

# CS-E5885 Modeling biological networks

## Stochastic biochemical kinetics

Harri Lähdesmäki

Department of Computer Science  
Aalto University

January 16, 2024

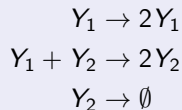
# Contents

- ▶ 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)



- ▶ Reaction equations capture the key interactions, but are insufficient to determine the full stochastic dynamic behaviour of the system
- ▶ Need to know rates of all reactions

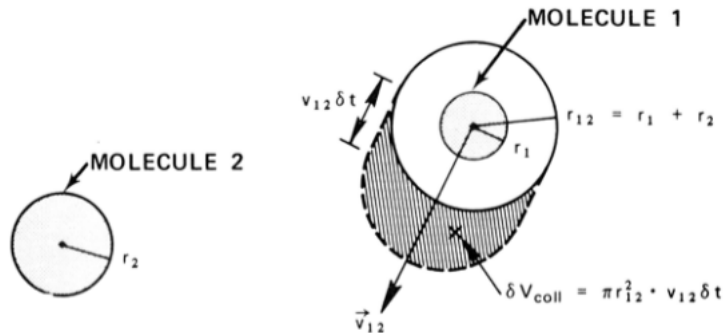
# Molecular approach to kinetics

- ▶ Consider a bi-molecular reaction



- ▶ 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 a (small) constant volume  $V$ , the contents are well stirred, and the temperature is constant

## Physical basis of the stochastic formulation



**Figure 1.** The “collision volume”  $\delta V_{\text{coll}}$  which molecule 1 will sweep out relative to molecule 2 in the next small time interval  $\delta t$ .

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, content is well-stirred, and temperature remains constant, then
  - ▶ Collision of the two molecules has a **constant hazard**
  - ▶ (Formal proof using statistical mechanical arguments)
- ▶ The above conclusion builds on the following arguments:
  - ▶  $P_1$  and  $P_2$  are uniformly (and independently) distributed over the volume
  - ▶ This distribution does not depend on time
  - ▶ The probability that the molecules are within reaction distance is also independent on time
  - ▶ In other words, collision of the two molecules has a **constant hazard**

## Molecular approach to kinetics: intuitive explanation

- 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: intuitive explanation

- 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)

$$\mathbb{P}(|P_1 - P_2| < r) = \mathbb{E}_{P_2} \left( \underbrace{\mathbb{P}(|P_1 - P_2| < r \mid P_2 = p_2)}_A \right)$$



## Molecular approach to kinetics: intuitive explanation

- ▶ 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)

$$\mathbb{P}(|P_1 - P_2| < r) = \mathbb{E}_{P_2} \left( \underbrace{\mathbb{P}(|P_1 - P_2| < r \mid P_2 = p_2)}_A \right)$$

- ▶ But the conditional probability  $A$  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: intuitive explanation

- ▶ 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)

$$\mathbb{P}(|P_1 - P_2| < r) = \mathbb{E}_{P_2} \left( \underbrace{\mathbb{P}(|P_1 - P_2| < r \mid P_2 = p_2)}_A \right)$$

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

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

- ▶ Using a similar reasoning: reactions that involve different molecule(s) are independent

## Molecular approach to kinetics: varying volume

- ▶ The results on the previous slides mean that if
  1. Molecules are uniformly distributed (time independent), and
  2. 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: varying volume

- ▶ The results on the previous slides mean that if
  1. Molecules are uniformly distributed (time independent), and
  2. 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**
- ▶ Given that molecules are within a reaction distance, they will not necessarily interact but do so with a probability independent of  $V$
- ▶ If the volume changes, then the reaction hazard is inversely proportional to the volume
- ▶ We will assume a fixed volume  $V$

## Mass-action stochastic kinetics

- ▶ Species  $P = (\mathcal{X}_1, \dots, \mathcal{X}_u)^T$  and reactions  $T = (R_1, \dots, R_v)^T$
- ▶ Qualitative structure of the reaction network is encoded as a Petri net  $N = (P, T, Pre, Post, M)$
- ▶  $\mathbf{x} = (x_1, \dots, x_u)^T$  is the current state of the system

# Mass-action stochastic kinetics

- ▶ Species  $P = (\mathcal{X}_1, \dots, \mathcal{X}_u)^T$  and reactions  $T = (R_1, \dots, R_v)^T$
- ▶ Qualitative structure of the reaction network is encoded as a Petri net  $N = (P, T, Pre, Post, M)$
- ▶  $\mathbf{x} = (x_1, \dots, x_u)^T$  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  $h_i(\mathbf{x}, c_i) dt$
- ▶ This model is called stochastic Petri net (SPN)

# Rate law: zeroth and first-order reactions

- ▶ Zeroth-order reactions



- ▶ In practice, products are not created from nothing though...

- ▶  $h_i(\mathbf{x}, c_i) = c_i$

- ▶ Can be used to model constant influx

# Rate law: zeroth and first-order reactions

- ▶ Zeroth-order reactions

- ▶  $R_i : \emptyset \xrightarrow{c_i} \mathcal{X}$
- ▶ In practice, products are not created from nothing though...
- ▶  $h_i(\mathbf{x}, c_i) = c_i$
- ▶ Can be used to model constant influx

- ▶ First-order reactions

- ▶  $R_i : \mathcal{X}_j \xrightarrow{c_i} ?$
- ▶  $c_i$  represents the hazard of a particular molecule of  $\mathcal{X}_j$
- ▶ There are  $x_j$  molecules of  $\mathcal{X}_j$ , thus

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

- ▶ Represents a spontaneous change of a molecule into one or more molecules



## Rate law: second-order reactions

- ▶ Second-order reactions

- ▶  $R_i : \mathcal{X}_j + \mathcal{X}_k \xrightarrow{c_i} ?$
- ▶  $c_i$  represents the hazard that a particular pair of molecules  $\mathcal{X}_j$  and  $\mathcal{X}_k$  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: second-order reactions

- ▶ Second-order reactions

- ▶  $R_i : \mathcal{X}_j + \mathcal{X}_k \xrightarrow{c_i} ?$
- ▶  $c_i$  represents the hazard that a particular pair of molecules  $\mathcal{X}_j$  and  $\mathcal{X}_k$  will react
- ▶ There are  $x_j x_k$  different pairs molecules, thus

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

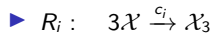
- ▶ For a special type of second-order reaction  $R_i : 2\mathcal{X}_j \xrightarrow{c_i} ?$

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

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

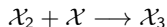
## Rate law: higher-order reactions

- ▶ Higher-order reactions



- ▶  $h_i(\mathbf{x}, c_i) = c_i \binom{x}{3} = c_i \frac{x!}{(x-3)!3!} = c_i \frac{x(x-1)(x-2)}{6}$

- ▶ In most cases it is likely to be more realistic to model the process as the pair of second-order reactions



## Poisson process perspective

- ▶ Reaction hazards depend only on the current state of the system, i.e., reactions have constant hazard for fixed  $\mathbf{x}$
- ▶ Thus, using the results for Poisson processes, we can conclude that
  - ▶ In the absence of any other reactions, the time to reaction  $i$  would be distributed as  $\text{Exp}(h_i(\mathbf{x}, c_i))$
  - ▶ In a given reaction system with  $v$  reactions, 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)

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

## Poisson process perspective

- ▶ Reaction hazards depend only on the current state of the system, i.e., reactions have constant hazard for fixed  $\mathbf{x}$
- ▶ Thus, using the results for Poisson processes, we can conclude that
  - ▶ In the absence of any other reactions, the time to reaction  $i$  would be distributed as  $\text{Exp}(h_i(\mathbf{x}, c_i))$
  - ▶ In a given reaction system with  $v$  reactions, 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)

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

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

# The Gillespie Algorithm: summary

## The Gillespie Algorithm

1. Initialize the system at  $t = 0$  with rate constants  $c_1, \dots, c_v$  and initial numbers of molecules for each species  $\mathbf{x} = (x_1, \dots, x_u)^T$
2. For each  $i = 1, \dots, v$  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  $t' \sim \text{Exp}(h_0(\mathbf{x}, \mathbf{c}))$
5. Set  $t := t + t'$
6. Simulate the reaction index  $j$  as a discrete random quantity with probabilities  $h_i(\mathbf{x}, c_i)/h_0(\mathbf{x}, \mathbf{c})$ ,  $i = 1, \dots, v$
7. Update  $\mathbf{x} := \mathbf{x} + S^{(j)}$ , where  $S^{(j)}$  is the  $j$ th column of  $S$
8. Output  $\mathbf{x}$  and  $t$
9. If  $t < t_{\max}$ , then goto step 2

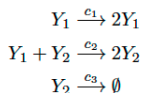
# The Gillespie algorithm: stochastics aspects

- ▶ To summarize: we have that
  1. Reaction hazards remain constant until the next reaction, i.e.:
    - ▶ Homogeneous Poisson process until the next reaction, or alternatively
    - ▶ Homogeneous continuous time Markov process until the next reaction
  2. Each reaction is an independent random event with exponential waiting time (in the absence of other reactions)
  3. The next reaction has an exponential waiting time
    - ▶ Rate (or hazard) is the sum of rates of all reactions
  4. 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

# Stochastic Petri nets example

## ► Lotka-Volterra example

Here we will use the usual equations



leading to stochastic rate laws

$$\begin{aligned}h_1(y, c_1) &= c_1 y_1 \\ h_2(y, c_2) &= c_2 y_1 y_2 \\ h_3(y, c_3) &= c_3 y_2.\end{aligned}$$

The SPN corresponding to this system could be written

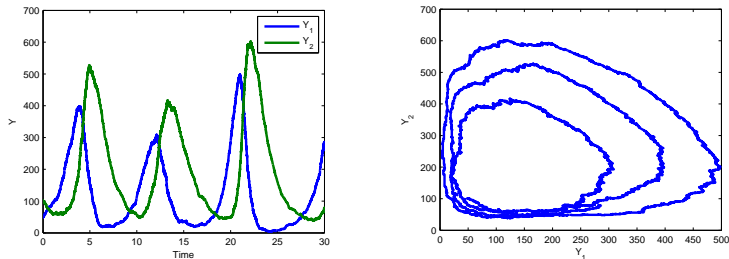
$$\begin{aligned}N &= (P, T, Pre, Post, M, h, c), \quad P = (\text{Prey}, \text{Predator})^T, \\ T &= (\text{Prey reproduction}, \text{Predator-prey interaction}, \text{Predator death})^T, \\ Pre &= \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad Post = \begin{pmatrix} 2 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix}, \quad h(y, c) = (c_1 y_1, c_2 y_1 y_2, c_3 y_2)^T.\end{aligned}$$

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



## Stochastic Petri nets example (2)

### ► Lotka-Volterra example

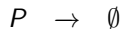
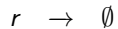
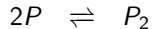
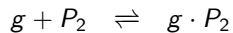


**Figure:** A single realization of a stochastic Lotka–Volterra process in time-space and phase-space for initial values  $Y_1(0) = 50$ ,  $Y_2(0) = 100$  and the stochastic rate constants are  $c_1 = 1$ ,  $c_2 = 0.005$ ,  $c_3 = 0.6$



# Prokaryotic auto-regulation: coupled reactions

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



# Prokaryotic auto-regulation: stochastic simulation

- ▶ Stochastic simulation of the auto-regulation model: first 5000 time units (right) and 250 time units (left)
- ▶ Changes in transcript, protein and protein dimer levels happen at approx. same time

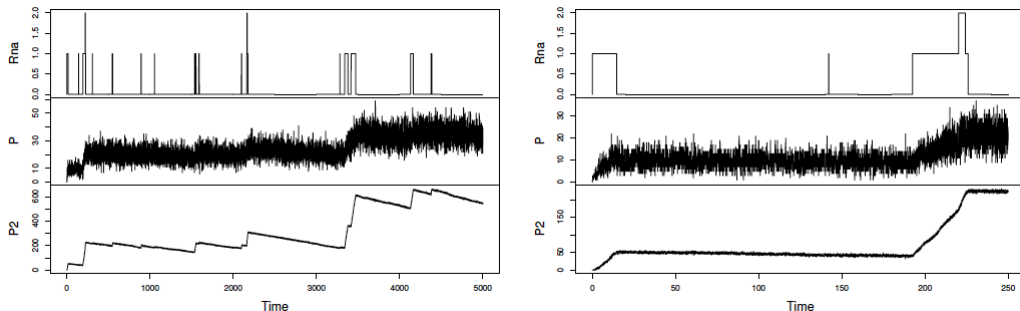


Figure: Figure 7.12 from (Wilkinson, 2011)

## Prokaryotic auto-regulation: stochastic simulation (2)

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

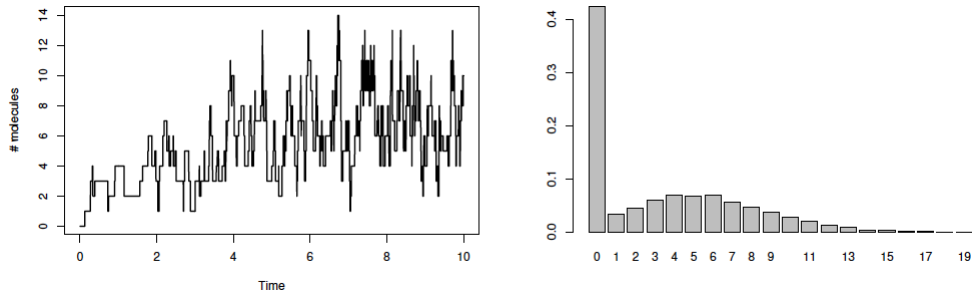
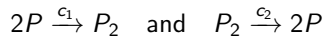


Figure: Figure 7.13 from (Wilkinson 2011)

## Analysis of reaction network realizations

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



## Analysis of reaction network realizations (2)

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

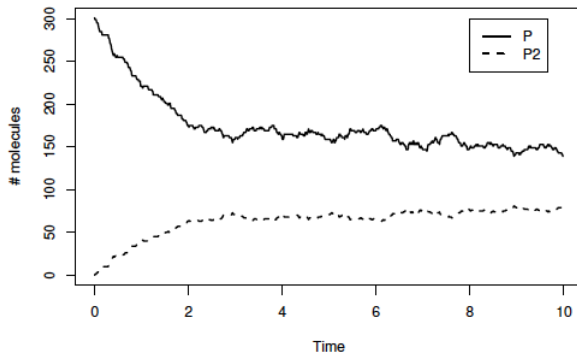


Figure: Figure 7.2b from (Wilkinson, 2011)

## 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

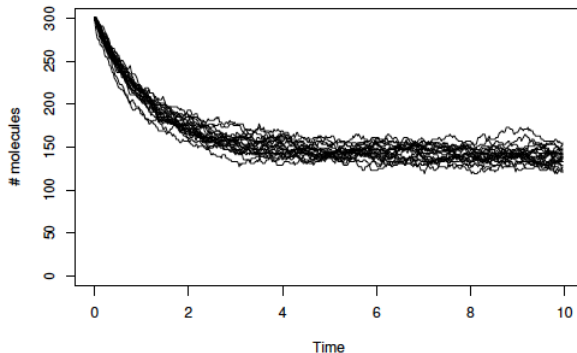


Figure: Figure 7.5 b) from (Wilkinson, 2001)



## 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

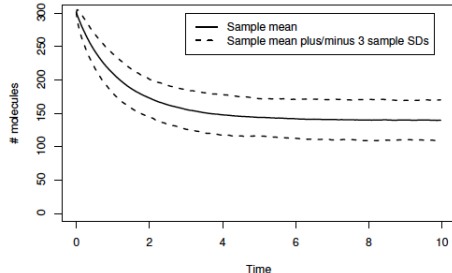


Figure: Figure 7.6 a) from (Wilkinson, 2011)

## Analysis of reaction network realizations (5)

- ▶ Full (estimated) density can also be shown for each time point
- ▶ Previous illustrations suggest that the system has reached a steady-state distribution

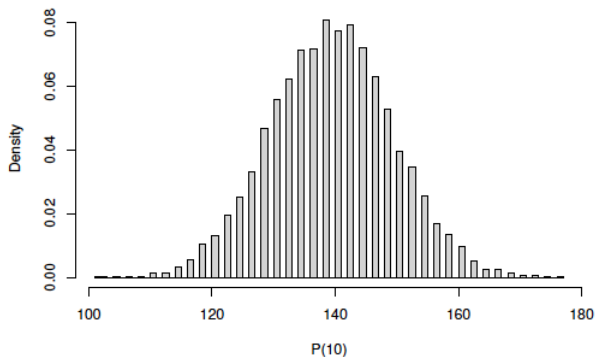


Figure: Figure 7.6 b) from (Wilkinson, 2011)

# Contents

- ▶ Mass-action stochastic kinetics:
- ▶ **Approximate simulation strategies**
- ▶ The master equation
- ▶ Connection between continuous Markov processes and ODEs

# Approximate simulation strategies

- ▶ Gillespie's algorithm has nice properties
  - ▶ Simulates every reaction event
  - ▶ Generates exact realizations of the underlying stochastic process
  - ▶ 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  $\tau$ -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 intervals  $\Delta t$

# 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 intervals  $\Delta t$
- ▶ Assumptions:
  - ▶ Time intervals  $\Delta t$  are sufficiently small so that reaction hazards  $h_i(\mathbf{x}, c_i)$  can be assumed to be approximately 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

# Poisson timestep method

## Poisson timestep algorithm

1. Initialize:  $t = 0$ , rate constants  $\mathbf{c}$ , state  $\mathbf{x}$ , stoichiometry  $S$
2. Calculate  $h_i(\mathbf{x}, c_i)$  for all  $i = 1, \dots, \nu$  and simulate  $\nu$ -dimensional reaction vector  $\mathbf{r} = (r_1, \dots, r_\nu)^T$ , 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

# Contents

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



# 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 be written as

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

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)

$$\begin{aligned}\frac{d}{dt}p(x_0, t_0, x, t) &= \sum_{\{x' \in \mathcal{M}\}} q_{x',x}p(x_0, t_0, x', t) \\&= \left[ \sum_{\{x' \in \mathcal{M} | x' \neq x\}} q_{x',x}p(x_0, t_0, x', t) \right] + q_{x,x}p(x_0, t_0, x, t) \\&= \left[ \sum_{\{x' \in \mathcal{M} | x' \neq x\}} q_{x',x}p(x_0, t_0, x', t) \right] - p(x_0, t_0, x, t) \sum_{\{x' \in \mathcal{M} | x' \neq x\}} q_{x,x'} \\&= \sum_{\{x' \in \mathcal{M} | x' \neq x\}} \left[ q_{x',x}p(x_0, t_0, x', t) - q_{x,x'}p(x_0, t_0, x, t) \right].\end{aligned}$$

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

## The master equation (3)

$$= \sum_{\{x' \in \mathcal{M} | x' \neq x\}} \left[ q_{x',x} p(x_0, t_0, x', t) - q_{x,x'} p(x_0, t_0, x, t) \right].$$

This is just Kolmogorov's forward equation rewritten more appropriately for a Markov process with general countable state space  $\mathcal{M}$ . The equation involves a sum over all possible transitions, but for a SPN, only a finite number of transition events are possible, corresponding to the  $v$  different reaction channels. Considering first the hazard  $q_{x,x'}$ , we note that starting from  $x$ , it is only possible to move to  $x + S^{(i)}$ ,  $i = 1, 2, \dots, v$ , and then by definition we have  $q_{x, x+S^{(i)}} = h_i(x, c_i)$ . Similarly, in order to get to  $x$ , the process must have come from one of  $x - S^{(i)}$ ,  $i = 1, 2, \dots, v$ , and in this case we have  $q_{x-S^{(i)}, x} = h_i(x - S^{(i)}, c_i)$ . Substituting these into the above equation gives the result.  $\square$

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

# Contents

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

## 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 value of the stochastic biochemical kinetic model is

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

## From stochastic kinetics to deterministic formulation (2)

$$\begin{aligned}\frac{\partial}{\partial t} \mathbb{E}(X_t) &= \frac{\partial}{\partial t} \sum_{x \in \mathcal{M}} x p(x, t) \\&= \sum_{x \in \mathcal{M}} x \frac{\partial}{\partial t} p(x, t) \\&= \sum_{x \in \mathcal{M}} x \sum_{i=1}^v \left[ h_i(x - S^{(i)}, c_i) p(x - S^{(i)}, t) - h_i(x, c_i) p(x, t) \right] \\&= \sum_{i=1}^v \left[ \sum_{x \in \mathcal{M}} x h_i(x - S^{(i)}, c_i) p(x - S^{(i)}, t) - \sum_{x \in \mathcal{M}} x h_i(x, c_i) p(x, t) \right] \\&= \sum_{i=1}^v \left[ \sum_{x \in \mathcal{M}} (x + S^{(i)}) h_i(x, c_i) p(x, t) - \sum_{x \in \mathcal{M}} x h_i(x, c_i) p(x, t) \right] \\&= \sum_{i=1}^v \left[ \mathbb{E} \left( (X_t + S^{(i)}) h_i(X_t, c_i) \right) - \mathbb{E} (X_t h_i(X_t, c_i)) \right] \\&= \sum_{i=1}^v \mathbb{E} \left( S^{(i)} h_i(X_t, c_i) \right) \\&= \sum_{i=1}^v S^{(i)} \mathbb{E} (h_i(X_t, c_i)).\end{aligned}$$

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

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

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

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

## From stochastic kinetics to deterministic formulation (3)

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

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

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

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

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

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

where  $S = [S^{(1)}, \dots, S^{(v)}]$  and  $\mathbf{h}(\mathbf{y}(t), \mathbf{c}) = (h_1(\mathbf{y}(t), \mathbf{c}), \dots, h_v(\mathbf{y}(t), \mathbf{c}))^T$



## 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