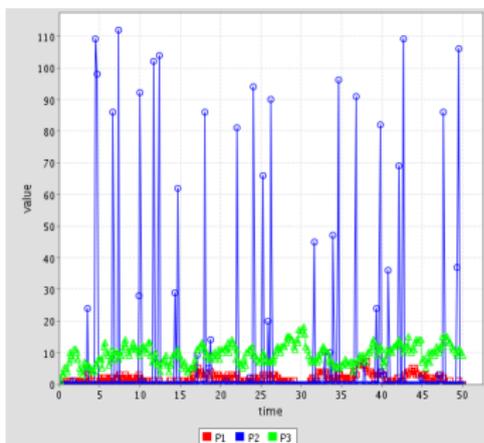
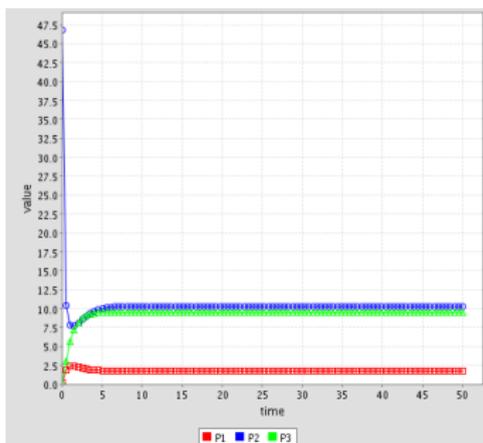
SSA

ODEs vs SSA

Third example: Negative feedback loop:



Starting with $G_i = 1$ and $P_i = 0$



ODEs

SSA

Computational cost of Gillespie's algorithm

The computational cost of this detailed stochastic simulation algorithm might be very high

- the key issue is that the time elapsing between two reactions can be very small

The algorithm becomes very inefficient when:

- there are large number of molecules
- kinetic constant are high

Computational cost is the main disadvantage of stochastic simulation with respect to ODEs

Computational cost of Gillespie's algorithm

Several approximated versions of Gillespie's algorithm aimed at reducing the computational cost have been proposed:

- Gibson and Bruck proposed the use of some efficient data structure to improve the choice of the reaction to happen at each step
- Gillespie proposed the τ -leap method: the key idea is to allow for more reactions to take place in a single (longer) time interval, under the condition that the propensities do not change too much in that interval.
- Hybrid simulation is a technique which combines ODEs with stochastic simulation: ODEs are applied to molecules occurring in big numbers, stochastic simulation to molecules occurring in small numbers