Module 11: Stochastic simulations of chemical reaction networks

CSE590: Molecular programming and neural computation. Slides courtesy of Niles Pierce (Caltech).

Gillespie's stochastic formulation: species and reaction notation

Consider *N* chemically active species $S_i \ (i = 1, N)$ each with population X_i in volume *V*

These species can interact via *M* types of unidirectional chemical reactions, $R_\mu \ (\mu = 1, M)$

For example

$$
\begin{aligned}
\emptyset &\rightarrow \text{products} \\
S_i &\rightarrow \text{products} \\
S_i + S_j &\rightarrow \text{products} \ (i \neq j) \\
2S_i &\rightarrow \text{products} \\
S_i + S_j + S_k &\rightarrow \text{products} \ (i \neq j \neq k \neq i) \\
S_i + 2S_j &\rightarrow \text{products} \ (i \neq j) \\
3S_i &\rightarrow \text{products}
\end{aligned}
$$

Gillespie's stochastic formulation: species and reaction notation

Consider *N* chemically active species $S_i \ (i = 1, N)$ each with population X_i in volume *V*

These species can interact via *M* types of unidirectional chemical reactions, $R_\mu \ (\mu = 1, M)$

For example

$$
\begin{aligned}
\emptyset &\rightarrow \text{products} \\
S_i &\rightarrow \text{products} \\
S_i + S_j &\rightarrow \text{products} \ (i \neq j) \\
2S_i &\rightarrow \text{products} \\
S_i + S_j + S_k &\rightarrow \text{products} \ (i \neq j \neq k \neq i) \\
S_i + 2S_j &\rightarrow \text{products} \ (i \neq j) \\
3S_i &\rightarrow \text{products}
\end{aligned}
$$

D.T. Gillespie, *J Comput Phys*, **22**, 403-434,1976
D.T. Gillespie, *J Phys Chem*, **81**, 2340-2361, 1977

Distinct reactant combinations

For each reaction type R_μ, let h_μ denote the number of distinct molecular reactant combinations given that the system is in state $(X_1, \ldots, X_N)$

Reactants for reaction R_μ		Distinct reactant combinations h_μ
$\emptyset$		1
S_i		X_i
$S_i + S_j$	$(i \neq j)$	$X_i X_j$
$2S_i$		$X_i(X_i - 1)/2$
$S_i + S_j + S_k$	$(i \neq j \neq k \neq i)$	$X_i X_j X_k$
$S_i + 2S_j$	$(i \neq j)$	$X_i X_j (X_j - 1)/2$
$3S_i$		$X_i(X_i - 1)(X_i - 2)/6$

Well-mixed requirement

Assumption: for each reaction type R_μ, the average probability (over all distinct reactant combinations h_μ) that a particular combination of reactant molecules in V will react in the next infinitesimal time interval *dt* may be expressed in the form

$$c_\mu dt$$

where c_μ is a "reaction parameter" representing a "reaction probability per unit time"

Validity: Based on collision theory, the assumption can be justified *if the reactive molecules are randomly distributed with a uniform distribution prior to each reaction*

Hence, the stochastic formulation is valid for "well-mixed" systems in which *nonreactive molecular collisions are much more frequent than inelastic reactive molecular collisions*

It then follows that the probability of a reaction of type R_μ occurring in infinitesimal time interval $(t, t+dt)$ is

$$a_\mu dt \qquad \text{with} \qquad a_\mu \equiv h_\mu c_\mu$$

given state $(X_1, \ldots, X_N)$ at time t

Reaction probability density function

We wish to simulate stochastic trajectories of the chemical system starting from given initial molecular populations based on the reaction probability density function $P(\tau, \mu)$ such that

$$P(\tau, \mu)d\tau$$

denotes the probability that, given the state $(X_1, \ldots, X_N)$ at time t, the next reaction in V will occur in the infinitesimal time interval $(t + \tau, t + \tau + d\tau)$ and will be an R_μ reaction

This may be calculated as the product

$$P(\tau, \mu)d\tau = P_0(\tau) \cdot a_\mu d\tau$$

$P_0(\tau)$ is the probability that, given state $(X_1, \ldots, X_N)$ at time *t*, no reaction will occur in time interval $(t, t + \tau)$

$a_\mu d\tau$ is the probability that, given state $(X_1, \ldots, X_N)$ at time $t + \tau$, an R_μ reaction will occur in infinitesimal time interval $(t + \tau, t + \tau + d\tau)$

Reaction probability density function

To find $P_0(\tau)$ note that the probability that *no reaction* will occur in *infinitesimal* time interval $(\tau, \tau + d\tau)$ is given by

$$1 - a_0 d\tau \qquad \text{with} \qquad a_0 \equiv \sum_{\mu=1}^{M} a_\mu$$

given state $(X_1, \ldots, X_N)$ at time τ

Hence

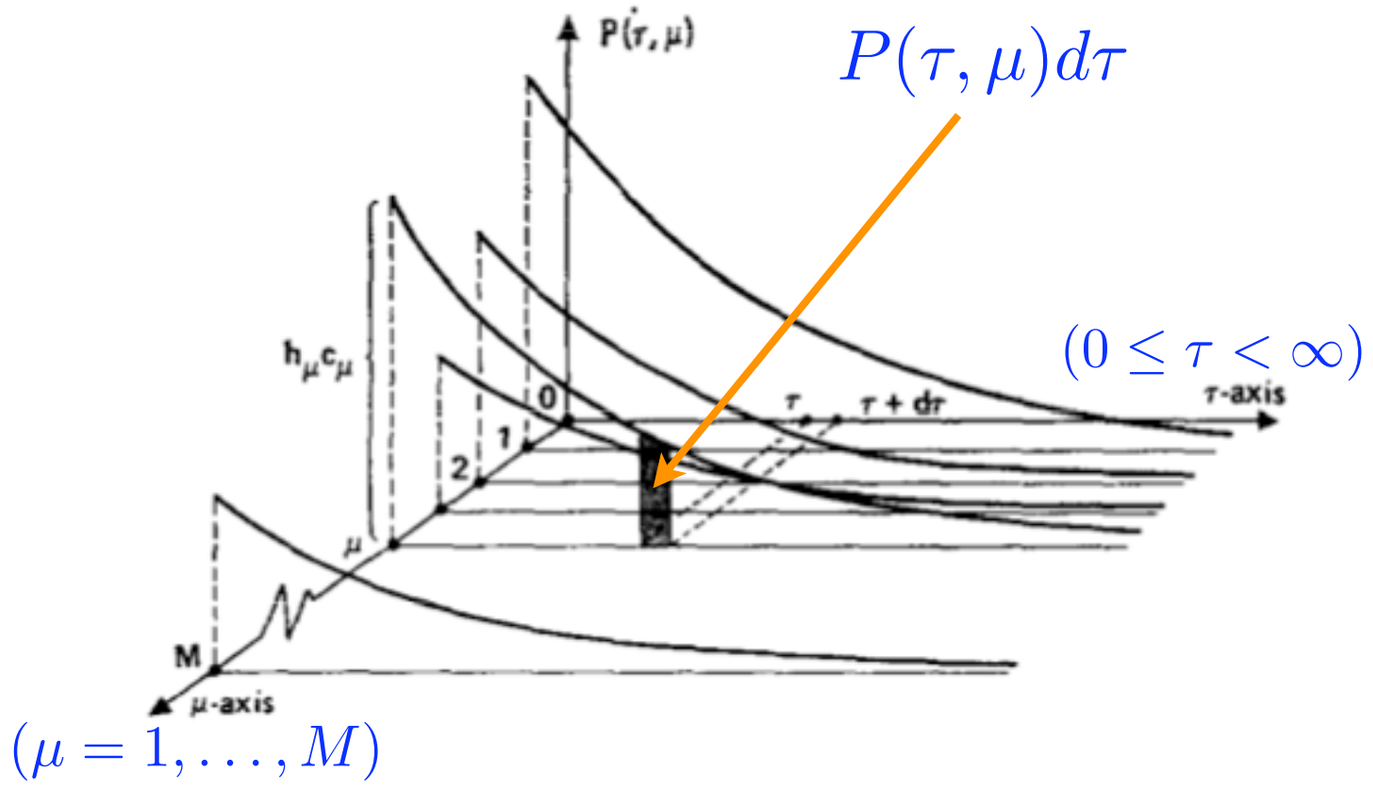
$$P_0(\tau + d\tau) = P_0(\tau) \cdot (1 - a_0 d\tau)$$

from which we deduce

$$P_0(\tau) = \exp(-a_0 \tau)$$

This yields the form of the desired reaction probability density function

$$P(\tau, \mu) = a_\mu \cdot \exp(-a_0 \tau)$$



Gillespie's stochastic simulation algorithm: overview

Step 0 (Initialization): Set $t = 0$, set the reaction parameters $(c_1, \ldots, c_M)$ and the initial molecular populations $(X_1, \ldots, X_N)$ and calculate $(a_1, \ldots, a_M)$

Step 1 (Monte Carlo): generate a random reaction time and type (τ, μ) according to probability density function $P(\tau, \mu)$

Step 2 (Update): advance t by τ, update the molecular populations $(X_1, \ldots, X_N)$ to reflect the occurrence of one R_μ reaction, update $(a_1, \ldots, a_M)$

Step 3 (Terminate): If $t > t_{\text{stop}}$ or all reactants have been consumed ($a_0 = 0$) then terminate, else go to Step 1

Gillespie's stochastic simulation algorithm: overview

Step 0 (Initialization): Set $t = 0$, set the reaction parameters $(c_1, \ldots, c_M)$ and the initial molecular populations $(X_1, \ldots, X_N)$ and calculate $(a_1, \ldots, a_M)$

Step 1 (Monte Carlo): generate a random reaction time and type (τ, μ) according to probability density function $P(\tau, \mu)$

Step 2 (Update): advance t by τ, update the molecular populations $(X_1, \ldots, X_N)$ to reflect the occurrence of one R_μ reaction, update $(a_1, \ldots, a_M)$

Step 3 (Terminate): If $t > t_{\text{stop}}$ or all reactants have been consumed ($a_0 = 0$) then terminate, else go to Step 1

D.T. Gillespie, *J Comput Phys*, **22**, 403-434, 1976
D.T. Gillespie, *J Phys Chem*, **81**, 2340-2361, 1977

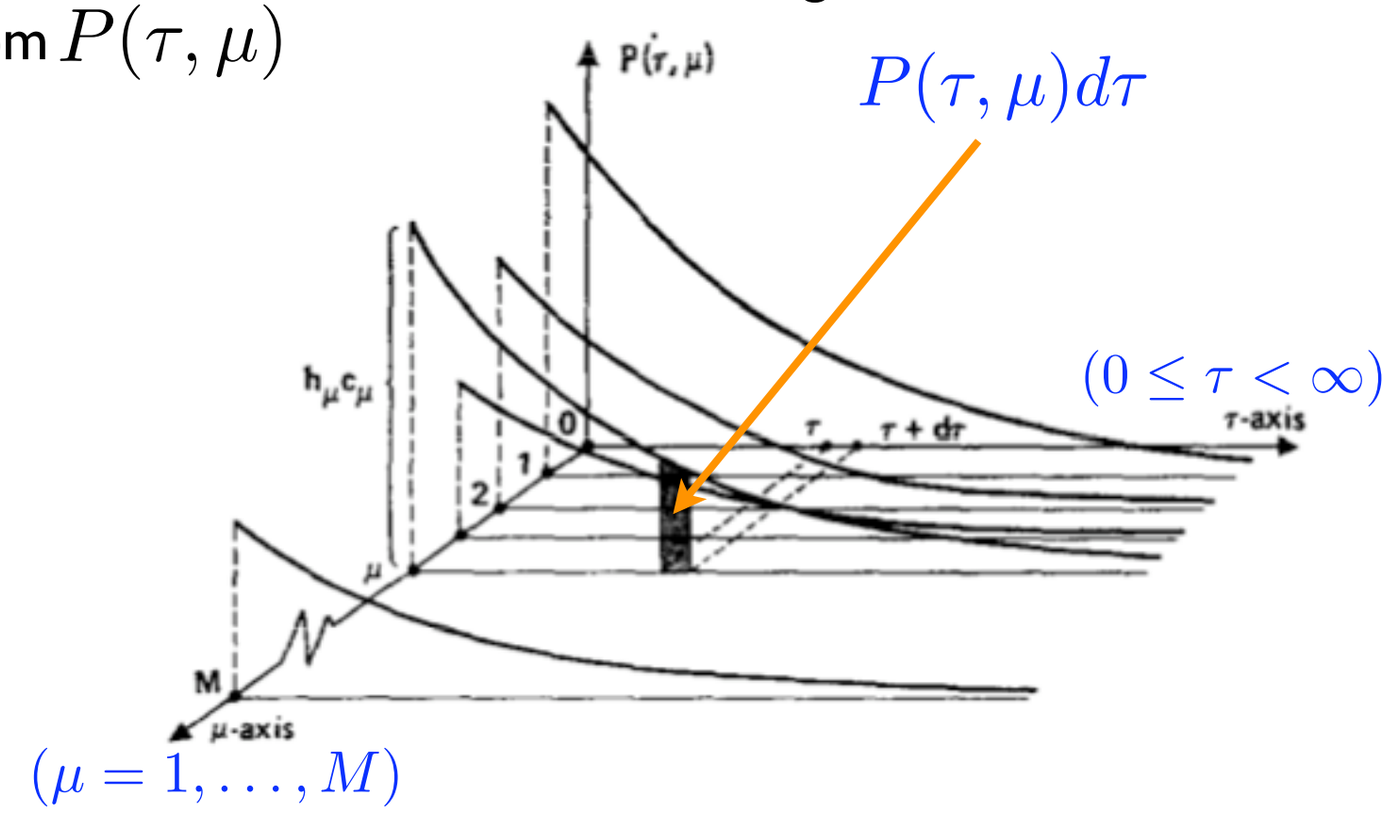
Implementing the Monte Carlo Step

A standard "unit-interval uniform random number generator" produces a sample r drawn from a uniform distribution on $[0, 1]$

We need to sample the two-variable probability density function

$$P(\tau, \mu) = a_\mu \cdot \exp(-a_0 \tau)$$

Let's figure out a way to use two calls to the standard random number generator to obtain one sample (τ, μ) drawn from $P(\tau, \mu)$



Implementing the Monte Carlo Step

Let

$$P(\tau, \mu) = P_1(\tau) \cdot P_2(\mu|\tau)$$

$P_1(\tau)d\tau$ is the probability that the next reaction (irrespective of type) will occur in the infinitesimal time interval $(t + \tau, t + \tau + d\tau)$

$P_2(\mu|\tau)$ is the probability that the next reaction will be an R_μ reaction given that it occurs at time τ

Implementing the Monte Carlo Step

Let

$$P(\tau, \mu) = P_1(\tau) \cdot P_2(\mu|\tau)$$

$P_1(\tau)d\tau$ is the probability that the next reaction (irrespective of type) will occur in the infinitesimal time interval $(t + \tau, t + \tau + d\tau)$

$P_2(\mu|\tau)$ is the probability that the next reaction will be an R_μ reaction given that it occurs at time τ

Recalling

$$P(\tau, \mu) = a_\mu \cdot \exp(-a_0\tau)$$

we find

$$P_1(\tau) = \sum_{\mu=1}^{M} P(\tau, \mu) = a_0 \exp(-a_0\tau), \quad (0 \leq \tau < \infty)$$

$$P_2(\mu|\tau) = P(\tau, \mu)/P_1(\tau) = a_\mu/a_0, \quad (\mu = 1, \ldots, M)$$

Implementing the Monte Carlo Step

Generate random time τ according to $P_1(\tau)$ and random reaction type μ according to $P_2(\mu|\tau)$ to obtain (τ, μ) distributed according to $P(\tau, \mu)$

Step 1a (pick the reaction time): generate a unit-interval uniformly distributed random number r_1 and set

$$\tau = (1/a_0)\ln(1/r_1)$$

Step 1b (pick the reaction type): generate a unit-interval uniformly distributed random number r_2 and set μ to be the integer for which

$$\sum_{\nu=1}^{\mu-1} a_\nu < r_2 a_0 \leq \sum_{\nu=1}^{\mu} a_\nu$$

D.T. Gillespie, *J Comput Phys*, **22**, 403-434,1976
D.T. Gillespie, *J Phys Chem*, **81**, 2340-2361, 1977

Comparing deterministic and stochastic formulations

Deterministic formulation

- The concentration of each molecular species is uniform throughout the reaction volume
- Concentrations vary continuously with time
- Each reaction is characterized by a "reaction rate constant"
- Fluctuation and correlation effects are neglected
- Valid in the dilute thermodynamic limit (i.e., when the number of molecules of each species and the reaction volume approach infinity such that molecular concentrations are finite, and not too large)

Stochastic formulation

- The molecules of each species are distributed randomly throughout the reaction volume with a uniform distribution
- The population of each molecular species changes discretely as reactions occur
- Each reaction is characterized by a "reaction probability per unit time"
- Fluctuation and correlation effects are included
- Valid for "well-mixed" systems (e.g., when elastic nonreactive molecular collisions vastly outnumber inelastic reactive molecular collisions)

The deterministic and stochastic formulations are equivalent in the thermodynamic limit

D.T. Gillespie, *J Comput Phys*, **22**, 403-434,1976
D.T. Gillespie, *J Phys Chem*, **81**, 2340-2361, 1977

Comparing reaction parameters to reaction rate constants

Suppose the stochastic formulation is described in terms of molecular populations X_i and the deterministic formulation is described in terms of molecular concentrations

$$x_i \equiv X_i/V$$

Consider a reaction of type R_μ, for example the bimolecular reaction:

$$S_i + S_j \rightarrow \text{products } (i \neq j)$$

What is the relationship between the stochastic "reaction parameter" c_μ and the more familiar deterministic "reaction rate constant" k_μ ?

Comparing reaction parameters to reaction rate constants

Recall that $a_\mu dt$ represents the probability of a reaction of type R_μ in the infinitesimal time interval $(t, t+dt)$ given populations $(X_1, \ldots, X_N)$ at time t

For the reaction type in question, the number of distinct reactant combinations is $h_\mu = X_i X_j$ so we have

$$a_\mu dt = h_\mu c_\mu dt = X_i X_j c_\mu dt$$

If we average over an ensemble of reaction volumes containing stochastically identical systems the *average rate* at which R_μ reactions are occurring inside V is

$$\langle X_i X_j c_\mu \rangle = \langle X_i X_j \rangle \, c_\mu$$

so the *average reaction rate per unit volume* expressed in terms of *concentrations* is

$$\langle X_i X_j \rangle \, c_\mu / V = \langle x_i x_j \rangle \, V c_\mu$$

D.T. Gillespie, *J Comput Phys*, **22**, 403-434,1976
D.T. Gillespie, *J Phys Chem*, **81**, 2340-2361, 1977

Comparing reaction parameters to reaction rate constants

The *reaction rate constant* is conventionally defined to be the *average reaction rate per unit volume* divided by the *product of the average reactant concentrations*

$$k_\mu = \langle x_i x_j \rangle V c_\mu / \langle x_i \rangle \langle x_j \rangle$$

The deterministic formulation imposes

$$\langle x_i x_j \rangle = \langle x_i \rangle \langle x_j \rangle$$

thus neglecting *fluctuations* $(i = j)$ and *correlations* $(i \neq j)$

We therefore find for this particular reaction type R_μ

$$k_\mu = V c_\mu$$

This relationship changes for each reaction type depending on the form of h_μ, representing the distinct reactant combinations for reaction type R_μ

D.T. Gillespie, *J Comput Phys*, **22**, 403-434,1976
D.T. Gillespie, *J Phys Chem*, **81**, 2340-2361, 1977