Stochastic Simulation.
(and Gillespie's algorithm)

Alberto Policriti

Dipartimento di Matematica e Informatica
Istituto di Genomica Applicata

Quote of the day

D.T. Gillespie *Exact Stochastic Simulation of Coupled Chemical Reactions* J. of Physical Chemistry, 81(25), 1977

There are two formalisms for mathematically describing the time behavior of a spatially homogeneous chemical system: the deterministic approach regards the time evolution as a continuous, wholly predictable process which is governed by a set of coupled, ordinary differential equations (the "reaction-rate equations"); the stochastic approach regards the time evolution as a kind of random-walk process which is governed by a single differential-difference equation (the "master equation"). Fairly simple kinetic theory arguments show that the stochastic formulation of chemical kinetics has a firmer physical basis than the deterministic formulation, but unfortunately the stochastic master equation is often mathematically intractable. There is, however, a way to make exact numerical calculations within the framework of the stochastic formulation without having to deal with the master equation directly. [...]

Outline

Exponential distribution

An exponential distribution models the time of occurrence of a (simple) random event.

It is given by a random variable T, with values in $[0, \infty)$, with density

$$f(t) = \lambda e^{-\lambda t},$$

where λ is the rate of the exponential distribution.
The probability of the event happening within time t is

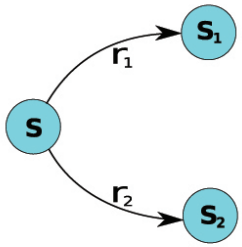
$$P(T \leq t) = 1 - e^{-\lambda t}.$$

Mean: $E[T] = \frac{1}{\lambda}$ Variance: $VAR[T] = \frac{1}{\lambda^2}$

λ is the average density of frequency of events per unit of time.

Continuous Time Markov Chains

What happens if we have more than one event competing?



In this case, there is a race condition between events: the fastest event is executed and modifies globally the state of the system.

Continuous Time Markov Chains

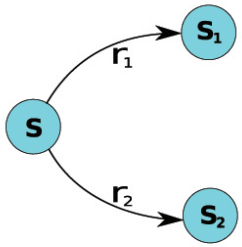
The is a discrete set of states, connected by transitions each with an associated rate of an exponential distribution.

In each state, transitions compete in a race condition: the fastest one determines the new state and the time elapsed.

In the new state, the race condition is started anew (memoryless property).

Continuous Time Markov Chains

What happens if we have more than one event competing?



In this case, there is a race condition between events: the fastest event is executed and modifies globally the state of the system.

Continuous Time Markov Chains

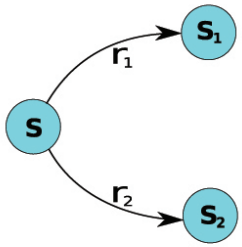
The is a discrete set of states, connected by transitions each with an associated rate of an exponential distribution.

In each state, transitions compete in a race condition: the fastest one determines the new state and the time elapsed.

In the new state, the race condition is started anew (memoryless property).

Continuous Time Markov Chains

What happens if we have more than one event competing?



In this case, there is a race condition between events: the fastest event is executed and modifies globally the state of the system.

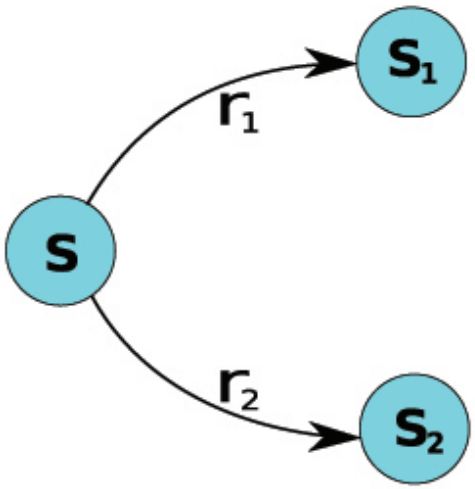
Continuous Time Markov Chains

The is a discrete set of states, connected by transitions each with an associated rate of an exponential distribution.

In each state, transitions compete in a race condition: the fastest one determines the new state and the time elapsed.

In the new state, the race condition is started anew (memoryless property).

Continuous Time Markov Chains



Equivalent Characterization

- In each state, we select the next state according to a *probability distribution* obtained normalizing rates (from S to S_1 with prob. $\frac{r_1}{r_1+r_2}$).
- The time spent in a state is given by an exponentially distributed random variable, with rate given by the *sum of outgoing transitions* from the actual node $(r_1 + r_2)$.

Stochastic model of a chemical system

We have a set of chemical substances $S_1, \ldots, S_N$ contained in a volume V, with

$$X_i \quad = \quad \text{number of molecules of species } i,$$

subject to a set of chemical reactions

$$R_1, \ldots, R_M,$$

where each R_j is of the form

$$R_j : X_{j_1} + X_{j_2} \to X_{j'_1} + \ldots + X_{j'_{p_j}},$$

$$R_j : X_{j_1} \to X_{j'_1} + \ldots + X_{j'_{p_j}},$$

$$R_j : \emptyset \to X_{j'_1} + \ldots + X_{j'_{p_j}},$$

The system is supposed to be in thermal equilibrium

Stochastic model of a chemical system

Key assumption

Each reaction R_j has associated a specific probability rate constant c_j:

$c_j dt$ = probability that a randomly chosen combination of R_j reactant molecules inside V at time t will react according to R_j in the next infinitesimal time interval $[t, t + dt)$.

Key observation

The next reaction that will happen depends only on the current configuration of the system (number of molecules), not on past history (memoryless property).

Stochastic model of a chemical system

Key assumption

Each reaction R_j has associated a specific probability rate constant c_j:

$c_j dt$ = probability that a randomly chosen combination of R_j reactant molecules inside V at time t will react according to R_j in the next infinitesimal time interval $[t, t + dt)$.

Key observation

The next reaction that will happen depends only on the current configuration of the system (number of molecules), not on past history (memoryless property).

Deriving kinetic parameters

Let's focus on a bimolecular reaction...

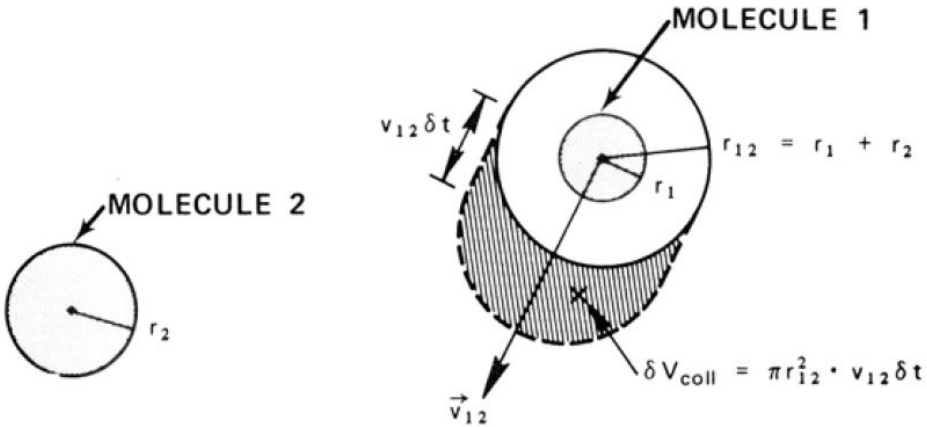


Figure 1. The "collision volume" δV_{coll} which molecule 1 will sweep out relative to molecule 2 in the next small time interval δt.

Problem

"... it is *physically meaningless* to talk about "the number of molecules whose centers lie inside δV_{coll}" in the required limit of vanishingly small δt. ..."

The stochastic model

Under the hypothesis of thermal equilibrium, molecules are uniformly distributed in space, and velocities follow a Boltzmann distribution.

The collision volume swept *on average* is

$$\delta V = \pi r_{12}^2 \langle v_{12} \rangle \delta t.$$

The collision probability is therefore

$$\frac{\delta V}{V} = \frac{\pi r_{12}^2 \langle v_{12} \rangle \delta t}{V}.$$

The stochastic model

Reaction probability

$$p_j \quad =_{def} \quad \text{probability that a } \textit{colliding} \text{ pair of } R_j \text{ reactant molecules will chemically react according to } R_j.$$

The basic rate of reaction c_j is therefore

$$c_j = V^{-1}\pi r_{12}^2 \langle v_{12} \rangle p_j.$$

Rate functions

c_j gives the rate of reaction for a single pair of molecules involved in R_j.

To determine the global rate of reaction R_j, we need to count how many pairs of reacting molecules we have. We do this with the rate function $h_j(c_j, \mathbf{X})$.

Reactants of different species

$$X_{j_1} + X_{j_2} \to \blacksquare \quad h_j(c_j, \mathbf{X}) = c_j X_{j_1} X_{j_2}$$

Reactants of the same species

$$2X_{j_1} \to \blacksquare \quad h_j(c_j, \mathbf{X}) = c_j \frac{X_{j_1}(X_{j_1}-1)}{2}$$

Chemical Master Equation

From rate functions, we can derive a differential equations saying how the probability of being in different states (of having different number of molecules) varies over time. It is called the chemical master equation:

$$\frac{dP(\mathbf{X},t)}{dt} = \sum_{j=1}^{M} \Bigg(\underbrace{h_j(c_j, \mathbf{X}-\nu_j)P(\mathbf{X}-\nu_j,t)}_{\text{A reaction } R_j \text{ happened in time } [t,t+dt]} - \underbrace{h_j(c_j,\mathbf{X})P(\mathbf{X},t)}_{\text{No reaction happened in } [t,t+dt]} \Bigg)$$

Pro

This equation is *everything we need to know* about the stochastic process.

Cons

This equation is *very difficult to solve*, even numerically.

The approach of Gillespie

The central notion becomes the following definition of reaction probability density function:

$$P(\tau, j) \quad =_{def} \quad \begin{array}{l}\text{probability that, given the state } \mathbf{X} = \\ (X_1, \ldots, X_N) \text{ at time } t, \text{ the } \textit{next} \text{ reac-} \\ \text{tion in } V \text{ will occur in the infinitesimal} \\ \text{time interval } (t+\tau, t+\tau+dt), \textit{and} \text{ will} \\ \text{be an } R_j \text{ reaction}\end{array}$$

The approach of Gillespie

Explicit form of $P(\tau, j)$

$$P(\tau, j) = \underbrace{h_0(\mathbf{X}) e^{-h_0(\mathbf{X})\tau}}_{\text{time elapsed}} \cdot \underbrace{\frac{h_j(c_j, \mathbf{X})}{h_0(\mathbf{X})}}_{\text{next reaction}}$$

Intuitively...

The equation says that the next reaction is chosen with probability $\frac{h_j}{h_0}$, while the time elapsed to see this reaction happen is exponentially distributed with rate h_0.

This stochastic process is a Continuous Time Markov Chain.

Numerically simulating $P(\tau, \mu)$

Numerically simulating $P(\tau, \mu)$

A random number generator can be used to draw random pairs (τ, μ) whose probability density function is $P(\tau, \mu)$.
Given r_1 and r_2 randomly generated, determine τ and μ such that:

$$\tau = (1/h_0)\log(1/r_1) \qquad \Sigma_{\nu=1}^{j-1} h_\nu < r_2 h_0 \leq \Sigma_{\nu=1}^{j} h_\nu$$

The method

A general Monte Carlo technique called inversion method: x will be randomly drawn with probability density function $P(x)$ if $x = F^{-1}(r)$ with r randomly drawn with uniform probability density function in $[0, 1]$ and F is the probability distribution function ($\int_{-\infty}^{x} P(y)dy$).

The algorithm

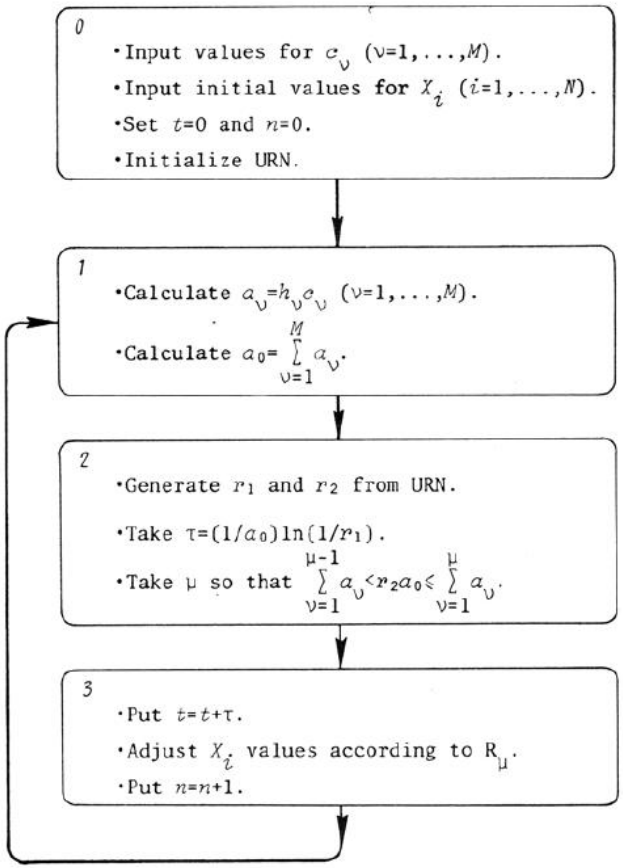
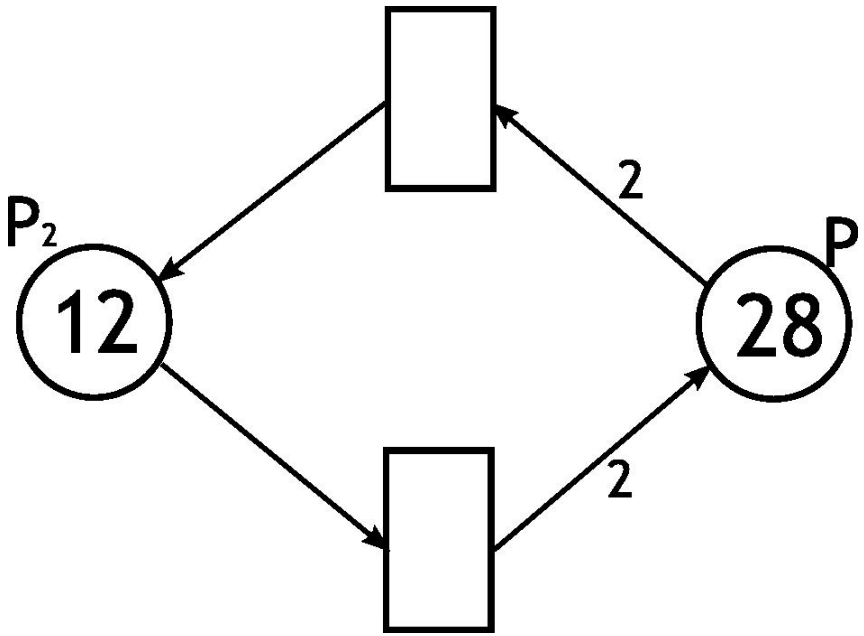


Figure 2. Schematic of the stochastic simulation algorithm.

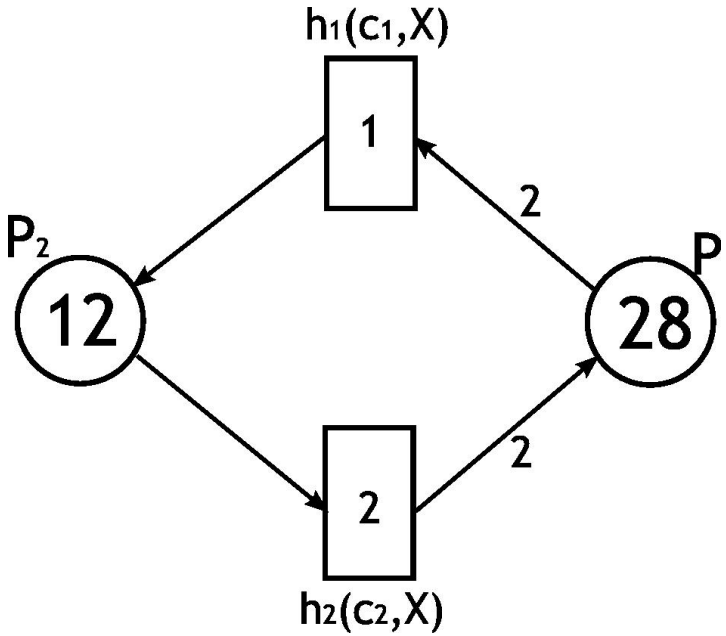
Gillespie Algorithm and Petri Nets



$$2P \leftrightarrow P_2$$

$h_1(1, \mathbf{X}) = 1 * (27 * 28)/2 = 378$
$h_2(2, \mathbf{X}) = 2 * 12 = 24$
$h_0(X) = 378 + 24 = 402$
$p_1 = 0.94$
$p_2 = 0.06$
$\text{time} = 1/402$

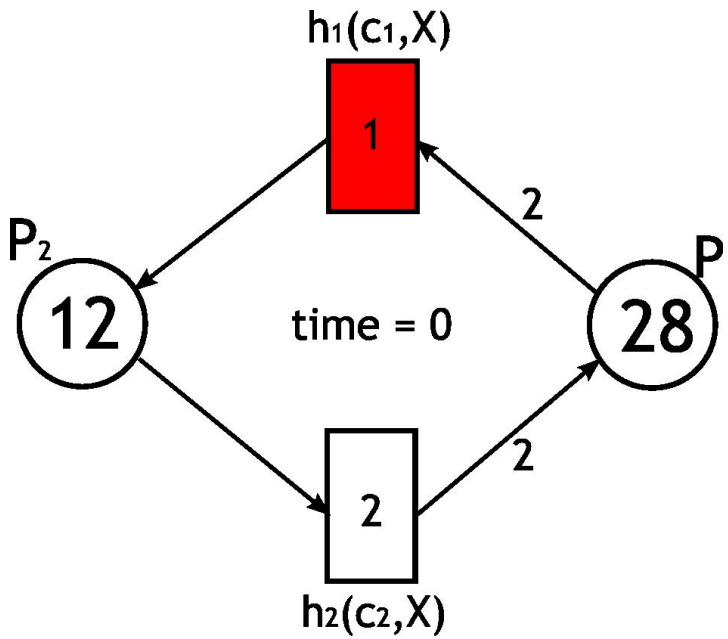
Gillespie Algorithm and Petri Nets



$$2P \leftrightarrow P_2$$

$h_1(1, \mathbf{X}) = 1 * (27 * 28)/2 = 378$
$h_2(2, \mathbf{X}) = 2 * 12 = 24$
$h_0(X) = 378 + 24 = 402$
$p_1 = 0.94$
$p_2 = 0.06$
$\text{time} = 1/402$

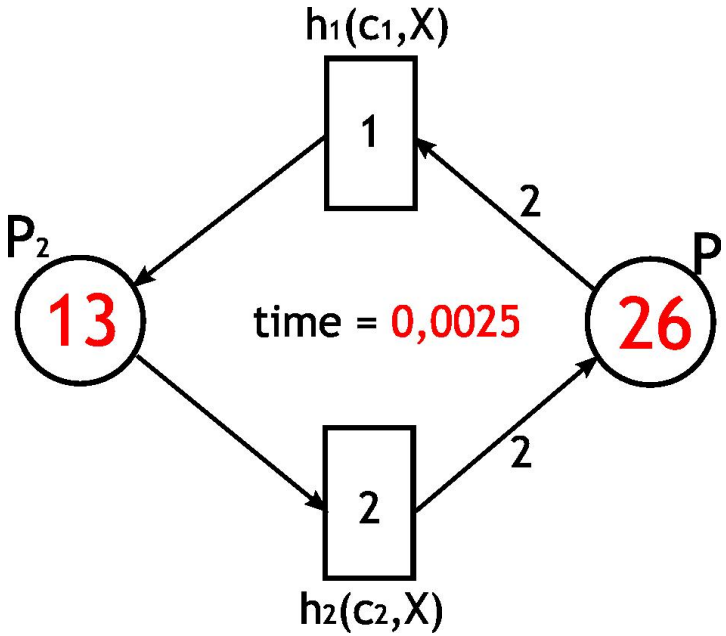
Gillespie Algorithm and Petri Nets



$$2P \leftrightarrow P_2$$

$h_1(1, \mathbf{X}) = 1 * (27 * 28)/2 = 378$
$h_2(2, \mathbf{X}) = 2 * 12 = 24$
$h_0(X) = 378 + 24 = 402$
$p_1 = 0.94$
$p_2 = 0.06$
$\text{time} = 1/402$

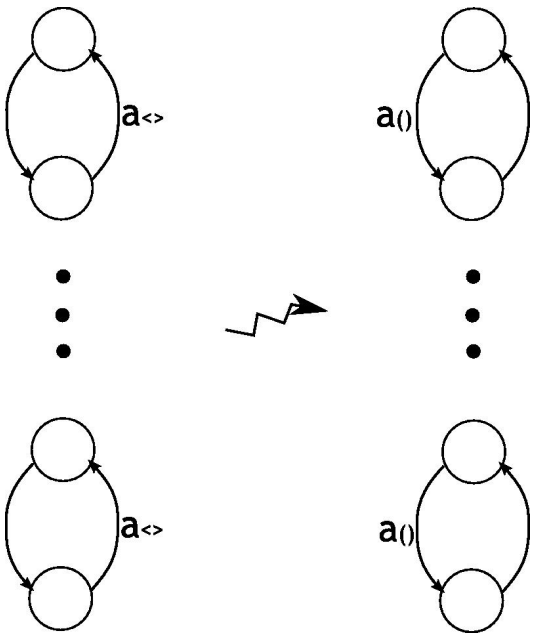
Gillespie Algorithm and Petri Nets



$$2P \leftrightarrow P_2$$

$h_1(1, \mathbf{X}) = 1 * (27 * 28)/2 = 378$
$h_2(2, \mathbf{X}) = 2 * 12 = 24$
$h_0(X) = 378 + 24 = 402$
$p_1 = 0.94$
$p_2 = 0.06$
$\text{time} = 1/402$

Gillespie Algorithm and π-calculus



Each channel has a basic rate λ associated to it.

The global rate of a channel depends on how many agents are ready to communicate on it.

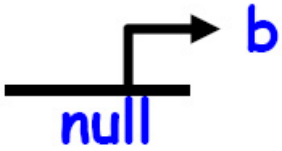
In this example:

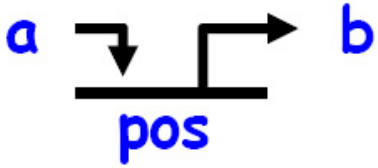
$$\lambda MN$$

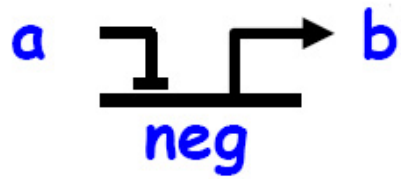
The functions h_i are determined implicitly by the semantics of the language. Gillespie can be used to simulate stochastic π-calculus as well!!!

An example: genetic regulatory networks

Genes as logical gates







Repressilator