Stochastic Biomodelling

Andrzej Mizera (andrzej.mizera@uni.lu)
Faculty of Science, Technology and Communication
University of Luxembourg

Course content

1. Introduction
   - Stochasticity in biological processes
   - Deterministic vs stochastic biomodelling
2. Prerequisite:
   - Crash course on probability theory
3. Stochastic modelling of chemical kinetics: the chemical master equation (CME)
4. Stochastic simulation of the CME – Gillespie's direct method algorithm
5. Practicals:
   - Implementing the Gillespie's algorithm in MATLAB and investigating its characteristics on various biochemical systems
   - Comparing the obtained simulation results with the solutions in the deterministic formulation

Some words on what is the role of mathematical modelling in Systems Biology

Stochastic modelling in Systems Biology

Stochastic simulation of the CME – Gillespie's direct method algorithm

Towards numerical simulations

- The chemical master equation is exact and mathematically elegant

- Difficult to use it for numerical simulations
  - it can be analytically solved only for the simplest reactions
  - it describes the evolution of the probability of all states in time
    - it does not give the transitions from state to state
  - the differential equations for the time evolution of the molecular populations $X_i(t)$ may be written, but they involve the expected values of higher powers X_i^n and thus lead to infinite systems of ODEs

- Solution: Gillespie's stochastic simulation algorithm (SSA), 1976, 1977

Gillespie's SSA: foundations

- Assume we have m reactions R_1, R_2, …, R_m and n molecular species S_1, S_2, …, S_n
- Given that the system is in state $(X_1,\ldots,X_n)$ at time t, we need to answer two questions in order to simulate the evolution of the system
  - when will the next reaction occur?
  - which reaction will it be?
- We combine the answers to these 2 questions in the following joint probability distribution:
  - $P(\tau,\mu)\, d\tau$ = the probability that, given the state $(X_1,\ldots,X_n)$ at time t, the next reaction will occur in the infinitesimal time interval $(t+\tau, t+\tau+d\tau)$ AND it will be reaction R_μ
  - note that if we thought about the probability of a reaction occurring exactly at time $t+\tau$, then the probability would be 0
- Strategy: based on CME, deduce the analytical expression of $P(\tau,\mu)$

Gillespie's SSA: foundations

- Given the state $(X_1,\ldots,X_n)$ at time t, we need to compute

  $P(\tau,\mu)\, d\tau$ = the probability that, given the state $(X_1,\ldots,X_n)$ at time t, the next reaction will occur in the infinitesimal time interval $(t+\tau,t+\tau+d\tau)$ AND it will be reaction R_μ

- Let h_μ be the number of distinct combinations of reactants for reaction R_μ in the state at time $t+\tau$ (same state as at time t)
  - then, as observed for the CME, the probability that reaction R_μ will occur in the infinitesimal time interval $(t+\tau,t+\tau+d\tau)$ is $c_\mu h_\mu d\tau$

- Let $P_0(\tau)$ be the probability that no reaction occurs in the time interval $(t,t+\tau)$
  - Then $P(\tau,\mu)\, d\tau = P_0(\tau)\cdot h_\mu c_\mu d\tau$

Gillespie's SSA: foundations

$P(\tau,\mu)\, d\tau$ = the probability that, given the state $(X_1,\ldots,X_n)$ at time t, the next reaction will occur in the infinitesimal time interval $(t+\tau,t+\tau+d\tau)$ AND it will be reaction R_μ

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

We need to compute $P_0(\tau)$, the probability that no reaction occurs in the time interval $(t,t+\tau)$

- Careful because the time interval $(t,t+\tau)$ may not necessarily be infinitesimal!
- $P_0(\tau+d\tau)$, where $d\tau$ is infinitesimally small: no reaction occurs in the interval $(t,t+\tau+d\tau)$ if and only if no reaction occurs in $(t,t+\tau)$ AND no reaction occurs in the infinitesimal interval $(t+\tau,t+\tau+d\tau)$
  - Thus, $P_0(\tau+d\tau)=P_0(\tau)(1-\sum_\mu h_\mu c_\mu d\tau)$, i.e., $dP_0(\tau)/d\tau=-P_0(\tau)\sum_\mu h_\mu c_\mu$
  - It follows that $P_0(\tau)=\exp(-\sum_\mu h_\mu c_\mu \tau)$
- Finally, we obtain that

$$\mathbf{P(\tau,\mu)= h_\mu c_\mu e^{-\alpha\tau}},$$

for all $\tau\geq 0$ and $\mu=1,\ldots,n$, where $\alpha=\sum_\mu h_\mu c_\mu$

Gillespie's SSA

To simulate numerically the time evolution of our system starting from the given initial state:

- Generate a pair (τ,μ) according to the probability density function $P(\tau,\mu)$
  - waiting time is exponentially distributed
  - estimate reaction probabilities in the current state, choose one
- Adjust the molecular levels according to reaction R_μ (decrease the level of reactants, increase the level of the output species)
- Advance time to $t+\tau$
- Iterate the procedure

Gillespie's SSA

We need to generate a pair (τ,μ) according to the probability density function $P(\tau,\mu)=h_\mu c_\mu e^{-\alpha\tau}$, where $\alpha=\sum_\mu h_\mu c_\mu$

- we first generate the time point τ such that the next reaction (any kind of reaction!) occurs in the infinitesimal time interval $(t+\tau,t+\tau+d\tau)$
  - the corresponding probability density function is $P(\tau)=\sum_\mu P(\tau,\mu)=\alpha e^{-\alpha\tau}$
    - To do this, generate a random number r_1 in (0,1) and let τ_0 be such that $P(\tau<\tau_0)=r_1$: $$P(\tau<\tau_0)=\int_{-\infty}^{\tau_0} P(\tau)d\tau = e^{-\alpha\tau_0}$$
    - Thus, $\tau_0=1/\alpha \ln(1/r_1)$ is the time point we will consider
- we then select the reaction R_μ according to their relative probabilities of being triggered in the current step: $P(\mu)= P(\tau,\mu)/\sum_\nu P(\tau,\nu)= h_\mu c_\mu/\alpha$
  - To do this, generate a random number r_2 in (0,1) and let μ_0 be such that $P(\mu\leq\mu_0)=r_2$
  - We consider the distribution $F(m)=\sum_{i\leq m}P(i)$ and choose μ_0 such that $F(\mu_0\text{-}1)<r_2\leq F(\mu_0)$: $$\frac{1}{\alpha}\sum_{\nu=1}^{\mu_0-1} h_\nu c_\nu < r_2 \leq \frac{1}{\alpha}\sum_{\nu=1}^{\mu_0} h_\nu c_\nu$$

Gillespie's SSA: summary

- This is the only ***exact*** simulation algorithm of the chemical master equation
  - it is essentially just a reformulation of CME
  - the crucial point is that there is no time slicing (as in the numerical simulation of ODEs): jump to the next time point according to the correct probability distribution
  - It is also referred to as the ***first reaction method***

- Many variants of Gillespie's SSA exist
  - some offer speedups (e.g., ***next reaction method***, also known as the ***Gibson-Bruck algorithm***)
  - some are reformulations for various special cases, such as for hybrid models, involving both continuous and discrete variables