Systems Biology: Stochastic Approach

A. Borri[1] P. Palumbo[1]

[1]IASI-CNR, Rome, Italy

1st Introductory Course on Systems Biology,
IASI-CNR, Rome, Italy
March 6, 2014

The micro-world is inherently discrete and stochastic (noisy)

"An approach to the chemical kinetics of spatially homogeneous systems which is somewhat more broadly applicable than the deterministic formulation is the stochastic formulation.

Here the reaction constants are viewed not as reaction rates but as reaction probabilities per unit time, and the temporal behavior of a chemically reacting system takes the form of a Markovian random walk in the n-dimensional space of the molecular populations of the n species."

"From a physical point of view, the stochastic formulation of chemical kinetics is superior to the deterministic formulation: the stochastic approach is always valid whenever the deterministic approach is valid, and is sometimes valid when the deterministic approach is not."

Daniel Gillespie, 1976.

The micro-world is inherently discrete and stochastic (noisy)

> *"The time evolution of a spatially homogeneous mixture of chemically reacting molecules is often modeled using a stochastic formulation, which takes into account the inherent randomness of thermal molecular motion."*

Abhyudai Singh, 2011.

Outline

Probability Space

Definition (Probability Space)

A probability space is a triple $(\Omega, \mathcal{F}, P)$, where:

- Ω is the set of all possible outcomes (sample space);
- $\mathcal{F}$ is a collection of subsets of Ω (events), satisfying the properties of a σ-algebra
- $P : \mathcal{F} \to [0, 1]$ is a function assigning a probability to the events of $\mathcal{F}$.

Properties of $\mathcal{F}$:

- $\mathcal{F}$ is non-empty;
- $A \in \mathcal{F}$ implies the complementary event $A^c \in \mathcal{F}$.
- $\{A_i\} \subseteq \mathcal{F} \implies \cup_i A_i \in \mathcal{F}$.

Properties of P:

- $P(\Omega) = 1$;
- $P(A) = 1 - P(A^c)$;
- $A_j \cap A_k = \emptyset \quad j \neq k \implies P(\cup_i A_i) = \sum_i P(A_i)$

Conditional Probabilities

Definition (Conditional Probability)

Conditional probability of A given B:

$$P(A|B) := \frac{P(A \cap B)}{P(B)}.$$

Theorem (Law of total probability)

Let $\{B_i\}$ be a finite (countable) partition of Ω:

$$B_j \cap B_k = \emptyset \qquad j \neq k.$$

Then, for any event $A \in \mathcal{F}$, one has:

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i).$$

Random Variable

Definition (Discrete Random Variable)

A discrete random variable $\mathcal{X}$ is an object defined by:

- a finite/countable set of possible values (state space) X;
- a probability density function over this set $P : X \to [0, 1]$ satisfying $\sum_{x \in X} P(x) = 1$.

Expectation of a random variable:

$$\langle \mathcal{X} \rangle = \sum_{x \in X} x \cdot P(x)$$

Expectation of a function of random variable:

$$\langle f(\mathcal{X}) \rangle = \sum_{x \in X} f(x) \cdot P(x)$$

The value $\langle \mathcal{X}^m \rangle$ is called m-th moment of $\mathcal{X}$.

Stochastic Process

Definition (Stochastic Process)

Given a totally ordered set (time) T, a stochastic (random) process is a collection

$$\{\mathcal{X}(t), t \in T\}$$

s.t. $\mathcal{X}(t)$ is a random variable for any t.

A stochastic process is called:

- discrete-time, if T is a finite/countably infinite set;
- continuous-time, if T is a uncountably infinite set.

The probability density $P(t, x)$ expresses the probability that the process assumes the value x at time t, for any $t \in T$ and any $x \in X$:

$$P(t, x) \qquad t \in T, x \in X$$

Stochastic Process

We can define, for any n, any choice of successive times $t_1 < t_2 < ... < t_n$ and any $x_1, ...x_n$ in X the following:

- Joint Probability density:

$$P(t_1, x_1; t_2, x_2; ...; t_{n-1}, x_{n-1}; t_n, x_n)$$

- Conditional Probability densities:

$$P(t_n, x_n | t_1, x_1; t_2, x_2; ...; t_{n-1}, x_{n-1}) = \frac{P(t_1, x_1; t_2, x_2; ...; t_{n-1}, x_{n-1}; t_n, x_n)}{P(t_1, x_1; t_2, x_2; ...; t_{n-1}, x_{n-1})}$$

Markov Property

Definition (Markov Process)

A Markov Process is defined as a stochastic process with the property that, for any n and any choice of successive times $t_1 < t_2 < ... < t_n$, one has:

$$P(t_n, x_n | t_1, x_1; t_2, x_2; ...; t_{n-1}, x_{n-1}) = P(t_n, x_n | t_{n-1}, x_{n-1}).$$

A Markov Process is *memoryless*, i.e. the conditional probability distribution of future values depends just on the value assumed at the most recent time (*state* of the process).

In a stationary Markov Process, the transition probability just depends on the time shift. Hence, we write:

$$\tilde{w}_{\Delta t}(x_n | x_{n-1}) := P(t_n, x_n | t_{n-1}, x_{n-1}) \qquad \text{where } \Delta t = t_n - t_{n-1}.$$

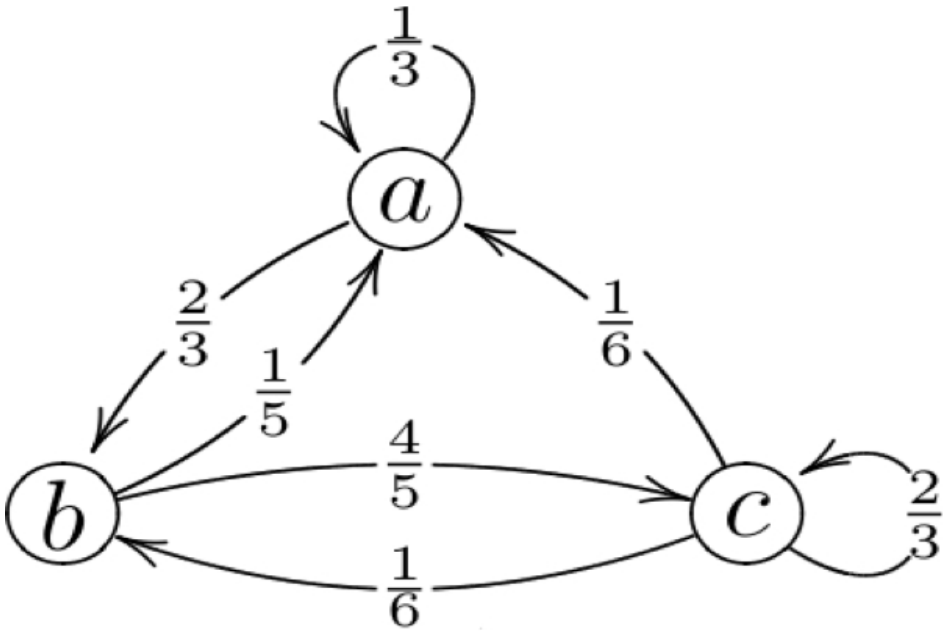
Discrete-Time Markov Chains (DTMC)

We assume integer times: $T = \mathbb{Z}$ and we define the one-step transition probabilities:

$$w(x_n|x_{n-1}) := \tilde{w}_{\Delta t}(x_n|x_{n-1})|_{\Delta t=1}.$$

Note that $w(\cdot|x_{n-1})$ defines a conditional density:

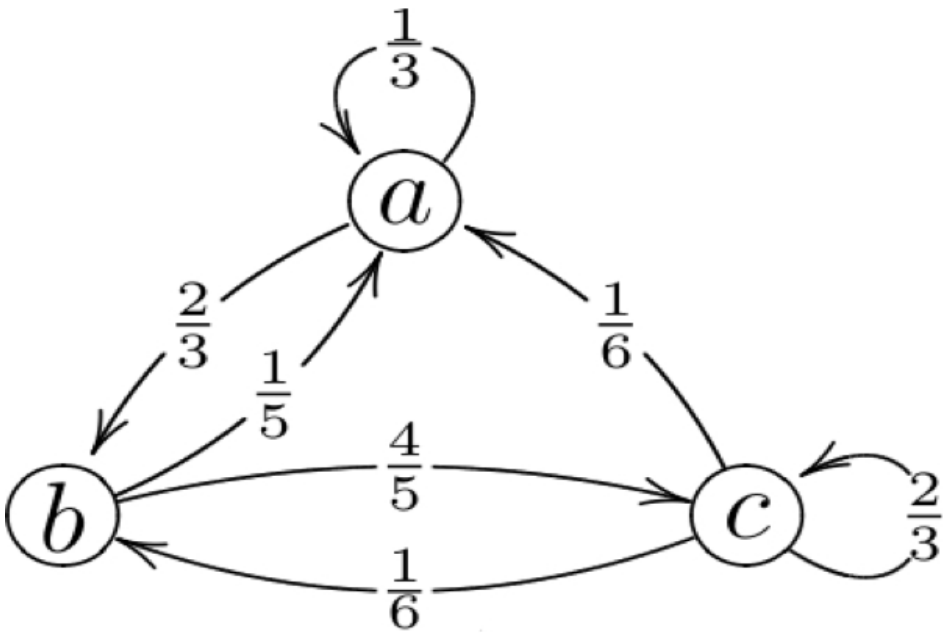
$$\sum_{x_n \in X} w(x_n|x_{n-1}) = 1.$$



Discrete-Time Markov Chains (DTMC)

We write the variation of probability of being in state x_i:

$$P_i(t+1) = P_i(t) - \overbrace{\sum_{j\neq i} \underbrace{w(x_j|x_i)}_{\substack{\text{conditional probability} \\ \text{of a transition from } x_i}} P_i(t)}^{\text{Law of total probability}} + \overbrace{\sum_{j\neq i} \underbrace{w(x_i|x_j)}_{\substack{\text{conditional probability} \\ \text{of a transition to } x_i}} P_j(t)}^{\text{Law of total probability}}$$



Continuous-Time Markov Chains (CTMC)

- We assume real times: $T = \mathbb{R}$. Since there is no “basic step", we define the *propensity* or *transition probability per unit time*:

$$w(x_n|x_{n-1}) := \lim_{\Delta t \to 0} \frac{P(t + \Delta t, x_n|t, x_{n-1})}{\Delta t} \qquad x_n \neq x_{n-1}.$$

- We assume the limit exists and is independent of t (time-homogeneous process).
- It means that the probability that a transition from x_{n-1} to x_n occurs in the interval $[t, t + dt)$ is $w(x_n|x_{n-1})dt$.
- Note that $w(\cdot|x_{n-1})$ is not a probability and does not define a probability density:

$$\sum_{x_n \in X} w(x_n|x_{n-1}) \neq 1.$$

CTMC applied to chemical reactions

- Consider the following system (or network) of M (bio)chemical reactions in the general form:

$$\begin{cases} a^l_{11} Y_1 + \cdots + a^l_{1n} Y_n \longrightarrow a^r_{11} Y_1 + \cdots + a^r_{1n} Y_n \\ \vdots \\ a^l_{M1} Y_1 + \cdots + a^l_{Mn} Y_n \longrightarrow a^r_{M1} Y_1 + \cdots + a^r_{Mn} Y_n \end{cases} \tag{1}$$

  where $Y_1, \ldots, Y_n$ are the species involved.
- The number $v_{\mu j} = a^r_{\mu j} - a^l_{\mu j}$ is called *stoichiometric coefficient* of species j in reaction μ, for all species $j = 1, \ldots, n$ and for all reactions $\mu = 1, \ldots, M$.
- The stoichiometry vector for each reaction μ is defined as:

$$v_\mu := [v_{\mu 1} \cdots v_{\mu n}]'.$$

- We define as $s(t)$ the state of the system at time t, with j-th component $s_j(t) \in \mathbb{N}_0$ denoting the number of copies of the j-th species at time t.

Non-redundant representation of chemical networks

- We denote by n_E the number of distinct elements that form the n species.
- Closed processes: the total number of copies of each element is conserved, no matter what species it may be a part of.
- This imposes n_E mass-balance constraints on the system, which can be written in matrix form as:

$$F \cdot s(t) = b, \tag{2}$$

where $b \in \mathbb{N}^{n_E}$ is the vector collecting the total mass, expressed in numbers of copies of each element, and $F \in \mathbb{N}_0^{n_E \times n}$ is a mass-balance matrix whose (i, j)-th entry F_{ij} is the number of copies of element i present in one copy of species j.

- Because of the n_E mass-balance equations in (2), the state vector $s(t)$ is redundant, because only $q = n - n_E$ state variables are independent, assuming $n > n_E$ and $rank(F) = n_E$. This leads to the concept of *independent species*.

Non-redundant representation of chemical networks

Definition (Independent species)

Consider M chemical reactions (1) involving n_E elements and $n > n_E$ species, and let $F = [F_1 \cdots F_n]$ be the matrix of mass-balance constraints $Fs(t) = b$ with $rank(F) = n_E$. Let $\{Y_{r_1}, \ldots, Y_{r_q}\}$, with $\{r_1, \ldots, r_M\} \subseteq \{1, \ldots, n\}$, be a subset of the species in (1), with $q = n - n_E$. $\{Y_{r_1}, \ldots, Y_{r_q}\}$ is a set of *independent species* for (1) if the matrix $\bar{F}$, obtained by erasing columns $r_1, \ldots, r_q$ from F, is nonsingular.

Proposition

Consider M chemical reactions involving n_E elements and $n > n_E$ species, with the mass-balance constraints $Fs(t) = b$ and $rank(F) = n_E$. Let $\{Y_{r_1}, \ldots, Y_{r_q}\}$, with $\{r_1, \ldots, r_q\} \subseteq \{1, \ldots, n\}$, be a set of independent species for the M chemical reactions. Then the reduced state vector $x(t) = [s_{r_1}(t) \cdots s_{r_q}(t)]^T$ is a non-redundant representation of vector $s(t)$ at any time t.

Toy Example 1

Consider the following closed system of $M = 4$ reactions:

$$\begin{cases} 2H_2 + O_2 & \rightleftharpoons & 2H_2O \\ CaCO_3 & \rightleftharpoons & CaO + CO_2, \end{cases}$$

in which we have $n = 6$ species ($Y_1 = H_2$, $Y_2 = O_2$, $Y_3 = H_2O$, $Y_4 = CaCO_3$, $Y_5 = CaO$, $Y_6 = CO_2$) and $n_E = 4$ elements (H, O, Ca, C). The matrix F is

$$F = \begin{bmatrix} 2 & 0 & 2 & 0 & 0 & 0 \\ 0 & 2 & 1 & 3 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

The number of independent species is $q = n - n_E = 2$. One possible choice is $\{H_2O, CaO\}$ ($r_1 = 3$, $r_2 = 5$), because the matrix

$$\bar{F} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 3 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

Toy Example 2

Consider a closed system in which $n = 5$ species react according to:

$$\begin{cases} 2H_2 + O_2 & \rightleftharpoons & 2H_2O \\ H_2O + CO_2 & \rightleftharpoons & H_2CO_3. \end{cases}$$

One can write $n_E = 3$ constraints by the mass balance of the elements H, O, C present in the above $M = 4$ reactions.

Since $q = n - n_E = 2$, the system has a 2-dimensional reduced state vector. Two possible choices of independent species are the sets $\{H_2, H_2O\}$ and $\{H_2, CO_2\}$.

Example 1 (Cyclin-dependent kinase)

Let us consider the following $M = 7$ biochemical reactions:

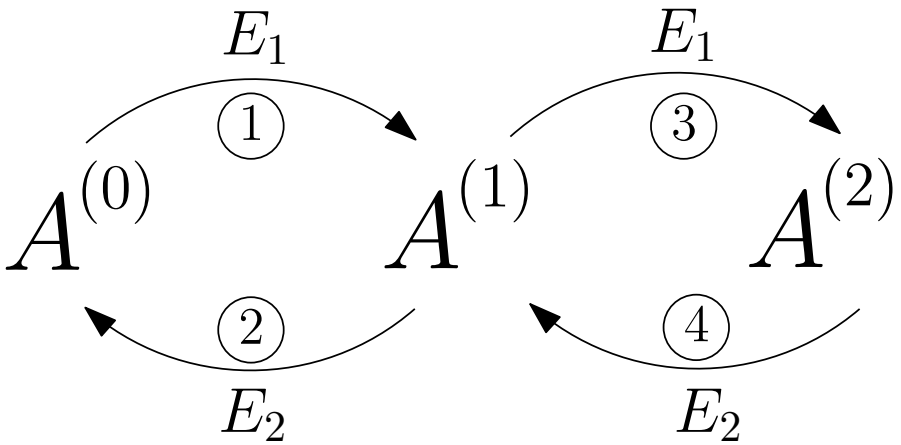
$$\begin{cases} \mathrm{Cdk} + \mathrm{Cln} & \rightleftharpoons & \mathrm{CdkCln} \\ \mathrm{CdkCln} + \mathrm{I} & \rightleftharpoons & \mathrm{CdkClnI} \\ \mathrm{CdkClnI} & \longrightarrow & \mathrm{CdkClnI}_p \\ \mathrm{CdkClnI}_p & \rightleftharpoons & \mathrm{CdkCln} + \mathrm{I}_p, \end{cases}$$

which represent a very general framework involving *cyclin-dependent kinases* (*Cdk*), a family of protein kinases playing a crucial role in regulating the cell cycle.

We have $n = 7$ chemical species, $n_E = 4$ elements, but only $n_C = 3$ mass-balance constraints for the three elementary species *Cdk*, *Cln* and *I*, being *I* and I_p two different forms of the same chemical player. The reduced state is 4-dimensional and a choice of independent species is, for example, given by $Y_1 = Cdk$, $Y_2 = I$, $Y_3 = CdkClnI$ and $Y_4 = I_p$.

Example 2 (Double phosphorylation)

It can be described by $M = 4$ biochemical reactions:



and one mass constraint on the substrate:

$$[A^{(0)}] + [A^{(1)}] + [A^{(2)}] = A_{tot}$$

by neglecting the mass constraint on the phosphate.

We have $n = 3$ species and $n_C = 1$ mass-balance constraints, hence the reduced state is 2-dimensional and a choice of independent species is, for example, given by $Y_1 = [A^{(0)}]$, $Y_2 = [A^{(2)}]$.

CTMC applied to chemical reactions

- n reactants (species), M reactions. Assume independent species.
- $x \in \mathbf{Z}^n$ is the state, i.e. the number of molecules of each species.
- Each reaction $\mu \in \{1, ..., M\}$ is defined by the stoichiometry vector v_μ:

$$\text{Reaction } \mu: \qquad x \to x + v_\mu$$

- For a system containing M distinct reaction types, there are:
  - at most M reaction events that will take the system from x_i to some other state $x_j \neq x_i$;
  - at most M reaction events that will bring the system from $x_k \neq x_i$ to x_i.
- Each reaction has an infinitesimal probability of occurring in the next infinitesimal time step of length dt; this state-dependent quantity is given by the propensity function $w_\mu(x) := w(x + v_\mu|x)$.

$$\text{Probability of reaction } \mu \text{ occurring in } [t, t + dt): \qquad w_\mu(x)dt$$

Continuous-Time Markov Chains (CTMC)

Define the following quantities:

- $P_i(t)$ is the probability that the system is in state $x_i \in Z^n$ at time t;
- $P_i^\mu(t)$ is the probability that the system is in state $x_i^\mu := x_i - v_\mu \in Z^n$ at time t, for $\mu \in \{1, ..., M\}$.

Assumption

At most one reaction occurs in $[t, t + dt)$.

We write the variation of probability of being in state x_i at time t:

$$P_i(t+dt) = P_i(t) - \overbrace{\sum_{\mu=1}^{M} \underbrace{w_\mu(x_i)dt}_{\substack{\text{conditional probability} \\ \text{of reaction } \mu \text{ from } x_i}} P_i(t)}^{\text{Law of total probability}} + \overbrace{\sum_{\mu=1}^{M} \underbrace{w_\mu(x_i^\mu)dt}_{\substack{\text{conditional probability} \\ \text{of reaction } \mu \text{ to } x_i}} P_i^\mu(t)}^{\text{Law of total probability}}$$

hence:

$$\dot{P}_i(t) = -\sum_{\mu=1}^{M} w_\mu(x_i)P_i(t) + \sum_{\mu=1}^{M} w_\mu(x_i^\mu)P_i^\mu(t).$$

The Chemical Master Equation (CME)

Master Equation:

$$\dot{P}_i(t) = -\sum_{\mu=1}^{M} w_\mu(x_i) P_i(t) + \sum_{\mu=1}^{M} w_\mu(x_i^\mu) P_i^\mu(t).$$

> *"The name Master Equation first appeared in a paper (Nordsieck, 1940) in which it actually had the role of a general equation from which all other results were derived. It then got stuck onto a special type of equation, namely the above probability balance equation."*

N.G. Van Kampen.

Example 3 (Genetic Toggle Switch)

A genetic toggle switch can be formalized by the following set of 4 reactions:

$$\begin{cases} \text{Reaction } \mu = 1 & \emptyset \longrightarrow S_1 & v_1 = [1 \quad 0]'; \\ \text{Reaction } \mu = 2 & S_1 \longrightarrow \emptyset & v_2 = [-1 \quad 0]'; \\ \text{Reaction } \mu = 3 & \emptyset \longrightarrow S_2 & v_3 = [0 \quad 1]'; \\ \text{Reaction } \mu = 4 & S_2 \longrightarrow \emptyset & v_4 = [0 \quad -1]'. \end{cases}$$

We define a CME by means of the following transition probabilities per unit time:

$$w_1(x) = \frac{16}{1 + x_2}$$
$$w_2(x) = x_1$$
$$w_3(x) = \frac{50}{1 + x_1^{2.5}}$$
$$w_4(x) = x_2$$

The Chemical Master Equation (CME)

$$\dot{P}_i(t) = -\sum_{\mu=1}^{M} w_\mu(x_i) P_i(t) + \sum_{\mu=1}^{M} w_\mu(x_i^\mu) P_i^\mu(t).$$

The Master Equation can be restated in the notation of Markov Processes, independently of the reactions:

$$\dot{P}(x_i, t) = -\sum_{x_j \in X \setminus \{x_i\}} w(x_j|x_i) P(x_i, t) + \sum_{x_j \in X \setminus \{x_i\}} w(x_i|x_j) P(x_j, t).$$

Note that the Master Equation for state x_i is a *linear ordinary differential equation (ODE)* depending (in general) on all the states x_j. Assume $|X| < +\infty$. We can define the vector $\mathcal{P}(t)$ collecting all probabilities $P(x_i, t)$, for all states $x_i \in X$, to get a linear system in the form:

$$\dot{\mathcal{P}}(t) = G\mathcal{P}(t),$$

whose solution is:

$$\mathcal{P}(t) = e^{Gt}\mathcal{P}(0).$$

In the following, we study the properties of G.

Properties of CMEs

From the CME, the matrix G has the following expression:

$$[G]_{ij} = \begin{cases} -\sum_{j \neq i} w(x_j|x_i) & i = j \\ w(x_i|x_j) & i \neq j \end{cases}$$

P1) G is a Metzler matrix (namely all the off-diagonal components are nonnegative), hence the CME system can be regarded as a positive linear dynamical system.

P2) $\sum_i G_{ij} = G_{jj} + \sum_{i \neq j} G_{ij} = -\sum_{i \neq j} w(x_i|x_j) + \sum_{i \neq j} w(x_i|x_j) = 0$, hence G is singular and admits a non-trivial null space; in matrix form, we write:

$$\mathbf{1}^T G = \mathbf{0}^T.$$

P3) all eigenvalues of G have nonpositive real part, ensuring the convergence of the dynamics to the null space;

P4) $\mathbf{1}^T e^{Gt} = \mathbf{1}^T$ for all t, i.e e^{Gt} is a *column-stochastic* matrix. This ensures that $\mathcal{P}(t)$ is a probability vector (nonnegative entries which add up to 1) at any time t, provided that the initial condition $\mathcal{P}(0)$ is a probability vector.

Stationary solution of CME

A stationary distribution $\mathcal{P}_{ss}$ of the CME is a probability vector that satisfies the following steady-state conditions:

$$\begin{cases} G\mathcal{P} & = \mathbf{0} \\ \mathbf{1}^T\mathcal{P} & = 1 \\ \mathcal{P} & \geq \mathbf{0} \end{cases}$$

Since G is singular, there always exists a non-trivial solution. When is it unique?

Definition

The digraph associated to a CTMC is a weighted digraph (V, E, A), where:

- each vertex in V is a discrete state of the process;
- the weight A_{ij} is the propensity of reaching state v_j from the state v_i;
- The set of edges E is uniquely defined by A and includes all the links (v_i, v_j) with non-zero weight.

Stationary solution of CME

Proposition

- *The stationary distribution of a CTMC is unique if and only if* $rank(G) = dim(G) - 1$*, namely if and only if the digraph associated with the Markov process has a globally reachable vertex.*
- *Under this assumption,* 0 *is a simple eigenvalue of* G*, with left eigenvector* $\mathbf{1}^T$*, and the other eigenvalues of* G *have negative real part.*
- *The stationary distribution is unique and is given by* $\mathcal{P}_{ss} = \frac{u_0}{\mathbf{1}^T u_0}$*, where* u_0 *is the right eigenvector corresponding to the eigenvalue* 0*.*
- *The second smallest eigenvalue of* G *(also called algebraic connectivity of the digraph associated with the Markov process) is related to the convergence speed to the stationary distribution.*

From the CME to ODEs

Let $\mathcal{X}(t)$ be a discrete-valued continuous-time Markov Process at time t. Define

$$\mathbf{x}(t) = \langle \mathcal{X}(t) \rangle = \sum_{x \in X} x \cdot P(x,t)$$

$$\begin{aligned}
\dot{\mathbf{x}}(t) &= \sum_{x \in X} x \cdot \frac{d}{dt} P(x,t) \\
&= \sum_{x \in X} \sum_{x' \in X \setminus \{x\}} x \cdot w(x|x') P(x',t) - \sum_{x \in X} \sum_{x' \in X \setminus \{x\}} x \cdot w(x'|x) P(x,t) \\
&= \sum_{x \in X} \sum_{x' \in X \setminus \{x\}} (x' - x) \cdot w(x'|x) P(x,t)
\end{aligned}$$

Definition (Jump moment)

$$a_v(x) = \sum_{x' \in X \setminus \{x\}} (x' - x)^v \cdot w(x'|x) \qquad v \in \mathbf{N}$$

From the CME to ODEs

As a consequence, one has:

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \sum_{x \in X} \sum_{x' \in X \setminus x} (x' - x) \cdot w(x'|x) P(x,t) \\ &= \sum_{x \in X} a_1(x) P(x,t) = \langle a_1(\mathcal{X}(t)) \rangle\end{aligned}$$

The expectation operator $\langle \cdot \rangle$ is linear, but $a_1(\cdot)$ is nonlinear in general, hence:

$$\langle a_1(\mathcal{X}(t)) \rangle \neq a_1(\langle \mathcal{X}(t) \rangle)$$

Anyway, $a_1(\cdot)$ can be expanded around $\langle \mathcal{X}(t) \rangle$, to have:

$$\begin{aligned}a_1(\mathcal{X}(t)) = a_1(\langle \mathcal{X}(t) \rangle) + a_1'(\langle \mathcal{X}(t) \rangle)(\mathcal{X}(t) - \langle \mathcal{X}(t) \rangle) \\ + \sum_{k=2}^{+\infty} a_1^{(k)}(\langle \mathcal{X}(t) \rangle) \frac{(\mathcal{X}(t) - \langle \mathcal{X}(t) \rangle)^k}{k!}\end{aligned}$$

From the CME to ODEs

We apply the expectation operator, to get:

$$\langle a_1(\mathcal{X}(t))\rangle = a_1(\langle \mathcal{X}(t)\rangle) + \sum_{k=2}^{+\infty} a_1^{(k)}(\langle \mathcal{X}(t)\rangle)\frac{\langle(\mathcal{X}(t) - \langle \mathcal{X}(t)\rangle)^k\rangle}{k!}$$

The macroscopic approach consists in ignoring higher-order moments around the average:

$$\langle a_1(\mathcal{X}(t))\rangle \simeq a_1(\langle \mathcal{X}(t)\rangle)$$

hence, by definition of $\mathbf{x}$, one gets the deterministic equation:

$$\dot{\mathbf{x}}(t) = \langle a_1(\mathcal{X}(t))\rangle \simeq a_1(\langle \mathcal{X}(t)\rangle) = a_1(\mathbf{x}(t)).$$

From ODEs to CMEs: a heuristic

Given an ODE system

$$\begin{cases} \dot{p}(t) &= \bar{\alpha} + \frac{k_1 p(t)^2}{\Gamma_1 + p(t)^2 + \Gamma_2 m(t)} - \delta p(t) \\ \dot{m}(t) &= \beta + k_2 p(t) - \gamma m(t). \end{cases}$$

Basic idea: we turn a systems of n ODEs, representing the reaction-rate equations of a given reaction network (deterministic approach) into $M = 2n$ reactions, by using a one-step assumption.

One-step CTMC: each species can increase/decrease of one unit per time.

Goal: Write the CMEs such that the initial ODEs represent the *macroscopic equation* approximating the dynamics of the mean value of the process.

Drawback: the procedure is not unique.

Example 4 (miRNA toggle switch)

A microRNA-protein toggle switch:

$$\begin{cases} \dot{p}(t) &= \bar{\alpha} + \frac{k_1 p(t)^2}{\Gamma_1 + p(t)^2 + \Gamma_2 m(t)} - \delta p(t) \\ \dot{m}(t) &= \beta + k_2 p(t) - \gamma m(t), \end{cases}$$

where $p(t)$ and $m(t)$ represent the E2F-Myc complex and the miRNA cluster concentrations, respectively.
Species $Y_1 = S$, Species $Y_2 = R$. We can formalize the following set of 4 reactions:

$$\begin{cases} \text{Reaction } \mu = 1 & \emptyset & \longrightarrow & S & v_1 = [1 \quad 0]'; \\ \text{Reaction } \mu = 2 & S & \longrightarrow & \emptyset & v_2 = [-1 \quad 0]'; \\ \text{Reaction } \mu = 3 & \emptyset & \longrightarrow & R & v_3 = [0 \quad 1]'; \\ \text{Reaction } \mu = 4 & R & \longrightarrow & \emptyset & v_4 = [0 \quad -1]'. \end{cases}$$

Example 4 (miRNA toggle switch)

A microRNA-protein toggle switch:

$$\begin{cases} \dot{p}(t) &= \overbrace{\bar{\alpha} + \frac{k_1 p(t)^2}{\Gamma_1 + p(t)^2 + \Gamma_2 m(t)}}^{\emptyset \to S} - \overbrace{\delta p(t)}^{S \to \emptyset} \\ \dot{m}(t) &= \underbrace{\beta + k_2 p(t)}_{\emptyset \to R} - \underbrace{\gamma m(t)}_{R \to \emptyset}, \end{cases}$$

We define a CME by means of the following transition probabilities per unit time:

$$w_1(x) = \bar{\alpha} + \frac{k_1 x_1^2}{\Gamma_1 + x_1^2 + \Gamma_2 x_2}$$
$$w_2(x) = \delta x_1$$
$$w_3(x) = \beta + k_2 x_1$$
$$w_4(x) = \gamma x_2$$

Goal: verify that the original system of ODEs is the macroscopic equation associated to the CME.

The concept of exact simulation

- The Gillespie algorithm allows a discrete and stochastic simulation of a system by simulating explicitly every reaction.
- A Gillespie realization is a random walk that exactly represents the distribution of the master equation.

In which sense is the simulation exact?

Theorem (Law of Large Numbers)

The average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed.

Exact simulation of a (Discrete-time) Markov Chain

1: **Init:**
   T (time horizon);
   x (current state);
2: **for** ($t = 1$ to T) **do**
3: Draw r from the continuous distribution $\mathcal{U}(0,1)$;
   Choose μ s.t. $\sum_{i=0}^{\mu-1} w_i(x) < r \leq \sum_{i=0}^{\mu} w_i(x)$;
   $x := x + v_\mu$ (with $v_0 = 0$);
4: **end for**

Algorithm 1: Simulation of Markov Chains.

Exact simulation of a Continuous-time Markov Chain: SSA

1: **Init:**
   T (time horizon);
   $count_{\max}$ (max number of reactions to observe);
   x (current state);
   $t = 0$ (current time);
   $count = 0$ (reaction counter) ;
2: **while** $(t \leq T) \wedge (count \leq count_{\max})$ **do**
3: $\bar{w} := \sum_{i=1}^{M} w_i(x)$;
   Draw r_1, r_2 from the continuous distribution $\mathcal{U}(0,1)$;
   $\tau := (1/\bar{w}) \ln(1/r_1)$;
   Choose μ s.t. $\sum_{i=1}^{\mu-1} w_i(x) < r_2 \bar{w} \leq \sum_{i=1}^{\mu} w_i(x)$;
   $t := t + \tau$;
   $count := count + 1$;
   $x := x + v_\mu$;
4: **end while**

Algorithm 2: Gillespie Stochastic Simulation Algorithm (SSA).

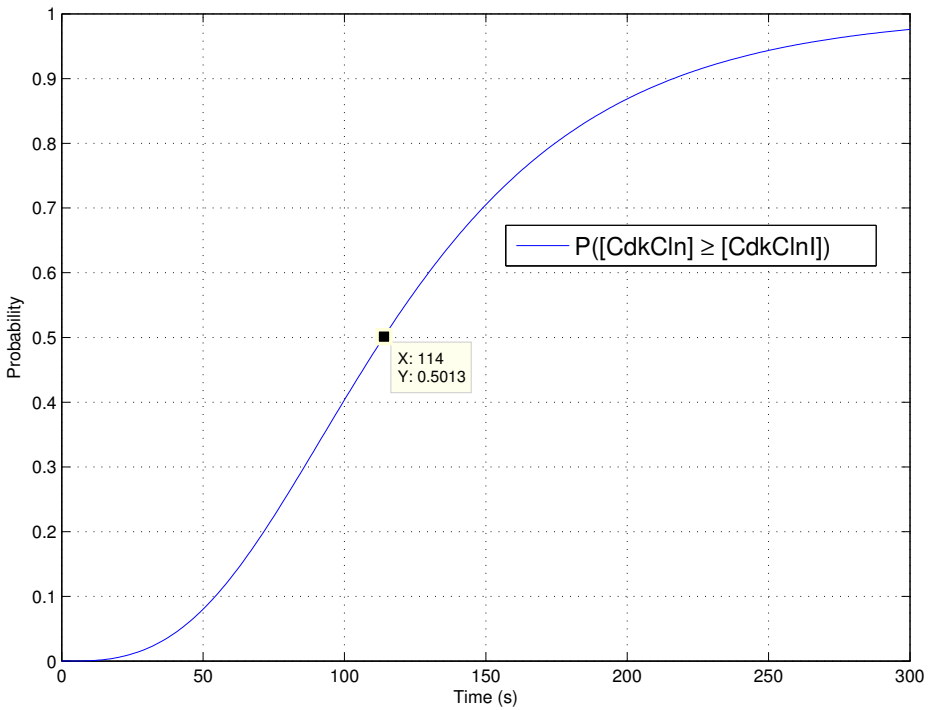
Stochastic Simulation of Ergodic Processes

- Markov processes with a unique stationary distribution are also *ergodic*.
- In ergodic processes, the statistical properties can be deduced from a single, sufficiently long sample (realization) of the process. In particular, the statistical stationary distribution is obtainable as the limit, as t goes to infinity, of the temporal distribution of the *dwell times* over the states of the process.
- Main idea: exploit the ergodicity of CTMCs to obtain the unique stationary distribution by means of a single *very long* run of the Gillespie Algorithm.
- Advantages in terms of time complexity (by avoiding the Monte Carlo iteration, i.e. $count_{\max} = 1$).

Example 1 (Cyclin-dependent kinase)

$M = 7$ biochemical reactions:

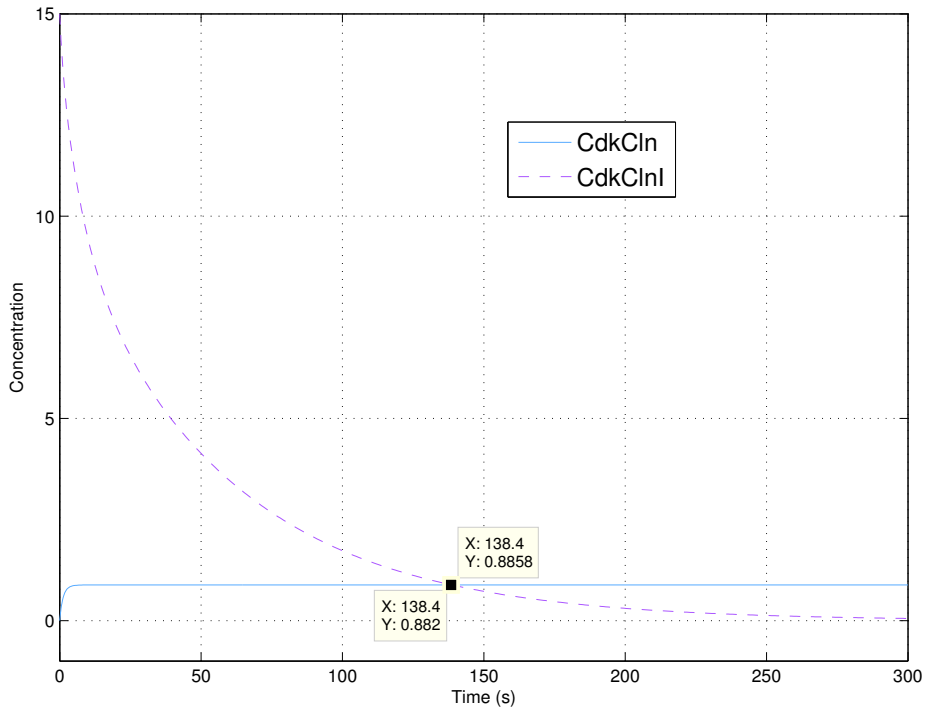
$$\begin{cases} \mathrm{Cdk} + \mathrm{Cln} & \rightleftharpoons & \mathrm{CdkCln} \\ \mathrm{CdkCln} + \mathrm{I} & \rightleftharpoons & \mathrm{CdkClnI} \\ \mathrm{CdkClnI} & \longrightarrow & \mathrm{CdkClnI}_p \\ \mathrm{CdkClnI}_p & \rightleftharpoons & \mathrm{CdkCln} + \mathrm{I}_p, \end{cases}$$



Example 1 (Cyclin-dependent kinase)

$M = 7$ biochemical reactions:

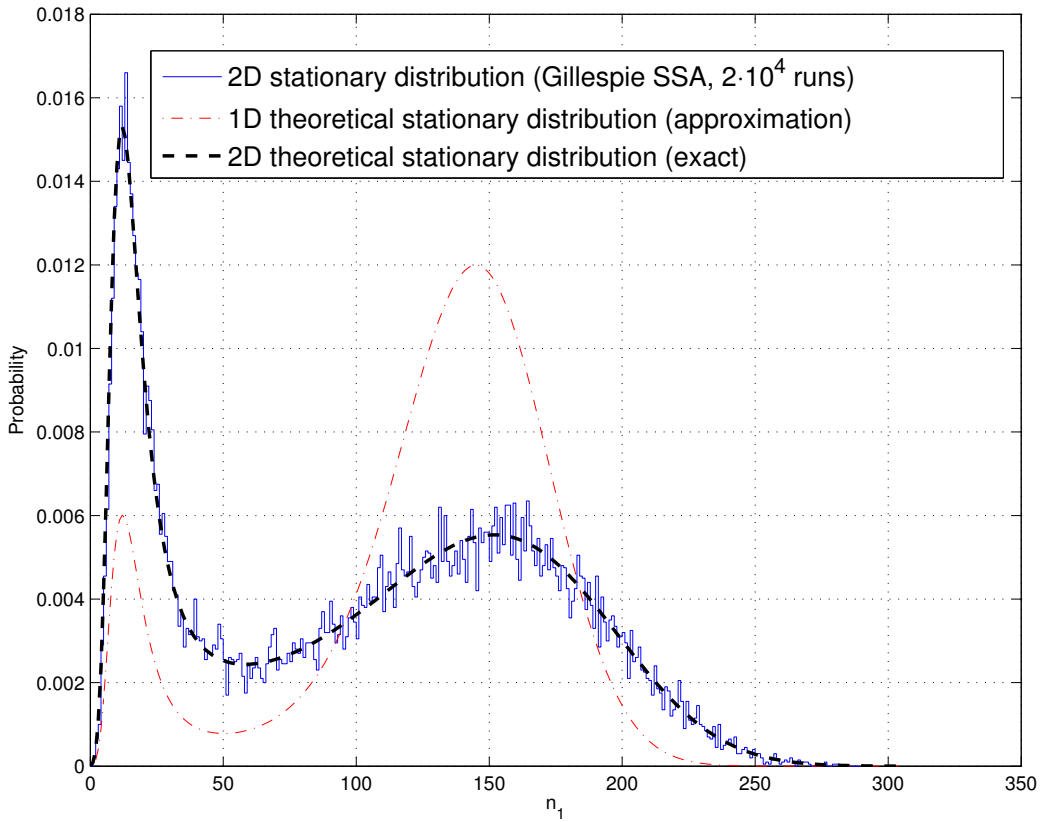
$$\begin{cases} \mathrm{Cdk} + \mathrm{Cln} & \rightleftharpoons & \mathrm{CdkCln} \\ \mathrm{CdkCln} + \mathrm{I} & \rightleftharpoons & \mathrm{CdkClnI} \\ \mathrm{CdkClnI} & \longrightarrow & \mathrm{CdkClnI}_p \\ \mathrm{CdkClnI}_p & \rightleftharpoons & \mathrm{CdkCln} + \mathrm{I}_p, \end{cases}$$



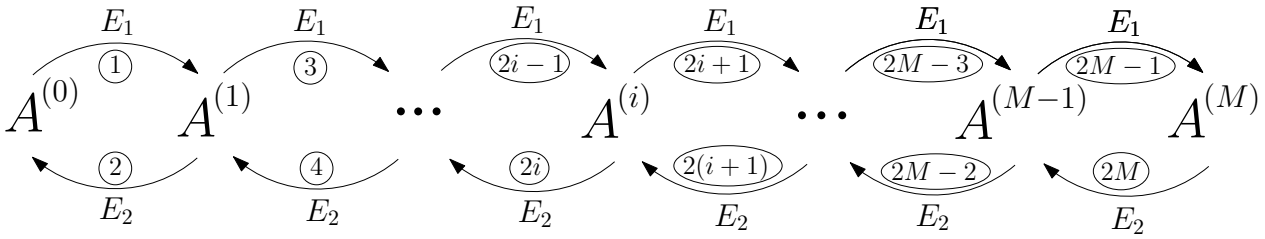
Example 4 (miRNA toggle switch)

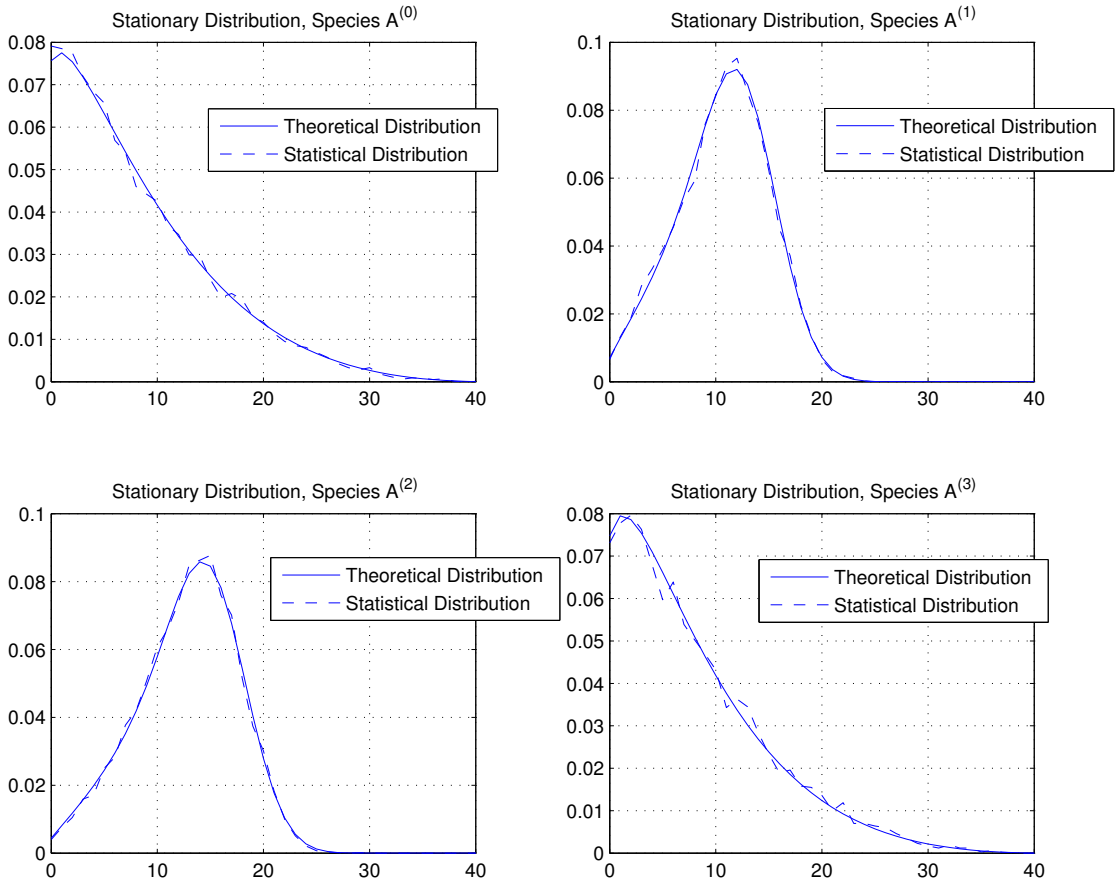
MicroRNA-protein toggle switch:

$$\begin{cases} \dot{p}(t) &= \bar{\alpha} + \frac{k_1 p(t)^2}{\Gamma_1 + p(t)^2 + \Gamma_2 m(t)} - \delta p(t) \\ \dot{m}(t) &= \beta + k_2 p(t) - \gamma m(t), \end{cases}$$



Example 5 (M-phosphorylation)





Advanced Topics

Some interesting topics of research in Stochastic Systems Biology:

- Approximation of the moment dynamics of chemical reactions
  - Moment Closure
- Finite-state CME approximation of infinite CTMC
  - Finite State Projection (FSP)
  - BCMP Closure
- Grid-based reduction of high-dimensional CMEs
  - Adaptive Coarse-Graining
- Uncountably infinite-state Markov Processes
  - Stochastic Differential Equations
- Enhanced Simulation Algorithms
  - τ-leaping

References

1. N.G. van Kampen, *Stochastic Processes in Physics and Chemistry*, North Holland, Third edition, 2007.
2. D. T. Gillespie, Exact Stochastic Simulation of Coupled Chemical Reactions, *The Journal of Physical Chemistry* 81(25), 234–2361, 1977.
3. T. Gardner, C. Cantor and J. Collins, Construction of a toggle switch in Escherichia coli, *Nature* 403(6767), 339–342, 2000.
4. E. Giampieri, D. Remondini, L. de Oliveira, G. Castellani and P. Lió, Stochastic analysis of a miRNA-protein toggle switch, *Molecular Biosystems*, 7(10), 2796-2803, 2011.
5. B. Munsky and M. Khammash, The finite state projection algorithm for the solution of the chemical master equation, *J. Chem. Phys.*, 124, 044104, 2006.
6. A. Singh and J. P. Hespanha, Approximate Moment Dynamics for Chemically Reacting Systems, *IEEE Trans. Automat. Contr.* 56(2), 414-418, 2011.