Chapter 4

Stochastic kinetics

4.1 Introduction

Chemical systems are inherently stochastic, as reactions depend on random (thermal) motion. Deterministic models represent an aggregate behavior of the system. They are accurate in much of classical chemistry, where the numbers of molecules are usually expressed in multiples of Avogadro's number, which is $\approx 6 \times 10^{23}$.[1] In such cases, basically by the law of large numbers, the mean behavior is a good description of the system. The main advantage of deterministic models is that they are comparatively easier to study than probabilistic ones. However, they may be inadequate when the "copy numbers" of species, i.e. the numbers of units (ions, atoms, molecules, individuals) are very small, as is often the case in molecular biology when looking at single cells: copy numbers are small for genes (usually one or a few copies), mRNA's (in the tens), ribosomes and RNA polymerases (up to hundreds) and certain proteins may be at low abundances as well. Analogous situations arise in other areas, such as the modeling of epidemics (where the "species" are individuals in various classes), if populations are small. This motivates the study of stochastic models.

We assume that temperature and volume Ω are constant, and the system is well-mixed.

We consider a chemical reaction network consisting of m reactions which involve the n species

$$S_i,\ i \in \{1, 2, \dots n\}\,.$$

The reactions $\mathcal{R}_j$, $j \in \{1, 2, \dots, m\}$ are specified by combinations of reactants and products:

$$\mathcal{R}_j: \quad \sum_{i=1}^{n} a_{ij} S_i \to \sum_{i=1}^{n} b_{ij} S_i \tag{4.1}$$

where the a_{ij} and b_{ij} are non-negative integers, the *stoichiometry coefficients*[2], and the sums are understood informally, indicating combinations of elements. The integer $\sum_{i=1}^{n} a_{ij}$ is the *order* of the reaction $\mathcal{R}_j$. One allows the possibility or zero order, that is, for some reactions j, $a_{ij} = 0$ for all i. This is the case when there is "birth" of species out of the blue, or more precisely, a species is created by what biologists call a "constitutive" process, such as the production of an mRNA molecule by a

[1] There is this number of atoms in 12g of carbon-12. A "mole" is defined as the amount of substance of a system that contains an Avogadro number of units.

[2] In Greek, *stoikheion* = element, so "measure of elements"

gene that is always active. Zeroth order reactions may also be used to represent inflows to a system from its environment. Similarly, also allowed is the possibility that, for some reactions j, $b_{ij} = 0$ for all i. This is the case for reactions that involve degradation, dilution, decay, or outflows.

The data in (4.1) serves to specify the stoichiometry of the network. The $n \times m$ *stoichiometry matrix* $\Gamma = \{\gamma_{ij}\}$ has entries:

$$\gamma_{ij} \;=\; b_{ij} - a_{ij}\,, \quad i = 1, \ldots, n\,, \quad j = 1, \ldots, m\,. \tag{4.2}$$

Thus, γ_{ij} counts the net change in the number of units of species S_i each time that reaction $\mathcal{R}_j$ takes place.

We will denote by γ_j the jth column of Γ:

$$\gamma_j = b_j - a_j$$

where[3]

$$a_j = (a_{1j}, \ldots, a_{nj})' \quad \text{and} \quad b_j = (b_{1j}, \ldots, b_{nj})'$$

and assume that no $\gamma_j = 0$ (that is, every reaction changes at least one species).

Stoichiometry information is not sufficient, by itself, to completely characterize the behavior of the network: one must also specify the *rates* at which the various reactions take place. This can be done by specifying "propensity" or "intensity" functions.

We will consider deterministic as well as stochastic models, and propensities take different forms in each case. To help readability, we will use the symbol ρ^σ, possibly subscripted, to indicate stochastic propensities, and $\rho^\#$ and ρ^c to indicate deterministic propensities (for numbers of elements or for concentrations, respectively).

[3] prime indicates transpose

4.2 Stochastic models of chemical reactions

Stochastic models of chemical reaction networks are described by a column-vector Markov stochastic process $X = (X_1, \ldots, X_n)'$ which is indexed by time $t \geq 0$ and takes values in $\mathbb{Z}^n_{\geq 0}$. Thus, $X(t)$ is a $\mathbb{Z}^n_{\geq 0}$-valued random variable, for each $t \geq 0$. Abusing notation, we also write $X(t)$ to represent an outcome of this random variable on a realization of the process. The interpretation is:

$$X_i(t) \;=\; \text{number of units of species } i \text{ at time } t\,.$$

One is interested in computing the probability that, at time t, there are k_1 units of species 1, k_2 units of species 2, k_3 units of species 3, and so forth:

$$p_k(t) = \mathbb{P}\left[X(t) = k\right]$$

for each $k \in \mathbb{Z}^n_{\geq 0}$. We call the vector k the *state* of the process at time t.

Arranging the collection of all the $p_k(t)$'s into an infinite-dimensional vector, after an arbitrary order has been imposed on the integer lattice $\mathbb{Z}^n_{\geq 0}$, we have that $p(t) = (p_k)_{k \in \mathbb{Z}^n_{\geq 0}}$ is the discrete probability density (also called the "probability mass function") of $X(t)$.

Biological systems are often studied at "steady state", that is to say after processes have had time to equilibrate. In that context, it is of interest to study the *stationary* (or "equilibrium") density π obtained as the limit as $t \to \infty$ (provided that the limit exists) of $p(t)$. Its entries are the steady state probabilities of being in the state k:

$$\pi_k = \lim_{t \to \infty} p_k(t)$$

for each $k \in \mathbb{Z}^n_{\geq 0}$.

All these probabilities will, in general, depend upon the initial distribution of species, that is, on the $p_k(0)$, $k \in \mathbb{Z}^n_{\geq 0}$, but under appropriate conditions studied in probability theory (ergodicity), the steady state density π will be independent of the initial density.

Also interesting, and often easier to compute, are statistical objects such as the expectation or mean (i.e, the average over all possible random outcomes) of the numbers of units of species at time t:

$$\mathbb{E}\left[X(t)\right] = \sum_{k \in \mathbb{Z}^n_{\geq 0}} p_k(t) k$$

which is a column vector whose entries are the means

$$\mathbb{E}\left[X_i(t)\right] \;=\; \sum_{k \in \mathbb{Z}^n_{\geq 0}} p_k(t) k_i \;=\; \sum_{\ell=0}^{\infty} \ell \sum_{\{k \in \mathbb{Z}^n_{\geq 0}, k_i = \ell\}} p_k(t) \;=\; \sum_{\ell=0}^{\infty} \ell\, p^{(i)}_\ell(t)$$

of the $X_i(t)$'s, where the vector $(p^{(i)}_0(t), p^{(i)}_1(t), p^{(i)}_2(t), \ldots)$ is the marginal density of $X_i(t)$. Also of interest, to understand variability, are the matrix of second moments at time t:

$$\mathbb{E}\left[X(t)X(t)'\right]$$

whose (i,j)th entry is $\mathbb{E}\left[X_i(t)X_j(t)\right]$ and the (co)variance matrix at time t:

$$\mathbb{V}\text{ar}\left[X(t)\right] \;=\; \mathbb{E}\left[\left(X(t) - \mathbb{E}\left[X(t)\right]\right)\left(X(t) - \mathbb{E}\left[X(t)\right]\right)'\right] \;=\; \mathbb{E}\left[X(t)X(t)'\right] - \mathbb{E}\left[X(t)\right]\mathbb{E}\left[X(t)\right]'$$

whose (i,j)th entry is the covariance of $X_i(t)$ and $X_j(t)$, $\mathbb{E}\left[X_i(t)X_j(t)\right] - \mathbb{E}\left[X_i(t)\right]\mathbb{E}\left[X_j(t)\right]$.

4.3 The Chemical Master Equation

A *Chemical Master Equation (CME)* (also known in mathematics as a *Kolmogorov forward equation*) is a system of linear differential equations for the p_k's, of the following form. Suppose given m functions

$$\rho_j^\sigma : \mathbb{Z}^n_{\geq 0} \to \mathbb{R}_{\geq 0}\,, \quad j = 1, \ldots, m\,, \quad \text{with } \rho_j^\sigma(0) = 0\,.$$

These are the *propensity functions* for the respective reactions $\mathcal{R}_j$. As we'll discuss later, the intuitive interpretation is that $\rho_j^\sigma(k)dt$ is the probability that reaction $\mathcal{R}_j$ takes place, in a short interval of length dt, provided that the state was k at the begining of the interval. The CME is:

$$\boxed{\frac{dp_k}{dt} \;=\; \sum_{j=1}^m \rho_j^\sigma(k-\gamma_j)\, p_{k-\gamma_j} \;-\; \sum_{j=1}^m \rho_j^\sigma(k)\, p_k\,, \quad k \in \mathbb{Z}^n_{\geq 0}} \tag{4.3}$$

where, for notational simplicity, we omitted the time argument "t" from p, and where we make the convention that $\rho_j^\sigma(k-\gamma_j) = 0$ unless $k \geq \gamma_j$ (coordinatewise inequality). There is one equation for each $k \in \mathbb{Z}^n_{\geq 0}$, so this is an infinite system of linked equations. When discussing the CME, we will assume that an initial probability vector $p(0)$ has been specified, and that there is a unique solution of (4.3) defined for all $t \geq 0$.

Exercise. Suppose that $p(t)$ satisfies the CME. Show that if $\sum_{k\in\mathbb{Z}^n_{\geq 0}} p_k(0) = 1$ then $\sum_{k\in\mathbb{Z}^n_{\geq 0}} p_k(t) = 1$ for all $t \geq 0$. (Hint: first, using that $\rho_j^\sigma(k-\gamma_j) = 0$ unless $k \geq \gamma_j$, observe that, for each $j \in \{1, \ldots, m\}$:

$$\sum_{k\in\mathbb{Z}^n_{\geq 0}} \rho_j^\sigma(k-\gamma_j) p_{k-\gamma_j} \;=\; \sum_{k\in\mathbb{Z}^n_{\geq 0}} \rho_j^\sigma(k) p_k$$

and use this to conclude that $\sum_{k\in\mathbb{Z}^n_{\geq 0}} p_k(t)$ must be constant. You may use without proof that the derivative of $\sum_{k\in\mathbb{Z}^n_{\geq 0}} p_k(t)$ with respect to time is obtained by term-by-term differentiation.) □

A different CME results for each choice of propensity functions, a choice that is dictated by physical chemistry considerations. Later, we discuss the special case of mass-action kinetics propensities.

Approximating the derivative $\frac{dp_k}{dt}$ by $\frac{1}{h}\,[p_k(t+h) - p_k(t)]$, (4.3) means that:

$$p_k(t+h) \;=\; \sum_{j=1}^m \rho_j^\sigma(k-\gamma_j)\, h\, p_{k-\gamma_j}(t) \;+\; \left(1 - \sum_{j=1}^m \rho_j^\sigma(k)\, h\right) p_k(t) \;+\; o(h)\,. \tag{4.4}$$

This equation allows an intuitive interpretaion of the CME, as follows:

The probability of being in state k at the end of the interval $[t, t+h]$ is the sum of the probabilities of the following $m+1$ events:

- *for each possible reaction $\mathcal{R}_j$, the reaction $\mathcal{R}_j$ happened, and the final state is k, and*
- *no reaction happened, and the final state is k.*

We will justify this interpretation after developing some theory. The discussion will also explain why, for small enough h, the probability that more than one reaction occurs in the interval $[t, t+h]$ is $o(h)$.

We will also introduce the n-column vector:

$$f^\sigma(k) \;:=\; \sum_{j=1}^{m} \rho_j^\sigma(k)\,\gamma_j \;=\; \Gamma R^\sigma(k) \qquad\qquad k \in \mathbb{Z}^n_{\geq 0}$$

where $R^\sigma(k) = (\rho_1^\sigma(k), \ldots, \rho_m^\sigma(k))'$.

Interpreting $\rho_j^\sigma(k)h$ as the probability that reaction $\mathcal{R}_j$ takes place during an interval of length h (if the current state is k), one may then interpret $f^\sigma(k)h$ as the expected change of state during such an interval (since γ_j quantifies the size of the jump if the reaction is $\mathcal{R}_j$). Thus, $f^\sigma(k)$ may be thought of as the rate of change of the state, if the state is k.

When studying steady-state properties, we will not analyze convergence of the random variables $X(t)$ as $t \to \infty$. We will simply *define* a (not necessarily unique) steady state distribution $\pi = (\pi_k)$ of the process as any solution of the equations

$$\boxed{\sum_{j=1}^{m} \rho_j^\sigma(k-\gamma_j)\,\pi_{k-\gamma_j} \;-\; \sum_{j=1}^{m} \rho_j^\sigma(k)\,\pi_k \;=\; 0\,, \quad k \in \mathbb{Z}^n_{\geq 0}}\,.$$

4.3.1 Propensity functions for mass-action kinetics

We first introduce some additional notations. For each $j \in \{1, \ldots, m\}$,

$$A_j = \sum_{i=1}^{n} a_{ij}$$

is the total number of units of all species participating in one reaction of type $\mathcal{R}_j$, the order of $\mathcal{R}_j$.

For each $k = (k_1, \ldots, k_n)' \in \mathbb{Z}^n_{\geq 0}$, we let (recall that a_j denotes the vector $(a_{1j}, \ldots, a_{nj})'$):

$$\binom{k}{a_j} = \prod_{i=1}^{n} \binom{k_i}{a_{ij}}$$

where $\binom{k_i}{a_{ij}}$ is the usual combinatorial number $k_i!/(k_i - a_{ij})!a_{ij}!$, which we define to be zero if $k_i < a_{ij}$.

The most commonly used propensity functions, and the ones best-justified from elementary physical principles, are *ideal mass action kinetics* propensities, defined as follows:

$$\rho^\sigma_{j,\Omega}(k) \;=\; \frac{c_j}{\Omega^{A_j - 1}} \binom{k}{a_j}, \quad j = 1, \ldots, m\,. \tag{4.5}$$

The subscript Ω is used for emphasis, even though Ω is a constant, when we want to emphasize how the different rates depend on the volume, but it is omitted when there is no particular interest in the dependence on Ω. The m non-negative constants $c_1, \ldots, c_m$ are arbitrary, and they represent quantities related to the shapes of the reactants, chemical and physical information, and temperature.

4.3.2 Some examples

We will illustrate our subsequent discussions with a few simple but extremely important examples.

mRNA production and degradation

Consider the chemical reaction network consisting of the two reactions $0 \to M$ (formation) and $M \to 0$ (degradation), also represented as:

$$0 \xrightarrow{\alpha} M \xrightarrow{\beta} 0 \tag{4.6}$$

where α and β are the respective rates and mass-action kinetics is assumed. The symbol "0" is used to indicate an empty sum of species.

The application we have in mind is that in which M indicates number of mRNA molecules, and the formation process is transcription from a gene G which is assumed to be at a constant level of activity. (Observe that one could alternatively model the transcription process by means of a reaction "$G \to G + M$" instead of "$0 \to M$", where G would indicate the activity level of the gene G. Since G is neither "created" nor "destroyed" in the reactions, including it in the model is redundant. Of course, if we wanted also to include in our model temporal changes in the activation of G, then a more complicated model would be called for.)

The stoichiometry matrix and propensities are:[4]

$$G \;=\; (1 \quad -1)\,, \;\; \rho_1^\sigma(k) = \alpha\,, \; \rho_2^\sigma(k) = \beta k \tag{4.7}$$

so that

$$f^\sigma(k) \;=\; \alpha - \beta k\,. \tag{4.8}$$

The CME becomes:

$$\frac{dp_k}{dt} \;=\; \alpha p_{k-1} + (k+1)\beta p_{k+1} - \alpha p_k - k\beta p_k \tag{4.9}$$

where, recall, the convention is that a term is zero if the subscript is negative. Observe that here $k \in K = \mathbb{Z}_{\geq 0}$ is just a non-negative integer.

We later discuss how to solve the CME for this example. For now, we limit ourselves to a discussion of its steady-state solution.

In general, let π be the steady-state probability distribution obtained by setting $\frac{dp}{dt} = 0$. Under appropriate technical conditions, not discussed here, there is a unique such distribution, and it holds that $\pi_k = \lim_{t\to\infty} p_k(t)$ for each $k \in \mathbb{Z}^n_{\geq 0}$ and every solution $p(t)$ of the CME for an initial condition that is a probability density ($\sum_k p_k(0) = 1$). We may interpret π as the probability distribution of a random variable $X(\infty)$ obtained as the limit of $X(t)$ as $t \to \infty$.

In this example, by definition the numbers π_k satisfy:

$$\alpha\pi_{k-1} + (k+1)\beta\pi_{k+1} - \alpha\pi_k - k\beta\pi_k = 0\,, \quad k = 0, 1, 2, \ldots \tag{4.10}$$

(the first term is not there if $k = 0$). It is easy to solve recursively for π_k, $k \geq 1$ in terms of π_0, and then use the condition $\sum_k \pi_k(0) = 1$ to find π_0; there results that

$$\pi_k = e^{-\lambda}\frac{\lambda^k}{k!} \tag{4.11}$$

[4] Volume dependence is assumed to be already incorporated into α, in this and other examples.

where $\lambda = \frac{\alpha}{\beta}$. In other words, the steady state probability distribution is Poisson distributed with parameter λ.

Exercise. Show, using induction on k, that indeed (4.11) solves (4.10).

Bursts of mRNA production

In an often-studied variation of the above model, mRNA is produced in "bursts" of $r > 1$ (assumed to be a fixed integer) transcripts at a time. This leads to the reactions

$$0 \xrightarrow{\alpha} r\,M\,, \quad M \xrightarrow{\beta} 0 \tag{4.12}$$

with stoichiometry matrix and propensities:

$$\Gamma \;=\; (r \quad -1), \quad \rho_1^\sigma(k) = \alpha\,, \quad \rho_2^\sigma(k) = \beta k \tag{4.13}$$

so that

$$f^\sigma(k) \;=\; r\alpha - \beta k\,. \tag{4.14}$$

The form of f^σ is exactly the same as in the non-bursting case: the only difference is that the rate α has to be redefined as $r\alpha$. This will mean that the deterministic chemical equation representation is the same as before (up to this redefinition), and, as we will see, the mean of the stochastic process will also be the same (up to redefinition of α). Interestingly, however, we will see that the "noisiness" of the system can be lowered by a factor of up to $1/2$.

Exercise. Write the CME for the bursting model.

A simple dimerization example

Here is another simple example. Suppose that a molecule of A can be produced at constant rate α and degrades when dimerized:

$$0 \xrightarrow{\alpha} A\,, \quad A + A \xrightarrow{\beta} 0 \tag{4.15}$$

which leads to

$$\Gamma = (1 \quad -2)\,, \quad \rho_1^\sigma(k) = \alpha\,, \quad \rho_2^\sigma(k) = \frac{\beta k(k-1)}{2} \tag{4.16}$$

and

$$f^\sigma(k) = \alpha - \beta k(k-1) = \alpha + \beta k - \beta k^2\,. \tag{4.17}$$

Exercise. Write the CME for the dimerization model.

A model of transcription and translation

One of the most-studied models of gene expression is as follows. We consider the reactions for mRNA production and degradation (4.6):

$$0 \xrightarrow{\alpha} M \xrightarrow{\beta} 0$$

together with:

$$M \xrightarrow{\theta} M + P\,, \quad P \xrightarrow{\delta} 0 \tag{4.18}$$

where P represents the protein translated from M. Now

$$\Gamma \;=\; \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}\,, \quad \rho_1^\sigma(k) = \alpha\,,\;\; \rho_2^\sigma(k) = \beta k_1\,,\;\; \rho_3^\sigma(k) = \theta k_1\,,\;\; \rho_4^\sigma(k) = \delta k_2\,. \tag{4.19}$$

where $k = (k_1, k_2)$ is a vector that counts mRNA and protein numbers respectively, and (writing "(M, P)" instead of $k = (k_1, k_2)$):

$$f^\sigma(M, P) = \begin{pmatrix} \alpha - \beta M \\ \theta M - \delta P \end{pmatrix} . \tag{4.20}$$

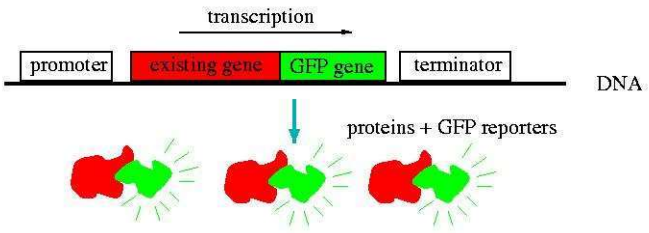
Observe that P does not affect M, so the behavior of M will be the same as in the transcription model, and in particular the steady-state distribution of M is Poisson. However, P depends on M, making the problem much more interesting.

Exercise. Write the CME for the transcription/translation model. (Remember that now "k" is a vector (k_1, k_2).)

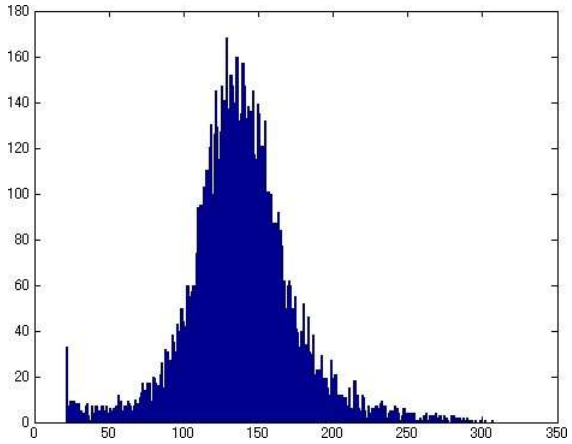
Remark on FACS: Experimentally estimating the probability distribution of protein numbers

Suppose that we wish to know at what rate a certain gene X is being transcribed under a particular set of conditions in which the cell finds itself. Fluorescent proteins may be used for that purpose. For instance, *green fluorescent protein (GFP)* is a protein with the property that it fluoresces in green when exposed to UV light. It is produced by the jellyfish *Aequoria victoria*, and its gene has been isolated so that it can be used as a *reporter gene*. The GFP gene is inserted (cloned) into the chromosome, adjacent to or very close to the location of gene X, so both are controlled by the same promoter region. Thus, gene X and GFP are transcribed simultaneously and then translated. and so by measuring the intensity of the GFP light emitted one can estimate how much of X is being expressed.

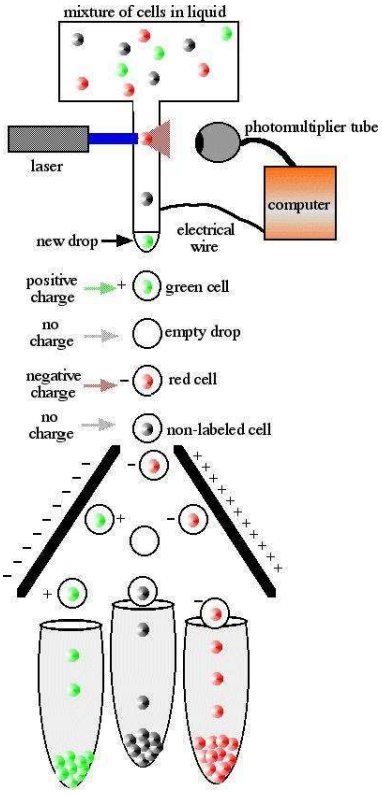
Fluorescent protein methods are particularly useful when combined with *flow cytometry*.[5] Flow Cytometry devices can be used to sort individual cells into different groups, on the basis of characteristics such as cell size, shape, or amount of measured fluorescence, and at rates of up to thousands of cells per second. In this manner, it is possible, for instance, to classify the strength of gene expression in individual cells in a population, perhaps under different sets of conditions.



fluorescent protein construct



cell count versus intensity



[5] FACS = "fluorescence-activated cell sorting".

4.4 Theoretical background, algorithms, and discussion

The abstract theoretical mathematical background for the CME is as follows.

4.4.1 Markov Processes

Suppose that $\{X(t)\}, t \in [0,\infty)$ is a stochastic process, that is to say a collection of jointly distributed random variables, each of which takes values in a fixed countable set K ($K = \mathbb{Z}^n_{\geq 0}$ in our case).[6]

From now on, we assume that the process is a *continuous-time stationary Markov chain*, meaning that it satisfies the following properties:[7]

- **[Markov]** For any two non-negative real numbers t, h, any function $x : [0,s] \to K$, and any $k \in K$,

$$\mathbb{P}\left[X(t+h) = k \mid X(s) = x(s),\, 0 \leq s \leq t\right] \;=\; \mathbb{P}\left[X(t+h) = k \mid X(t) = x(t)\right].$$

  This property means that $X(t)$ contains all the information necessary in order to estimate the future values $X(T), T \geq t$: additional values from the past do not help to get a better prediction.

- **[Stationarity]** The conditional or *transition* probabilities $\mathbb{P}\left[X(s) = \ell \,|X(t) = k\right]$ depend only on the difference $t - s$. This property, also called homogeneity, means that the probabilities do not change over time.

- **[Differentiability]** With $p_{\ell k}(h) := \mathbb{P}\left[X(t+h) = \ell \,|X(t) = k\right]$ and $p_k(t) := \mathbb{P}\left[X(t) = k\right]$ for every $\ell, k \in K$ and all $t, h \geq 0$, the functions $p_{\ell k}(h)$ and $p_k(t)$ are differentiable in h, t.

Note the following obvious facts:

- $\sum_{\ell \in K} p_{\ell k}(h) = 1$ for every $k \in K$ and $h \geq 0$.
- $p_{\ell k}(0) = \begin{cases} 0 & \text{if } \ell \neq k \\ 1 & \text{if } \ell = k\,. \end{cases}$

Take any $t, h > 0$, and any $\ell \in K$. Then

$$\begin{aligned} p_\ell(t+h) \;=\; \mathbb{P}\left[X(t+h) = \ell\right] \;&=\; \sum_{k \in K} \mathbb{P}\left[X(t+h) = \ell \;\&\; X(t) = k\right] \\ &=\; \sum_{k \in K} \mathbb{P}\left[X(t+h) = \ell \,|\, X(t) = k\right] \,\times\, \mathbb{P}\left[X(t) = k\right] \end{aligned}$$

[6] The more precise notation would be "$X_t(\omega)$", where ω is an element of the outcome space, but we adopt the standard convention of not showing ω. We also do not specify the sample space nor the sigma-algebra of measurable sets which constitute events to which a probability is assigned. If one imposes the requirement that, with probability one, sample paths are continuous from the right and have well-defined limits from the left, a suitable sample space can then be taken to be a space of piecewise constant mappings from $\mathbb{R}_{\geq 0}$ to K.

[7] A subtle fact, usually not mentioned in textbooks, is that conditional probabilities are not always well-defined: "$\mathbb{P}\left[A|B\right] = \mathbb{P}\left[A\&B\right]/\mathbb{P}\left[B\right]$" makes no sense if $\mathbb{P}\left[B\right] = 0$. However, for purposes of our discussions, one may define $\mathbb{P}\left[A|B\right]$ arbitrarily in that case, and no arguments will be affected.

because the events $\{X(t) = k\}$ are mutually exclusive for different k. In other words:

$$p_\ell(t+h) \;=\; \sum_{k\in K} p_{\ell k}(h) p_k(t)\,. \tag{4.21}$$

Similarly. we have[8] the *Chapman-Kolmogorov* equation for the process:

$$p_{\ell\tilde{\ell}}(t+h) \;=\; \sum_{k\in K} p_{\ell k}(t)\, p_{k\tilde{\ell}}(h)\,. \tag{4.22}$$

4.4.2 The jump time process: how long do we wait until the next reaction?

Suppose that $X(t_0) = k$, and consider a time interval $I = [t_0, t_0+h]$. If $X(t) \neq k$ for some $t \in I$, one says that a "change of state" or an "event" has occurred during the interval, or, for chemical networks, that "a reaction has occurred".

For each $k \in K$ and $h \geq 0$, let:

$$\begin{array}{rcl} C_k(h) & := & \mathbb{P}\,[\text{no reaction occurred on } [t_0, t_0+h] \mid X(t_0) = k] \\ & = & \mathbb{P}\,[X(t) = k \;\forall t \in [t_0, t_0+h] \mid X(t_0) = k] \end{array}$$

(the definition is independent of the particular t_0, by homogeneity). The function $C_k(h)$ is non-increasing on h, and $C_k(0) = 1$. Consider any two $h_1 \geq 0$ and $h_2 \geq 0$. We claim that

$$C_k(h_1+h_2) \;=\; C_k(h_1) C_k(h_2)\,.$$

Indeed, using the shorthand notation "$X(a,b) = k$" to mean that "$X(t) = k$ for all $t \in [a,b]$", we have:

$$\begin{array}{l} \mathbb{P}\,[X(t_0, t_0+h_1+h_2) = k \mid X(t_0) = k] \\ \quad = \;\; \mathbb{P}\,[X(t_0, t_0+h_1+h_2) = k \;\;\&\;\; X(t_0) = k] \,/\, \mathbb{P}\,[X(t_0) = k] \\ \quad = \;\; \mathbb{P}\,[X(t_0, t_0+h_1+h_2) = k] \,/\, \mathbb{P}\,[X(t_0) = k] \\ \quad = \;\; \mathbb{P}\,[X(t_0, t_0+h_1) = k \;\;\&\;\; X(t_0+h_1, t_0+h_1+h_2) = k] \,/\, \mathbb{P}\,[X(t_0) = k] \\ \quad = \;\; \mathbb{P}\,[X(t_0, t_0+h_1) = k] \;\times\; \mathbb{P}\,[X(t_0+h_1, t_0+h_1+h_2) = k \mid X(t_0, t_0+h_1) = k] \,/\, \mathbb{P}\,[X(t_0) = k] \\ \quad = \;\; \mathbb{P}\,[X(t_0, t_0+h_1) = k] \;\times\; \mathbb{P}\,[X(t_0+h_1, t_0+h_1+h_2) = k \mid X(t_0+h_1) = k] \,/\, \mathbb{P}\,[X(t_0) = k] \\ \quad = \;\; (\mathbb{P}\,[X(t_0, t_0+h_1) = k] \,/\, \mathbb{P}\,[X(t_0) = k]) \;\times\; \mathbb{P}\,[X(t_0+h_1, t_0+h_1+h_2) = k \mid X(t_0+h_1) = k] \\ \quad = \;\; C_k(h_1)\, C_k(h_2) \end{array}$$

(we used the formula $\mathbb{P}\,[A\&B] = \mathbb{P}\,[A] \times \mathbb{P}\,[B|A]$ which comes from the definition of conditional probabilities, as well as the Markov property).

Thus, if we define $c_k(h) = \ln C_k(h)$, we have that $c_k(h_1+h_2) = c_k(h_1) + c_k(h_2)$, that is, c_k is an additive function. Notice that the functions C_k, and hence also c_k, are monotonic. Therefore each c_k is linear: $c_k(h) = -\lambda_k h$, for some number $\lambda_k \geq 0$.[9] (The negative sign because $c_k(h)$ is the logarithm of a probability, which is a number ≤ 1.) We conclude that $C_k(h) = e^{-\lambda_k h}$.

[8] Prove this as an exercise.

[9] Read the Wikipedia article on "Cauchy's functional equation".

In summary:

$$\mathbb{P}\left[\text{no reaction takes place on } [t_0, t_0+h] \mid X(t_0)=k\right] \;=\; e^{-\lambda_k h}$$

from which it follows that

$$\mathbb{P}\left[\text{at least one reaction takes place on } [t_0, t_0+h] \mid X(t_0)=k\right] \;=\; 1-e^{-\lambda_k h} \;=\; \lambda_k h + o(h)\,.$$

A central role in both theory and numerical algorithms is played by the following random variable:

$$\boxed{\mathcal{T}_k \;:=\; \text{time until the next reaction ("event") will occur, if the current state is } X(t_0)=k}\,.$$

That is,

if $X(t_0)=k$, an outcome $\mathcal{T}_k = h$ means that the next reaction occurs at time t_0+h.

Observe that, because of the stationary Markov property, $\mathcal{T}_k$ depends only on the current state k, and not on the current time t_0,

If the current time is t_0, then these two events:

- "the next reaction occurs at some time $> h$"
- "no reaction occurs during the interval $[t_0, t_0+h]$"

are the same. Thus:

$$\mathbb{P}\left[\mathcal{T}_k > h\right] = e^{-\lambda_k h}$$

which means that:

$$\boxed{\textit{the variable } \mathcal{T}_k \textit{ is exponentially distributed with parameter } \lambda_k}\,.$$

Starting from state k, the time to wait until the Nth subsequent reaction takes place is:

$$\mathcal{T}_{k^{(1)}} + \mathcal{T}_{k^{(2)}} + \ldots + \mathcal{T}_{k^{(N)}}$$

where $k^{(1)} = k$, $k^{(2)}$ is the state reached after the first reaction, $k^{(3)}$ is the state reached after the second reaction (starting from state $k^{(2)}$), and so forth. Note that the choice of which particular "waiting time" random variable $\mathcal{T}_\ell$ is used at each step depends on the past state sequence.

If two or more reactions happen during an interval $[t_0, t_0+h]$, then $\mathcal{T}_{k^{(1)}} + \mathcal{T}_{k^{(2)}} + \ldots + \mathcal{T}_{k^{(N)}} \leq h$ for some N and some sequence of states, so in particular $\mathcal{T}_k + \mathcal{T}_\ell \leq h$ for some ℓ. Observe that

$$\mathbb{P}\left[\mathcal{T}_k + \mathcal{T}_\ell \leq h\right] \;\leq\; \mathbb{P}\left[\mathcal{T}_k \leq h \,\&\, \mathcal{T}_\ell \leq h\right] \;=\; \mathbb{P}\left[\mathcal{T}_k \leq h\right] \times \mathbb{P}\left[\mathcal{T}_\ell \leq h\right] \;=\; (\lambda_k h + o(h))(\lambda_\ell h + o(h)) \;=\; o(h)$$

because the variables $\mathcal{T}$ are conditioned on the initial state, and are therefore independent.[10] The probability that ≥ 2 reactions happen is upper bounded by $\sum_\ell \mathbb{P}\left[\mathcal{T}_k + \mathcal{T}_\ell \leq h\right]$, where the sum is taken over all those states ℓ that can be reached from k after one reaction. We assume from now on that:

jumps from any given state k can only take place to one of a **finite** *number of possible* states ℓ (4.23)

[10] This step in the argument needs to be made more rigorous: one should specify the joint sample space for the $\mathcal{T}$'s.

(as is the case with chemical networks). Thus this sum is finite, and so we can conclude:

$$\mathbb{P}\left[\geq 2 \text{ reactions happen on the interval } [t_0, t_0+h] \mid X(t_0)=k\right] \;=\; o(h)\,.$$

Note that

$$\begin{aligned} 1-e^{-\lambda_k h} &= \mathbb{P}\left[\text{some reaction happens on } [t_0,t_0+h] \mid X(t_0)=k\right] \\ &= \mathbb{P}\left[\text{exactly one reaction happens on } [t_0,t_0+h] \mid X(t_0)=k\right] \\ &\qquad + \mathbb{P}\left[\geq 2 \text{ reactions happen on } [t_0,t_0+h] \mid X(t_0)=k\right] \quad \big(=o(h)\big) \end{aligned}$$

and thus

$$\mathbb{P}\left[\text{exactly one reaction happens on } [t_0,t_0+h] \mid X(t_0)=k\right] \;=\; 1-e^{-\lambda_k h}+o(h) \;=\; \lambda_k h + o(h)\,.$$

For any two states $k \neq \ell$, and any interval $[t_0, t_0+h]$, $p_{\ell k}(h) = \mathbb{P}\left[X(t+h)=\ell \,|X(t)=k\right]$ is the sum of

$$\mathbb{P}\left[\text{there is a jump from } k \text{ to } \ell \text{ in the interval } [t_0,t_0+h]\right]$$

plus

$$\mathbb{P}\left[\text{there is no (direct) jump from } k \text{ to } \ell\text{, but there is a sequence of jumps that take } k \text{ into } \ell\right]$$

and, as the probability of ≥ 2 jumps is $o(h)$, this last probability is $o(h)$. Thus:

$$p_{\ell k}(h) \;=\; \mathbb{P}\left[\text{there is a jump from } k \text{ to } \ell \text{ in the interval } [t_0,t_0+h]\right] + o(h)\,.$$

Assumption (4.23) then implies that

$$p_{\ell k}(h) \;=\; o(h) \quad \text{for all but a finite number of states } \ell\,.$$

4.4.3 Propensities

A key role in Markov process theory is played by the *infinitesimal transition probabilities* defined as follows:

$$q_{\ell k} \;:=\; \left.\frac{dp_{\ell k}(h)}{dh}\right|_{h=0}$$

Since $p_{\ell k}(h) = o(h)$ for all but a finite number of states ℓ, it follows that, for each k, there are only a finite number of nonzero $q_{\ell k}$'s. In general, $p_{\ell k}(h) = p_{\ell k}(0) + q_{\ell k}h + o(h)$, so, since $p_{\ell k}(0) = 0$ if $\ell \neq k$ and $= 1$ if $\ell = k$:

$$p_{\ell k}(h) \;=\; \begin{cases} hq_{\ell k}+o(h) & \text{if } \ell \neq k \\ 1+hq_{kk}+o(h) & \text{if } \ell = k. \end{cases}$$

Recall that λ_k is the parameter for the exponentially distributed random variable $\mathcal{T}_k$ that gives the time of the next reaction provided that the present state is k. We claim that:

$$q_{kk} \;=\; -\lambda_k \quad \text{for all } k\,.$$

Indeed, $p_{kk}(h) := \mathbb{P}\left[X(t+h)=k \,|X(t)=k\right]$, and this event is the union of the mutually exclusive events "no reaction happened" (which has probability $e^{-\lambda_k h}$) and "two or more reactions happened,

and the end state is again k". This second event has probability $o(h)$, because the probability that more than one reaction happens (even if the final state is different) is already $o(h)$. Thus: $p_{kk}(h) = e^{-\lambda_k h} + o(h)$, which gives $(dp_{kk}/dt)(0) = -\lambda_k$, as claimed.

Note also that, since $\sum_{\ell \in K} p_{\ell k}(h) = 1$ for all h, taking $d/dh|_{h=0}$ gives:

$$\sum_{\ell \in K} q_{\ell k} \;=\; 0 \quad \text{or, equivalently,} \quad q_{kk} = -\sum_{\ell \neq k} q_{\ell k} \tag{4.24}$$

and hence also $\lambda_k = \sum_{\ell \neq k} q_{\ell k}$, for every k.

Recall that the Chapman-Kolmogorov equation (4.22) says that $p_{\ell\tilde{\ell}}(t+h) = \sum_{k \in K} p_{\ell k}(t)\, p_{k\tilde{\ell}}(h)$ for all t, h. By definition, $q_{\ell k} = (dp_{\ell k}/dh)(0)$, so taking the derivative with respect to h and evaluating at $h = 0$, we arrive at the *forward Kolmogorov differential equation*

$$\frac{dp_{\ell\tilde{\ell}}}{dt} \;=\; \sum_{k \in K} p_{\ell k}\, q_{k\tilde{\ell}} \tag{4.25}$$

which is an equation relating conditional probabilities through the infinitesimal transitions. Similarly, the corresponding equation on probabilities (4.21) is $p_k(t+h) = \sum_{\ell \in K} p_{k\ell}(h) p_\ell(t)$, which leads under differentiation to:

$$\frac{dp_k}{dt} \;=\; \sum_{\ell \in K} q_{k\ell}\, p_\ell \,. \tag{4.26}$$

This differential equation is often also called the forward Kolmogorov equation, *and it is exactly the same as the CME* (4.3) $\frac{dp_k}{dt} = \sum_{j=1}^m \rho_j^\sigma(k-\gamma_j)\, p_{k-\gamma_j} - \sum_{j=1}^m \rho_j^\sigma(k)\, p_k$, where

the propensities $\rho_j^\sigma(k)$ are, by definition, the infinitesimal transition probabilities $q_{\ell k}$.

More precisely, consider the m reactions $\mathcal{R}_j$, which produce the stoichiometry changes $k \mapsto k + \gamma_j$ respectively. We define $\rho_j^\sigma(k) = q_{\ell k}$ for $\ell = k + \gamma_j$, $j = 1, \ldots, m$. So:

$$q_{k\ell} \;=\; \begin{cases} \rho_j^\sigma(\ell) & \text{if } \ell = k - \gamma_j \text{ for some } j \in \{1, \ldots, m\} \\ -\sum_{\ell \neq k} q_{\ell k} = -\sum_{j=1}^m \rho_j^\sigma(k) & \text{if } \ell = k \quad \text{(recall (4.24))} \\ 0 & \text{otherwise .} \end{cases}$$

Since $\lambda_k = -q_{kk}$,

$$\lambda_k \;=\; \sum_{\ell \neq k} q_{\ell k} = \sum_{j=1}^m \rho_j^\sigma(k) \,. \tag{4.27}$$

4.4.4 Interpretation of the Master Equation and propensity functions

Since, by definition of the $q_{\ell k}$'s, $p_{k+\gamma_j,k}(h) \;=\; q_{k+\gamma_j,k} h + o(h) \;=\; \rho_j^\sigma(k) h + o(h)$ and $p_{kk}(h) \;=\; 1 + q_{kk} h + o(h) = 1 - \sum_{j=1}^m \rho_j^\sigma(k) h + o(h)$,

$$\mathbb{P}\left[X(t+h) = k + \gamma_j \,|\, X(t) = k\right] \;=\; \rho_j^\sigma(k)\, h \;+\; o(h) \;\approx\; \rho_j^\sigma(k)\, h$$

and

$$\mathbb{P}\left[X(t+h) = k \,|\, X(t) = k\right] \;=\; 1 - \sum_{j=1}^m \rho_j^\sigma(k)\, h \;+\; o(h) \;\approx\; 1 - \sum_{j=1}^m \rho_j^\sigma(k)\, h \,.$$

Since the probability that more than one reaction occurs on an interval of length h is $o(h)$, the probability that $X(t+h) = k + \gamma_j$ is approximately the same as that of $\mathcal{R}_j$ happening in the interval. This justifies the intepretation of the propensity of the reaction $\mathcal{R}_j$ as:

$$\rho_j^\sigma(k)h \;\approx\; \text{probability that the reaction } \mathcal{R}_j \text{ will take place, during a time interval } [t, t+h] \text{ of (short) duration } h, \text{ if the state was } k \text{ at time } t.$$

In other words, ρ_j^σ is the rate at which the reaction $\mathcal{R}_j$ "fires". This rate depends, obviously, on how many units of the various reactants are present (k). Furthermore, with this interpretation,

$$\begin{aligned}
&\rho_j^\sigma(k)\, h\, p_k(t) \;\approx\\
&\mathbb{P}\left[\text{reaction } \mathcal{R}_j \text{ takes place during interval } [t, t+h] \;\middle|\; \text{state was } k \text{ at time } t\right] \times \mathbb{P}\left[\text{state was } k \text{ at time } t\right]\\
&\quad = \;\mathbb{P}\left[\text{state was } k \text{ at time } t \;\&\; \text{reaction } \mathcal{R}_j \text{ takes place during interval } [t, t+h]\right],
\end{aligned}$$

and so

$$\sum_{j=1}^{m} \rho_j^\sigma(k)\, h\, p_k(t)$$

is the probability that the state at time t is k and *some* reaction takes place during the time interval $[t, t+h]$. (Implicitly assuming that these events are mutually exclusive, i.e. at most one reaction can happen, if the time interval is very short.)

Therefore, the second term in (4.4):

$$\begin{aligned}
&\left(1 - \sum_{j=1}^{m} \rho_j^\sigma(k)h\right) p_k(t) \;=\; p_k(t) - \sum_{j=1}^{m} \rho_j^\sigma(k)\, h\, p_k(t)\\
\approx\;& \mathbb{P}\left[\{\text{initial state was } k\} \setminus \{\text{initial state was } k \text{ and some reaction happens during interval } [t, t+h]\}\right]\\
=\;& \mathbb{P}\left[\text{initial state was } k \text{ and no reaction happens during interval } [t, t+h]\right]\\
=\;& \mathbb{P}\left[\text{final state is } k \text{ and no reaction happens during interval } [t, t+h]\right]
\end{aligned}$$

where the last equality is true because the events:

$$\text{no reaction happened and the } \textit{initial} \text{ state was } k$$

and

$$\text{no reaction happened and the } \textit{final} \text{ state is } k$$

are the same.

On the other hand, regarding the m first terms in (4.4), note that the event:

$$\text{reaction } \mathcal{R}_j \text{ happened and the final state is } k$$

is the same as the event:

$$\text{reaction } \mathcal{R}_j \text{ happened and the initial state was } k - \gamma_j,$$

and the probability of this last event is $\approx \rho_j^\sigma(k - \gamma_j)\, h\, p_{k-\gamma_j}(t)$

In summary, we are justified in interpreting (4.4) as asserting that the probability of being in state k at the end of the interval $[t, t+h]$ is the sum of the probabilities of the following $m+1$ events:

- for each possible reaction $\mathcal{R}_j$, the reaction $\mathcal{R}_j$ happened, and the final state is k, and
- no reaction happened, and the final state is k.

4.4.5 The embedded jump chain

The exponentially distributed variable $\mathcal{T}$ tells what is the waiting time until the next reaction. In order to understand the behavior of the system as a sequence of jumps, one needs, in addition, a random variable that specifies *which* reaction takes place next (or, more generally for Markov processes, to which state is the next transition), *given that* a transition happens.

For each $\ell \neq k$ and h, let $\alpha_{\ell k}(h)$ be the probability that the state is ℓ at time $t+h$, assuming that the initial state is k and that some reaction has happened. If k is not an absorbing state, that is, if transitions out of k are possible, an elementary calculation with conditional probabilities (using that $X(t+h)=\ell$ implies $X(t+h)\neq k$) shows that:[11]

$$\alpha_{\ell k}(h) \;=\; \mathbb{P}\left[X(t+h)=\ell \mid X(t)=k \;\&\; X(t+h)\neq k\right] \;=\; \frac{\mathbb{P}\left[X(t+h)=\ell \,|\, X(t)=k\right]}{\mathbb{P}\left[X(t+h)\neq k \,|\, X(t)=k\right]}\,.$$

Ideally, one would like to compute this expression, but the transition probabilities are hard to obtain. However,

$$\lim_{h\to 0}\alpha_{\ell k}(h) \;=\; \lim_{h\to 0}\frac{p_{\ell k}(h)}{1-p_{kk}(h)} \;=\; \lim_{h\to 0}\frac{hq_{\ell k}+o(h)}{1-(1+hq_{kk}+o(h))} \;=\; -\frac{q_{\ell k}}{q_{kk}} \;=\; \frac{q_{\ell k}}{\sum_{\widetilde{\ell}\neq k} q_{\widetilde{\ell}k}} \;=:\; d_\ell^{(k)}\,.$$

(If k is an absorbing state, the denominators are zero, but in that case we know that $\alpha_{\ell k}(h)=0$ for all $\ell\neq k$.)

Although in principle only an approximation, it was proved by J.L. Doob[12] that the discrete probability distribution $d_\ell^{(k)}$ (for any fixed k, over all $\ell\neq k$), together with the process $\mathcal{T}_k$, characterize a process with the same probability distribution as the original $X(t)$. By itself, the matrix D with entries $d_\ell^{(k)}$ is the transition matrix for the discrete-time *embedded Markov chain* or *jump chain* of the process. This discrete chain provides a complete statistical description of the possible sequences of states visited, except that it ignores the actual times at which jumps occur. It is very helpful in theoretical developments, especially in the classification of states ("recurrent", "transient", etc.) of the continuous process.

4.4.6 The stochastic simulation algorithm (SSA)

To understand the behavior of the process $X(t)$, one could attempt to solve the CME (with a known initial $p(0)$) and compute the probability vector $p(t)$. For most problems, this is a computationally very difficult task, starting with the fact that $p(t)$ is an infinite vector. Thus, it is often useful to *simulate* sample paths of the process. Statistics, such as means and variances, can then be obtained by averaging the results of several such simulations.

The naïve approach to simulation is to discretize time into small intervals, and iterate on intervals, randomly deciding at each instant whether a reaction happens. This is not at all an efficient way to proceed: if the discretization is too fine, no reactions will take place in most intervals, and the iteration step is wasted; if the discretization is too gross, we miss fast behaviors. Luckily, there is a far better way to proceed. The basic method[13] for simulating sample paths of CME's is the *stochastic simulation*

[11] The calculation is: $\mathbb{P}\left[A|B\&C\right] = \frac{\mathbb{P}\left[A\&B\&C\right]}{\mathbb{P}\left[B\&C\right]} = \frac{\mathbb{P}\left[A\&B\right]}{\mathbb{P}\left[B\&C\right]} == \frac{\mathbb{P}\left[A|B\right]\mathbb{P}\left[B\right]}{\mathbb{P}\left[C|B\right]\mathbb{P}\left[B\right]} = \frac{\mathbb{P}\left[A|B\right]}{\mathbb{P}\left[C|B\right]}$.

[12] "Markoff chains - Denumerable case," *Transactions of the American Mathematical Society* **58**(1945): 455-473.

[13] There are many variants that are often more efficient to implement, but the basic idea is always the same.

algorithm. Also known as the *kinetic Monte Carlo algorithm*[14], it has been probably known and used for a long time, at least since J.L. Doob's work cited earlier, but in its present form was introduced independently by A.B. Bortz, M.H. Kalos, and J.L. Lebowitz[15] and by D.T. Gillespie[16] (the SSA is often called the "Gillespie algorithm" in the systems biology field).

The method is very simple: if the present state is k, first use the random variable $\mathcal{T}_k$ to compute the next-reaction time, and then pick the particular reaction according to the discrete distribution $d_j^{(k)}$, where we are writing $d_j^{(k)}$ instead of $d_{k+\gamma_j}^{(k)}$, for each $j \in \{1,\ldots,m\}$ (all other $d_\ell^{(k)} = 0$). With the notations for propensities used in the CME, we have, for each $J \in \{1,\ldots,m\}$:

$$d_J^{(k)} \;=\; \frac{q_{k+\gamma_J,k}}{\sum_{\widetilde{\ell}=k+\gamma_j} q_{\widetilde{\ell},k}} \;=\; \frac{\rho_J^\sigma(k)}{\sum_{j=1}^m \rho_j^\sigma(k)} \;=\; \frac{\rho_J^\sigma(k)}{\lambda_k}\,.$$

Generating samples of the exponential random variable $\mathcal{T}_k$ is easy provided that a uniform (pseudo) random number generator is available, like the "rand" function in MATLAB. In general, if U is a uniformly distributed random variable on $[0,1]$, that is, $\mathbb{P}[U<p]=p$ for $p\in[0,1]$, then $\mathcal{T}=-\frac{\ln U}{\lambda}$ is an exponentially distributed random variable with parameter λ, because:

$$\mathbb{P}[\mathcal{T}>t] \;=\; \mathbb{P}\left[-\frac{\ln U}{\lambda}>t\right] \;=\; \mathbb{P}\left[U<e^{-\lambda t}\right] \;=\; e^{-\lambda t}\,.$$

Here is the pseudo-code for the SSA:

Initialization:

1. inputs: state k, maximal simulation time $T_{\max}$
2. set current simulation time $t := 0$.

Iteration:

1. compute $\rho_j^\sigma(k)$, for each reaction $\mathcal{R}_j$, $j=1,\ldots,m$
2. compute $\lambda := \sum_{j=1}^m \rho_j^\sigma(k)$
3. if $\lambda = 0$, stop (state is an absorbing state, no further transitions are possible)
4. generate two uniform random numbers r_1, r_2 in $[0,1]$
5. compute $T := -\frac{1}{\lambda}\ln r_1$
6. if $t+T > T_{\max}$, stop
7. find the index J such that $\frac{1}{\lambda}\sum_{j=1}^{J-1}\rho_j^\sigma(k) \le r_2 < \frac{1}{\lambda}\sum_{j=1}^{J}\rho_j^\sigma(k)$
8. update $k := k+\gamma_J$
9. update $t := t+T$.

Note that, in step 7, the probability that a particular $j=J$ is picked is the same as the length of the interval $[\frac{1}{\lambda}\sum_{j=1}^{J-1}\rho_j^\sigma(k), \frac{1}{\lambda}\sum_{j=1}^{J}\rho_j^\sigma(k)]$, which is $\frac{1}{\lambda}\rho_J^\sigma(k) = d_J^{(k)}$.

[14] In general, "Monte Carlo" methods are algorithms that rely on repeated random sampling to compute their results.

[15] "New algorithm for Monte-Carlo simulations of Ising spin systems," *J. Comput. Phys.* **17**(1975): 10-18.

[16] "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Physics* **22**(1976): 403-434.

Of course, one will also want to add code to store the sequence of states k and the jump times T, so as to plot sample paths. Note that, in MATLAB, if v is an array with the numbers $\rho_j^\sigma(k)$, then the command "J = find(cumsum(v)>sum($r_2 * v$))" provides the index J.

Exercise. (1) Implement the SSA in your favorite programming system (MATLAB, Maple, Mathematica). (2) Take the mRNA/protein model described earlier, pick some parameters, and an initial state; now plot many sample paths, averaging to get means and variances as a function of time, as well as steady state means and variances. (3) Compare the latter with the numbers obtained by using theory as described later. □

Remark: An equivalent way to generate the next reaction in the SSA ("direct method") is as follows (the "first reaction method", also discussed by Gillespie): generate m independent exponential random variables $\mathcal{T}_j$, $j = 1, \ldots, m$, with parameters $\lambda_k^{(j)} = \rho_j^\sigma(k)$ respectively –we think of $\mathcal{T}_j$ as indicating when the reaction j would next take place– and pick the "winner" (smallest $\mathcal{T}_j$) as the time (and index) of the next reaction. The same result obtains, because of the following general mathematical fact:[17] if $\mathcal{T}_1, \ldots, \mathcal{T}_m$ are independent exponentially distributed random variables with rate parameters $\mu_1, \ldots, \mu_m$ respectively, then $\mathcal{T} = \min_j \mathcal{T}_j$ is also exponentially distributed, with parameter $\mu = \sum_j \mu_j$. This fact is simple to prove:

$$\mathbb{P}[\mathcal{T} > t] \;=\; \mathbb{P}[\mathcal{T}_1 > t \,\&\, \ldots \,\&\, \mathcal{T}_m > t_m] \;=\; \prod_j \mathbb{P}[\mathcal{T}_j > t] \;=\; \prod_j e^{-\mu_j t} \;=\; e^{-\mu t}\,.$$

Moreover, it is also true that the index J of the variable which achieves the minimum –i.e., the "next reaction"– is a discrete random variable with is distributed according to the law $\mathbb{P}[J = p] = \mu_p/(\sum_j \mu_j)$.

From a computational point of view, the first reaction method would appear to be less efficient than the direct method, as m random variables have to be generated at each step (compared to just two for the direct method). However, the first reaction method has one advantage: since any given reaction will typically affect only a small number of species, there is no need to re-compute propensities for those indices j for which $\rho_j^\sigma(k)$ has not changed. This observation, together with the use of an indexed priority queue data structure, and a re-use of previously-generated $\mathcal{T}_j$'s, leads to a more efficient algorithm, the "next reaction method" due to M.A. Gibson and J. Bruck.[18]

4.4.7 Interpretation of mass-action kinetics

We explain now, through an informal discussion, how the formula (4.5): $\rho_{j,\Omega}^\sigma(k) = \frac{c_j}{\Omega^{A_j - 1}} \binom{k}{a_j}$ ($j = 1, \ldots, m$) is derived.

Suppose that the state of the system at time t is $k = (k_1, \ldots, k_n)'$, and we consider an interval of length $0 < h \ll 1$. What is the probability of a reaction $\mathcal{R}_j$ taking place in the interval $[t, t+h]$?

For this reaction to even have a chance of happening, the first requirement is that some subset $\mathcal{S}$ consisting of

$a_{1,j}$ units of species S_1, $a_{2,j}$ units of species S_2, $a_{3,j}$ units of species S_3, $\ldots$, $a_{n,j}$ units of species S_n

[17] While formally this provides the same numbers, it is not clear *a priori* why reaction times should be independent!

[18] "Efficient exact stochastic simulation of chemical systems with many species and many channels," *J. Phys. Chem. A* **104**(2000): 1876-1889.

come together in some small volume Ω_0 (Ω_0 depends on the physical chemistry of the problem). For the purpose of this discussion, let us call such an event a "collision" and a set of this form a "reactant set" for reaction $\mathcal{R}_j$.

The system is assumed to be "well-mixed", in the sense that species move randomly and fast, thus giving every possible reactant set an equal chance to have a collision.

The basic assumption of mass-action kinetics is that the probability $\rho_j^\sigma(k)h$ that some collision will happen, during a short interval $[t, t+h]$, is proportional to:

- the length h of the interval;
- the probability that a fixed reactant set has a collision; and
- the number of ways in which a reactant set can be picked, if the state is k.

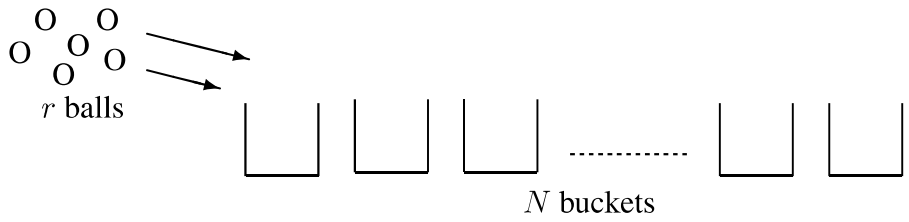
This model implicitly assumes that, if $\Omega_0 \ll \Omega$ (the total volume), then the chance that more than one collision will happen during a short period is much smaller than the probability of just one collision.

There are $\binom{k}{a_j} = \prod_{i=1}^n \binom{k_i}{a_{ij}}$ possible reactant subsets, if the state is k.

Next, we will argue that the probability of a collision, for any one given reactant set $\mathcal{S}$, is $\left(\frac{\Omega_0}{\Omega}\right)^{r-1}$, where $r = A_j$ is the cardinality of $\mathcal{S}$ (the order of the reaction).

From here, one obtains the formula for $\rho_j^\sigma(k)$. (The constant Ω_0 is absorbed into the proportionality constant c_j, which also includes other biophysical information, such as the probability that a reaction takes place when a collision happens, which in turn depends on the collision energy exceeding a threshold value and on the temperature. The Arrhenius equation gives the dependence of the rate constant on the absolute temperature T as $k = Ae^{-E/RT}$, were E is the "activation energy" and R is the gas constant.)

Suppose that $N = \frac{\Omega}{\Omega_0}$ is an integer. (This is a mild hypothesis, if $\Omega \gg \Omega_0$.) Then, the probability of having a collision, for a given reactant set $\mathcal{S}$, is the probability that r balls all land in the same bucket (an "urn" in probability theory) when assigned uniformly at random to one of N buckets.



We need to show that this probability is $\left(\frac{1}{N}\right)^{r-1}$. Indeed, the probability that all balls end up in the first bucket is $\left(\frac{1}{N}\right)^r$ (each ball has probability $1/N$ of landing in bucket 1, and the events are independent). The probability that all balls end up in the second bucket is also $\left(\frac{1}{N}\right)^r$, and similarly for all other buckets.

Since the events "all balls land in bucket i" and "all balls land in bucket j" are mutually exclusive for $i \neq j$, the probability of success is $N \times \left(\frac{1}{N}\right)^r = \left(\frac{1}{N}\right)^{r-1}$, which is what we wanted to prove.

The main examples are:

(0) zeroth-order reactions, in which an isolated species is created by means of a process which involve precursors which are not explicitly made a part of the model and which are well-distributed in space; in this case $A_j = 0$ and $\rho_j^\sigma(k)$ is independent of k, so it is just a constant, proportional to the volume;

(1) first-order or monomolecular reactions, in which a single unit of species i is degraded, diluted, decays, flows out, or gets transformed into one or more species; in this case $A_j = 1$ and exactly one a_{ij} is equal to 1 and the rest are zero, so $\rho_j^\sigma(k) = c_j k_i$ (since $\Omega^0 = 1$);

(2) homogeneous second-order (bimolecular) reactions involving two different species S_i and S_ℓ, one unit of each; now there two entries a_{ij} and $a_{\ell j}$ equal to 1, and the rest are zero, $A_j = 2$, and $\rho_j^\sigma(k) = \frac{1}{\Omega} c_j k_i k_\ell$;

(3) homogeneous second-order (bimolecular) reactions involving two units of the same species S_i; now $A_j = 2$ and exactly one a_{ij} is equal to 2 and the rest are zero, so $\rho_j^\sigma(k) = \frac{1}{\Omega} c_j \frac{k_i(k_i-1)}{2}$.

It is frequently argued that at most mono and bimolecular reactions are possible in the real world, since the chance of three or more molecules coming together in a small volume is vanishingly small. In this case, reactions involving multiple species would really consist of a sequence of more elementary bimolecular reactions, involving short-lived, intermediate, species. However, multi-species reactions might still make sense, either as an approximation of a more complicated sequence that occurs very fast, or if molecules are very large compared to the volume, of if the model is one that involves non-chemical substances (for example, in population biology).

4.5 Moment equations and fluctuation-dissipation formula

We next see how to obtain equations for the derivatives of the mean $\mathbb{E}[X_i(t)]$ and the covariance $\mathbb{V}\text{ar}[X(t)]$ of $X(t)$, assuming that the probability density of $X(t)$ is given by a CME as in (4.3). No special form needs to be assumed for the propensities, for these theoretical considerations to be valid, but in examples we use mass-action kinetics.

We provide first a very general computation, which we will later specialize to first and second moments. Suppose for this purpose that we have been given a function M which will be, in our two examples, a vector- of matrix-valued function defined on the set of non-negative integers. More abstractly, we take $M : \mathbb{Z}^n_{\geq 0} \to \mathcal{V}$, where $\mathcal{V}$ is any vector space. For first moments (means), we have $\mathcal{V} = \mathbb{R}^n$ and $M(k) = k$. For second moments, $\mathcal{V} = \mathbb{R}^{n\times n}$, the space of all $n \times n$ matrices, and $M(k) = kk'$.[19]

The first goal is to find a useful expression for the time derivative of $\mathbb{E}[M(X(t))]$. The definition of expectation gives:[20]

$$\mathbb{E}[M(X(t))] \;=\; \sum_{k\in\mathbb{Z}^n_{\geq 0}} p_k(t)\, M(k)$$

because $\mathbb{P}[X(t) = k] = p_k(t)$. We have:

$$\frac{d}{dt}\mathbb{E}[M(X(t))] \;=\; \sum_{k\in\mathbb{Z}^n_{\geq 0}} \frac{dp_k}{dt}(t)M(k) \;=\; \sum_{k\in\mathbb{Z}^n_{\geq 0}} \left(\sum_{j=1}^m \rho_j^\sigma(k-\gamma_j)\, p_{k-\gamma_j} - \sum_{j=1}^m \rho_j^\sigma(k)\, p_k\right) M(k)\,.$$

Note this equality, for each fixed j:

$$\sum_{k\in\mathbb{Z}^n_{\geq 0}} p_{k-\gamma_j}(t)\rho_j^\sigma(k-\gamma_j)M(k) \;=\; \sum_{k\in\mathbb{Z}^n_{\geq 0}} p_k(t)\rho_j^\sigma(k)M(k+\gamma_j)$$

(by definition, $\rho_j^\sigma(k-\gamma_j) = 0$ unless $k \geq \gamma_j$, so one may perform a change of variables $\widetilde{k} = k - \gamma_j$). There results:

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[M(X(t))] \;&=\; \sum_{k,j} p_k(t)\rho_j^\sigma(k)M(k+\gamma_j) \;-\; \sum_{k,j} p_k(t)\rho_j^\sigma(k)M(k)\\ &=\; \sum_{k\in\mathbb{Z}^n_{\geq 0}} p_k(t) \sum_{j=1}^m \rho_j^\sigma(k)\,[M(k+\gamma_j) - M(k)]\;.\end{aligned}$$

Let us define, for any $\gamma \in \mathbb{Z}^n_{\geq 0}$, the new function $\Delta_\gamma M$ given by $(\Delta_\gamma M)(k) := M(k+\gamma) - M(k)$. With these notations,

$$\frac{d}{dt}\mathbb{E}[M(X(t))] \;=\; \mathbb{E}\left[\sum_{j=1}^m \rho_j^\sigma(X(t))\,\Delta_\gamma M(X(t))\right]\,. \tag{4.28}$$

Note that this is *not* an ordinary differential equation for $\mathbb{E}[M(X(t)]$, because the right-hand side is not, generally, a function of $\mathbb{E}[M(X(t))]$. In some cases, however, various approximations result in differential equations, as discussed below.

[19] As usual, prime indicates transpose, so this is the product of a column vector by a row vector, which is a rank 1 matrix if $k \neq 0$.

[20] Note that this is a deterministic function, not depending on the random outcomes of the process.

Remark. Suppose that M is a polynomial of degree δ_M and that the propensities are polynomials of degree $\leq \delta_\rho$ (the maximal order of reactions, in the mass action case). Then $\Delta_\gamma M$ is a polynomial of degree $\delta_M - 1$, so the monomials appearing inside the expectation have degree $\leq \delta_\rho + \delta_M - 1$. This means that $\frac{d}{dt}\mathbb{E}[M(X(t))]$ depends on moments of order $\leq \delta_\rho + \delta_M - 1$. Thus, if all reactions have order at most 1, a system of differential equations can obtained for the set of moments of up to any fixed order: the derivative of each moment depends only on equal and lower-order ones, not higher moments. On the other hand, if some reactions have order larger than 1, then $\delta_\rho + \delta_M - 1 > \delta_M$, so in general no clsoed set of equations is available for any finite subset of moments.

4.5.1 Means

For the mean $\mathbb{E}[X(t)]$, we have $M(k) = k$, so $\Delta_j M(k) = k + \gamma_j - k = \gamma_j$ (a constant function), and thus:

$$\sum_{j=1}^{m} \rho_j^\sigma(X(t))\,\Delta_\gamma M(X(t)) \;=\; \sum_{j=1}^{m} \rho_j^\sigma(X(t))\,\gamma_j \;=\; f^\sigma(X(t))$$

where, recall, we defined the n-column vector:

$$f^\sigma(k) \;:=\; \sum_{j=1}^{m} \rho_j^\sigma(k)\,\gamma_j \qquad\qquad k \in \mathbb{Z}^n_{\geq 0}\,.$$

With these notations, Equation (4.28) specializes to:

$$\boxed{\frac{d}{dt}\mathbb{E}[X(t)] \;=\; \mathbb{E}[f^\sigma(X(t))]}\,. \tag{4.29}$$

Recall that $f^\sigma(k)$ can also be written in the form

$$f^\sigma(k) \;=\; \Gamma R^\sigma(k) \tag{4.30}$$

where $R^\sigma(k) = (\rho_1^\sigma(k), \ldots, \rho_m^\sigma(k))'$ and Γ is the stoichiometry matrix.

For mass-action kinetics, the function f^σ is basically the same one[21] that is used in the deterministic differential equation model for the corresponding chemical network. Thus, it is a common mistake to think that the deterministic equation represents an equation that is satisfied by the mean $\mu(t) = \mathbb{E}[X(t)]$, that is to say, to believe that $d\mu/dt = f^\sigma(\mu)$. However, the precise formula is (4.29). Since expectation of a nonlinear function is generally not the same as the nonlinear function of the expectation[22], (4.29) is, in general, very different from $(d/dt)\mathbb{E}[X(t)] = f^\sigma(\mathbb{E}[X(t)])$. One important exception, which permits the replacement $\mathbb{E}[f^\sigma(X(t))] = f^\sigma(\mathbb{E}[X(t)])$, is that in which f^σ is an affine function (linear + constant), that is to say if all propensities are affine, which for mass-action kinetics means that all the reactions involve zero or at most one reactant:

$$\boxed{\frac{d}{dt}\mathbb{E}[X(t)] \;=\; f^\sigma(\mathbb{E}[X(t)]) \qquad \text{if all reactions are mass-action of order 0 or 1}}\,. \tag{4.31}$$

[21] There is just a very minor difference, discussed later, having to do with replacing terms such as "$x(x-1)$" in a second-order homodimerization reaction by the simpler expression x^2.

[22] Example: $\mathbb{E}[X^2] \neq \mathbb{E}[X]^2$; in fact, the variance of X is precisely the concept introduced in order to quantify the difference between these two quantities!

On the other hand, even for reactions of arbitrary order, one might expect that Equation (4.31) holds at least approximately provided that the variance of $X(t)$ is small, so that $X(t)$ is almost deterministic. More precisely, one has the following argument.

Let us assume that the function f^σ, which is defined only for non-negative integer vectors, can be extended to a differentiable function, also written as $f^\sigma(x)$, that is defined for all non-negative real numbers x. This is the case with all propensities that are used in practice, such as those arising from mass-action kinetics. Thus, around each vector ξ, we may expand $f^\sigma(x)$ to first-order around $x = \xi$:

$$f^\sigma(x) \;=\; f^\sigma(\xi) \;+\; J(\xi)(x-\xi) \;+\; g_\xi(x-\xi) \tag{4.32}$$

where $J(x)$ is the Jacobian matrix of f^σ evaluated at $x = \xi$ and where g_ξ is a vector function which is $o(|x-\xi|)$. When f is second-order differentiable, the entries g^i_ξ of the vector g_ξ can be expressed as:

$$g^i_\xi(x) \;=\; \frac{1}{2}\,(x-\xi)'\,H_i(\xi)\,(x-\xi) \;+\; o(|x-\xi|^2)$$

where $H_i(\xi)$ is the Hessian of the ith component of the vector field f^σ (the matrix of second order partial derivatives) evaluated at $x = \xi$.

For notational simplicity, let us write μ for means: $\mu(t) \;=\; \mathbb{E}\,[X(t)]$. In the particular case that $\xi = \mu(t)$ and $x = X(t)$ (along a sample path), we have:

$$f^\sigma(X(t)) \;=\; f^\sigma(\mu(t)) \;+\; J(\mu(t))\,(X(t)-\mu(t)) \;+\; g_{\mu(t)}(X(t)-\mu(t))\,. \tag{4.33}$$

Now, $J(\mu(t))$ is a deterministic function, so, since the expectation operator is linear,

$$\mathbb{E}\,[J(\mu(t))\,(X(t)-\mu(t))] \;=\; J(\mu(t))\,(\mathbb{E}\,[X(t)]-\mu(t)) \;=\; J(\mu(t))\,(\mathbb{E}\,[X(t)]-\mathbb{E}\,[X(t)]) \;=\; 0\,.$$

Since also $f^\sigma(\mu(t))$ is deterministic, it follows that:

$$\frac{d}{dt}\mathbb{E}\,[X(t)] \;=\; \mathbb{E}\,[f^\sigma(X(t))] \;=\; f^\sigma(\mathbb{E}\,[X(t)]) \;+\; G(t)$$

where

$$G(t) = \mathbb{E}\left[g_{\mu(t)}(X(t)-\mu(t))\right]\,. \tag{4.34}$$

This term involves central moments (covariances, etc.) of order ≥ 2.

4.5.2 Variances

For the matrix of second order moments $\mathbb{E}\,[X(t)X(t)']$, we have $M(k) = kk'$, so

$$\Delta_j M(k) \;=\; (k+\gamma_j)(k+\gamma_j)' - kk' \;=\; k\gamma_j' + \gamma_j k' + \gamma_j\gamma_j'$$

and so Equation (4.28) $\frac{d}{dt}\mathbb{E}\,[M(X(t))] = \mathbb{E}\left[\sum_{j=1}^m\, \rho_j^\sigma(X(t))\,\Delta_\gamma M(X(t))\right]$ specializes to:

$$\frac{d}{dt}\mathbb{E}\,[X(t)X(t)'] \;=\; \mathbb{E}\left[\sum_{j=1}^m \rho_j^\sigma(X(t))\,X(t)\,\gamma_j'\right] + \mathbb{E}\left[\sum_{j=1}^m \rho_j^\sigma(X(t))\,\gamma_j\,X(t)'\right] + \sum_{j=1}^m \mathbb{E}\left[\rho_j^\sigma(X(t))\right]\gamma_j\gamma_j' \tag{4.35}$$

(note that the $\rho_j^\sigma(X(t))$'s are scalar, and that $X(t)$ and the γ_j's are vectors). Since we had defined $f^\sigma(k) = \sum_{j=1}^m \rho_j^\sigma(k)\,\gamma_j$, the second term in this sum can be written as $\mathbb{E}\,[f^\sigma(X(t))\,X(t)']$. Similarly, the first term is $\mathbb{E}\,[X(t)\,f^\sigma(X(t))']$. The last term can be written in the following useful form.

We introduce the $n \times n$ *diffusion matrix*[23] $B(k) = (B_{pq}(k))$ which has the following entries:

$$B_{pq}(k) \;=\; \sum_{j=1}^{m} \rho_j^\sigma(k)\,\gamma_{pj}\,\gamma_{qj}\,, \qquad p,q = 1,\ldots,n\,, \tag{4.36}$$

where γ_{pj} is the pth row of the column vector γ_j, that is to say the (p,j)th entry of the stoichiometry matrix Γ, so that $\gamma_{pj}\gamma_{qj}$ is the (p,q)th entry of the matrix $\gamma_j\gamma_j'$. Note that B is an $n \times n$ symmetric matrix. In summary, we can write (4.35) as follows:

$$\boxed{\frac{d}{dt}\mathbb{E}\,[X(t)X(t)'] \;=\; \mathbb{E}\,[X(t)\,f^\sigma(X(t))'] \,+\, \mathbb{E}\,[f^\sigma(X(t))\,X(t)'] \,+\, \mathbb{E}\,[B(X(t))]}\,. \tag{4.37}$$

One interpretation of the entries $\mathbb{E}\left[B_{pq(X(t)}\right]$ is as follows. The product $\gamma_{pj}\gamma_{qj}$ is positive provided both species S_p and S_q change with the same sign (both increase or both decrease) when the reaction $\mathcal{R}_j$ fires. The product is negative if one species increases but the other decreases, when $\mathcal{R}_j$ fires. The absolute value of this product is large if at least one of these two species jumps by a large amount. Finally, the expected value of the coefficient $\rho_j^\sigma(k)$ quantifies the rate at which the corresponding reaction takes place. In this manner, $\mathbb{E}\left[B_{pq(X(t)}\right]$ contributes toward an instantaneous change in the correlation between species S_p and S_q.

An equation for the derivative of the variance is easily obtained from here. By definition, $\mathbb{V}\text{ar}\,[X(t)] = \mathbb{E}\,[X(t)X(t)'] - \mathbb{E}\,[X(t)]\,\mathbb{E}\,[X(t)]'$, so we need to compute the derivative of this last term. For a vector function $v = v(t)$, $(d/dt)vv' = v(dv/dt)' + (dv/dt)v'$, so with $dv/dt = \frac{d}{dt}\mathbb{E}\,[X(t)] = \mathbb{E}\,[f^\sigma(X(t))]$ from (4.29),

$$\boxed{\frac{d}{dt}\mathbb{V}\text{ar}\,[X(t)] \;=\; \mathbb{E}\,[(X(t)-\mu(t))\,f^\sigma(X(t))'] \,+\, \mathbb{E}\,[f^\sigma(X(t))\,(X(t)-\mu(t))'] \,+\, \mathbb{E}\,[B(X(t))]} \tag{4.38}$$

where we wrote $\mu(t) = \mathbb{E}\,[X(t)]$ for clarity.

Exercise. Show that an alternative way of writing the third term in the right-hand side of (4.38) is as follows:

$$\Gamma\,\text{diag}\,(\mathbb{E}\,[\rho_1^\sigma(X(t))]\,,\ldots,\mathbb{E}\,[\rho_m^\sigma(X(t))])\;\Gamma' \tag{4.39}$$

(where "diag $(r_1,\ldots,r_m)$" means a diagonal matrix with entries r_i in the diagonal). □

The first-order Taylor expansion of f^σ, $f^\sigma(X(t)) = f^\sigma(\mu(t)) + J(\mu(t))(X(t)-\mu(t)) + g_{\mu(t)}(X(t)-\mu(t))$, given in (4.33), can be substituted into the term $\mathbb{E}\,[f^\sigma(X(t))\,(X(t)-\mu(t))']$ in the formula (4.38) for the covariance, giving (dropping the arguments "t" for readability):

$$\begin{aligned}\mathbb{E}\,[f^\sigma(X)(X-\mu)'] \;&=\; \mathbb{E}\,[f^\sigma(\mu)(X-\mu)'] \,+\, \mathbb{E}\,[J(\mu)(X-\mu)(X-\mu)'] \,+\, \mathbb{E}\,[g_\mu(X-\mu)(X-\mu)']\\ &=\; J(\mu)\mathbb{V}\text{ar}\,[X] \,+\, \mathbb{E}\,[g_\mu(X-\mu)(X-\mu)']\end{aligned}$$

[23] Normally, "diffusion" is interpreted in a spatial sense. Here it is thought of, instead, as diffusion in "concentration space".

where we used that $f^\sigma(\mu(t))$ and $J(\mu(t))$ are deterministic and that $\mathbb{E}[X-\mu]=0$. Similarly,

$$\mathbb{E}[(X-\mu)f^\sigma(X)'] \;=\; \mathbb{V}\text{ar}[X]\,J(\mu)' \;+\; \mathbb{E}[(X-\mu)\,g_\mu(X-\mu)']$$

(the covariance matrix is symmetric, so there is no need to transpose it). Therefore,

$$\frac{d}{dt}\mathbb{V}\text{ar}[X(t)] \;=\; \mathbb{V}\text{ar}[X(t)]\,J(\mu(t))' \;+\; J(\mu(t))\mathbb{V}\text{ar}[X(t)] \;+\; \mathbb{E}[B(X(t))] \;+\; \alpha(t) \tag{4.40}$$

where $\alpha(t)=\mathbb{E}\left[(X(t)-\mu(t))\,g_{\mu(t)}(X(t)-\mu(t))' \;+\; (X(t)-\mu(t))\,g_{\mu(t)}(X(t)-\mu(t))'\right]$. Dropping the term $\alpha(t)$, one has the *fluctuation-dissipation* formula:

$$\boxed{\frac{d}{dt}\mathbb{V}\text{ar}[X(t)] \;\approx\; \mathbb{V}\text{ar}[X(t)]\,J(\mu(t))' \;+\; J(\mu(t))\mathbb{V}\text{ar}[X(t)] \;+\; \mathbb{E}[B(X(t))] \qquad \text{(FD)}}\;. \tag{4.41}$$

If the higher-order moments of $X(t)$ are small, one may be justified in making this approximation, because $\alpha(t)$ is $o(|X(t)-\mu(t)|^2)$, while the norm of the covariance matrix is $O(|X(t)-\mu(t)|^2)$.

Equation (4.41) is sometimes called the *mass fluctuation kinetics* equation, and the term "fluctuation-dissipation" is used for a slightly different object, as follows. Suppose that we expand $B(x)$ as a Taylor series around the mean $\mathbb{E}[X(t)]$. Arguing as earlier, we have that $\mathbb{E}[B(X(t))] \;=\; B(\mathbb{E}[X(t)]) \;+\; o(|X(t)-\mu(t)|)$. This suggests replacing the last term in (FD) by $B(\mathbb{E}[X(t)])$.

4.5.3 Reactions or order ≤ 1 or ≤ 2

The special case in which f^σ is a polynomial of degree two is arguably the most general that often needs to be considered. (Recall the discussion about reactions of order >2.) In this case, the function g_ξ in (4.32) is a vector field that is quadratic on the coordinates of $X(t)-\mu(t)$, with *constant* coefficients, because the Hessian of a quadratic polynomial is constant. The expectations of such expressions are the covariances $\mathbb{C}\text{ov}[X_i(t),X_j(t)]$ (variances if $i=j$). So, $G(t)$ is a linear function L of the n^2 entries of $\mathbb{V}\text{ar}[X(t)]$. The linear function L can be easily computed from the second derivatives of the components of f^σ. Similarly, as the entries of the diffusion matrix (4.36) are polynomials of degree equal to the largest order of the reactions, when all reactions have order ≤ 2 the term $\mathbb{E}[B(X(t))]$ is an affine linear function of the entries of $\mathbb{E}[X(t)]$ and $\mathbb{V}\text{ar}[X(t)]$, which we write as $H_0+H_1\mathbb{E}[X(t)]+H_2\mathbb{V}\text{ar}[X(t)]$. Thus:

For mass-action kinetics and all reactions of order at most 2, the fluctuation-dissipation equation says that the mean $\mu(t)=\mathbb{E}[X(t)]$ ***and covariance matrix*** $\Sigma(t)=\mathbb{V}\textbf{\textit{ar}}[X(t)]$ ***satisfy***

$$d\mu/dt \;=\; f^\sigma(\mu)+L\Sigma \tag{4.42a}$$

$$d\Sigma/dt \;\approx\; \Sigma\,J(\mu)' \;+\; J(\mu)\,\Sigma \;+\; H_0+H_1\mu+H_2\Sigma \tag{4.42b}$$

(where the "approximate" sign indicates that α, which involves third-order moments, because $g_{\mu(t)}$ is quadratic, was dropped). Moreover, the function $J(\mu(t))$ is linear in $\mu(t)$.

The FD formula is *exact* for zero- and first-order mass-action reactions, because in that case the Hessian and thus $g_{\mu(t)}$ are zero, so also $\alpha(t)\equiv 0$. Moreover, in this last case the entries $B_{pq}(k)=\sum_{j=1}^m \rho_j^\sigma(k)\gamma_{pj}\gamma_{qj}$ of the diffusion matrix are also affine, so that the last term is just $B(\mathbb{E}[X(t)])$. It is worth emphasizing this fact:

For mass-action kinetics and all reactions of order zero or one, the mean $\mu(t) = \mathbb{E}[X(t)]$ ***and covariance matrix*** $\Sigma(t) = \mathbb{V}$***ar***$[X(t)]$ ***are solutions of the coupled system of differential equations***

$$\dot{\mu} = f^\sigma(\mu) \tag{4.43a}$$

$$\dot{\Sigma} = \Sigma J' + J\Sigma + B(\mu) \tag{4.43b}$$

and in this case J does not depend on μ, because J is a constant matrix, being the Jacobian of an affine vector field. Also,

$$B(\mu) = \Gamma \,\mathrm{diag}\,(\rho_1^\sigma(\mu), \ldots, \rho_m^\sigma(\mu))\ \Gamma' \tag{4.44}$$

in the case of order ≤ 1.

Note that (4.43) is a set of $n + n^2$ linear differential equations. Since covariances are symmetric, however, one can equally well restrict to the equations on the diagonal and upper-triangular part of Σ, so that it is sufficient to solve $n + n(n+1)/2$ equations.

The term "fluctuation-dissipation" is used because the first two terms for Σ may be though of as describing a "dissipation" of initial uncertainty, while the last term can be thought of as a "fluctuation" due to future randomness. To understand the dissipation component, let's discuss what would happen if the fluctuation term were not there. Then (FD) is a linear differential equation on $\mathbb{V}\mathrm{ar}[X(t)]$ (a "Lyapunov equation" in control theory). Given the initial variance $\mathbb{V}\mathrm{ar}[X(0)]$, a solution can be computed. This solution is identically zero when $X(0)$ is perfectly known (that is, $p(0)$ has exactly one nonzero entry), because $\mathbb{V}\mathrm{ar}[X(0)] = 0$ in that case. But even for nonzero $\mathbb{V}\mathrm{ar}[X(0)]$, and under appropriate stability conditions one would have that $\mathbb{V}\mathrm{ar}[X(t)] \to 0$ as $t \to \infty$. If a matrix J has eigenvalues with negative real part, then the operator $P \mapsto PJ' + JP$ on symmetric matrices has all eigenvalues also with negative real part.[24] So if $\mu(t)$ is approximately constant and the linearization of the differential equation for the mean is stable, the equation for the variance will be, too. Since in general the matrices $J(\mu(t))$ depend on t, this argument is not quite correct, but it provides the basic intuition for the term "dissipation".

[24]This is because the eigenvalues of this operator are the sums of pairs of eigenvalues of J; see e.g. the author's control theory textbook.

4.6 Generating functions

We next discuss how to use generating functions in order to (1) find solutions of the CME, or at least (2) find differential equations satisfied by moments. Often only simple problems can be solved explicitly with this technique, but it is nonetheless a good source of theoretical insight.

We assume that $p(t)$, an infinite vector function of time indexed by $k \in K = \mathbb{Z}^n_{\geq 0}$, is a solution of the CME (4.3):

$$\frac{dp_k}{dt} \;=\; \sum_{j=1}^{m} \rho_j^\sigma(k-\gamma_j)\, p_{k-\gamma_j} \;-\; \sum_{j=1}^{m} \rho_j^\sigma(k)\, p_k\,.$$

The *(probability) generating function* $P(z,t)$ is a scalar-valued function of time $t \geq 0$ and of n auxilliary variables $z = (z_1, \ldots, z_n)$ (which may be thought of as complex variables), defined as follows:

$$P(z,t) \;:=\; \mathbb{E}\left[z^X\right] \;=\; \sum_{k\in K} p_k(t)\, z^k \tag{4.45}$$

where we denote $z^k := z_1^{k_1} \ldots z_n^{k_n}$ and $z_i^0 = 1$. As the $p_k(t)$'s are non-negative and add up to one, the series is convergent for $z = 1$ (we write the vector $(1, \ldots, 1)$ as "1" when clear from the context):

$$P(1,t) \;=\; 1 \qquad \text{for all } t \geq 0. \tag{4.46}$$

Moments of arbitrary order can be computed once that P is known. For example,

$$\left.\frac{\partial P(z,t)}{\partial z}\right|_{z=1} \;=\; \mathbb{E}\left[X(t)\right]\,,$$

where we interpret the above partial derivative as the vector $\left(\left.\frac{\partial P(t,z)}{\partial z_1}\right|_{z=1}, \ldots, \left.\frac{\partial P(t,z)}{\partial z_m}\right|_{z=1}\right)'$. Also,

$$\left.\frac{\partial^2 P(z,t)}{\partial z_i \partial z_j}\right|_{z=1} \;=\; \begin{cases} \mathbb{E}\left[X_i(t)X_i(t)\right] & \text{if } i \neq j \\ \mathbb{E}\left[X_i(t)^2\right] - \mathbb{E}\left[X_i(t)\right] & \text{if } i = j\,. \end{cases}$$

Note that $\mathbb{V}\text{ar}\left[X(t)\right]$ can be computed from these formulas.

Exercise. Prove the above two formulas. □

We remark that there are other power series than are often associated to P, especially the *moment generating function*[25]

$$M(\theta,t) \;:=\; \mathbb{E}\left[e^{\theta X}\right] = \sum_{k\in K} p_k(t)\, e^{\theta k}$$

where we define $e^q = e^{q_1} \ldots e^{q_n}$.

Of course, actually computing $P(z,t)$ from its definition is not particularly interesting, since the whole purpose of using generating functions is to gain information about the unknown $p_k(t)$'s. The idea, instead, is to use the knowledge that $p(t)$ satisfies an infinite system of ordinary differential equations in order to obtain a *finite* set of *partial* differential equations for P. Sometimes these PDE's can be solved, and other times just the form of the PDE will be enough to allow computing ODE's for moments. We illustrate both of these ideas next, through examples.

[25]The terminology arises from the fact that the coefficients of the Taylor expansions of P and M, at $z = 0$ and $\theta = 0$, give the probabilities and moments, respectively.

Let us start with the mRNA example given by the reactions in (4.6), $0 \xrightarrow{\alpha} M \xrightarrow{\beta} 0$, for which (cf. (4.7)-(4.9)) $G = (1, -1)$, $\rho_1^\sigma(k) = \alpha$, $\rho_2^\sigma(k) = \beta k$, $f^\sigma(k) = \alpha - \beta k$, and the CME is

$$\frac{dp_k}{dt} \;=\; \alpha p_{k-1} + (k+1)\beta p_{k+1} - \alpha p_k - k\beta p_k \,.$$

Let us compute now a PDE for $P(z,t)$. For simplicity, from now on we will write $\frac{\partial}{\partial t}P$ as P_t and $\frac{\partial}{\partial z}P$ as P_z.

By definition, $P(z,t) = \sum_{k=0}^{\infty} p_k(t) z^k$, so

$$P_t \;=\; \sum_{k=0}^{\infty} \frac{dp_k}{dt}\, z^k \;=\; \alpha \sum_{k=1}^{\infty} p_{k-1} z^k + \beta \sum_{k=0}^{\infty} (k+1) p_{k+1} z^k - \alpha \sum_{k=0}^{\infty} p_k z^k - \beta \sum_{k=1}^{\infty} k p_k z^k \qquad (4.47)$$

where we started the first sum at 1 because of the convention that $p_{-1} = 0$, and the last at 1 because for zero we have a factor $k = 0$. The third sum in the right-hand side is just P; the rest are:

$$\sum_{k=1}^{\infty} p_{k-1} z^k \;=\; z \sum_{k=1}^{\infty} p_{k-1} z^{k-1} \;=\; z \sum_{k=0}^{\infty} p_k z^k \;=\; zP$$

$$\sum_{k=0}^{\infty} (k+1)\, p_{k+1} z^k \;=\; \sum_{k=1}^{\infty} k\, p_k z^{k-1} \;=\; P_z$$

$$\sum_{k=1}^{\infty} k\, p_k z^k \;=\; z \sum_{k=1}^{\infty} k\, p_k z^{k-1} \;=\; z P_z \,.$$

Thus, P satisfies:

$$P_t \;=\; \alpha z P + \beta P_z - \alpha P - \beta z P_z \qquad (4.48)$$

which can also be written as

$$P_t \;=\; (z-1)\,(\alpha P - \beta P_z) \;. \qquad (4.49)$$

To obtain a unique solution, we need to impose an initial condition, specifying $p(0)$, or equivalently $P(z,0)$.

(Recall from Equation (4.46) that we also have the boundary condition $P(1,t) = 1$ for all t, because $p(t)$ is a probability distribution.)

Let us say that we are interested in the solution that starts with $M = 0$: $p_0(0) = \mathbb{P}[M(0) = 0] = 1$ and $p_k(0) = \mathbb{P}[M(0) = k] = 0$ for all $k > 0$. This means that $P(z,0) = \sum_{k=1}^{\infty} p_k(0) z^k = 1$.

Equation (4.49) is a first-order PDE for P. Generally speaking, such PDE's can be solved by the "method of characteristics." Here we simply show that the following guess, which satisfies $P(1,t) = 0$ and $P(z,0) = 1$:

$$P(z,t) \;=\; e^{\frac{\alpha}{\beta}\left(1 - e^{-\beta t}\right)(z-1)} \qquad (4.50)$$

is a solution.[26] Indeed, note that, with this definition,

$$P_t(z,t) \;=\; \alpha e^{-\beta t}(z-1)\, P(t,z) \qquad (4.51)$$

[26] To be added: solution by characteristics and proof of uniqueness.

$$P_z(z,t) \;=\; \frac{\alpha}{\beta}(1-e^{-\beta t})\,P(t,z) \tag{4.52}$$

so:

$$P_t \;=\; (z-1)\,\alpha e^{-\beta t}P \;=\; (z-1)\,\left[\alpha P - \beta\frac{\alpha}{\beta}(1-e^{-\beta t})P\right] \;=\; (z-1)\,(\alpha P - \beta P_z)$$

as claimed.

Once that we have obtained the formula (4.50) for $P(z,t)$, we can expand it in a Taylor series in order to obtain $p_k(t)$. For example, for $k=1$ we have:

$$\mathbb{P}\,[X(t)=1] \;=\; p_1(t) \;=\; P_z(0,t) \;=\; \frac{\alpha}{\beta}(1-e^{-\beta t})P(0,t) \;=\; \frac{\alpha}{\beta}(1-e^{-\beta t})e^{-\frac{\alpha}{\beta}\left(1-e^{-\beta t}\right)}\,.$$

We can also compute moments, for example

$$\mu(t) \;=\; \mathbb{E}\,[X(t)] \;=\; P_z(1,t) \;=\; \frac{\alpha}{\beta}\left(1-e^{-\beta t}\right)\,P(1,t) \;=\; \frac{\alpha}{\beta}\left(1-e^{-\beta t}\right)\,.$$

As mentioned above, even without solving the PDE for P, one may obtain ODE's for moments from it. For example we have:[27]

$$\begin{aligned}\dot\mu \;&=\; \frac{\partial}{\partial t}\frac{\partial}{\partial z}\bigg|_{z=1} P \;=\; \frac{\partial}{\partial z}\bigg|_{z=1} P_t \;=\; \frac{\partial}{\partial z}\bigg|_{z=1} (z-1)\,(\alpha P-\beta P_z)\\ &=\; (\alpha P-\beta P_z) + (z-1)\,(\alpha P_z - \beta P_{zz})\,|_{z=1}\\ &=\; \alpha-\beta P_z(1,t) \;=\; \alpha-\beta\mu\,.\end{aligned}$$

Since every reaction has order 0 or 1, this equation for the mean is the same as the deterministic equation satisfied by concentrations.

Exercise. Use the PDE for P to obtain an ODE for the variance, following a method similar to that used for the mean.

Still for the mRNA example, let us compute the generating function $Q(z)$ of the steady state distribution π obtained by setting $\frac{dp}{dt} = 0$. At steady state, that is setting $P_t = 0$, we have that $(z-1)\,(\alpha Q - \beta Q_z) = 0$, so $\alpha Q - \beta Q_z = 0$, or equivalently $Q_z = \lambda Q$, where $\lambda = \frac{\alpha}{\beta}$. Thus, $Q(z) = ce^{\lambda z}$ for some constant c. Since π is a probability distribution, $Q(1) = 1$, and so $c = e^{-\lambda}$, and thus we conclude:

$$Q(z) \;=\; e^{-\lambda}e^{\lambda z} \;=\; e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}z^k\,.$$

Therefore, since by definition $Q(z) = \sum_{k=0}^{\infty} q_k z^k$, it follows that

$$q_k \;=\; e^{-\lambda}\frac{\lambda^k}{k!}$$

and we yet again have recovered the fact that the steady-state distribution is that of a Poisson random variable with parameter λ.

[27] Using $\frac{\partial}{\partial t}\frac{\partial}{\partial z} = \frac{\partial}{\partial z}\frac{\partial}{\partial t}$.

4.7 Examples computed using the fluctuation-dissipation formula

Consider again the mRNA example given by the reactions in (4.6), $0 \xrightarrow{\alpha} M \xrightarrow{\beta} 0$, for which (cf. (4.7)-(4.9)) $G = (1, -1)$, $\rho_1^\sigma(k) = \alpha$, $\rho_2^\sigma(k) = \beta k$, $f^\sigma(k) = \alpha - \beta k$. Since the reactions are of order 0 and 1, the FD formula is exact, so that the mean and variance $\mu(t)$ and $\Sigma(t)$ satisfy (4.43): $\dot{\mu} = f^\sigma(\mu)$, $\dot{\Sigma} = \Sigma J' + J\Sigma + B(\mu)$. Here both μ and Σ are scalar variables. The Jacobian of f^σ is $J = -\beta$. The diffusion term is

$$B(\mu) \;=\; \sum_{j=1}^{2} \rho_j^\sigma(\mu)\gamma_{1j}\gamma_{1j} \;=\; \alpha 1^2 + \beta\mu(-1)^2 \;=\; \alpha + \beta\mu\,,$$

so that the FD equations become:

$$\dot{\mu} \;=\; \alpha - \beta\mu \tag{4.53a}$$

$$\dot{\Sigma} \;=\; -2\beta\Sigma + \alpha + \beta\mu\,. \tag{4.53b}$$

Note that the equation for the mean is the same that we derived previously using the probability generating function. There is a unique steady state for this equation, given by $\mu = \alpha/\beta = \lambda$ (the parameter of the Poisson random variable $X(\infty)$) and, solving $-2\beta\Sigma + \alpha + \beta\mu = 0$:

$$\Sigma = \frac{\alpha + \beta\mu}{2\beta} = \frac{\alpha}{\beta} = \lambda$$

which is, of course, consistent with the property that the variance and mean of a Poisson random variable are the same.

Exercise. Derive the variance equation from the probability generating function, and show that the same result is obtained.

Exercise. Solve explicitely the linear differential equations (4.53). (Use matrix exponentials, or variation of parameters.)

One measure of how "noisy" a scalar random variable X is, is the ratio between its standard deviation $\sigma = \sqrt{\Sigma}$ and its mean, called the *coefficient of variation*:

$$\mathbf{cv}\,[X] \;:=\; \frac{\sigma\,[X]}{\mathbb{E}\,[X]}$$

(only defined if $\mathbb{E}\,[X] \neq 0$).

This number may be small even if the variance is large, provided that the mean is large. It represents a "relative noise" and is a "dimensionless" number, thus appropriate, for example, when comparing objects measured in different units.[28]

For a Poisson random variable X with parameter λ, $\mathbb{E}\,[X] = \lambda$ and $\sigma\,[X] = \sqrt{\lambda}$, so $\mathbf{cv}\,[X] = 1/\sqrt{\lambda}$.

Next, we return to the mRNA bursting example given by the reactions in (4.12), $0 \xrightarrow{\alpha} rM$, $M \xrightarrow{\beta} 0$, for which (cf. (4.13)-(4.14)) $G = (r, -1)$, $\rho_1^\sigma(k) = \alpha$, $\rho_2^\sigma(k) = \beta k$, $f^\sigma(k) = r\alpha - \beta k$. Since the

[28] Related to the CV, but not dimensionless, is the "Fano factor" defined as $\frac{\sigma^2(X)}{\mathbb{E}[X]}$.

reactions are of order ≤ 1, the FD formula is exact. We have that $J = -\beta$ and $B(\mu) = \alpha r^2 + \beta\mu$, so that:

$$\dot{\mu} = f(\mu) = \alpha r - \beta\mu \tag{4.54a}$$
$$\dot{\Sigma} = -2\beta\Sigma + B(\mu) = -2\beta\Sigma + \alpha r^2 + \beta\mu\,. \tag{4.54b}$$

In particular, at steady state we have:

$$\mu = \frac{\alpha r}{\beta} = \lambda r$$
$$\Sigma = \frac{\alpha r^2 + \beta\frac{\alpha r}{\beta}}{2\beta} = \frac{\alpha r^2 + \alpha r}{2\beta} = \lambda\frac{r(r+1)}{2}$$

where we again denote $\lambda = \frac{\alpha}{\beta}$. Thus,

$$\mathbf{cv}\,[M]^2 = \lambda\frac{r(r+1)}{2}\Big/\lambda^2 r^2 = \frac{r+1}{2r}\frac{1}{\lambda}$$

which specializes to $1/\lambda$ in the Poisson case (no bursting, $r = 1$). Note that noise, as measured by the CV, is lower when r is higher, but never lower than $1/2$ of the Poisson rate.

This example is a typical one in which experimental measurement of means (or of the deterministic model) does not allow one to identify a parameter (r in this case), but the parameter can be identified from other statistical information: r (as well as λ) can be recovered from μ and Σ.

Next, we return to the dimerization example given by the reactions in (4.15), $0\xrightarrow{\alpha}A$, $A + A\xrightarrow{\beta}0$, for which (cf. (4.16)-(4.17)) $G = (1, -2)$, $\rho_1^\sigma(k) = \alpha$, $\rho_2^\sigma(k) = \frac{\beta k(k-1)}{2}$, $f^\sigma(k) = \alpha + \beta k - \beta k^2$. Some reactions are now of order 2, and the FD formula is not exact. In fact,

$$\dot{\mu} = \mathbb{E}\left[f^\sigma(X(t))\right] = \alpha + 2\beta m - \beta\mathbb{E}\left[X(t)^2\right] = \alpha + \beta\mu - \beta(\Sigma + \mu^2) = \alpha + \beta\mu - \beta\mu^2 - \beta\Sigma$$

shows that the mean depends on the variance

Exercise. Obtain an equation for $\dot{\Sigma}$ (which will depend on moments of order three).

Finally, we study in some detail the *transcription/translation* model (4.6)-(4.18):

$$0\xrightarrow{\alpha}M\xrightarrow{\beta}0\,,\quad M\xrightarrow{\theta}M+P\,,\quad P\xrightarrow{\delta}0\,.$$

We had from (4.19)-(4.20) that

$$\Gamma = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}\,,\quad \rho_1^\sigma(k) = \alpha\,,\ \rho_2^\sigma(k) = \beta k_1\,,\ \rho_3^\sigma(k) = \theta k_1\,,\ \rho_4^\sigma(k) = \delta k_2\,.$$

and (writing "(M, P)" instead of $k = (k_1, k_2)$):

$$f^\sigma(M,P) = \begin{pmatrix} \alpha - \beta M \\ \theta M - \delta P \end{pmatrix}\,.$$

Since *all reactions are of order at most one, the FD formula is exact.* There are 5 differential equations: 2 for the means and 3 (omiting one by symmetry) for the covariances. For means we have:

$$\dot{\mu}_M = \alpha - \beta\mu_M \tag{4.55a}$$
$$\dot{\mu}_P = \theta\mu_M - \delta\mu_P\,. \tag{4.55b}$$

Now, using the formula $\mathbb{E}\left[B(X(t))\right] = \Gamma\,\mathrm{diag}\left(\mathbb{E}\left[\rho_1^\sigma(X(t))\right],\ldots,\mathbb{E}\left[\rho_m^\sigma(X(t))\right]\right)\Gamma'$ (see (4.39)) for the expectation of the diffusion term, we obtain that $B(\mu)$ equals:

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \alpha & & & \\ & \beta\mu_M & & \\ & & \theta\mu_M & \\ & & & \delta\mu_P \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} \alpha+\beta\mu_M & 0 \\ 0 & \theta\mu_M+\delta\mu_P \end{pmatrix}.$$

Also,

$$J \;=\; \text{Jacobian of } \begin{pmatrix} \alpha-\beta M \\ \theta M-\delta P \end{pmatrix} \;=\; \begin{pmatrix} -\beta & 0 \\ \theta & -\delta \end{pmatrix}.$$

It follows that the variance part of the FD equation

$$\dot\Sigma \;=\; \Sigma J' + J\Sigma + B$$

is (omitting the symmetric equation for Σ_{PM}):

$$\dot\Sigma_{MM} \;=\; -2\beta\,\Sigma_{MM} + \alpha + \beta\mu_M \tag{4.56a}$$

$$\dot\Sigma_{PP} \;=\; -2\delta\,\Sigma_{PP} + 2\,\theta\Sigma_{MP} + \theta\mu_M + \delta\mu_P \tag{4.56b}$$

$$\dot\Sigma_{MP} \;=\; \theta\,\Sigma_{MM} - (\beta+\delta)\,\Sigma_{MP}\,. \tag{4.56c}$$

In particular, at steady state we have the following mean number of proteins:

$$\mu_P \;=\; \frac{\alpha\theta}{\beta\delta} \tag{4.57}$$

and the following squared coefficient of variation for protein numbers:

$$\mathbf{cv}\left[P\right]^2 \;=\; \frac{\Sigma_{PP}}{\mu_P^2} \;=\; \frac{(\theta+\beta+\delta)\beta\delta}{\alpha\theta(\beta+\delta)} \;=\; \frac{1}{\mu_P} + \frac{1}{\mu_M}\frac{\delta}{\beta+\delta}\,. \tag{4.58}$$

Exercise. Prove the above formula for the CV. Show also that $\Sigma_{MP} = \frac{\theta\alpha}{\beta(\beta+\delta)}$.

The first term in (4.58) is usually referred to as the "*intrinsic noise*" of transcription, in the sense that this is what the cv would be, if M was constant (so that P would be a Poisson process).

The second is term is usually referred to as the "*extrinsic noise*" of transcription, due to mRNA variability.

Notice that the total noise is bounded from below by the intrinsic noise, and from above by the sum of the intrinsic noise and the mRNA noise, in the following sense:

$$\frac{1}{\mu_P} \;\le\; \mathbf{cv}\left[P\right]^2 \;\le\; \frac{1}{\mu_P} + \frac{1}{\mu_M}$$

(the second inequality because $\frac{\delta}{\beta+\delta} < 1$).

Also, note that even if the mean protein number $\mu_P \gg 1$, the second term, $\frac{1}{\mu_M}\frac{\delta}{\beta+\delta}$, may be large, so that extrinsic noise may dominate even in "large" systems.

Moreover, even accounting for much faster mRNA than protein degradation: $\beta \gg \delta$, which implies $\frac{\delta}{\beta+\delta} \ll 1$, this term may well be large if $\mu_M \ll 1$.

Yet another way to rewrite the total protein noise is as follows:

$$\mathbf{cv}\,[P]^2 \;=\; \frac{1}{\mu_P}\left[1+\frac{b}{1+\eta}\right]$$

where $\eta = \frac{\theta}{\beta}$ is the ratio of mRNA to protein lifetimes, and $b = \theta/\beta$ is the *burst factor* of the translation/transcription process. The number η is typically very small, in which case we have the approximation $\mathbf{cv}\,[P]^2 \approx \frac{1+b}{\mu_P}$. Since b is typically much larger than one, this means that the noise in P is much larger than would be expected for a Poisson random variable ($1/\mu_P$).[29]

Exercise. Give an argument to justify why the burst factor may be thought of as the average number of proteins produced per transcript (i.e, during an mRNAs' lifetime). (The argument will be similar to the one used in the context of epidemics.)

[29] According to M. Thattai and A. Van Oudenaarden, "Intrinsic noise in gene regulatory networks," Proc. Natl Acad. Sci. USA 98, 8614-8619, 2001, which is one of the foundational papers in the field, 'typical values for b are 40 for lacZ and 5 for lacI'.

4.8 Conservation laws and stoichiometry

Suppose that $\nu \in \ker \Gamma'$, i.e. its transpose ν' is in the left nullspace of the stoichiometry matrix Γ, $\nu'\Gamma = 0$. For differential equation models of chemical reactions, described as $\dot{x} = \Gamma R(x)$, it is clear that $\nu' x(t)$ is constant, because $d(\nu' x)/dt = \nu'\Gamma R(x) = 0$. A similar invariance property holds for solutions of the CME. The basic observation is as follows.

Suppose that $\nu'\gamma_j = 0$ and that $c \in \mathbb{Z}$. Then

$$\sum_{\nu' k = c} \rho_j^\sigma(k - \gamma_j)\, p_{k-\gamma_j}(t) \;=\; \sum_{\nu' k = c} \rho_j^\sigma(k)\, p_k(t)\,,$$

where the sums are being taken over all those $k \in \mathbb{Z}^n_{\geq 0}$ such that $\nu' k = c$ (recall the convention that $\rho_j^\sigma(\ell) = 0$ if $\ell \notin \mathbb{Z}^n_{\geq 0}$). This is clear by a change of variables $\ell = k - \gamma_j$, since $\nu' k = c$ if and only if $\nu(k - \gamma_j) = 0$.

Therefore, for any $\nu \in \ker \Gamma'$ it follows that (dropping arguments t):

$$\frac{d}{dt} \sum_{\nu' k = c} \dot{p}_k \;=\; \sum_{j=1}^{m} \sum_{\nu' k = c} \left[\rho_j^\sigma(k - \gamma_j)\, p_{k-\gamma_j} - \rho_j^\sigma(k)\, p_k\right] \;=\; \sum_{j=1}^{m} 0 \;=\; 0\,.$$

So $\sum_{\nu' k = c} p_k(t)$ is constant.

Suppose that the initial state $X(0)$ is known to satisfy $\nu' X(0) = c$. In other words, $\sum_{\nu' k = c} p_k(0) = 1$. It then follows that $\sum_{\nu' k = c} p_k(t) = 1$, which means that, with probability one, $\nu' X(t) = c$ for each $t \geq 0$. This invariance property is an analogue of the one for deterministic systems.

The limit $\pi = p(\infty)$, if it exists, of the distribution vector satisfies the constraint $\sum_{\nu' k = c} \pi_k = 1$. This constraint depends on the initial conditions, through the number c. Steady-state solutions of the CME are highly non-unique when there are conservation laws. To deal with this problem, the usual approach consists of reducing the space by expressing redundant species in terms of a subset of "independent" species, as follows.

Consider a basis $\nu_1, \ldots, \nu_s$ of $\ker \Gamma'$. If it is known that $\nu_i' X(0) = c_i$ for $i = 1, ..., s$, then the above argument says that $\sum_{\nu_i' k = c_i} p_k(t) = 1$ and $\nu_i' X(t) = c_i$ for each $i = 1, ..., s$ and each $t \geq 0$. This fact typically allows one to reduce the Markov chain to a smaller subset.

The simplest example is that of the reaction network

$$A \xrightarrow{\mu} B\,, \quad B \xrightarrow{\nu} A\,,$$

for which we have:

$$\Gamma \;=\; \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}\,, \quad \rho_1^\sigma(k) = \mu k_1\,, \;\; \rho_2^\sigma(k) = \nu k_2\,.$$

We pick $s = 1$ and $\nu = (1, 1)'$.

Suppose that, initially, there is just one unit of A, that is, $X(0) = (A(0), B(0))' = (1, 0)'$. Thus $\nu' X(0) = 1$, from which it follows that $A(t) + B(t) = \nu' X(t) = 1$ for all $t \geq 0$ (with probability one), or equivalently, that $\sum_{k_1 + k_2 = 1} p_k(t) = 1$ for all $t \geq 0$.

Since $p_k(t) = 0$ if either $k_1 < 0$ or $k_2 < 0$, this amounts to saying that $p_{(1,0)'}(t) + p_{(0,1)'}(t) = 1$ for all $t \geq 0$, and $p_k(t) = 0$ for all other k.

If we are only interested in the initial condition $X(0) = (1,0)'$, there is no need to compute $p_k(t)$ except for these two k's. The finite Markov chain with the two states $(1,0)'$ and $(0,1)'$ carries all the information that we care for. Moreover, since $p_{(0,1)'}(t) = 1 - p_{(1,0)'}(t)$, it is enough to consider the differential equation for $p(t) = p_{(1,0)'}(t)$:

$$\begin{aligned}\dot{p} \;=\; & \rho_1^\sigma\left(\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}-1\\1\end{pmatrix}\right)p_{\left(\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}-1\\1\end{pmatrix}\right)} \;+\; \rho_2^\sigma\left(\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}1\\-1\end{pmatrix}\right)p_{\left(\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}1\\-1\end{pmatrix}\right)} \\ & -\rho_1^\sigma\begin{pmatrix}1\\0\end{pmatrix}p_{\begin{pmatrix}1\\0\end{pmatrix}} \;-\; \rho_2^\sigma\begin{pmatrix}1\\0\end{pmatrix}p_{\begin{pmatrix}1\\0\end{pmatrix}}\,.\end{aligned}$$

Since $\rho_1^\sigma(2,-1)' = \rho_2^\sigma(1,0)' = 0$, $\rho_2^\sigma(0,1)' = \nu$, and $\rho_1^\sigma(1,0)' = \mu$ and $p_{(0,1)'} = 1 - p$, we conclude that

$$\dot{p} \;=\; (1-p)\nu \;-\; p\mu\,, \quad p(0) = 1\,,$$

so

$$p(t) \;=\; \frac{\nu}{\mu+\nu} \;+\; e^{-(\mu+\nu)t}\left(1 - \frac{\nu}{\mu+\nu}\right)\,.$$

In particular, at steady state,

$$p_{\begin{pmatrix}1\\0\end{pmatrix}}(\infty) \;=\; \frac{\nu}{\mu+\nu}\,, \quad p_{\begin{pmatrix}0\\1\end{pmatrix}}(\infty) \;=\; \frac{\mu}{\mu+\nu}\,,$$

i.e., the steady-state distribution is Bernoulli with parameter $\frac{\nu}{\mu+\nu}$.

Exercise. Suppose that, in the reaction network $A \xrightarrow{\mu} B$, $B \xrightarrow{\nu} A$, we know that initially, there are just r units of A, that is, $X(0) = (A(0), B(0))' = (r,0)'$. Show how to reduce the CME to a Markov chain on $s+1$ states, and that the steady-state probability distribution is a binomial distribution.

Exercise. The example of $A \xrightarrow{\mu} B$, $B \xrightarrow{\nu} A$ with $X(0) = (A(0), B(0))' = (r,0)'$ can be thought of as follows: A is the inactive form of a gene, and B is its active form. There are a total of r copies of the same gene, and the activity of each switches randomly and independently. Suppose that we now consider transcription and translation, where transcription is only possible when one of these copies of the gene is active. This leads to the following system:

$$A \xrightarrow{\mu} B\,,\; B \xrightarrow{\nu} A\,,\; B \xrightarrow{\alpha} M + B\,,\; M \xrightarrow{\beta} 0\,,\; M \xrightarrow{\theta} M + P\,,\; P \xrightarrow{\delta} 0\,.$$

1. Write down the CME for this system.
2. Assuming only one copy of the gene, $r = 1$, compute (using the FD method or generating functions) the steady-state mean and standard deviation of M.
3. Optional (very tedious computation): again with $r = 1$, use the FD formula to compute the steady-state mean and standard deviation of P.
4. Optional: repeat the calculations with an arbitrary copy number r.

4.9 Relations to deterministic equations, and approximations

In this section, we briefly discuss various additional topics, in an informal fashion. All propensities are mass-action type now.

4.9.1 Deterministic chemical equations

The mean of the state $X(t)$ satisfies the differential equation (4.29): $\frac{d}{dt}\mathbb{E}[X(t)] = \mathbb{E}[f^\sigma(X(t))]$. This suggests the approximation

$$\frac{d}{dt}\mathbb{E}[X(t)] \;\approx\; f^\sigma(\mathbb{E}[X(t)])\,, \tag{4.59}$$

which is an equality when the reactions have order 0 or 1. This would also be an equality of the variability of $X(t)$ were small. However, in general, the variance of $X(t)$ is large, of the order of the volume Ω in which the reaction takes place, as we discuss later.

On the other hand, if we consider the *concentration* $Z(t) = X(t)/\Omega$, this quantity has variance of order $\Omega/\Omega^2 = 1/\Omega$. So, for concentrations, and assuming that Ω is large, it makes sense to expect that the analog of (4.59) will be very accurate.

Now, to get a well-defined meaning of concentrations $Z(t) = X(t)/\Omega$ as $\Omega \to \infty$, $X(t)$ must also be very large. (Since otherwise $Z(t) = X(t)/\Omega \approx 0$.) This is what one means by a "thermodynamic limit" in physics.

What equation is satisfied by $\mathbb{E}[Z(t)]$? To be precise, let us consider the stochastic process $Z(t) = \frac{X(t)}{\Omega}$ that describes concentrations as opposed to numbers of units. Equation (4.29) said that $\frac{d}{dt}\mathbb{E}[X(t)] = \mathbb{E}[f^\sigma(X(t))]$. Therefore,

$$\frac{d}{dt}\mathbb{E}[Z(t)] \;=\; \frac{1}{\Omega}\frac{d}{dt}\mathbb{E}[X(t)] \;=\; \frac{1}{\Omega}\mathbb{E}[f^\sigma(X(t))] \;=\; \mathbb{E}\left[\frac{1}{\Omega}f^\sigma(\Omega Z(t))\right]. \tag{4.60}$$

The numbers $Z(t)$, being concentrations, should be expected to satisfy some sort of equation that does not in any way involve volumes. Thus, we want to express the right-hand side of (4.60) in a way that does not involve Ω-dependent terms. Unfortunately, this is not possible without appealing to an approximation. To illustrate the problem, take a homodimerization reaction, which will contribute terms of the form $\frac{1}{\Omega}k(k-1)$ to the vector field f^σ. Then the right-hand side of (4.60) will involve an expression

$$\frac{1}{\Omega^2}\left(\Omega Z(t)\right)\left(\Omega Z(t)-1\right) \;=\; \left(\frac{\Omega Z(t)}{\Omega}\right)\left(\frac{\Omega Z(t)}{\Omega}\right)\left(1-\frac{1}{Z(t)\Omega}\right) \;=\; Z(t)^2\left(1-\frac{1}{Z(t)\Omega}\right)$$

Thus, we need to have $Z(t)\,\Omega \gg 1$ in order to eliminate Ω-dependence. This is justified provided that $\Omega \to \infty$ and $Z(t) \not\to 0$. More generally, the discussion is as follows.

The right-hand side of (4.60) involves $\frac{1}{\Omega}f^\sigma$, which is built out of terms of the form $\frac{1}{\Omega}\rho_j^\sigma$, where the propensities for mass-action kinetics are $\rho_j^\sigma(k) = \frac{c_j}{\Omega^{A_j-1}}\binom{k}{a_j}$ for each $j \in \{1,\dots,m\}$.

The combinatorial numbers $\binom{k}{a_j} = \prod_{i=1}^n \binom{k_i}{a_{ij}}$ can be approximated as follows. For each $j \in \{1,\dots,m\}$, using the notation $a_j! = \prod_{i=1}^n a_{ij}!$, we have:

$$\binom{k}{a_j} \;=\; \frac{k^{a_j}}{a_j!}\left[1+O\left(\frac{1}{k}\right)\right]. \tag{4.61}$$

For example, since

$$\begin{aligned}\binom{k_1}{3} &= \frac{1}{3!}\,k_1(k_1-1)(k_1-2) \;=\; \frac{k_1^3}{3!}\left[1+\frac{1}{k_1}P\left(\frac{1}{k_1}\right)\right]\\ \binom{k_2}{2!} &= \frac{1}{2}\,k_2(k_2-1) \;=\; \frac{k_2^2}{2!}\left[1+\frac{1}{k_2}\,Q\right]\end{aligned}$$

with $P(x) = -3+2x$, and $Q=-1$, then if $n=2$ and $a_j=(3,2)'$ (that is, the reaction $\mathcal{R}_j$ consumes three units of S_1 and two of S_2), we have that

$$\binom{k_1}{3}\times\binom{k_2}{2} \;=\; \frac{k_1^3k_2^2}{3!2!}\left[1+\frac{1}{k_1}P+\frac{1}{k_2}Q+\frac{1}{k_1k_2}PQ\right]\,.$$

Let us introduce the following functions:

$$\rho_j^c(s) \;=\; \frac{c_j}{a_j!}\,s^{a_j}$$

where, for each $s\in\mathbb{R}^n_{\geq 0}$ with components s_i,

$$s^{a_j} = \prod_{i=1}^n s_i^{a_{ij}}$$

(with the convention that $s^0=1$ for all s).

Observe that, with our notations,

$$\frac{k^{a_j}}{\Omega_j^A} \;=\; \left(\frac{k}{\Omega}\right)^{a_j}\,.$$

So, we consider the approximation:

$$\frac{1}{\Omega}\rho_j^\sigma(k) \;=\; \frac{c_j}{\Omega^{A_j}}\binom{k}{a_j} \;=\; \frac{c_j}{\Omega^{A_j}}\frac{k^{a_j}}{a_j!}\left[1+O\left(\frac{1}{k}\right)\right] \;=\; \rho_j^c\left(\frac{k}{\Omega}\right)\left[1+O\left(\frac{1}{k}\right)\right] \;\approx\; \rho_j^c\left(\frac{k}{\Omega}\right)\,, \tag{4.62}$$

which is valid if

both $k\to\infty$ and $\Omega\to\infty$ in such a way that the ratio k/Ω remains constant.

This type of limit is often referred to as a "thermodynamic limit". It is interpreted as saying that both the copy numbers and volume are large, but the *concentrations* or *densities* are not. Another way to think of this is by thinking of a larger and larger volume in which a population of particles remains at constant density (so that the number of particles scales like the volume). For purposes of this discussion, let us just agree to say that "in the thermodynamic approximation" will mean whenever the approximation has been performed.

Recall from Equation (4.30) that $f^\sigma(k)=\Gamma R^\sigma(k)$, where $R^\sigma(k)=(\rho_1^\sigma(k),\ldots,\rho_m^\sigma(k))'$ and Γ is the stoichiometry matrix. Let $R(x)$ be defined for any non-negative real vector s as follows:

$$R(x) \;:=\; (\rho_1^c(s),\ldots,\rho_m^c(s))' \tag{4.63}$$

and let

$$f(s) \;:=\; \Gamma\,R(s)\,. \tag{4.64}$$

Under the thermodynamic approximation for (4.62),

$$\frac{1}{\Omega} f^\sigma(k) \;\approx\; f\left(\frac{k}{\Omega}\right) .$$

That is to say, (4.60) becomes

$$\frac{d}{dt}\mathbb{E}\left[Z(t)\right] \;\approx\; \mathbb{E}\left[f(Z(t))\right] . \tag{4.65}$$

We achieved our goal of writing an (approximate) expression that is volume-independent, for the rate of change of the mean concentration

Provided that the variance of $Z(t)$ is small compared to its mean, then we may approximate $\mathbb{E}\left[f(Z(t))\right] \approx f(\mathbb{E}\left[Z(t)\right])$ and write

$$\frac{d}{dt}\mathbb{E}\left[Z(t)\right] \;\approx\; f(\mathbb{E}\left[Z(t)\right]) .$$

This argument motivates the form of the *deterministic chemical reaction equation*[30] which is (using dot for time derivative, and omitting the time argument):

$$\dot{s} \;=\; f(s) \;=\; \Gamma R(s) . \tag{4.66}$$

Observe that we may also write this deterministic equation as an equation on the abundances $x(t)$ of the species, where $x(t) = \Omega s(t)$. The equation is:

$$\dot{x} \;=\; f^{\#}(x) \;=\; \Gamma R^{\#}(x) \tag{4.67}$$

where

$$R^{\#}(x) \;:=\; \left(\rho_1^{\#}(x), \ldots, \rho_m^{\#}(x)\right)' \tag{4.68}$$

and

$$\rho_j^{\#}(x) \;=\; \Omega \rho_j^c\left(\frac{x}{\Omega}\right) \;=\; \frac{c_j}{\Omega^{A_j - 1}} \frac{x^{a_j}}{a_j!} .$$

The only difference with the expression for concentrations is that now there is a denominator which depends on volume.

Both forms of deterministic equations are used in the literature, usually not distinguishing among them. *They both may be written in the same form*, using rates "$\rho(u) = k_j u^{a_j}$" after collecting all constants into k_j, and the only difference is the expression of k_j in terms of the volume. For problems in which the deterministic description is used, and if one is not interested in the stochastic origin of the reaction constants k_j, this is all unimportant. In fact, in practice the coefficients k_j are often estimated by fitting to experimental data, using a least-squares or maximum-likelihood method. In that context, *the physical origin of the coefficients, and their volume dependence or lack thereof, plays no role.*

4.9.2 Unit Poisson representation

We next discuss an integral representation which is extremely useful in the theoretical as well as in the intuitive understanding of the behavior of the process $X(t)$.

[30] Also called a "mean field equation" in physics

To motivate this representation, note first that a vector $x(t)$ is a solution of the deterministic differential equation $\dot{x}(t) = f(x(t))$ with initial condition $x(0) = x_0$ if and only if $x(t) = x_0 + \int_0^t f(x(\tau))\,d\tau$ for all t. This reformulation as an integral equation is merely a statement of the Fundamental Theorem of Calculus, and in fact is a key step in proving existence theorems for differential equations (by looking for fixed points of the operator $x \mapsto x_0 + \int_0^t f(x(\tau))\,d\tau$ in function space).

Specialized to the chemical reaction case, using abundances x, $f(x) = f^{\#}(x) = \Gamma\, R^{\#}(x)$, the integral equation reads:

$$x(t) \;=\; =\; x(0) \;+\; \sum_{j=1}^{m} \gamma_j\, y_j(t)\,, \qquad \text{where } y_j(t) = \int_0^t \rho_j^{\#}(x(\tau))\,d\tau\,. \tag{4.69}$$

The quantity $y_j(t)$ may be thought of as the number of reactions of $\mathcal{R}_j$ that have taken place until time t, because each such reaction adds γ_j to the state. As $\dot{y}_j(t) = \rho_j^{\#}(x(t))$, $\rho_j^{\#}$ can be interpreted as the rate at which the reaction $\mathcal{R}_j$ takes place.

We now turn to the stochastic model. The random state $X(t)$ at time t is obtained from a sequence of jumps:

$$X(t) \;=\; X(0) + W_1 + \ldots + W_N\,.$$

Collecting all the terms W_v that correspond to events in which $\mathcal{R}_j$ fired, and keeping in mind that, every time that the reaction $\mathcal{R}_j$ fires, the state changes by $+\gamma_j$, there results:

$$X(t) \;=\; X(0) \;+\; \sum_{j=1}^{m} \gamma_j\, \widetilde{Y}_j(t)\,, \tag{4.70}$$

where $\widetilde{Y}_j$ counts *how many times* the reaction j has taken place from time 0 until time t. The stochastic Equation (4.70) is a counterpart of the deterministic Equation (4.69). Of course, $\widetilde{Y}_j(t)$ depends on the past history $X(\tau)$, $\tau < t$. The following *Poisson representation* makes that dependence explicit:

$$X(t) \;=\; X(0) \;+\; \sum_{j=1}^{m} \gamma_j\, Y_j\left(\int_0^t \rho_j^{\sigma}(X(\tau))\,d\tau\right)\,, \tag{4.71}$$

where the Y_j's are m independent and identically distributed ("IID") Poisson processes with unit rate. This most beautiful formula is exact and requires no approximations[31]. Here we simply provide an intuitive idea of why one may expect such a formula to hold. The intuitive idea is based on an argument as the one used to derive the SSA.

If $k = X_{v-1}(t_{v-1})$ is the state right after the $(v-1)$st jump, then the *time* until the next jump is given by the variable $\mathcal{T}_k$, which is exponential with the parameter in Equation (4.27), $\lambda_k = \sum_{j=1}^m \rho_j^{\sigma}(k)$. If the state k does not change much, then these distributions do not depend strongly on k, and we can say that reactions occur at times that are separated by an exponentially distributed random variable $\mathcal{T}$ with rate λ. From basic probability theory, we know that this means that the total number of reactions during an interval of length t is Poisson distributed with parameter $t\lambda$. That is to say, there is a Poisson process Y with rate λ that counts how many reactions happen in any given interval.

[31] For details, including proofs, see S.N. Ethier and T.G. Kurtz, *Markov processes: Characterization and convergence*, John Wiley & Sons, New York, 1986.

The random choice of *which* reaction takes place is distributed according to the probabilities

$$\mathbb{P}\left[\text{next reaction is } \mathcal{R}_j\right] \;=\; d_j^{(k)} \;=\; \frac{\rho_j^\sigma(k)}{\sum_{j=1}^m \rho_j^\sigma(k)}\,.$$

If the reaction events form a Poisson process with parameter λ, and if at each time the reaction to be used is picked according to a discrete distribution with $p_j = d_j = \rho_j^\sigma/\lambda$ (we drop "k" since we are assuming that it is approximately constant), then the events "$\mathcal{R}_j$ fires" form a "thinning" of the Poisson stream and hence are known, again from elementary probability theory, to be themselves Poisson distributed, with parameter $d_j\lambda = \rho_j^\sigma$.

This means, putting back the k now, that the number of reactions of type $\mathcal{R}_j$ that occur are distributed according to $\widetilde{Y}_j(t) = Y_j(\rho_j^\sigma(k)t)$, where the Y_j are independent unit Poisson processes[32] (independence also assumes that k is approximately constant during the interval). Now, if we break up a long interval into small intervals of length dt, in each of which we assume that k is constant (somewhat analogous to making an approximation of an integral using a rectangle rule), we have that the total $\widetilde{Y}_j(t)$ is a sum of Poisson random variables, one for each sub-interval, with rates $\rho_j^\sigma(k)dt$. A sum of (independent) Poisson random variables with rates $\mu_1, \ldots, \mu_\nu$ is Poisson with rate $\mu_1 + \ldots + \mu_\nu$, and, as the intervals get smaller, this sum approximates the integral $\int_0^t \rho_j^\sigma(X(\tau))\,d\tau$, if the $\mu_i = \rho_j^\sigma(X(\tau_i))$. This results in the formula (4.71), though of course the argument is not at all rigorous as given.

4.9.3 Diffusion approximation

A *stochastic differential equation* (SDE) is an ordinary differential equation with noise terms in its right-hand side, so that its solution is random.[33] The Markov jump process $X(t)$ is not the solution of an SDE, since by definition, it is discrete-valued.[34] However, there is an SDE whose solutions give a so-called *diffusion approximation* of $X(t)$.[35] The diffusion approximation is useful when numbers of species are "large enough". (But not so large that the equation becomes basically deterministic and so there is no need for stochastics to start with.) It arises as a normal approximation of a Poisson process. We very roughly outline the construction, as follows.

We consider the formula (4.71), which works on any interval $[t, t+h]$:

$$X(t+h) \;=\; X(t) \;+\; \sum_{j=1}^{m} \gamma_j\, Y_j\left(\int_t^{t+h} \rho_j^\sigma(X(\tau))\,d\tau\right)$$

where the Y_j's are IID unit Poisson random processes.

In general, under appropriate conditions ($\lambda \gg 1$), if a variable Y is Poisson with parameter λ, then it is well approximated by a normal random variable N with mean λ and variance λ (this is a special

[32] Saying that Z is a Poisson process with rate λ is the same as saying that $Z(t) = Y(\lambda t)$, where Y is a unit-rate Poisson process.

[33] In physics, SDE's are called *Langevin equations*.

[34] Of course, there is an ODE associated to $X(t)$, namely the CME. But the CME is a *deterministic* differential equation for the *probability distribution* of $X(t)$, not for the sample paths of $X(t)$.

[35] As if things were not confusing enough already, there is yet another (deterministic) differential equation that enters the picture, namely the *Fokker-Planck Equation* (FPE), which, describes the evolution of the probability distribution of the state of the SDE, just like the CME describes the evolution of the probability distribution of the state $X(t)$. The FPE is a PDE (enough acronyms?), because the probability the state of the SDE is a continuous variable, hence requiring a variable for space as well as time.

case of the Central Limit Theorem). Equivalently,

$$Y \;\approx\; \lambda \,+\, \sqrt{\lambda}\, N_0\,,$$

where N_0 is an $\mathcal{N}(0,1)$ random variable.

We make this approximation in the above formula. We denote the random variable N_0 as "$N_j(t)$" to indicate the fact that we have a different one for each j and for each interval $[t, t+h]$ where the approximation is made. Note that, given the initial state $X(t)$, the changes in the interval $[t, t+h]$ are independent of changes in previous intervals; thus the $N_j(t)$ are independent of previous values. Using that $f = \sum_j \gamma_j \rho_j^\sigma$:

$$\begin{aligned} X(t+h) \;&\approx\; X(t) \,+\, \sum_{j=1}^{m} \gamma_j \left[\left(\int_t^{t+h} \rho_j^\sigma(X(\tau))\, d\tau \right) + \sqrt{\left(\int_t^{t+h} \rho_j^\sigma(X(\tau))\, d\tau \right)} N_j(t) \right] \\ &\approx\; X(t) \,+\, f(X(t))h \,+\, \sum_{j=1}^{m} \gamma_j \sqrt{\rho_j^\sigma(X(t))}\, \sqrt{h}\, N_j(t)\,. \end{aligned}$$

The expressions $\sqrt{h}\, N_j(t)$ correspond to increments on time h of a Brownian motion. Thus (dividing by h and letting $h \to 0$), formally we obtain:

$$dX(t) \;\approx\; f(X(t))\, dt \,+\, \sum_{j=1}^{m} \gamma_j \sqrt{\rho_j^\sigma(X(t))}\, B_j(t)$$

where the B_t are independent standard Brownian motion processes.[36]

4.9.4 Relation to deterministic equation

We next sketch why, in the thermodynamic limit, the solution $s(t)$ of the deterministic equation for concentrations provides a good approximation of the mean $\mathbb{E}\,[X(t)]$.

We consider a thermodynamic limit, and let $Z(t) = X(t)/\Omega$. Then:

$$X(t) \;=\; X(0) \,+\, \sum_{j=1}^{m} \gamma_j Y_j \left(\int_0^t \rho_j^\sigma(\Omega Z(\tau))\, ds \right) \;=\; X(0) \,+\, \sum_{j=1}^{m} \gamma_j Y_j \left(\Omega \int_0^t \rho_j^c(Z(\tau))\, ds \right)\,.$$

On any fixed time interval, $Z(\tau)$ is bounded (assuming that there is a well-defined behavior for the densities in the thermodynamic limit), so that the variance of each $Y_j(\ldots)$ is $O(\Omega t)$ (if Y is a unit random process, the variance of $Y(\lambda t)$ is λt), and hence so is the variance of $X(t)$. On a bounded time interval, we may drop the "t" and just say that $\mathbb{V}\mathrm{ar}\,[X(t)] = O(\Omega)$

Now,

$$\frac{d}{dt}\mathbb{E}\,[Z(t)] = \frac{d}{dt}\mathbb{E}\,[X(t)/\Omega] = \frac{1}{\Omega}\frac{d}{dt}\mathbb{E}\,[X(t)] = \frac{1}{\Omega} f^\sigma(\mathbb{E}\,[X(t)]) + \frac{M}{\Omega} \approx f(\mathbb{E}\,[Z(t)]) + \frac{M}{\Omega}$$

[36] For technical reasons, one does not write the derivative form of the equation. The problem is that dB_j/dt is not well-defined as a function because B is highly irregular.

where "M" represents terms that involve central moments of $X(t)$ of order ≥ 2 (recall (4.34)). Moreover, M comes from a Taylor expansion of f^σ, and the nonlinear terms in f^σ (corresponding to all the reactions of order > 1) all have at least a factor $1/\Omega$. Thus, M is of order $O((1/\Omega) \times \mathbb{V}\text{ar}\,[X(t)])$. Since, by the previous discussion, $\mathbb{V}\text{ar}\,[X(t)] = O(\Omega)$, it follows that $M = O(1)$. We conclude that, in the thermodynamic limit,

$$\frac{d}{dt}\mathbb{E}\,[Z(t)] \;\approx\; f(\mathbb{E}\,[Z(t)]) \,+\, O\left(\frac{1}{\Omega}\right) \;\approx\; f(\mathbb{E}\,[Z(t)])\,,$$

which is (with equality) the deterministic equation.

Note, also, that $\mathbb{V}\text{ar}\,[Z(t)] = \frac{1}{\Omega^2}\mathbb{V}\text{ar}\,[X(t)] = O(1/\Omega)$. In other words, the "noise" in concentrations, as measured by their standard deviations, scales as $1/\sqrt{\Omega}$.

We close this section with a citation to a precise theorem of Kurtz[37] that provides one rigorous version of the above arguments. It says roughly that, on each finite time interval $[0, T]$, and for every $\varepsilon > 0$,

$$\mathbb{P}\,[\,\forall\, 0 \leq t \leq T\,,\; |Z(t) - s(t)| < \varepsilon\,] \;\approx\; 1$$

if Ω is large, where $Z = X/\Omega$ and $s(t)$ is the solution of the deterministic equation, assuming that $X(0) = s(0)$ (deterministic initial condition) and that the solution of $s(t)$ exists on this interval. In other words, "almost surely" the sample paths of the process, normalized to concentrations, are almost identical to the solution of the deterministic system. Of course, $X(0) = s(0)$ means $Z(0) = \Omega s(0)$, which makes no sense as $\Omega \to \infty$. So the precise statement is as follows:

Suppose that $X_\Omega(t)$ is a sample path of the process with volume Ω (that is, this is the volume that appears in the propensities), for each Ω. If $\frac{1}{\Omega}X_V(0) \to s(0)$, then:

$$\lim_{\Omega\to\infty} \mathbb{P}\left[\sup_{0\leq t\leq T}\left|\frac{1}{\Omega}X_V(t) - s(t)\right| \geq \varepsilon\right] = 0$$

for all $T \geq 0$ and all $\varepsilon > 0$.

It is important to realize that, on longer time intervals (or we want a smaller error ε, the required Ω might need to be larger.

[37] See the previously cited book. Originally from T.G. Kurtz, "The relationship between stochastic and deterministic models for chemical reactions," *The Journal of Chemical Physics* **57**(1972): 2976-2978.

4.10 Problems for stochastic kinetics

1. Suppose that $p(t)$ satisfies the CME. Show that if $\sum_{k\in\mathbb{Z}^n_{\geq 0}} p_k(0) = 1$ then $\sum_{k\in\mathbb{Z}^n_{\geq 0}} p_k(t) = 1$ for all $t \geq 0$. (Hint: first, using that $\rho^\sigma_j(k-\gamma_j) = 0$ unless $k \geq \gamma_j$, observe that, for each $j \in \{1,\ldots,m\}$:
$$\sum_{k\in\mathbb{Z}^n_{\geq 0}} \rho^\sigma_j(k-\gamma_j)p_{k-\gamma_j} \;=\; \sum_{k\in\mathbb{Z}^n_{\geq 0}} \rho^\sigma_j(k)p_k$$
and use this to conclude that $\sum_{k\in\mathbb{Z}^n_{\geq 0}} p_k(t)$ must be constant. You may use without proof that the derivative of $\sum_{k\in\mathbb{Z}^n_{\geq 0}} p_k(t)$ with respect to time is obtained by term-by-term differentiation.)

2. Show, using induction on k, that, as claimed in the notes, $\pi_k = e^{-\lambda}\frac{\lambda^k}{k!}$, where $\lambda = \frac{\alpha}{\beta}$, solves
$$\alpha\pi_{k-1} + (k+1)\beta\pi_{k+1} - \alpha\pi_k - k\beta\pi_k = 0\,, \quad k = 0,1,2,\ldots$$
(the first term is not there if $k = 0$).

3. Write the CME for the bursting model.

4. Write the CME for the dimerization model.

5. Write the CME for the transcription/translation model. (Remember that now "k" is a vector (k_1, k_2).)

6. This is a problem regarding the SSA.

   Implement the SSA in your favorite programming system (MATLAB, Maple, Mathematica).

   (b) Take the mRNA/protein model described in the notes, pick some parameters, and an initial state; now plot many sample paths, averaging to get means and variances as a function of time, as well as steady state means and variances.

   (c) Compare the latter with the numbers obtained by using theory as described in the notes.

7. Show that an alternative way of writing the diffusion term in the FD equation is as follows:
$$\Gamma\,\mathrm{diag}\left(\mathbb{E}\left[\rho^\sigma_1(X(t))\right],\ldots,\mathbb{E}\left[\rho^\sigma_m(X(t))\right]\right)\,\Gamma'$$
(where "diag $(r_1,\ldots,r_m)$" means a diagonal matrix with entries r_i in the diagonal).

8. Prove that, for the probability generating function P:
$$\left.\frac{\partial^2 P(z,t)}{\partial z_i \partial z_j}\right|_{z=1} = \begin{cases} \mathbb{E}\left[X_i(t)X_i(t)\right] & \text{if } i \neq j \\ \mathbb{E}\left[X_i(t)^2\right] - \mathbb{E}\left[X_i(t)\right] & \text{if } i = j\,. \end{cases}$$

9. For the mRNA example, derive the variance equation from the probability generating function, and show that the same result is obtained as in the notes.

10. For the mRNA example, solve explicitely the FD differential equations shown in the notes. (You may use matrix exponentials and variation of parameters, Laplace transforms, or whatever method you prefer.)

11. For the dimerization example, obtain an equation for $\dot{\Sigma}$ (which will depend on moments of order three).

12. For the transcription/translation example:

    (a) prove this formula for the squared coefficient of variation for protein numbers:

$$\mathbf{cv}\,[P]^2 \;=\; \frac{\Sigma_{PP}}{\mu_P^2} \;=\; \frac{(\theta+\beta+\delta)\beta\delta}{\alpha\theta(\beta+\delta)} \;=\; \frac{1}{\mu_P} + \frac{1}{\mu_M}\frac{\delta}{\beta+\delta}\,.$$

    (b) Show that $\Sigma_{MP} = \frac{\theta\alpha}{\beta(\beta+\delta)}$.

13. Suppose that, in the reaction network $A \xrightarrow{\mu} B$, $B \xrightarrow{\nu} A$, we know that initially, there are just r units of A, that is, $X(0) = (A(0), B(0))' = (r, 0)'$. Show how to reduce the CME to a Markov chain on $s+1$ states, and that the steady-state probability distribution is a binomial distribution.

14. The example of $A \xrightarrow{\mu} B$, $B \xrightarrow{\nu} A$ with $X(0) = (A(0), B(0))' = (r, 0)'$ can be thought of as follows: A is the inactive form of a gene, and B is its active form. There are a total of r copies of the same gene, and the activity of each switches randomly and independently. Suppose that we now consider transcription and translation, where transcription is only possible when one of these copies of the gene is active. This leads to the following system:

$$A \xrightarrow{\mu} B\,,\; B \xrightarrow{\nu} A\,,\; B \xrightarrow{\alpha} M + B\,,\; M \xrightarrow{\beta} 0\,,\; M \xrightarrow{\theta} M + P\,,\; P \xrightarrow{\delta} 0\,.$$

    (a) Write down the CME for this system.

    (b) Assuming only one copy of the gene, $r = 1$, compute (using the FD method or generating functions) the steady-state mean and standard deviation of M.

    (c) Optional (very tedious computation): again with $r = 1$, use the FD formula to compute the steady-state mean and standard deviation of P.

    (d) Optional: repeat the calculations with an arbitrary copy number r.