

Dynamical Systems and Chaos
Part II: Biology Applications

**Lecture 12: Stochastic dynamical
systems.**

Ilya Potapov
Mathematics Department, TUT
Room TD325

- ▶ **Deterministic differential equations (ODEs)**
 - ▶ In many cases, the time-evolution of a chemically reacting system can be treated as continuous-state, deterministic process.
 - ▶ ODEs characterize the time-evolution of chemical species as continuous deterministic process.
 - ▶ Qualitative method. Can be used when concentrations are large.

▶ Deterministic differential equations (ODEs)

- ▶ Each equation expresses the time-rate-of-change of the continuous molecular concentration of one chemical species as a function of the molecular concentrations of all the species.
- ▶ Rate constants are viewed as "reaction rates".
- ▶ Fast to simulate. Produce the same result for the same initial conditions.
- ▶ $X_1 + X_2 \xrightarrow{k_1} X_3$,
 $v = \frac{d[X_3]}{dt} = -\frac{d[X_1]}{dt} = -\frac{d[X_2]}{dt} = k_1[X_1][X_2]$

$$d\mathbf{X}(t) = \mathbf{S}\mathbf{v}(\mathbf{K}, \mathbf{X}(t))dt$$

- ▶ \mathbf{X} is the vector of variables ($[X_1]$, $[X_2]$, and $[X_3]$).
- ▶ \mathbf{S} is the stoichiometric matrix.
- ▶ \mathbf{v} is the vector of reaction rates.
- ▶ \mathbf{K} is the matrix of rate constants.

▶ Stochastic modeling

- ▶ A chemical reaction occurs whenever two or more molecules of appropriate kinds collide in an appropriate way.
- ▶ Collisions in a system of molecules in thermal equilibrium occur in a random way.
- ▶ **Gillespie stochastic simulation algorithm (SSA)** characterizes the time-evolution of chemical species as discrete-state stochastic process. SSA numerically simulates the Markov process that the **Chemical Master Equation** describes analytically.
 - ▶ The time-evolution of chemical species takes the form of a Markovian random walk in the M -dimensional space of molecular populations.
 - ▶ Exact method to solve chemical reaction systems.
 - ▶ It is always valid when the deterministic approach is valid and sometimes valid when the deterministic approach is not valid.
 - ▶ Needs to be used when small numbers of molecules.
 - ▶ Time consuming to solve larger systems.
 - ▶ There are several ways to speed up the SSA, but we discuss only the methods based on stochastic differential equations.

▶ **Stochastic modeling**

- ▶ **Stochastic differential equations (SDEs)** characterize the time-evolution of chemical species as continuous-state stochastic process.
 - ▶ Approximate method.
 - ▶ Cannot be used when concentrations are small, good for larger concentrations when stochasticity is still important.
 - ▶ M chemical species $\rightarrow M$ differential equations.

Fluctuations in dynamical systems

Noise exists in all dissipative systems. Thus, variables describing the state of the system with noise are random entities. Evolution of the system is a **stochastic process**.

Intrinsic noise

The system usually consists of many micro-particles. Macro-level variables of the system is always subject to heat fluctuations. Finite number of variables in a mathematical model and quantum nature of micro-particles are also sources of noise.

Extrinsic noise

On the other hand, the dynamical system describes a real process at some desired level. The system is usually separated from the surrounding environment. In this case, extrinsic noise influence the system (external stochastic force).

Two fundamental approaches

1. Stochastic differential equations

$$\dot{x}_\xi = f(x_\xi, \xi(t)),$$

where $\xi(t)$ is sequence of randomly generated numbers. Thus, evolution operator becomes random in time: mapping $x_\xi \rightarrow x_\xi(t + dt)$ depends on concrete choice of $\xi(t)$ and, hence, $x_\xi(t)$ is a stochastic process.

The same initial conditions give different resulting trajectory $x_\xi(t)$. \Rightarrow one should consider *statistical ensemble* which is determined by the perturbing source $\xi_j(t)$, $j = 1, 2, \dots, N$. For $N \rightarrow \infty$ statistical ensemble determines the stochastic process.

Two fundamental approaches

2. Probability and probability density distributions

One can measure realizations of a stochastic process x_1, x_2, \dots, x_n in time moments t_1, t_2, \dots, t_n and then take joint probability $p(x_1, t_1; x_2, t_2; \dots; x_n, t_n)$ of their appearance. Evolution of the system is based on deterministic equations with probabilities and probability densities as variables.

$$\frac{\partial}{\partial t} p = \mathbf{L} p$$

\mathbf{L} is linear evolution operator. We, as most of others, will consider \mathbf{L} which “does not remember previous states” (Markovian process).

Chemical Master Equation (CME)

Notation needed to formulate CME:

- ▶ The function $\mathbf{X}: \mathbb{R}_+ \rightarrow \mathbb{Z}^M$ describes the M variables (numbers of molecules of each chemical species).
- ▶ The system is in state $\mathbf{X}(t) = \mathbf{x}$ at time t meaning that there is \mathbf{x} numbers of molecules of each chemical species at time t .
- ▶ Reaction constants are viewed as "reaction probabilities per unit time" instead of "reaction rates":
 - ▶ c_μ characterizes reaction R_μ , $\mu \in \{1, \dots, N\}$:

$c_\mu dt :=$ the average probability that a randomly selected combination of reaction R_μ reactants in V at time t will react accordingly in the infinitesimal time interval $[t, t + dt)$.

The value for c_μ can be found using molecular collision rate and Maxwell's velocity. However, the relationship between deterministic rate constant k and stochastic reaction constant c is rather simple.

Other entities needed are

$h_\mu(\mathbf{x}) :=$ the number of possible reactant combinations of a reaction R_μ in V at time t with $\mathbf{X}(t) = \mathbf{x}$.

and

$\nu_\mu :=$ the state-change vector (stoichiometric vector) defining the change in the state $\mathbf{X}(t) = \mathbf{x}$ caused by a reaction R_μ .

The propensity function a_μ is defined as

$a_\mu(c_\mu, \mathbf{x})dt :=$ the probability that a reaction R_μ in V will occur in the infinitesimal time interval $[t, t + dt)$ given that the system is in the state $\mathbf{X}(t) = \mathbf{x}$ at time t ,

meaning that

$$a_\mu(c_\mu, \mathbf{x}) := c_\mu h_\mu(\mathbf{x}).$$

The stochastic formulation proceeds by considering the **probability function** $P(\mathbf{X}(t) = \mathbf{x} \mid \mathbf{X}(t_0) = \mathbf{x}_0)$ marked here briefly as $P(\mathbf{x}, t \mid \mathbf{x}_0, t_0)$:

$$P(\mathbf{x}, t \mid \mathbf{x}_0, t_0) := \text{the probability to be in the state } \mathbf{X}(t) = \mathbf{x} \\ \text{in } V \text{ at time } t \text{ given that the system is} \\ \text{in the state } \mathbf{X}(t_0) = \mathbf{x}_0 \text{ at time } t_0. \quad (1)$$

In the infinitesimal time interval $[t, t + dt)$ either zero or one reaction occurs. Thus, there exists $N + 1$ distinct, mutually exclusive routes (N reactions) from the state $\mathbf{X}(t) = \mathbf{x} - \nu_\mu$ ($\mu \in \{1, \dots, N\}$) or $\mathbf{X}(t) = \mathbf{x}$ (in case of no reaction) to the state $\mathbf{X}(t + dt) = \mathbf{x}$.

The probability that any of these routes takes place can be calculated by summing their distinct probabilities.

The probability $P(\mathbf{x}, t + dt | \mathbf{x}_0, t_0)$ can be written as follows

$$P(\mathbf{x}, t + dt | \mathbf{x}_0, t_0) = P(\mathbf{x}, t | \mathbf{x}_0, t_0)P(\text{no reaction in } [t, t + dt)) \\ + \sum_{\mu=1}^N P(\mathbf{x} - \boldsymbol{\nu}_\mu, t | \mathbf{x}_0, t_0)P(\text{one reaction } R_\mu \text{ in } [t, t + dt)).$$

Since in the first term no reaction is occurring in the infinitesimal time interval $[t, t + dt)$, it means that the system has to be in state \mathbf{x} already at time t . Thus, in the first term, the probability that the system was in the state $\mathbf{X}(t_0) = \mathbf{x}_0$ at t_0 and is in the state $\mathbf{X}(t) = \mathbf{x}$ at t is multiplied by the probability that no reaction occurs in the infinitesimal time interval $[t, t + dt)$.

The second term is the sum of N routes where one reaction R_μ occurs in the infinitesimal time interval $[t, t + dt)$. The probability that the system was in the state $\mathbf{X}(t_0) = \mathbf{x}_0$ at t_0 and is in the state $\mathbf{X}(t) = \mathbf{x} - \boldsymbol{\nu}_\mu$ at t is multiplied by the probability that one reaction R_μ (which will change state from $\mathbf{x} - \boldsymbol{\nu}_\mu$ to \mathbf{x}) occurs in the infinitesimal time interval $[t, t + dt)$.

It can be seen that

$$P(\text{no reaction in } [t, t + dt]) = 1 - \sum_{\mu=1}^N a_{\mu}(c_{\mu}, \mathbf{x})dt$$

and

$$P(\text{one reaction } R_{\mu} \text{ in } [t, t + dt]) = a_{\mu}(c_{\mu}, \mathbf{x} - \boldsymbol{\nu}_{\mu})dt.$$

Now we can write

$$\begin{aligned} P(\mathbf{x}, t + dt | \mathbf{x}_0, t_0) &= P(\mathbf{x}, t | \mathbf{x}_0, t_0) \left(1 - \sum_{\mu=1}^N a_{\mu}(c_{\mu}, \mathbf{x})dt \right) \\ &+ \sum_{\mu=1}^N P(\mathbf{x} - \boldsymbol{\nu}_{\mu}, t | \mathbf{x}_0, t_0) a_{\mu}(c_{\mu}, \mathbf{x} - \boldsymbol{\nu}_{\mu})dt. \end{aligned}$$

The **Chemical Master Equation** is obtained

$$\begin{aligned}\frac{\partial P(\mathbf{x}, t | \mathbf{x}_0, t_0)}{\partial t} &\equiv \lim_{dt \rightarrow 0} \frac{P(\mathbf{x}, t + dt | \mathbf{x}_0, t_0) - P(\mathbf{x}, t | \mathbf{x}_0, t_0)}{dt} \\ &= \sum_{\mu=1}^N [a_{\mu}(c_{\mu}, \mathbf{x} - \boldsymbol{\nu}_{\mu})P(\mathbf{x} - \boldsymbol{\nu}_{\mu}, t | \mathbf{x}_0, t_0) \\ &\quad - a_{\mu}(c_{\mu}, \mathbf{x})P(\mathbf{x}, t | \mathbf{x}_0, t_0)].\end{aligned}$$

Chemical master equation describes the time evolution of the probability of a chemical system to occupy each one of the discrete set of states.

The chemical master equation is usually intractable, both analytically and numerically. The **Gillespie stochastic simulation algorithm** is a computer algorithm for numerical simulation of the process that chemical master equation describes.

Simple example

The CME for reaction



can be given with the help of (x is the number of molecules A)

$$x + 1 \rightarrow x \rightarrow x - 1, \quad a_{\mu}(x) = cx, \quad \nu_{\mu} = -1.$$

The CME has the form

$$\begin{aligned} \frac{\partial P(x, t | x_0, t_0)}{\partial t} &= \sum_{\mu=1}^1 (a_{\mu}(x - \nu_{\mu})P(x - \nu_{\mu}, t | x_0, t_0) \\ &\quad - a_{\mu}(x)P(x, t | x_0, t_0)) \\ &= c(x + 1)P(x + 1, t | x_0, t_0) - cxP(x, t | x_0, t_0). \end{aligned}$$

Linear system is a special case:

Reaction $A \xrightarrow{c} \emptyset$ has the CME

$$\frac{\partial P(x, t | x_0, t_0)}{\partial t} = c(x + 1)P(x + 1, t | x_0, t_0) - cxP(x, t | x_0, t_0).$$

Using mathematical induction, it is possible to show that

$$P(x, t | x_0, t_0) = \frac{x_0!}{x!(x_0 - x)!} e^{-cxt} (1 - e^{-ct})^{x_0 - x}.$$

Mean is

$$\begin{aligned} E(X) &= \sum_{x=0}^{x_0} xP(x, t | x_0, t_0) = \sum_{x=0}^{x_0} x \frac{x_0!}{x!(x_0 - x)!} e^{-cxt} (1 - e^{-ct})^{x_0 - x} = \\ &= x_0 e^{-ct} \sum_{x=1}^{x_0} \frac{(x_0 - 1)!}{(x - 1)!(x_0 - 1 - (x - 1))!} (e^{-ct})^{x-1} (1 - e^{-ct})^{x_0 - 1 - (x-1)} = \\ &= x_0 e^{-ct} \quad (\text{This is solution of the deterministic ODE}). \end{aligned}$$

Gillespie Stochastic Simulation Algorithm (SSA)

Gillespie stochastic simulation algorithm is a way to simulate the system at state $\mathbf{X}(t)$ to state $\mathbf{X}(t + \tau)$ in a stochastic manner by randomly selecting

1. **what is the next reaction time moment τ and**
2. **which reaction R_μ that happens at the time interval $(t + \tau, t + \tau + d\tau)$.**

This problem can be solved using **reaction probability density function**.

In the SSA, the probability function $P(\mathbf{x}, t | \mathbf{x}_0, t_0)$ defined in Equation 1 is not used but instead a joint probability density function $p(\tau, \mu | \mathbf{x}, t)$ of two random variables, τ and μ .

The joint probability density function is defined as

$p(\tau, \mu | \mathbf{x}, t)d\tau$:= the probability that the next reaction in V will occur in the differential time interval $[t + \tau, t + \tau + d\tau)$ and will be a reaction R_μ given that the system is in the state $\mathbf{X}(t) = \mathbf{x}$ at time t , when $\tau \in [0, \infty)$ and $\mu \in \{1, \dots, N\}$

$p(\tau, \mu | \mathbf{x}, t)d\tau$ can be defined as **the probability of no reaction occurring in the time interval $[t, t + \tau)$, denoted by $P_0(\tau | \mathbf{x}, t)$, multiplied by the probability that R_μ will occur in the infinitesimal time interval $[t + \tau, t + \tau + d\tau)$, denoted by $a_\mu d\tau$:**

$$p(\tau, \mu | \mathbf{x}, t)d\tau = P_0(\tau | \mathbf{x}, t)a_\mu(c_\mu, \mathbf{x})d\tau.$$

Next, the formula for $P_0(\tau | \mathbf{x}, t)$ is solved. The probability that no reaction occurs in $[t, t + \tau + d\tau)$ is

$$P_0(\tau + d\tau | \mathbf{x}, t) = P_0(\tau | \mathbf{x}, t) \left(1 - \sum_{\mu=1}^N a_{\mu}(c_{\mu}, \mathbf{x}) d\tau \right),$$

where the first term is the probability that no reaction occurs in $[t, t + \tau)$ and the second term is the probability that no reaction occurs in $[t + \tau, t + \tau + d\tau)$. It is obtained that

$$\begin{aligned} \frac{dP_0(\tau | \mathbf{x}, t)}{d\tau} &= \lim_{d\tau \rightarrow 0} \frac{P_0(\tau + d\tau | \mathbf{x}, t) - P_0(\tau | \mathbf{x}, t)}{d\tau} \\ &= -a_0(c_{\mu}, \mathbf{x}) P_0(\tau | \mathbf{x}, t), \end{aligned}$$

where $a_0(c_{\mu}, \mathbf{x}) = \sum_{\mu=1}^N a_{\mu}(c_{\mu}, \mathbf{x})$. The solution with the initial condition $P_0(\tau = 0 | \mathbf{x}, t) = 1$ is

$$P_0(\tau | \mathbf{x}, t) = e^{-a_0(c_{\mu}, \mathbf{x})\tau}.$$

$$p(\tau, \mu | \mathbf{x}, t) d\tau = P_0(\tau | \mathbf{x}, t) a_\mu(c_\mu, \mathbf{x}) d\tau.$$

Finally it is concluded that the joint probability function has the form

$$p(\tau, \mu | \mathbf{x}, t) = \begin{cases} a_\mu(c_\mu, \mathbf{x}) e^{-a_0(c_\mu, \mathbf{x})\tau} & \text{if } \tau \in [0, \infty) \text{ and} \\ & \mu \in \{1, \dots, N\}, \\ 0 & \text{otherwise,} \end{cases}$$

where $a_0(c_\mu, \mathbf{x}) = \sum_{\mu=1}^N a_\mu(c_\mu, \mathbf{x})$.

The joint probability density function can be given in form $p_{\tau,\mu}(\tau, \mu) = p_{\tau}(\tau)p_{\mu|\tau}(\mu | \tau)$, where $p_{\tau}(\tau)d\tau$ is the probability that the next reaction will occur in $[t + \tau, t + \tau + d\tau)$ regardless of which reaction it might be and $p_{\mu|\tau}(\mu | \tau)$ is the probability that the next reaction will be R_{μ} given that it occurs at time $t + \tau$:

$$\begin{aligned}
 p_{\tau}(\tau) &= \begin{cases} \sum_{\mu=1}^N p_{\tau,\mu}(\tau, \mu) & \text{if } \tau \in [0, \infty), \\ 0 & \text{otherwise,} \end{cases} \\
 &= \begin{cases} \sum_{\mu=1}^N a_{\mu} e^{-a_0 \tau} & \text{if } \tau \in [0, \infty), \\ 0 & \text{otherwise,} \end{cases} \quad (2) \\
 &= \begin{cases} a_0 e^{-a_0 \tau} & \text{if } \tau \in [0, \infty), \\ 0 & \text{otherwise,} \end{cases}
 \end{aligned}$$

and

$$\begin{aligned} p_{\mu|\tau}(\mu|\tau) &= \begin{cases} \frac{p_{\tau,\mu}(\tau,\mu)}{\sum_{\mu'=1}^N p_{\tau,\mu}(\tau,\mu')} & \text{if } \mu \in \{1, \dots, N\}, \\ 0 & \text{otherwise,} \end{cases} \\ &= \begin{cases} \frac{a_{\mu}e^{-a_0\tau}}{\sum_{\mu'=1}^N a_{\mu'}e^{-a_0\tau}} & \text{if } \mu \in \{1, \dots, N\}, \\ 0 & \text{otherwise,} \end{cases} \quad (3) \\ &= \begin{cases} \frac{a_{\mu}}{a_0} & \text{if } \mu \in \{1, \dots, N\}, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Each probability density function depends only on one of the two parameters, $p_{\mu|\tau}(\mu|\tau) = p_{\mu}(\mu)$.

There are several ways to implement the exact SSA. In **the direct method**, with two random numbers r_1 and r_2 uniformly distributed on the interval $[0, 1]$, the time interval $[t, t + \tau)$ is determined according to Equation 2 and the reaction R_μ is determined according to Equation 3.

The probability distribution function F_τ is

$$\begin{aligned} F_\tau(\tau | \mathbf{x}, t) &= \int_{-\infty}^{\tau} p_\tau(\tau' | \mathbf{x}, t) d\tau' = \int_0^{\tau} p_\tau(\tau' | \mathbf{x}, t) d\tau' \\ &= \int_0^{\tau} a_0 e^{-a_0 \tau'} d\tau' = 1 - e^{-a_0(c_\mu, \mathbf{x})\tau}. \end{aligned}$$

Because $F_\tau(\tau) \in [0, 1]$, it is possible to mark $r_1 = F_\tau(\tau) = 1 - e^{-a_0 \tau}$. By replacing the random number $1 - r_1$ by the statistically equivalent random number r_1 , $r_1 = e^{-a_0 \tau}$ is obtained. Now $\ln(r_1) = -a_0 \tau$, $\tau = -\ln(r_1)/a_0$, and finally

$$\tau = \frac{1}{a_0(c_\mu, \mathbf{x})} \ln \left(\frac{1}{r_1} \right).$$

Using Equation 3, the probability distribution function $F_{\mu|\tau}$ is

$$\begin{aligned} F_{\mu|\tau}(\mu | \mathbf{x}, t) &= \sum_{\mu'=-\infty}^{\mu} p_{\mu|\tau}(\mu' | \mathbf{x}, t) = \sum_{\mu'=1}^{\mu} p_{\mu|\tau}(\mu' | \mathbf{x}, t) \\ &= \frac{1}{a_0(c_{\mu}, \mathbf{x})} \sum_{\mu'=1}^{\mu} a_{\mu'}(c_{\mu}, \mathbf{x}). \end{aligned}$$

Taking for μ the value which satisfies

$$F_{\mu|\tau}(\mu - 1) < r_2 \leq F_{\mu|\tau}(\mu),$$

$$\sum_{\mu'=1}^{\mu-1} p_{\mu|\tau}(\mu' | \mathbf{x}, t) < r_2 \leq \sum_{\mu'=1}^{\mu} p_{\mu|\tau}(\mu' | \mathbf{x}, t)$$

is obtained, and

$$\sum_{\mu'=1}^{\mu-1} a_{\mu'}(c_{\mu}, \mathbf{x}) < r_2 a_0(c_{\mu}, \mathbf{x}) \leq \sum_{\mu'=1}^{\mu} a_{\mu'}(c_{\mu}, \mathbf{x}).$$

Gillespie Algorithm (direct method)

1. Set $t = 0$. Specify initial values for $\mathbf{X} = [X_1, \dots, X_M]$ and values for c_μ and stoichiometric vectors $\boldsymbol{\nu}_\mu$, $\mu = 1, \dots, N$.
2. Calculate a_μ , $\mu = 1, \dots, N$, and $a_0 = \sum_{\mu=1}^N a_\mu$.
3. Generate random pair (τ, μ) according to $p(\tau, \mu | \mathbf{x}, t)$:
Generate two random numbers r_1 and r_2 uniformly distributed on the interval $[0, 1]$.
Take $\tau = \frac{1}{a_0} \ln \frac{1}{r_1}$.
Take μ so that $\sum_{\mu'=1}^{\mu-1} a_{\mu'} < r_2 a_0 \leq \sum_{\mu'=1}^{\mu} a_{\mu'}$.
4. Update of the system by

$$\mathbf{X}(t + \tau) = \mathbf{X}(t) + \boldsymbol{\nu}_\mu. \quad (4)$$

5. Go to step 2 or end the simulation.

D. T. Gillespie, *Journal of Computational Physics*, vol. 22, no. 4, pp. 403-434, 1976.

By carrying out the above procedure one obtains one possible realization of the stochastic process.

In order to get statistically complete picture of the time-evolution of the system, one must carry out several independent simulations with the same initial conditions and calculate **sample mean** and **sample standard deviation**.

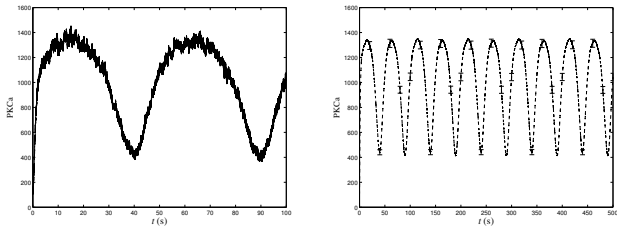


Figure: Left: one realization. Right: sample mean and standard deviation of 1000 realizations (T. Manninen et al., *Computational Biology and Chemistry*, vol. 30, pp. 280–291, 2006).

Difference between ODE and SSA model construction:

- ▶ ODE: concentrations $[X]$ in $M = \text{mol}/l$, $[v_\mu] = M/s$
- ▶ SSA: $[X]$ in numbers, $[a_\mu] = [c_\mu] = 1/s$
- ▶ $X = N_A V [X]$, where N_A is the Avogadro's constant and V is volume.
- ▶ $\emptyset \xrightarrow{k_\mu} S$:
 - ▶ Reaction rate: $v_\mu = k_\mu$, $[k_\mu] = M/s$
 - ▶ Propensity: $h_\mu = 1$, $a_\mu = c_\mu$, $c_\mu = N_A V k_\mu$
- ▶ $S_1 \xrightarrow{k_\mu} S_2$:
 - ▶ $v_\mu = k_\mu [X_1]$, $[k_\mu] = 1/s$
 - ▶ $h_\mu = X_1$, $a_\mu = c_\mu X_1$, $c_\mu = k_\mu$

- ▶ $S_1 + S_2 \xrightarrow{k_\mu} S_3$:
 - ▶ $v_\mu = k_\mu[X_1][X_2]$, $[k_\mu] = 1/(Ms)$
 - ▶ $h_\mu = X_1X_2$, $a_\mu = c_\mu X_1X_2$, $c_\mu = k_\mu/(N_A V)$
- ▶ $2S_1 \xrightarrow{k_\mu} S_2$:
 - ▶ $v_\mu = k_\mu[X_1]^2$, $d[X_1]/dt = -2k_\mu[X_1]^2$, $[k_\mu] = 1/(Ms)$
 - ▶ $h_\mu = X_1(X_1 - 1)/2$, $a_\mu = c_\mu X_1(X_1 - 1)/2$, $c_\mu = 2k_\mu/(N_A V)$

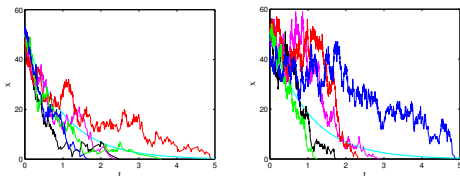


Figure: Five realizations of stochastic linear birth-death process together with the deterministic solution for two different (λ, μ) combinations, i.e., left: $\lambda = 3$, $\mu = 4$ and right: $\lambda = 7$, $\mu = 8$, each $\lambda - \mu = -1$ and $x_0 = 50$.

Stochastic Differential Equation (SDE) Models

If during any time interval $[t, t + \tau)$ all of the reactions can occur often, but none of the propensity functions a_μ changes significantly, then the discrete-state Markov process can be approximated by a continuous-state Markov process described by **the chemical Langevin equation (CLE)**.

The number of times a reaction can happen can be defined as

$K_\mu(c_\mu, \mathbf{x}, \tau)$:= the number of times a reaction R_μ occurs in the time interval $[t, t + \tau]$ given that the system is in the state $\mathbf{X}(t) = \mathbf{x}$ at time t .

The update of the system in Equation 4 then becomes

$$\mathbf{X}(t + \tau) = \mathbf{X}(t) + \sum_{\mu=1}^N K_\mu(c_\mu, \mathbf{x}, \tau) \boldsymbol{\nu}_\mu.$$

$K_\mu(c_\mu, \mathbf{x}, \tau)$ is of course a *random variable*. To compute it for arbitrary $\tau > 0$ would be quite as difficult as solving the master equation. But we can obtain excellent *approximation* to $K_\mu(c_\mu, \mathbf{x}, \tau)$ rather easily if we impose the following conditions:

1) Require τ to be *small* enough that the change in the state during $[t, t + \tau]$ will be so slight that none of the propensity functions changes its value “appreciably”, i.e.

$$a_\mu(\mathbf{x}(t')) \cong a_\mu(\mathbf{x}(t)), \quad \forall t' \in [t, t + \tau], \quad \forall \mu \in [1, N]$$

Practically: all reactant molecule populations are sufficiently large compared to 1.

So we can rewrite the “update” of the system:

$$\mathbf{X}(t + \tau) = \mathbf{X}(t) + \sum_{\mu=1}^N \nu_\mu \mathcal{P}_\mu(a_\mu(c_\mu, \mathbf{x}), \tau),$$

where $\mathcal{P}_\mu(a_\mu(c_\mu, \mathbf{x}), \tau)$ is Poisson random variable.

2) Require τ to be *large* enough that the expected number of occurrences of each reaction channel R_μ in $[t, t + \tau]$ be much larger than 1, i.e.

$$\langle \mathcal{P}_\mu(a_\mu(c_\mu, \mathbf{x}), \tau) \rangle = a_\mu(c_\mu, \mathbf{x}) \times \tau \gg 1, \quad \forall \mu \in [1, N].$$

Obviously, this condition runs counter to condition **1**.

Practically: $a_\mu(c_\mu, \mathbf{x})$ is proportional to $h_\mu(\mathbf{x})$ and thus to one or more components of \mathbf{x} , thus, sufficiently large molecular populations can make the condition **2** to hold even when τ is small.

This condition allows us to *approximate* each Poisson random variable $\mathcal{P}_\mu(a_\mu(c_\mu, \mathbf{x}), \tau)$ by a *normal* random variable with the same mean and variance.

So the update of the system:

$$\mathbf{X}(t + \tau) = \mathbf{X}(t) + \sum_{\mu=1}^N \nu_\mu \mathcal{N}_\mu(a_\mu(c_\mu, \mathbf{x})\tau, a_\mu(c_\mu, \mathbf{x})\tau),$$

where $\mathcal{N}(m, \sigma^2)$ denotes the normal random variable with mean m and variance σ^2 .

Further modifying the update (using the linear combination theorem for normal random variables:

$\mathcal{N}(n, \sigma^2) = m + \sigma\mathcal{N}(0, 1)$):

$$\begin{aligned}\mathbf{X}(t + \tau) &= \mathbf{X}(t) + \sum_{\mu=1}^N \boldsymbol{\nu}_{\mu} \mathcal{N}_{\mu}(a_{\mu}(c_{\mu}, \mathbf{x})\tau, a_{\mu}(c_{\mu}, \mathbf{x})\tau) = \\ &= \mathbf{X}(t) + \sum_{\mu=1}^N \boldsymbol{\nu}_{\mu} a_{\mu}(c_{\mu}, \mathbf{X}(t))\tau + \sum_{\mu=1}^N \boldsymbol{\nu}_{\mu} \sqrt{a_{\mu}(c_{\mu}, \mathbf{X}(t))\tau} \mathcal{N}_{\mu}(0, 1) = \\ &= \mathbf{X}(t) + \sum_{\mu=1}^N \boldsymbol{\nu}_{\mu} a_{\mu}(c_{\mu}, \mathbf{X}(t))\tau + \sum_{\mu=1}^N \boldsymbol{\nu}_{\mu} \sqrt{a_{\mu}(c_{\mu}, \mathbf{X}(t))\tau} \mathcal{N}_{\mu}(0, \tau).\end{aligned}$$

Some implications

- ▶ Condition **1** requires that any of the propensity functions does not change appreciably. Thus, all reactions occurring in $[t, t + \tau]$ will be essentially *independent* of each other. Hence, we can approximate each $K_\mu(c_\mu, \mathbf{x}, \tau)$ with *statistically independent* Poisson random variable.
- ▶ Notice that in replacing the *integer Poisson* random variables by the *real normal* random variables (by condition **2**), we in effect convert X_j from discretely changing integer variables to continuously changing real variables.
- ▶ Notice that the normal random variables \mathcal{N}_μ will be *statistically independent* as a consequence of statistical independence of the Poisson's \mathcal{P}_μ . That is why one can use Brownian motion as an approximation to \mathcal{N}_μ .

The chemical Langevin equation

$$d\mathbf{X}(t) = \sum_{\mu=1}^N \nu_{\mu} a_{\mu}(c_{\mu}, \mathbf{X}(t)) dt + \sum_{\mu=1}^N \nu_{\mu} \sqrt{a_{\mu}(c_{\mu}, \mathbf{X}(t))} d\mathbf{W}_{\mu}(t),$$

where $d\mathbf{W}(\tau) \sim \mathcal{N}(0, \tau) \equiv \sqrt{\tau} \mathcal{N}(0, 1)$ is the Brownian motion.

$W(0) = 0$, the increment $W(t) - W(s) \sim \mathcal{N}(0, t - s)$ for all $0 \leq s < t \leq T$, and for all $0 \leq s < t < u < v \leq T$, the increments $W(t) - W(s)$ and $W(v) - W(u)$ are independent.

D. T. Gillespie, *J. Chem. Phys.*, vol. 113, no. 1, pp. 297–306, 2000.

General Stochastic differential equation model

$$\boxed{dx(t) = f(x(t))dt + g(x(t))dW(t)}$$

can be simulated using, for example, **the Euler-Maruyama method**

$$x(t + \Delta t) = x(t) + f(x(t))\Delta t + g(x(t))\sqrt{\Delta t}\mathcal{N}(0, 1)$$

where Δt is a small numerical integration (simulation) time step.

Higher order **Milstein method** can also be applied:

$$x(t + \Delta t) = x(t) + f(x(t))\Delta t + g(x(t))\sqrt{\Delta t}\mathcal{N}(0, 1) + \frac{1}{2}g(x(t))g'(x(t)) (\mathcal{N}^2(0, 1) - 1) \Delta t.$$

Summary

- ▶ Mean field approximation (ODE approach) does not work for small number of molecules.
- ▶ All molecular systems get affected by the thermal fluctuations (intrinsic noise).
- ▶ There are two fundamental approaches to stochastic systems: Chemical Master Equation (CME) and stochastic differential equations.
- ▶ CME is usually intractable, because each state of the system needs a separate ODE.
- ▶ Gillespie stochastic simulation algorithm (SSA) numerically simulates the Markov process that the CME describes analytically.