CS-E5885 Modeling biological networks
Chemical and biochemical kinetics

Harri Lähdesmäki

Department of Computer Science
Aalto University

January 18, 2022

Outline

- ▶ Introduction
- ▶ Mass-action stochastic kinetics:
  - ▶ Rate laws
  - ▶ Gillespie algorithm
- ▶ Approximate simulation strategies
- ▶ The master equation
- ▶ Connection between continuous Markov processes and ODEs

- ▶ Reading (see references at the end):
  - ▶ Sections 6 and 8.3 from (Wilkinson, 2011)

Introduction

- ▶ Set of coupled chemical reactions

Lotka–Volterra (reaction equations)

$$Y_1 \rightarrow 2Y_1$$
$$Y_1 + Y_2 \rightarrow 2Y_2$$
$$Y_2 \rightarrow \emptyset$$
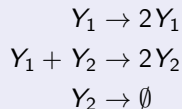
- ▶ Reaction equations capture the key interactions, but insufficient to determine the full dynamic behaviour of the system
- ▶ Need to know rates of all reactions

Molecular Approach to Kinetics

- Consider a bi-molecular reaction

$$X + Y \rightarrow ?$$

- Meaning: a molecule X is able to react with a molecule Y if they happen to collide with sufficient energy while moving around (Brownian motion)

- Question: what is the probability of a X-Y collision occurring in some volume V in any infinitesimal time interval?

- Assumptions: the container has constant volume V, the contents are well stirred and the temperature is constant
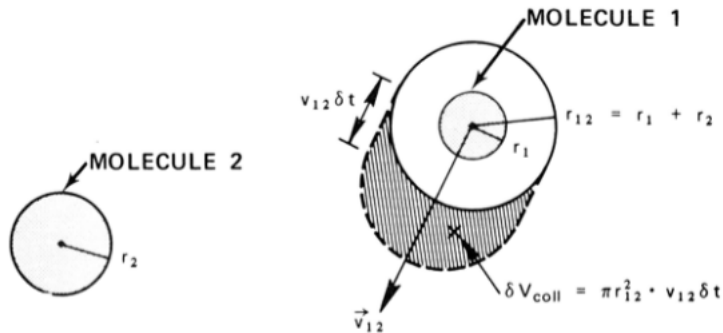
Physical basis of the stochastic formulation

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

Figure: from (Gillespie, 1977)

Molecular Approach to Kinetics

- Let P_1 and P_2 be the molecules' position in space
- If the volume of the container is fixed and temperature remains constant, then
    - P_1 and P_2 are uniformly (and independently) distributed over the volume
    - This distribution does not depend on time
    - Formal proof using statistical mechanical arguments
- $\rightarrow$ The probability that the molecules are within reaction distance is also independent on time
    - That is, collision of the two molecules has a constant hazard

Molecular Approach to Kinetics (2)

- Consider next the case of time-varying volume V
- For a region of d with volume v' we have

$$\mathbb{P}(P_i \in d) = \frac{v'}{V}, \quad i = 1, 2$$

Molecular Approach to Kinetics (2)

- ▶ Consider next the case of time-varying volume V
- ▶ For a region of d with volume v' we have

$$\mathbb{P}(P_i \in d) = \frac{v'}{V}, \quad i = 1, 2$$

- ▶ The probability that X and Y are within a reacting distance r can be computed as (by Proposition 3.11)

$$\mathbb{P}(|P_1 - P_2| < r) = \mathbb{E}_{P_2}(\mathbb{P}(|P_1 - P_2| < r | P_2 = p_2))$$

Molecular Approach to Kinetics (2)

- ▶ Consider next the case of time-varying volume V
- ▶ For a region of d with volume v' we have

$$\mathbb{P}(P_i \in d) = \frac{v'}{V}, \quad i = 1, 2$$

- ▶ The probability that X and Y are within a reacting distance r can be computed as (by Proposition 3.11)

$$\mathbb{P}(|P_1 - P_2| < r) = \mathbb{E}_{P_2}(\mathbb{P}(|P_1 - P_2| < r | P_2 = p_2))$$

- ▶ But the conditional probability will be the same for any $P_2 = p_2$ (away from the boundary), so we end up with

$$\mathbb{P}(|P_1 - p_2| < r) = \mathbb{P}(P_1 \in d) = \frac{4\pi r^3}{3V}$$

Molecular Approach to Kinetics (3)

- The results on the previous slides mean that if
  - molecules are uniformly distributed (time independent), and
  - the size of the container (V) does not change,

  then the probability that the molecules are within reaction distance is also independent on time
- In other words, reaction/collision hazard is constant

Molecular Approach to Kinetics (3)

- The results on the previous slides mean that if
  - molecules are uniformly distributed (time independent), and
  - the size of the container (V) does not change,

  then the probability that the molecules are within reaction distance is also independent on time

- In other words, reaction/collision hazard is constant
- If the volume changes, then the reaction hazard is inversely proportional to the volume
- Given that molecules are within a reaction distance, they will not necessarily interact but do so with a probability independent of V
- We will assume a fixed volume V

Mass-Action Stochastic Kinetics

- Species $P = (X_1, \ldots, X_u)'$ and reactions $T = (R_1, \ldots, R_v)'$
- Qualitative structure of the reaction network is encoded as a Petri net $N = (P, T, Pre, Post, M)$
- $\mathbf{x} = (x_1, \ldots, x_u)$ is the current state of the system

Mass-Action Stochastic Kinetics

▶ Species $P = (X_1, \ldots, X_u)'$ and reactions $T = (R_1, \ldots, R_v)'$

▶ Qualitative structure of the reaction network is encoded as a Petri net
$N = (P, T, Pre, Post, M)$

▶ $\mathbf{x} = (x_1, \ldots, x_u)$ is the current state of the system

▶ Stochastic rate constant c_i and a rate law (hazard function) $h_i(\mathbf{x}, c_i)$ for each reaction R_i

▶ Interpretation of the rate law:

    ▶ Given the state $\mathbf{x}$ at time t, the probability that an R_i reaction will occur in an infinitesimal time interval $(t, t + dt]$ is given by $h_i(\mathbf{x}, c_i) \, dt$

Mass-Action Stochastic Kinetics

▶ Species $P = (X_1, \ldots, X_u)'$ and reactions $T = (R_1, \ldots, R_v)'$

▶ Qualitative structure of the reaction network is encoded as a Petri net
$N = (P, T, Pre, Post, M)$

▶ $\mathbf{x} = (x_1, \ldots, x_u)$ is the current state of the system

▶ Stochastic rate constant c_i and a rate law (hazard function) $h_i(\mathbf{x}, c_i)$ for each reaction R_i

▶ Interpretation of the rate law:
  ▶ Given the state $\mathbf{x}$ at time t, the probability that an R_i reaction will occur in an infinitesimal time interval $(t, t + dt]$ is given by $h_i(\mathbf{x}, c_i)\, dt$

▶ Recall the Poisson process: in the absence of any other reactions, the time to such a reaction would be distributed as $\mathrm{Exp}(h_i(\mathbf{x}, c_i))$

Rate Law

- Zeroth-order reactions
    - $R_i : \quad \emptyset \xrightarrow{c_i} X$
    - In practice, products are not created from nothing though...
    - $h_i(\mathbf{x}, c_i) = c_i$
    - Constant influx

Rate Law

- Zeroth-order reactions
  - $R_i : \quad \emptyset \xrightarrow{c_i} X$
  - In practice, products are not created from nothing though...
  - $h_i(\mathbf{x}, c_i) = c_i$
  - Constant influx
- First-order reactions
  - $R_i : \quad X_j \xrightarrow{c_i} ?$
  - c_i represents the hazard of a particular molecule of X_j
  - There are x_j molecules of X_j, thus
    $$h_i(\mathbf{x}, c_i) = c_i x_j$$
  - Representation a spontaneous change of a molecule into one or more molecules

Rate Law (2)

- Second-order reactions
  - $R_i: \quad X_j + X_k \xrightarrow{c_i}$?
  - c_i represents the hazard that a particular pair of molecules X_j and X_k will react
  - There are $x_j x_k$ different pairs molecules, thus

$$h_i(\mathbf{x}, c_i) = c_i x_j x_k$$

Rate Law (2)

- Second-order reactions
  - $R_i: \quad X_j + X_k \xrightarrow{c_i} ?$
  - c_i represents the hazard that a particular pair of molecules X_j and X_k will react
  - There are $x_j x_k$ different pairs molecules, thus
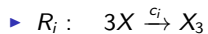
$$h_i(\mathbf{x}, c_i) = c_i x_j x_k$$

  - For a special type of second-order reaction $R_i: \quad 2X_j \xrightarrow{c_i} ?$

$$h_i(\mathbf{x}, c_i) = c_i \frac{x_j(x_j - 1)}{2}$$

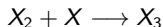- In the presence of a large pool of catalyst, second order reactions can be approximated by first-order reactions

Rate Law (3)

- Higher-order reactions
  - $R_i :\quad 3X \xrightarrow{c_i} X_3$
  - $h_i(\mathbf{x}, c_i) = c_i \binom{x}{3} = c_i \frac{x!}{(x-3)!3!} = c_i \frac{x(x-1)(x-2)}{6}$
  - In most cases it is likely to be more realistic to model the process as the pair of second-order reactions

$$2X \longrightarrow X_2$$
$$X_2 + X \longrightarrow X_3$$

The Gillespie Algorithm

▶ Previous slides show that time-evolution of a coupled chemical reaction system can be regarded as a stochastic process

▶ Reaction hazards depend only on the current state of the system
  - → Reaction hazards remain constant until the next reaction (i.e., homogeneous Poisson process until the next reaction)
  - → Each reaction is an independent random event with exponential waiting time (in the absence of other reactions)
  - → The next reaction has an exponential waiting time
  - → The overall time-evolution of the state of the reaction system can be modeled as a continuous-time, non-homogeneous Markov process with a discrete state space

The Gillespie Algorithm (2)

- ▶ Thus, using the results for Poisson processes, we can conclude that
  - ▶ In a given reaction system with v reactions and where the hazard for a reaction R_i is $h_i(\mathbf{x}, c_i)$, the combined hazard for any (i.e., the *first/next*) reaction to happen is

  $$h_0(\mathbf{x}, \mathbf{c}) \equiv \sum_{i=1}^{v} h_i(\mathbf{x}, c_i)$$

  and the time to the next reaction is distributed as $\text{Exp}(h_0(\mathbf{x}, \mathbf{c}))$
  - ▶ This first/next reaction will be a random type with probabilities (independent of the time to the next event)

  $$\pi_i = \frac{h_i(\mathbf{x}, c_i)}{h_0(\mathbf{x}, \mathbf{c})}$$

The Gillespie Algorithm (2)

- Thus, using the results for Poisson processes, we can conclude that
  - In a given reaction system with v reactions and where the hazard for a reaction R_i is $h_i(\mathbf{x}, c_i)$, the combined hazard for any (i.e., the *first/next*) reaction to happen is

  $$h_0(\mathbf{x}, \mathbf{c}) \equiv \sum_{i=1}^{v} h_i(\mathbf{x}, c_i)$$

  and the time to the next reaction is distributed as $\mathrm{Exp}(h_0(\mathbf{x}, \mathbf{c}))$
  - This first/next reaction will be a random type with probabilities (independent of the time to the next event)

  $$\pi_i = \frac{h_i(\mathbf{x}, c_i)}{h_0(\mathbf{x}, \mathbf{c})}$$

- Realizations of the time to the next reaction and the reaction type can be used for updating the state of the system
- This can be represented/is known as the Gillespie algorithm for chemical kinetics

The Gillespie Algorithm (3)

The Gillespie Algorithm

1. Initialize the system at $t = 0$ with rate constants $c_1, \ldots, c_v$ and initial numbers of molecules for each species $x_1, \ldots, x_u$
2. For each $i = 1, \ldots, v$ calculate $h_i(\mathbf{x}, c_i)$ based on the current state $\mathbf{x}$
3. Calculate $h_0(\mathbf{x}, \mathbf{c}) \equiv \sum_{i=1}^{v} h_i(\mathbf{x}, c_i)$, the combined reaction hazard
4. Simulate time to next event as a random quantity $t' \sim \mathrm{Exp}(h_0(\mathbf{x}, \mathbf{c}))$
5. Set $t := t + t'$
6. Simulate the reaction index j as a discrete random quantity with probabilities $h_i(\mathbf{x}, c_i)/h_0(\mathbf{x}, \mathbf{c})$, $i = 1, \ldots, v$
7. Update $\mathbf{x} := \mathbf{x} + S^{(j)}$, where $S^{(j)}$ is the jth column of S
8. If $t \geq T_{\max}$, output x and t, else goto step 2

Stochastic Petri nets (SPN)

- SPN is a convenient mathematical and graphical representation of a stochastic kinetic process with rate laws h_i and stochastic rate constants c_i

Figure: An example from p. 184 from (Wilkinson, 2011)

The Gillespie Algorithm (4)

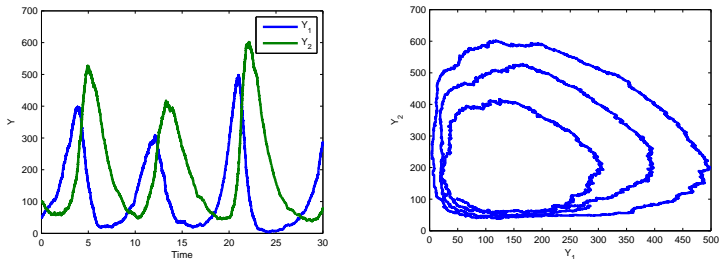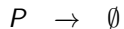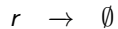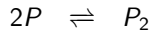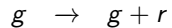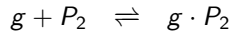▶ An example from (Wilkinson, 2011), Section 6.5

Figure: A single realization of a stochastic Lotka–Volterra process in time-space and phase-space for initial values $Y_1(0) = 50$, $Y_2(0) = 100$ and the stochastic rate constants are $c_1 = 1, c_2 = 0.005, c_3 = 0.6$

Prokaryotic auto-regulation

- An auto-regulatory model (see Fig. 1.7)

$$
\begin{aligned}
g + P_2 &\rightleftharpoons g \cdot P_2 \\
g &\rightarrow g + r \\
r &\rightarrow r + P \\
2P &\rightleftharpoons P_2 \\
r &\rightarrow \emptyset \\
P &\rightarrow \emptyset
\end{aligned}
$$

Prokaryotic auto-regulation (2)

Figure: Figure 1.7 from (Wilkinson, 2011)

Prokaryotic auto-regulation (3)

- Stochastic simulation of the auto-regulation model: first 5000 time units (right) and 250 time units (left)

Figure: Figure 7.12 from (Wilkinson, 2011)

- Jumps in protein dimer level co-inside with the transcript changes

Prokaryotic auto-regulation (4)

- Fluctuations are very significant during the first 10 time units
- Distribution of protein level P at time 10
  - Run many (10000) simulations and estimate the density

Figure: Figure 7.13 from (Wilkinson 2011)

- Almost 50% chance of having 0 proteins at time 10

Analysis of reaction network realizations

- Consider dimerization kinetics at very low concentrations
- Forward and backward equations:

$$2P \xrightarrow{c_1} P_2 \quad \text{and} \quad P_2 \xrightarrow{c_2} 2P$$
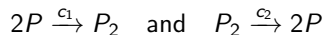
Analysis of reaction network realizations (2)

- Dynamics can be stimulated using the Gillespie algorithm
- A single realization looks as follows

Figure: Figure 7.2b form (Wilkinson, 2011)

Analysis of reaction network realizations (3)

- Fluctuations are due to stochasticity of the model, not measurement noise
- Simulation should be run many time in order to understand overall behavior of the system

Figure: Figure 7.5 b) from (Wilkinson, 2001)

Analysis of reaction network realizations (4)

▶ Multiple Gillespie runs (i.e., realisations of a stochastic process) can be summarized to represent distribution of component levels

▶ 3-standard deviations rule: if data was normally distributed, 99% of data points would lie within $\pm 3\sigma$ from the mean

  ▶ For visualization purposes, can be applied to each time point separately

Figure: Figure 7.6 a) from (Wilkinson, 2011)

Analysis of reaction network realizations (5)

- Full (estimated) density can also be shown for each time point

Figure: Figure 7.6 b) from (Wilkinson, 2011)

- Previous illustrations suggest that the system has reached a steady-state distribution, equilibrium distribution being the one shown above

Approximate simulation strategies

- Gillespie's algorithm has nice properties
  - Simulates every reaction event
  - Generates exact independent realizations
  - Is reasonably fast
- However, other simulation algorithms exist, some of which are
  - Faster yet still exact (e.g. next reaction method (NRM), the Gibson-Bruck algorithm)
  - Much faster, but approximative (e.g. time discretization methods, Gillespie's τ-leap method)
- We will look at the time discretization methods

Time discretization method

- ► All approximative methods are based on time discretization
  - ► Similar to fixed-time interval discretization approximation of continuous-time Markov chains
  - ► Essential idea: time axis is split into small fixed-size time interval

Time discretization method

- ▶ All approximative methods are based on time discretization
  - ▶ Similar to fixed-time interval discretization approximation of continuous-time Markov chains
  - ▶ Essential idea: time axis is split into small fixed-size time interval
- ▶ Assumptions:
  - ▶ Time intervals are sufficiently small so that reaction hazards can be assumed to be roughly constant over the interval
- ▶ Motivation of the approach:
  - ▶ A point process with a constant hazard is a homogeneous Poisson process
  - ▶ Based on the Poisson process properties, we assume that the number of reactions of a given type occurring in a short time interval has a Poisson distribution, independent of other reactions

Time discretization method

Poisson timestep method

1. Initialize: $t = 0$, rate constants $\mathbf{c}$, state $\mathbf{x}$, stoichiometry S
2. Calculate $h_i(\mathbf{x}, c_i)$ for all $i = 1, \ldots, v$ and simulate v-dimensional reaction vector $\mathbf{r}$, where

$$r_i \sim \text{Poisson}(h_i(\mathbf{x}, c_i)\Delta t)$$

3. Update the state according to $\mathbf{x} := \mathbf{x} + S\mathbf{r}$
4. Update $t := t + \Delta t$
5. Output t and $\mathbf{x}$
6. If $t < T_{\max}$, then return to step 2

The Master Equation

- The (chemical) master equation is a set of differential equations that fully determines the state evolution of a continuous-time discrete-state system

---

Proposition 6.1.

Kolmogorov's forward equations (5.8) for a SPN can written as

$$\frac{d}{dt}p(\mathbf{x}_0, t_0, \mathbf{x}, t) = \sum_{i=1}^{v}[h_i(\mathbf{x} - S^{(i)}, c_i)p(\mathbf{x}_0, t_0, \mathbf{x} - S^{(i)}, t)$$
$$- h_i(\mathbf{x}, c_i)p(\mathbf{x}_0, t_0, \mathbf{x}, t)], \quad \forall t_o \in \mathbb{R}, \mathbf{x}_0, \mathbf{x} \in \mathcal{M},$$

where $\mathcal{M}$ is a countable state space of the process. This set of differential equations is often referred to as the chemical master equation. Proof on page 195. □

- In general, a master equation approach to the analysis of stochastic kinetic models is not analytically tractable

The Master Equation (2)

Figure: Proof from p. 195 from (Wilkinson, 2011)

The Master Equation (3)

Figure: Proof from p. 195 from (Wilkinson, 2011)

From stochastic kinetics to deterministic formulation

▶ Using the master equation, it is relatively easy to show that the relationship between the continuous deterministic formulation and the expected valued of the stochastic kinetic model is

$$\frac{\partial}{\partial t}\mathrm{E}(X_t) = \sum_{i=1}^{v} S^{(i)}\mathrm{E}(h_i(X_t, c_i))$$

From stochastic kinetics to deterministic formulation (2)

Figure: Proof from p. 197 form (Wilkinson, 2011)

From stochastic kinetics to deterministic formulation (3)

- Recall the rate laws for the zero- and first-order reactions

$$h_i(x, c_i) = c_i \quad \text{and} \quad h_i(x, c_j) = c_i x_j$$

- Assuming zero- and first-order reactions, by linearity of expectation

$$
\begin{aligned}
\frac{\partial}{\partial t} \mathrm{E}(X_t) &= \sum_{i=1}^{v} S^{(i)} \mathrm{E}(h_i(X_t, c_i)) \\
&= \sum_{i=1}^{v} S^{(i)} h_i(\mathrm{E}(X_t), c_i)
\end{aligned}
$$

From stochastic kinetics to deterministic formulation (3)

- Recall the rate laws for the zero- and first-order reactions

$$h_i(x, c_i) = c_i \quad \text{and} \quad h_i(x, c_j) = c_i x_j$$

- Assuming zero- and first-order reactions, by linearity of expectation

$$
\begin{aligned}
\frac{\partial}{\partial t} \mathrm{E}(X_t) &= \sum_{i=1}^{v} S^{(i)} \mathrm{E}(h_i(X_t, c_i)) \\
&= \sum_{i=1}^{v} S^{(i)} h_i(\mathrm{E}(X_t), c_i)
\end{aligned}
$$

- Substituting $\mathbf{y}(t) = \mathrm{E}(X_t)$

$$\frac{d}{dt}\mathbf{y}(t) = \sum_{i=1}^{v} S^{(i)} h_i(\mathbf{y}(t), c_i) = Sh(\mathbf{y}(t), \mathbf{c}),$$

where $S = [S^1, \ldots, S^v]$ is the stoichiometric matrix

From stochastic kinetics to deterministic formulation (4)

- So, when all reactions are zero- and first-order, the deterministic solution is equal to the expected value of the stochastic kinetic model
  - A set of linear differential equations
- But the deterministic solution does not describe the expectation exactly for any system containing second- or higher-order reactions

Software for Simulating Stochastic Kinetic Networks

- Encode models in SBML and then import them into simulation software
- Simulators are (at least should be) memory efficient, accurate and fast
  - COPASI, Complex Pathway Simulator, http://www.copasi.org/

References

- Gillespie DT, Exact Stochastic Simulation of Coupled Chemical Reactions, *The Journal of Physical Chemistry*, 81(25): 2340-2361, 1977.
- Darren J. Wilkinson, *Stochastic Modelling for Systems Biology*, Chapman & Hall/CRC, 2011