Biomolecular Feedback Systems

Domitilla Del Vecchio
MIT

Richard M. Murray
Caltech

Version 1.0b, September 14, 2014

This is the electronic edition of *Biomolecular Feedback Systems*, available from
http://www.cds.caltech.edu/~murray/BFSwiki.

Printed versions are available from Princeton University Press,
http://press.princeton.edu/titles/10285.html.

Chapter 4
Stochastic Modeling and Analysis

In this chapter we explore stochastic behavior in biomolecular systems, building on our preliminary discussion of stochastic modeling in Section 2.1. We begin by reviewing methods for modeling stochastic processes, including the chemical master equation (CME), the chemical Langevin equation (CLE) and the Fokker-Planck equation (FPE). Given a stochastic description, we can then analyze the behavior of the system using a collection of stochastic simulation and analysis tools. This chapter makes use of a variety of topics in stochastic processes; readers should have a good working knowledge of basic probability and some exposure to simple stochastic processes.

4.1 Stochastic modeling of biochemical systems

Biomolecular systems are inherently noisy due to the random nature of molecular reactions. When the concentrations of molecules are high, the deterministic models we have used in the previous chapters provide a good description of the dynamics of the system. However, if the molecular counts are low then it is often necessary to explicitly account for the random nature of events. In this case, the chemical reactions in the cell can be modeled as a collection of stochastic events corresponding to chemical reactions between species. These include binding and unbinding of molecules (such as RNA polymerase and DNA), conversion of one set of species into another, and enzymatically controlled covalent modifications such as phosphorylation. In this section we will briefly survey some of the different representations that can be used for stochastic models of biochemical systems, following the material in the textbooks by Phillips et al. [78], Gillespie [32] and Van Kampen [53].

Statistical mechanics

At the core of many of the reactions and multi-molecular interactions that take place inside of cells is the chemical physics associated with binding between two molecules. One way to capture some of the properties of these interactions is through the use of statistical mechanics and thermodynamics.

As described briefly already in Chapter 2, the underlying representation for both statistical mechanics and chemical kinetics is to identify the appropriate mi-

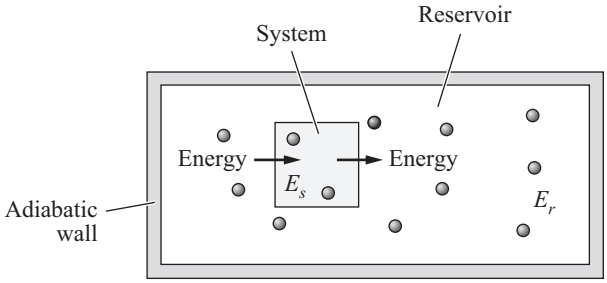

Figure 4.1: System in contact with a reservoir. While there is exchange of energy between the system and the reservoir, there is no exchange of energy between them and the rest of the world. Figure adapted from [78].

crostates of the system. A microstate corresponds to a given configuration of the components (species) in the system relative to each other and we must enumerate all possible configurations between the molecules that are being modeled.

In statistical mechanics, we model the configuration of the cell by the probability that the system is in a given microstate. This probability can be calculated based on the energy levels of the different microstates. Consider a setting in which our system is in contact with a reservoir (Figure 4.1). Let E_r represent the energy in the reservoir, E_s the energy in the system and $E_{\text{tot}} = E_r + E_s$ the total (conserved) energy. Given two different energy levels E_{q_1} and E_{q_2} for the system of interest, let $W_r(E_{\text{tot}} - E_{q_i})$ be the number of possible microstates of the reservoir with energy $E_r = E_{\text{tot}} - E_{q_i}$, $i = 1, 2$. The laws of statistical mechanics state that the ratio of probabilities of being in microstates q_1 and q_2 is given by the ratio of the number of possible states of the reservoir:

$$\frac{\mathbb{P}(E_{q_1})}{\mathbb{P}(E_{q_2})} = \frac{W_r(E_{\text{tot}} - E_{q_1})}{W_r(E_{\text{tot}} - E_{q_2})}. \tag{4.1}$$

Defining the entropy of the reservoir as $S_r = k_B \ln W_r$, where k_B is Boltzmann's constant, we can rewrite equation (4.1) as

$$\frac{W_r(E_{\text{tot}} - E_{q_1})}{W_r(E_{\text{tot}} - E_{q_2})} = \frac{e^{S_r(E_{\text{tot}} - E_{q_1})/k_B}}{e^{S_r(E_{\text{tot}} - E_{q_2})/k_B}}.$$

We now approximate $S_r(E_{\text{tot}} - E_s)$ in a Taylor series expansion around E_{tot}, under the assumption that $E_r \gg E_{q_i}$:

$$S_r(E_{\text{tot}} - E_s) \approx S_r(E_{\text{tot}}) - \frac{\partial S_r}{\partial E} E_s.$$

From the properties of thermodynamics, if we hold the volume and number of molecules constant, then we can define the temperature as

$$\left.\frac{\partial S}{\partial E}\right|_{V,N} = \frac{1}{T}$$

and we obtain

$$\frac{\mathbb{P}(E_{q_1})}{\mathbb{P}(E_{q_2})} = \frac{e^{-E_{q_1}/k_BT}}{e^{-E_{q_2}/k_BT}}.$$

This implies that

$$\mathbb{P}E_q \propto e^{-E_q/(k_BT)}$$

and hence the probability of being in a microstate q is given by

$$\mathbb{P}(q) = \frac{1}{Z} e^{-E_q/(k_BT)}, \tag{4.2}$$

where we have written E_q for the energy of the microstate and Z is a normalizing factor, known as the *partition function*, defined by

$$Z = \sum_{q \in Q} e^{-E_q/(k_BT)}.$$

In many situations we do not care about the specific microstate that a system is in, but rather whether the system is in any one of a number of microstates that all correspond to the same overall behavior of the system. For example, we will often not care whether a specific RNA polymerase is bound to a promoter, but rather whether *any* RNA polymerase is bound to that promoter. We call the collection of microstates that is of interest a *macrostate* (or sometimes *system state*). A macrostate is defined as a set of states $S \subset Q$ that correspond to a given condition that we wish to monitor. Given a macrostate S, the probability of being in that macrostate is

$$\mathbb{P}(S) = \frac{1}{Z} \sum_{q \in S} e^{-E_q/(k_BT)} = \frac{\sum_{q \in S} e^{-E_q/(k_BT)}}{\sum_{q \in Q} e^{-E_q/(k_BT)}}. \tag{4.3}$$

It is this probability that allows us, for example, to determine whether any RNA polymerase molecule is bound to a given promoter, averaged over many independent samples. We can then use this probability to determine the rate of expression of the corresponding gene.

Example 4.1 (Transcription factor binding)**.** Suppose that we have a transcription factor R that binds to a specific target region on a DNA strand (such as the promoter region upstream of a gene). We wish to find the probability P_{bound} that the transcription factor will be bound to this location as a function of the number of transcription factor molecules n_{R} in the system. If the transcription factor is a repressor, for example, knowing $P_{\text{bound}}(n_{\text{R}})$ will allow us to calculate the likelihood of transcription occurring.

To compute the probability of binding, we assume that the transcription factor can bind non-specifically to other sections of the DNA (or other locations in the cell) and we let N_{ns} represent the number of such sites. We let E_{bound} represent the free energy associated with R bound to its specified target region and E_{ns} represent

the free energy for R in any other non-specific location, where we assume that $E_{\text{bound}} < E_{\text{ns}}$. The microstates of the system consist of all possible assignments of the n_{R} transcription factors to either a non-specific location or the target region of the DNA. Since there is only one target site, there can be at most one transcription factor attached there and hence we must count all of the ways in which either zero or one molecule of R are attached to the target site.

If none of the n_{R} copies of R are bound to the target region then these must be distributed between the N_{ns} non-specific locations. Each bound protein has energy E_{ns}, so the total energy for any such configuration is $n_{\text{R}}E_{\text{ns}}$. The number of such combinations is $\binom{N_{\text{ns}}}{n_{\text{R}}}$, assuming the R's are indistinguishable, and so the contribution to the partition function from these microstates is

$$Z_{\text{ns}} = \binom{N_{\text{ns}}}{n_{\text{R}}} e^{-n_{\text{R}}E_{\text{ns}}/(k_BT)} = \frac{N_{\text{ns}}!}{n_{\text{R}}!(N_{\text{ns}}-n_{\text{R}})!} e^{-n_{\text{R}}E_{\text{ns}}/(k_BT)}.$$

For the microstates in which one molecule of R is bound at a target site and the other $n_{\text{R}} - 1$ molecules are at the non-specific locations, we have a total energy of $E_{\text{bound}} + (n_{\text{R}} - 1)E_{\text{ns}}$ and $\binom{N_{\text{ns}}}{(n_{\text{R}}-1)}$ possible such states. The resulting contribution to the partition function is

$$Z_{\text{bound}} = \frac{N_{\text{ns}}!}{(n_{\text{R}}-1)!(N_{\text{ns}}-n_{\text{R}}+1)!} e^{-(E_{\text{bound}}-(n_{\text{R}}-1)E_{\text{ns}})/(k_BT)}.$$

The probability that the target site is occupied is now computed by looking at the ratio of the Z_{bound} to $Z = Z_{\text{ns}} + Z_{\text{bound}}$. After some basic algebraic manipulations, it can be shown that

$$P_{\text{bound}}(n_{\text{R}}) = \frac{\left(\frac{n_{\text{R}}}{N_{\text{ns}}-n_{\text{R}}+1}\right)\exp[-(E_{\text{bound}}+E_{\text{ns}})/(k_BT)]}{1+\left(\frac{n_{\text{R}}}{N_{\text{ns}}-n_{\text{R}}+1}\right)\exp[-(E_{\text{bound}}+E_{\text{ns}})/(k_BT)]}.$$

If we assume that $N_{\text{ns}} \gg n_R$ then $N_{\text{ns}} - n_R + 1 \approx N_{\text{ns}}$, and we can write

$$P_{\text{bound}}(n_{\text{R}}) \approx \frac{kn_{\text{R}}}{1+kn_{\text{R}}}, \quad \text{where} \quad k = \frac{1}{N_{\text{ns}}}\exp[-(E_{\text{bound}}-E_{\text{ns}})/(k_BT)]. \tag{4.4}$$

As we would expect, this says that for very small numbers of repressors, P_{bound} is close to zero, while for large numbers of repressors, $P_{\text{bound}} \to 1$. The point at which we get a binding probability of 0.5 is when $n_{\text{R}} = 1/k$, which depends on the relative binding energies and the number of non-specific binding sites. ∇

Example 4.2 (Combinatorial promoter)**.** As mentioned in Section 2.3, a combinatorial promoter is a region of DNA in which multiple transcription factors can bind and influence the subsequent binding of RNA polymerase (RNAP). Combinatorial promoters appear in a number of natural and engineered circuits and represent a mechanism for creating switch-like behavior.

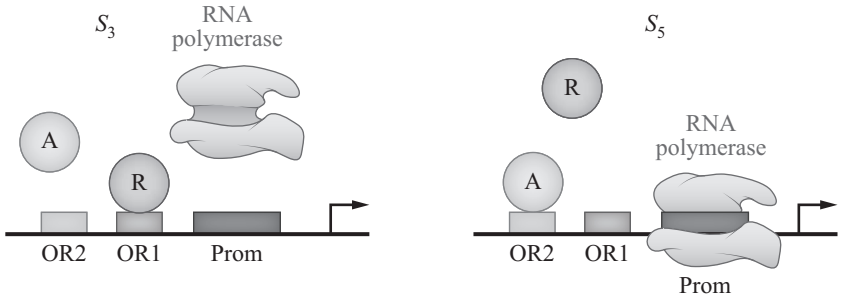

Figure 4.2: Two possible configurations of a combinatorial promoter where both an activator and a repressor can bind to specific operator sites. We show configurations S_3 and S_5 referring to Table 4.1.

One method to model a combinatorial promoter is to use the binding energies of the different combinations of proteins to the operator region, and then compute the probability of being in a given promoter state given the concentration of each of the transcription factors. Table 4.1 shows the possible states of a notional promoter that has two operator regions—one that binds a repressor protein R and another that binds an activator protein A.

As indicated in the table, the promoter has three (possibly overlapping) regions of DNA: OR1 and OR2 are binding sites for the repressor and activator proteins, respectively, and Prom is the location where RNA polymerase binds. (The individual labels are primarily for bookkeeping purposes and may not correspond to physically separate regions of DNA.)

To determine the probabilities of being in a given macrostate, we must compute the individual microstates that occur at given concentrations of repressor, activator and RNA polymerase. Each microstate corresponds to an individual set of molecules binding in a specific configuration. So if we have n_{R} repressor molecules,

Table 4.1: Configurations for a combinatorial promoter with an activator and a repressor. Each row corresponds to a specific macrostate of the promoter in which the listed molecules are bound to the target region. The relative energy of a state compared with the ground state provides a measure of the likelihood of that state occurring, with more negative numbers corresponding to more energetically favorable configurations.

State S_q	OR1	OR2	Prom	E_q (ΔG)	Comment
S_1	–	–	–	0	No binding (ground state)
S_2	–	–	RNAP	−5	RNA polymerase bound
S_3	R	–	–	−10	Repressor bound
S_4	–	A	–	−12	Activator bound
S_5	–	A	RNAP	−15	Activator and RNA polymerase

then there is one microstate corresponding to *each* different repressor molecule that is bound, resulting in n_R individual microstates. In the case of configuration S_5, where two different molecules are bound, the number of combinations is given by the product of the numbers of individual molecules, $n_\mathrm{A} \cdot n_\mathrm{RNAP}$, reflecting the possible combinations of molecules that can occupy the promoter sites. The overall partition function is given by summing up the contributions from each microstate:

$$Z = e^{-E_1/(k_B T)} + n_\mathrm{RNAP}\, e^{-E_2/(k_B T)} + n_\mathrm{R}\, e^{-E_3/(k_B T)} + n_\mathrm{A}\, e^{-E_4/(k_B T)} + n_\mathrm{A} n_\mathrm{RNAP}\, e^{-E_5/(k_B T)}. \quad (4.5)$$

The probability of a given macrostate is determined using equation (4.3). For example, if we define the promoter to be "active" if RNA polymerase is bound to the DNA, then the probability of being in this macrostate as a function of the various molecular counts is given by

$$\begin{aligned} P_\mathrm{active}(n_\mathrm{R}, n_\mathrm{A}, n_\mathrm{RNAP}) &= \frac{1}{Z}\left(n_\mathrm{RNAP}\, e^{-E_2/(k_B T)} + n_\mathrm{A}\, n_\mathrm{RNAP} e^{-E_5/(k_B T)}\right) \\ &= \frac{k_5\, n_\mathrm{A} + k_2}{1 + k_2 + k_3\, n_\mathrm{R} + (k_4 + k_5) n_\mathrm{A}}, \end{aligned}$$

where

$$k_q = e^{-(E_q - E_1)/(k_B T)}.$$

From this expression we see that if $n_\mathrm{R} \gg n_\mathrm{A}$ then P_active tends to 0 while if $n_\mathrm{A} \gg n_\mathrm{R}$ then P_active tends to 1, as expected.

∇

Chemical master equation (CME)

The statistical physics model we have just considered gives a description of the *steady state* properties of the system. In many cases, it is clear that the system reaches this steady state quickly and hence we can reason about the behavior of the system just by modeling the energy of the system. In other situations, however, we care about the transient behavior of a system or the dynamics of a system that does not have an equilibrium configuration. In these instances, we must extend our formulation to keep track of how quickly the system transitions from one microstate to another, known as the *chemical kinetics* of the system.

To model these dynamics, we return to our enumeration of all possible microstates of the system. Let $P(q,t)$ represent the probability that the system is in microstate q at a given time t. Here q can be any of the very large number of possible microstates for the system, which for chemical reaction systems we can represent in terms of a vector consisting of the number of molecules of each species that is present. We wish to write an explicit expression for how $P(q,t)$ varies as a function of time, from which we can study the stochastic dynamics of the system.

We begin by assuming we have a set of M reactions Rj, $j = 1, \ldots, M$, with ξ_j representing the change in state associated with reaction Rj. Specifically, ξ_j is given by the jth column of the stoichiometry matrix N (Section 2.1). The *propensity function* defines the probability that a given reaction occurs in a sufficiently small time step dt:

$$a_j(q,t)dt \quad = \quad \text{Probability that reaction Rj will occur between time } t \text{ and time } t+dt \text{ given that the microstate is } q.$$

The linear dependence on dt relies on the fact that dt is chosen sufficiently small. We will typically assume that a_j does not depend on the time t and write $a_j(q)dt$ for the probability that reaction j occurs in state q.

Using the propensity function, we can compute the distribution of states at time $t+dt$ given the distribution at time t:

$$\begin{aligned} P(q,t+dt) &= P(q,t)\prod_{j=1}^{M}\Big(1-a_j(q)dt\Big) + \sum_{j=1}^{M} P(q-\xi_j)a_j(q-\xi_j)dt \\ &= P(q,t) + \sum_{j=1}^{M}\Big(a_j(q-\xi_j)P(q-\xi_j,t) - a_j(q)P(q,t)\Big)dt + O(dt^2), \end{aligned} \tag{4.6}$$

where $O(dt^2)$ represents higher order terms in dt. Since dt is small, we can take the limit as $dt \to 0$ and we obtain the *chemical master equation* (CME):

$$\frac{\partial P}{\partial t}(q,t) = \sum_{j=1}^{M}\Big(a_j(q-\xi_j)P(q-\xi_j,t) - a_j(q)P(q,t)\Big). \tag{4.7}$$

This equation is also referred to as the *forward Kolmogorov equation* for a discrete state, continuous time random process.

Despite its complexity, the master equation does capture many of the important details of the chemical physics of the system and we shall use it as our basic representation of the underlying dynamics. As we shall see, starting from this equation we can then derive a variety of alternative approximations that allow us to answer specific questions of interest.

The key element of the master equation is the propensity function $a_j(q)$, which governs the rate of transition between microstates. Although the detailed value of the propensity function can be quite complex, its functional form is often relatively simple. In particular, for a unimolecular reaction of the form $\mathrm{A} \to \mathrm{B}$, the propensity function is proportional to the number of molecules of A that are present:

$$a_j(q) = k_j n_A. \tag{4.8}$$

This follows from the fact that the reaction associated with each molecule is independent and hence the likelihood of a reaction happening depends directly on the number of copies of A that are present.

Similarly, for a bimolecular reaction, we have that the likelihood of a reaction occurring is proportional to the product of the number of molecules of each type that are present (since this is the number of independent reactions that can occur) and inversely proportional to the volume Ω. Hence, for a reaction of the form $\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{C}$ we have

$$a_j(q) = \frac{k_j}{\Omega} n_A n_B. \tag{4.9}$$

The rigorous verification of this functional form is beyond the scope of this text, but roughly we keep track of the likelihood of a single reaction occurring between A and B and then multiply by the total number of combinations of the two molecules that can react ($n_{\mathrm{A}} \cdot n_{\mathrm{B}}$).

A special case of a bimolecular reaction occurs when $\mathrm{A} = \mathrm{B}$, so that our reaction is given by $\mathrm{A} + \mathrm{A} \rightarrow \mathrm{B}$. In this case we must take into account that a molecule cannot react with itself and that the molecules are indistinguishable, and so the propensity function is of the form

$$a_j(q) = \frac{k_j}{\Omega} \cdot \frac{n_A(n_A - 1)}{2}. \tag{4.10}$$

Here, $n_A(n_A - 1)/2$ represents the number of ways that two molecules can be chosen from a collection of n_A identical molecules.

Note that the use of the parameter k_j in the propensity functions above is intentional since it corresponds to the reaction rate parameter that is present in the reaction rate equation models we used in Chapter 2. The factor of Ω for bimolecular reactions models the fact that the propensity of a bimolecular reaction occurring depends explicitly on the volume in which the reaction takes place.

Although it is tempting to extend the formula for a bimolecular reaction to the case of more than two species being involved in a reaction, usually such reactions actually involve combinations of bimolecular reactions, e.g.:

$$\mathrm{A} + \mathrm{B} + \mathrm{C} \longrightarrow \mathrm{D} \qquad \Longrightarrow \qquad \mathrm{A} + \mathrm{B} \longrightarrow \mathrm{AB}, \quad \mathrm{AB} + \mathrm{C} \longrightarrow \mathrm{D}.$$

This more detailed description reflects the fact that it is extremely unlikely that three molecules will all come together at precisely the same instant. The much more likely possibility is that two molecules will initially react, followed by a second reaction involving the third molecule.

Example 4.3 (Repression of gene expression)**.** We consider a simple model of repression in which we have a promoter that contains binding sites for RNA polymerase and a repressor protein R. RNA polymerase only binds when the repressor is absent, after which it can undergo an isomerization reaction to form an open complex and initiate transcription (see Section 2.2). Once the RNA polymerase begins to create mRNA, we assume the promoter region is uncovered, allowing another repressor or RNA polymerase to bind.

The following reactions describe this process:

$$\begin{aligned}
\texttt{R1}&: \quad \mathrm{R} + \mathrm{DNA} \longrightarrow \mathrm{R{:}DNA},\\
\texttt{R2}&: \quad \mathrm{R{:}DNA} \longrightarrow \mathrm{R} + \mathrm{DNA},\\
\texttt{R3}&: \quad \mathrm{RNAP} + \mathrm{DNA} \longrightarrow \mathrm{RNAP{:}DNA^c},\\
\texttt{R4}&: \quad \mathrm{RNAP{:}DNA^c} \longrightarrow \mathrm{RNAP} + \mathrm{DNA},\\
\texttt{R5}&: \quad \mathrm{RNAP{:}DNA^c} \longrightarrow \mathrm{RNAP{:}DNA^o},\\
\texttt{R6}&: \quad \mathrm{RNAP{:}DNA^o} \longrightarrow \mathrm{RNAP} + \mathrm{DNA} + \mathrm{mRNA},
\end{aligned}$$

where $\mathrm{RNAP:DNA^c}$ represents the closed complex and $\mathrm{RNAP:DNA^o}$ represents the open complex, and reaction R6 lumps together start of transcription, elongation, mRNA creation, and termination. The states for the system depend on the number of molecules of each species that are present. If we assume that we start with n_R repressors and n_RNAP RNA polymerases, then the possible states (S) for our system are outlined below.

S	DNA	R	RNAP	R : DNA	$\mathrm{RNAP : DNA^c}$	$\mathrm{RNAP : DNA^o}$
q_1	1	n_R	n_RNAP	0	0	0
q_2	0	$n_\mathrm{R}-1$	n_RNAP	1	0	0
q_3	0	n_R	$n_\mathrm{RNAP}-1$	0	1	0
q_4	0	n_R	$n_\mathrm{RNAP}-1$	0	0	1

Note that we do not keep track of each individual repressor or RNA polymerase molecule that binds to the DNA, but simply keep track of whether they are bound or not.

We can now rewrite the chemical reactions as a set of transitions between the possible microstates of the system. Assuming that all reactions take place in a volume Ω, we use the propensity functions for unimolecular and bimolecular reactions to obtain:

$$\begin{aligned}
\texttt{R1}&: \quad \mathrm{q_1} \longrightarrow \mathrm{q_2}; \quad a_1(q_1) = (k_1/\Omega)n_\mathrm{R}, &\qquad \texttt{R4}&: \quad \mathrm{q_3} \longrightarrow \mathrm{q_1}; \quad a_4(q_3) = k_4,\\
\texttt{R2}&: \quad \mathrm{q_2} \longrightarrow \mathrm{q_1}; \quad a_2(q_2) = k_2, &\qquad \texttt{R5}&: \quad \mathrm{q_3} \longrightarrow \mathrm{q_4}; \quad a_5(q_3) = k_5,\\
\texttt{R3}&: \quad \mathrm{q_1} \longrightarrow \mathrm{q_3}; \quad a_3(q_1) = (k_3/\Omega)n_\mathrm{RNAP}, &\qquad \texttt{R6}&: \quad \mathrm{q_4} \longrightarrow \mathrm{q_1}; \quad a_6(q_4) = k_6.
\end{aligned}$$

The chemical master equation can now be written down using the propensity functions for each reaction:

$$\frac{d}{dt}\begin{pmatrix} P(q_1,t)\\ P(q_2,t)\\ P(q_3,t)\\ P(q_4,t)\end{pmatrix} = \begin{pmatrix} -(k_1/\Omega)n_\mathrm{R} - (k_3/\Omega)n_\mathrm{RNAP} & k_2 & k_4 & k_6\\ (k_1/\Omega)n_\mathrm{R} & -k_2 & 0 & 0\\ (k_3/\Omega)n_\mathrm{RNAP} & 0 & -k_4-k_5 & 0\\ 0 & 0 & k_5 & -k_6\end{pmatrix}\begin{pmatrix} P(q_1,t)\\ P(q_2,t)\\ P(q_3,t)\\ P(q_4,t)\end{pmatrix}.$$

The initial condition for the system can be taken as $P(q,0) = (1,0,0,0)$, corresponding to the state q_1. A simulation showing the evolution of the probabilities is shown in Figure 4.3.

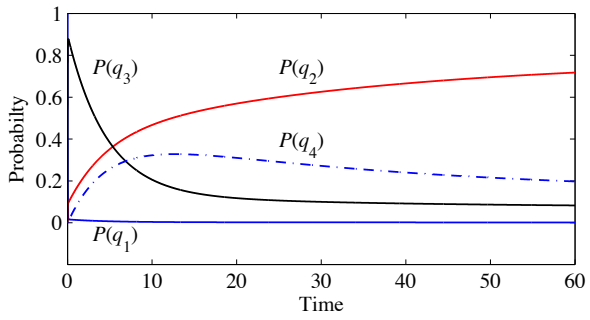

Figure 4.3: Numerical solution of chemical master equation for simple repression model.

The equilibrium solution for the probabilities can be solved by setting $\dot{P} = 0$, which yields:

$$P_e(q_1) = \frac{k_2 k_6 \Omega(k_4 + k_5)}{k_1 k_6 n_{\mathrm{R}}(k_4 + k_5) + k_2 k_3 n_{\mathrm{RNAP}}(k_5 + k_6) + k_2 k_6 \Omega(k_4 + k_5)},$$
$$P_e(q_2) = \frac{k_1 k_6 n_{\mathrm{R}}(k_4 + k_5)}{k_1 k_6 n_{\mathrm{R}}(k_4 + k_5) + k_2 k_3 n_{\mathrm{RNAP}}(k_5 + k_6) + k_2 k_6 \Omega(k_4 + k_5)},$$
$$P_e(q_3) = \frac{k_2 k_3 k_6 n_{\mathrm{RNAP}}}{k_1 k_6 n_{\mathrm{R}}(k_4 + k_5) + k_2 k_3 n_{\mathrm{RNAP}}(k_5 + k_6) + k_2 k_6 \Omega(k_4 + k_5)},$$
$$P_e(q_4) = \frac{k_2 k_3 k_5 n_{\mathrm{RNAP}}}{k_1 k_6 n_{\mathrm{R}}(k_4 + k_5) + k_2 k_3 n_{\mathrm{RNAP}}(k_5 + k_6) + k_2 k_6 \Omega(k_4 + k_5)}.$$

We see that the equilibrium distributions depend on the relative strengths of different combinations of the rate constants for the individual reactions in the system. For example, the probability that a repressor molecule is bound to the promoter is given by

$$P_{\mathrm{bound}}(n_{\mathrm{R}}) = P_e(q_2) = \frac{k_1 k_6 n_{\mathrm{R}}(k_4 + k_5)}{k_1 k_6 n_{\mathrm{R}}(k_4 + k_5) + k_2 k_3 n_{\mathrm{RNAP}}(k_5 + k_6) + k_2 k_6 \Omega(k_4 + k_5)},$$

which has a functional form similar to equation (4.4). Note that here the probability depends on the volume Ω because we used a different model for the diffusion of the repressor R (previously we assumed all repressors were non-specifically bound to DNA).

∇

Example 4.4 (Transcription of mRNA)**.** Consider the production of mRNA from a single copy of DNA. We have two basic reactions that can occur: mRNA can be produced by RNA polymerase transcribing the DNA and producing an mRNA strand, or mRNA can be degraded. We represent the microstate q of the system in terms of the number of mRNA's that are present, which we write as n for ease of

notation. The reactions can now be represented as $\xi_1 = +1$, corresponding to transcription, and $\xi_2 = -1$, corresponding to degradation. We choose as our propensity functions

$$a_1(n) = \alpha, \qquad a_2(n) = \delta n,$$

by which we mean that the probability that a gene is transcribed in time dt is αdt and the probability that a transcript is created in time dt is $\delta n dt$ (proportional to the number of mRNA's).

We can now write down the master equation as described above. Equation (4.6) becomes

$$\begin{aligned} P(n,t+dt) &= P(n,t)\Big(1 - \sum_{i=1,2} a_i(n)dt\Big) + \sum_{i=1,2} P(n-\xi_i,t)a_i(q-\xi_i)dt \\ &= P(n,t) - a_1(n)P(n,t) - a_2(n)P(n,t) \\ &\quad + a_1(n-1)P(n-1,t) + a_2(n+1)P(n+1,t) \\ &= P(n,t) + \alpha P(n-1,t)dt - (\alpha+\delta n)P(n,t)dt + \delta(n+1)P(n+1,t)dt. \end{aligned}$$

This formula holds for $n = 1, 2, \ldots$, with the $n = 0$ case satisfying

$$P(0,t+dt) = P(0,t) - \alpha P(0,t)dt + \delta P(1,t)dt.$$

Notice that we have an infinite number of equations, since n can be any positive integer.

We can write the differential equation version of the master equation by subtracting the first term on the right-hand side and dividing by dt:

$$\begin{aligned} \frac{d}{dt}P(n,t) &= \alpha P(n-1,t) - (\alpha+\delta n)P(n,t) + \delta(n+1)P(n+1,t), \qquad n = 1,2,\ldots \\ \frac{d}{dt}P(0,t) &= -\alpha P(0,t)dt + \delta P(1,t). \end{aligned}$$

Again, this is an infinite number of differential equations, although we could take some limit N and simply declare that $P(N,t) = 0$ to yield a finite number.

One simple type of analysis that can be done on this equation without truncating it to a finite number is to look for a steady state solution to the equation. In this case, we set $\dot{P}(n,t) = 0$ and look for a constant solution $P(n,t) = p_e(n)$. This yields an algebraic set of relations

$$\begin{array}{ll} 0 = -\alpha p_e(0) + \delta p_e(1) & \alpha p_e(0) = \delta p_e(1) \\ 0 = \alpha p_e(0) - (\alpha+\delta)p_e(1) + 2\delta p_e(2) & \alpha p_e(1) = 2\delta p_e(2) \\ 0 = \alpha p_e(1) - (\alpha+2\delta)p_e(2) + 3\delta p_e(3) \quad \Longrightarrow & \alpha p_e(2) = 3\delta p_e(3) \\ \vdots & \vdots \\ 0 = \alpha p_e(n-1) - (\alpha+\delta n)p_e(n) + \delta(n+1)p_e(n+1) & \alpha p(n-1) = n\delta p(n). \end{array}$$

Using this recursive expression to obtain $p(n)$ as a function of $p(0)$, we obtain

$$p(n) = \left(\frac{\alpha}{\delta}\right)^n \frac{1}{n!} p(0).$$

Further, using that $\sum_{n=0}^{\infty} p(n) = 1$, we have that

$$\sum_{n=0}^{\infty} \left(\frac{\alpha}{\delta}\right)^n \frac{1}{n!} p(0) = 1,$$

from which, considering that $\sum_{n=0}^{\infty} \left(\frac{\alpha}{\delta}\right)^n \frac{1}{n!} = e^{\alpha/\delta}$, we obtain $p(0) = e^{-\alpha/\delta}$, which finally leads to the Poisson distribution

$$p(n) = e^{\alpha/\delta} \frac{(\alpha/\delta)^n}{n!}.$$

The mean, variance and coefficient of variation (CV), given by the ratio between the standard deviation and the mean, are thus

$$\mu = \frac{\alpha}{\delta}, \qquad \sigma^2 = \frac{\alpha}{\delta}, \qquad CV = \frac{\sigma}{\mu} = \frac{1}{\sqrt{\mu}} = \sqrt{\frac{\delta}{\alpha}}.$$

The coefficient of variation is commonly used to quantify how noisy a process is since it provides a measure of the deviation relative to the mean value. Note that for fixed variance, the coefficient of variation increases if μ decreases. Thus as we have a small number of mRNA molecules present, we see higher variability in the (relative) mRNA concentration. ∇

Chemical Langevin equation (CLE)

The chemical master equation gives a complete description of the evolution of the probability distribution of a system, but it can often be quite cumbersome to work with directly. A number of approximations to the master equation are thus used to provide more tractable formulations of the dynamics. The first of these that we shall consider is known as the *chemical Langevin equation* (CLE).

To derive the chemical Langevin equation, we start by assuming that the number of molecules in the system is large and that we can therefore represent the system using a vector $X \in \mathbb{R}^n$, with X_i representing the (real-valued) number of molecules of species S_i. (Often X_i will be divided by the volume to give a real-valued concentration of species S_i.) In addition, we assume that we are interested in the dynamics on time scales in which individual reactions are not important and so we can look at how the system state changes over time intervals in which many reactions occur and hence the system state evolves in a smooth fashion.

Let $X(t)$ be the state vector for the system, where we assume now that the elements of X are real-valued rather than integer valued. We make the further approximation that we can lump together multiple reactions so that instead of keeping track of the individual reactions, we can average across a number of reactions over a time τ to allow the continuous state to evolve in continuous time. The resulting dynamics can be described by a stochastic process of the form

$$X_i(t+\tau) = X_i(t) + \sum_{j=1}^{M} \xi_{ij} a_j(X(t))\tau + \sum_{j=1}^{M} \xi_{ij} a_j^{1/2}(X(t)) \mathcal{N}_j(0,\tau),$$

where a_j are the propensity functions for the individual reactions, ξ_{ij} are the corresponding changes in the system states X_i and $\mathcal{N}_j$ are a set of independent Gaussian random variables with zero mean and variance τ.

If we assume that τ is small enough that we can use the derivative to approximate the previous equation (but still large enough that we can average over multiple reactions), then we can write

$$\frac{dX_i(t)}{dt} = \sum_{j=1}^{M} \xi_{ji} a_j(X(t)) + \sum_{j=1}^{M} \xi_{ji} a_j^{1/2}(X(t)) \Gamma_j(t) \quad =: A_i(X(t)) + \sum_{j=1}^{M} B_{ij}(X(t)) \Gamma_j(t), \tag{4.11}$$

where Γ_j are white noise processes (see Section 4.3). This equation is called the *chemical Langevin equation* (CLE).

Example 4.5 (Protein production)**.** Consider a simplified two-step model of protein production in which mRNA is produced by DNA and protein by mRNA. We do not model the detailed processes of isomerization and elongation of the mRNA and polypeptide chains. We can capture the state of the system by keeping track of the number of copies of DNA, mRNA, and protein, which we denote by X_D, X_m and X_P, respectively, so that $X = (X_D, X_m, X_P)$.

The simplified reactions with the corresponding propensity functions are given by

$$\begin{aligned}
&\texttt{R1}: && \text{DNA} \xrightarrow{\alpha} \text{mRNA} + \text{DNA}, && \xi_1 = (0,1,0), && a_1(X) = \alpha\, X_D,\\
&\texttt{R2}: && \text{mRNA} \xrightarrow{\delta} \phi, && \xi_2 = (0,-1,0), && a_2(X) = \delta\, X_m,\\
&\texttt{R3}: && \text{mRNA} \xrightarrow{\kappa} \text{mRNA} + \text{protein}, && \xi_3 = (0,0,1), && a_3(X) = \kappa\, X_m,\\
&\texttt{R4}: && \text{protein} \xrightarrow{\gamma} \phi, && \xi_4 = (0,0,-1), && a_4(X) = \gamma\, X_P.
\end{aligned}$$

Using these, we can write the Langevin equation as

$$\begin{aligned}
\frac{dX_m}{dt} &= \alpha X_D - \delta X_m + \sqrt{\alpha X_D}\Gamma_1(t) - \sqrt{\delta X_m}\Gamma_2(t),\\
\frac{dX_P}{dt} &= \kappa X_m - \gamma X_P + \sqrt{\kappa X_m}\Gamma_3(t) - \sqrt{\gamma X_p}\Gamma_4(t).
\end{aligned}$$

We can keep track of the species concentration by dividing the number of molecules by the volume Ω. Letting $m = X_m/\Omega$, $P = X_P/\Omega$, and $\alpha_0 = \alpha X_D/\Omega$, we obtain the final expression

$$\frac{d}{dt}\begin{pmatrix} m \\ P \end{pmatrix} = \begin{pmatrix} -\delta & 0 \\ \kappa & -\gamma \end{pmatrix}\begin{pmatrix} m \\ P \end{pmatrix} + \begin{pmatrix} \alpha_0 \\ 0 \end{pmatrix} + \frac{1}{\sqrt{\Omega}}\begin{pmatrix} \left(\sqrt{\alpha_0 + \delta m}\right)\Gamma_m \\ \left(\sqrt{\kappa m + \gamma P}\right)\Gamma_P \end{pmatrix},$$

where Γ_m and Γ_P are independent Gaussian white noise processes (note that here we have used that if Γ_1 and Γ_2 are independent identical Gaussian white noise processes, then $\sqrt{a}\Gamma_1 + \sqrt{b}\Gamma_2 = \sqrt{a+b}\Gamma$ with Γ a Gaussian white noise process identical to Γ_i). ∇

The Langevin equation formulation is particularly useful as it allows us to study the stochastic properties of the system by studying how the state responds to a (stochastic) input. Hence, a few of the tools available for studying input/output dynamic behavior can be employed (see Section 3.1, Section 3.2, and Section 4.3).

Fokker-Planck equations (FPE)

The chemical Langevin equation provides a stochastic ordinary differential equation that describes the evolution of the system state. A slightly different (but completely equivalent) representation of the dynamics is to model how the probability distribution $P(x,t)$ evolves in time. As in the case of the chemical Langevin equation, we will assume that the system state is continuous and write down a formula for the evolution of the density function $p(x,t)$. This formula is known as the Fokker-Planck equation (FPE) and is essentially an approximation to the chemical master equation.

Consider first the case of a random process in one dimension. We assume that the random process is in the same form as in the previous section:

$$\frac{dX(t)}{dt} = A(X(t)) + B(X(t))\Gamma(t). \tag{4.12}$$

The function $A(X)$ is called the *drift term* and $B(X)$ is the diffusion term. It can be shown that the probability density function for X, $p(x,t)$, satisfies the partial differential equation

$$\frac{\partial p}{\partial t}(x,t) = -\frac{\partial}{\partial x}(A(x,t)p(x,t)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(B^2(x,t)p(x,t)). \tag{4.13}$$

Note that here we have shifted to the probability density function since we are considering X to be a continuous state random process.

In the multivariate case, more care is required. Using the chemical Langevin equation (4.11), we define

$$D_{ij}(x,t) = \frac{1}{2}\sum_{k=1}^{M} B_{ik}(x,t)B_{jk}(x,t), \qquad i < j = 1,\ldots,M.$$

The Fokker-Planck equation now becomes

$$\frac{\partial p}{\partial t}(x,t) \;=\; -\sum_{i=1}^{M}\frac{\partial}{\partial x_i}(A_i(x,t)p(x,t)) \;+\; \sum_{i=1}^{M}\sum_{j=1}^{M}\frac{\partial^2}{\partial x_i \partial x_j}(D_{ij}(x,t)p(x,t)). \quad (4.14)$$

Note that the Fokker-Planck equation is very similar to the chemical master equation: both provide a description of how the probability distribution varies as a function of time. In the case of the Fokker-Planck equation, we regard the state as a continuous set of variables and we write a partial differential equation for how the probability density function evolves in time. In the case of the chemical master equation, we have a discrete state (microstates) and we write an ordinary differential equation for how the probability distribution (formally the probability mass function) evolves in time. Both formulations contain the same basic information, just using slightly different representations of the system and the probability of being in a given state.

Reaction rate equations (RRE)

As we already saw in Chapter 2, the reaction rate equations can be used to describe the dynamics of a chemical system in the case where there are a large number of molecules whose state can be approximated using just the concentrations of the molecules. We re-derive the results from Section 2.1 here, being more careful to point out what approximations are being made.

We start with the chemical Langevin equation (4.11), which has the form

$$\frac{dX_i(t)}{dt} = \sum_{j=1}^{M}\xi_{ji}a_j(X(t)) + \sum_{j=1}^{M}\xi_{ji}a_j^{1/2}(X(t))\Gamma_j(t).$$

While we have not derived this expression in detail, we note that the first term simply says that the value of the random variable X_i fluctuates according to possible reaction vectors ξ_{ji} scaled by the probability that reaction j occurs in time dt.

We are now interested in how the mean of the concentration X_i evolves. Writing $\langle X_i\rangle$ for the mean (taken over many different samples of the random process), the dynamics of the species at each point in time are given by

$$\frac{d\langle X_i(t)\rangle}{dt} = \sum_{j=1}^{M}\xi_{ji}\langle a_j(X(t))\rangle, \quad (4.15)$$

where the second term in the Langevin equation drops out under the assumption that the Γ_j's are independent processes with zero mean. We see that the reaction rate equations follow by defining $x_i = \langle X_i \rangle / \Omega$ and *assuming* that $\langle a_j(X(t)) \rangle = a_j(\langle X(t) \rangle)$. This relationship is true when a_j is linear (e.g., in the case of a unimolecular reaction), but is an approximation otherwise.

The formal derivation of the reaction rate equations from the chemical master equation and the chemical Langevin equation requires a number of careful assumptions (see the original work of Gillespie [34] for a full derivation). In particular, it requires that the chemical system is well-stirred (no spatial structure), that the molecular counts are sufficiently high that we can approximate concentrations with real numbers, and that the time scales of interest are appropriately long so that multiple individual reactions can be appropriately averaged, and yet at the same time sufficiently short so that we can approximate the derivative through a finite different approximation. As we have noted previously, most biological systems have significant spatial structure (thus violating the well-stirred assumption), but models based on that assumption are still very useful in many settings. The larger molecular count assumption is more critical in using the reaction rate equation and one must take care when molecular counts are in the single digits, for example.

4.2 Simulation of stochastic systems

Suppose that we want to generate a collection of sample trajectories for a stochastic system whose evolution is described by the chemical master equation (4.7):

$$\frac{d}{dt}P(q,t) = \sum_i a_i(q-\xi_i)P(q-\xi_i,t) - \sum_i a_i(q)P(q,t),$$

where $P(q,t)$ is the probability of being in a microstate q at time t (starting from q_0 at time t_0) and $a_i(q)$ is the propensity function for a reaction i starting at a microstate q and ending at microstate $q+\xi_i$. Instead of simulating the distribution function $P(q,t)$, we wish to simulate a specific instance $q(t)$ starting from some initial condition $q_0(t_0)$. If we simulate many such instances of $q(t)$, their distribution at time t should match $P(q,t)$.

The stochastic simulation algorithm

The stochastic simulation algorithm is a Monte Carlo procedure for numerically generating time trajectories of the number of molecules of the various species present in the system in accordance with the chemical master equation.

To illustrate the basic ideas that we will use, consider first a simple birth process in which the microstate is given by an integer $q \in \{0,1,2,\dots\}$ and we assume that the propensity function is given by

$$a(q)\,dt = \lambda\,dt, \qquad \xi = +1.$$

Thus the probability of transition increases linearly with the time increment dt (so birth events occur at rate λ, on average). If we assume that the birth events are independent of each other, then it can be shown that the number of arrivals in time τ is Poisson distributed with parameter $\lambda\tau$:

$$P(q(t+\tau) - q(t) = \ell) = \frac{(\lambda\tau)^\ell}{\ell!} e^{-\lambda\tau},$$

where τ is the difference in time and ℓ is the difference in count q. In fact, this distribution is a joint distribution in time τ and count ℓ. Setting $\ell = 1$, it can be shown that the time to the next reaction, T, follows an exponential distribution and hence has density function

$$p_T(\tau) = \lambda e^{-\lambda\tau}.$$

The exponential distribution has expectation $1/\lambda$ and so we see that the average time between events is inversely proportional to the reaction rate λ.

Consider next a more general case in which we have a countable number of microstates $q \in \{0, 1, 2, \dots\}$ and we let k_{ji} represent the transition probability between a microstate i and microstate j. The birth process is a special case given by $k_{i+1,i} = \lambda$ and all other $k_{ji} = 0$. The chemical master equation describes the joint probability that we are in state $q = i$ at a particular time t. We would like to know the probability that we transition to a new state $q = j$ at time $t + dt$. Given this probability, we can attempt to generate an instance of the variable $q(t)$ by first determining which reaction occurs and then when the reaction occurs.

Let $P(j, \tau) := P(j, t+\tau+d\tau \mid i, t+\tau)$ represent the probability that we transition from the state i to the state j in the time interval $[t+\tau, t+\tau+d\tau]$. For simplicity and ease of notation, we will take $t = 0$. Let $T := T_{j,i}$ be the time at which the reaction first occurs. We can write the probability that we transition to state j in the interval $[\tau, \tau + d\tau]$ as

$$P(j, \tau) = P(T > \tau)\, k_{ji}\, d\tau, \tag{4.16}$$

where $P(T > \tau)$ is the probability that no reaction occurs in the time interval $[0, \tau]$ and $k_{ji} d\tau$ is the probability that the reaction taking state i to state j occurs in the next $d\tau$ seconds (assumed to be independent events, giving the product of these probabilities).

To compute $P(T > \tau)$, define

$$\bar{k}_i = \sum_j k_{ji},$$

so that $(1 - \bar{k}_i) d\tau$ is the probability that no transition occurs from state i in the next $d\tau$ seconds. Then, the probability that no reaction occurs in the interval $[\tau, \tau + d\tau]$ can be written as

$$P(T > \tau + d\tau) = P(T > \tau)(1 - \bar{k}_i)\, d\tau. \tag{4.17}$$

It follows that

$$\frac{d}{d\tau}P(T > \tau) = \lim_{d\tau\to 0} \frac{P(T > \tau + d\tau) - P(T > \tau)}{d\tau} = -P(T > \tau)\,\bar{k}_i.$$

Solving this differential equation, we obtain

$$P(T > \tau) = e^{-\bar{k}_i \tau}, \tag{4.18}$$

so that the probability that no reaction occurs in time τ decreases exponentially with the amount of time that we wait, with rate given by the sum of all the reactions that can occur from state i.

We can now combine equation (4.18) with equation (4.16) to obtain

$$P(j, \tau) = P(j, \tau + d\tau \mid i, 0) = k_{ji}\, e^{-\bar{k}_i \tau}\, d\tau.$$

We see that this has the form of a density function in time and hence the probability that the next reaction is reaction j, independent of the time in which it occurs, is

$$P_{ji} = \int_0^\infty k_{ji} e^{-\bar{k}_i \tau}\, d\tau = \frac{k_{ji}}{\bar{k}_i}. \tag{4.19}$$

Thus, to choose the next reaction to occur from a state i, we choose between N possible reactions, with the probability of each reaction weighted by $k_{ji}/\bar{k}_i$.

To determine the time that the next reaction occurs, we sum over all possible reactions j to get the density function for the reaction time:

$$p_T(\tau) = \sum_j k_{ji} e^{-\bar{k}_i \tau} = \bar{k}_i e^{-\bar{k}_i \tau}.$$

This is the density function associated with an exponential distribution. To compute a time of reaction Δt that draws from this distribution, we note that the cumulative distribution function for T is given by

$$\int_0^{\Delta t} f_T(\tau)\, d\tau = \int_0^{\Delta t} \bar{k}_i e^{-\bar{k}_i \tau}\, d\tau = 1 - e^{-\bar{k}_i \Delta t}.$$

The cumulative distribution function is always in the range $[0, 1]$ and hence we can compute Δt by choosing a (uniformly distributed) random number r in $[0, 1]$ and then computing

$$\Delta t = \frac{1}{\bar{k}_i} \ln \frac{1}{1 - r}. \tag{4.20}$$

(This equation can be simplified somewhat by replacing $1 - r$ with r' and noting that r' can also be drawn from a uniform distribution on $[0, 1]$.)

Note that in the case of a birth process, this computation agrees with our earlier analysis. Namely, $\bar{k}_i = \lambda$ and hence the (only) reaction occurs according to an exponential distribution with parameter λ.

This set of calculations gives the following algorithm for computing an instance of the chemical master equation:

1. Choose an initial condition q at time $t = 0$.
2. Calculate the propensity functions $a_i(q)$ for each possible reaction i.
3. Choose the time for the reaction according to equation (4.20), where $r \in [0, 1]$ is chosen from a uniform distribution.
4. Use a weighted random number generator to identify which reaction will take place next, using the weights in equation (4.19).
5. Update q by implementing the reaction ξ and update the time t by Δt
6. If $T < T_{\text{stop}}$, go to step 2.

This method is sometimes called "Gillespie's direct method" [32, 33], but we shall refer to it here as the "stochastic simulation algorithm" (SSA). We note that the reaction number in step 4 can be computed by calculating a uniform random number on $[0, 1]$, scaling this by the total propensity $\sum_i a_i(q)$, and then finding the first reaction i such that $\sum_{j=0}^{i} a_j(q)$ is larger than this scaled random number.

4.3 Input/output linear stochastic systems

In many situations, we wish to know how noise propagates through a biomolecular system. For example, we may wish to understand how stochastic variations in RNA polymerase concentration affect gene expression. In order to analyze these cases, it is useful to make use of tools from stochastic control theory that allow analysis of noise propagation around a fixed operating point.

We begin with the chemical Langevin equation (4.11), which we can write as

$$\frac{dX(t)}{dt} = A(X(t)) + B(X(t))\Gamma(t).$$

The vector $X(t)$ consists of the individual random variables $X_i(t)$ representing the concentration of species S_i, the functions $A(X(t))$ and $B(X(t))$ are computed from the reaction vectors and propensity functions for the system, and Γ is a set of "white noise" processes. For the remainder of this chapter, we will assume that the function $A(X)$ is linear in X and that $B(X)$ is constant (by appropriately linearizing around the mean state, if needed). We will also rewrite Γ as W, to be more consistent with the literature of stochastic control systems.

Random processes

It will be useful in characterizing the properties of the vector $X(t)$ to treat it as a random process. We briefly review the basic definitions here, primarily to fix the terminology we will use in the rest of the section.

A *continuous-time random process* is a stochastic system characterized by the evolution of a random variable $X(t)$, $t \in [0, T]$. We are interested in understanding

how the (random) state of the system is related at separate times, i.e., how the two random variables $X(t_1)$ and $X(t_2)$ are related. We characterize the state of a random process using a (joint) time-varying probability density function p:

$$\mathbb{P}(\{x_{i,l} \leq X_i(t) \leq x_{i,u}\}) = \int_{x_{1,l}}^{x_{1,u}} \dots \int_{x_{n,l}}^{x_{n,u}} p_{X_1,\dots,X_n}(x;t) dx_n \dots dx_1.$$

Note that the state of a random process is not enough to determine the exact next state, but only the distribution of next states (otherwise it would be a deterministic process). We typically omit indexing of the individual states unless the meaning is not clear from context.

In general, the distributions used to describe a random process depend on the specific time or times that we evaluate the random variables. However, in some cases the relationship only depends on the difference in time and not the absolute times (similar to the notion of time invariance in deterministic systems, as described in Åström and Murray [1]). A process is *stationary* if the distribution is not changing and joint density functions only depend on the differences in times. More formally, $p(x, t+\tau) = p(x,t)$ for all τ, $p(x_i, x_j; t_1+\tau, t_2+\tau) = p(x_i, x_j; t_1, t_2)$, etc. In this case we can write $p(x_i, x_j; \tau)$ for the joint probability distribution. Stationary distributions roughly correspond to the steady state properties of a random process and we will often restrict our attention to this case.

Since each $X(t)$ is a random variable, we can define the mean and variance as $\mu(t)$ and $\sigma^2(t)$ at each time t:

$$\mu(t) := \mathbb{E}(X(t)) = \int_{-\infty}^{\infty} x\, p(x,t)\, dx,$$
$$\sigma^2(t) := \mathbb{E}((X(t) - \mu(t))^2) = \int_{-\infty}^{\infty} (x - \mu(t))^2\, p(x,t)\, dx,$$

where $\mathbb{E}(\cdot)$ is the expected value. To capture the relationship between the current state and the future state, we define the *correlation function* for a random process as

$$\rho(t_1, t_2) := \mathbb{E}(X[t_1]X[t_2]) = \int_{-\infty}^{\infty} x_1 x_2\, p(x_1, x_2; t_1, t_2)\, dx_1 dx_2.$$

These definitions can be extended to the vector case as well:

$$\mathbb{E}(X(t)) = \begin{pmatrix} \mathbb{E}(X_1(t)) \\ \vdots \\ \mathbb{E}(X_n(t)) \end{pmatrix} =: \mu(t),$$

$$\mathbb{E}((X(t) - \mu(t))(X(t) - \mu(t))^T) =$$
$$\begin{pmatrix} \mathbb{E}((X_1(t) - \mu_1(t))(X_1(t) - \mu_1(t))) & \dots & \mathbb{E}((X_1(t) - \mu_1(t))(X_n(t) - \mu_n(t))) \\ & \ddots & \vdots \\ & & \mathbb{E}((X_n(t) - \mu_n(t))(X_n(t) - \mu_n(t))) \end{pmatrix} =: \Sigma(t),$$

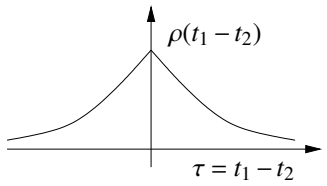

Figure 4.4: Correlation function for a first-order Markov process.

$$\mathbb{E}(X(t)X^T(s)) = \begin{pmatrix} \mathbb{E}(X_1(t)X_1(s)) & \dots & \mathbb{E}(X_1(t)X_n(s)) \\ & \ddots & \vdots \\ & & \mathbb{E}(X_n(t)X_n(s)) \end{pmatrix} =: R(t,s).$$

Note that the random variables and their statistical properties are all indexed by the time t (and s). The matrix $R(t,s)$ is called the *correlation matrix* for $X(t) \in \mathbb{R}^n$. If $t = s$ then $R(t,t)$ describes how the elements of x are correlated at time t (with each other) and in the case that the processes have zero mean, $R(t,t) = \Sigma(t)$. The elements on the diagonal of $\Sigma(t)$ are the variances of the corresponding scalar variables. A random process is uncorrelated if $R(t,s) = 0$ for all $t \neq s$. This implies that $X(t)$ and $X(s)$ are uncorrelated random events and is equivalent to $p_{X,Y}(x,y) = p_X(x)p_Y(y)$.

If a random process is stationary, then it can be shown that $R(t+\tau, s+\tau) = R(t,s)$ and it follows that the correlation matrix depends only on $t-s$. In this case we will often write $R(t,s) = R(s-t)$ or simply $R(\tau)$ where τ is the correlation time. The covariance matrix in this case is simply $R(0)$.

In the case where X is a scalar random process, the correlation matrix is also a scalar and we will write $r(\tau)$, which we refer to as the (scalar) correlation function. Furthermore, for stationary scalar random processes, the correlation function depends only on the absolute value of the correlation time, so $r(\tau) = r(-\tau) = r(|\tau|)$. This property also holds for the diagonal entries of the correlation matrix since $R_{ii}(s,t) = R_{ii}(t,s)$ from the definition.

Example 4.6 (Ornstein-Uhlenbeck process)**.** Consider a scalar random process defined by a Gaussian probability density function with $\mu = 0$,

$$p(x,t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}},$$

and a correlation function given by

$$r(t_1,t_2) = \frac{Q}{2\omega_0} e^{-\omega_0|t_2-t_1|}.$$

The correlation function is illustrated in Figure 4.4. This process is known as an *Ornstein-Uhlenbeck process* and it is a stationary process. ∇

Note on terminology. The terminology and notation for covariance and correlation varies between disciplines. The term covariance is often used to refer to both the relationship between different variables X and Y and the relationship between a single variable at different times, $X(t)$ and $X(s)$. The term "cross-covariance" is used to refer to the covariance between two random vectors X and Y, to distinguish this from the covariance of the elements of X with each other. The term "cross-correlation" is sometimes also used. Finally, the term "correlation coefficient" refers to the normalized correlation $\bar{r}(t,s) = \mathbb{E}(X(t)X(s))/\mathbb{E}(X(t)X(t))$.

We will also make use of a special type of random process referred to as "white noise." A *white noise process* $X(t)$ satisfies $\mathbb{E}(X(t)) = 0$ and $R(t,s) = W\delta(s-t)$, where $\delta(\tau)$ is the impulse function and W is called the *noise intensity*. White noise is an idealized process, similar to the impulse function or Heaviside (step) function in deterministic systems. In particular, we note that $r(0) = \mathbb{E}(X^2(t)) = \infty$, so the covariance is infinite and we never see this signal in practice. However, like the step and impulse functions, it is very useful for characterizing the response of a linear system, as described in the following proposition.

Linear stochastic systems with Gaussian noise

We now consider the problem of how to compute the response of a linear system to a random process. We assume we have a linear system described in state space as

$$\frac{dX}{dt} = AX + FW, \qquad Y = CX. \tag{4.21}$$

For simplicity, we take W and Y to be scalar random variables. Given an "input" W, which is itself a random process with mean $\mu(t)$, variance $\sigma^2(t)$ and correlation $r(t,t+\tau)$, what is the description of the random process Y?

Let W be a white noise process, with zero mean and noise intensity Q:

$$r(\tau) = Q\delta(\tau).$$

We can write the output of the system in terms of the convolution integral

$$Y(t) = \int_0^t h(t-\tau)W(\tau)\,d\tau,$$

where $h(t-\tau)$ is the impulse response for the system

$$h(t-\tau) = Ce^{A(t-\tau)}F.$$

We now compute the statistics of the output, starting with the mean:

$$\begin{aligned}\mathbb{E}(Y(t)) &= \mathbb{E}\Big(\int_0^t h(t-\eta)W(\eta)\,d\eta\Big) \\ &= \int_0^t h(t-\eta)\mathbb{E}(W(\eta))\,d\eta = 0.\end{aligned}$$

Note here that we have relied on the linearity of the convolution integral to pull the expectation inside the integral.

We can compute the covariance of the output by computing the correlation $r_Y(\tau)$ and setting $\sigma_Y^2 = r_Y(0)$. The correlation function for y is

$$\begin{aligned} r_Y(t,s) &= \mathbb{E}(Y(t)Y(s)) = \mathbb{E}\Big(\int_0^t h(t-\eta)W(\eta)\,d\eta \cdot \int_0^s h(s-\xi)W(\xi)\,d\xi\Big) \\ &= \mathbb{E}\Big(\int_0^t \int_0^s h(t-\eta)W(\eta)W(\xi)h(s-\xi)\,d\eta d\xi\Big), \end{aligned}$$

where we assume W is a scalar (otherwise $W(\xi)$ and $h(s-\xi)$ must be transposed). Once again linearity allows us to exchange expectation with the integral and

$$\begin{aligned} r_Y(t,s) &= \int_0^t \int_0^s h(t-\eta)\mathbb{E}(W(\eta)W(\xi))h(s-\xi)\,d\eta d\xi \\ &= \int_0^t \int_0^s h(t-\eta)Q\delta(\eta-\xi)h(s-\xi)\,d\eta d\xi \\ &= \int_0^t h(t-\eta)Qh(s-\eta)\,d\eta. \end{aligned}$$

Now let $\tau = s-t$ and write

$$\begin{aligned} r_Y(\tau) = r_Y(t,t+\tau) &= \int_0^t h(t-\eta)Qh(t+\tau-\eta)\,d\eta \\ &= \int_0^t h(\xi)Qh(\xi+\tau)\,d\xi \qquad \text{(setting } \xi = t-\eta). \end{aligned}$$

Finally, we let $t \to \infty$ (steady state)

$$\lim_{t\to\infty} r_Y(t,t+\tau) = \bar{r}_Y(\tau) = \int_0^\infty h(\xi)Qh(\xi+\tau)d\xi. \tag{4.22}$$

If this integral exists, then we can compute the second-order statistics for the output Y.

We can provide a more explicit formula for the correlation function r in terms of the matrices A, F and C by expanding equation (4.22). We will consider the general case where $W \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ and use the correlation matrix $R(t,s)$ instead of the correlation function $r(t,s)$. Define the *state transition matrix* $\Phi(t,t_0) = e^{A(t-t_0)}$ so that the solution of system (4.21) is given by

$$x(t) = \Phi(t,t_0)x(t_0) + \int_{t_0}^t \Phi(t,\lambda)FW(\lambda)d\lambda.$$

Proposition 4.1 (Stochastic response to white noise)**.** *Let* $\mathbb{E}(X(t_0)X^T(t_0)) = P(t_0)$ *and* W *be white noise with* $\mathbb{E}(W(\lambda)W^T(\xi)) = R_W\delta(\lambda-\xi)$*. Then the correlation matrix for* X *is given by*

$$R_X(t,s) = P(t)\Phi^T(s,t)$$

where $P(t)$ satisfies the linear matrix differential equation

$$\dot{P}(t) = AP + PA^T + FR_W F, \qquad P(0) = P_0.$$

The correlation matrix for the output Y can be computed using the fact that $Y = CX$ and hence $R_Y = CR_X C^T$. We will often be interested in the steady state properties of the output, which are given by the following proposition.

Proposition 4.2 (Steady state response to white noise)**.** *For a time-invariant linear system driven by white noise, the correlation matrices for the state and output converge in steady state to*

$$R_X(\tau) = R_X(t, t+\tau) = Pe^{A^T\tau}, \qquad R_Y(\tau) = CR_X(\tau)C^T$$

where P satisfies the algebraic equation

$$AP + PA^T + FR_W F^T = 0 \qquad P > 0. \tag{4.23}$$

Equation (4.23) is called the *Lyapunov equation* and can be solved in MATLAB using the function `lyap`.

Example 4.7 (First-order system)**.** Consider a scalar linear process

$$\dot{X} = -aX + W, \qquad Y = cX,$$

where W is a white, Gaussian random process with noise intensity σ^2. Using the results of Proposition 4.1, the correlation function for X is given by

$$R_X(t, t+\tau) = p(t)e^{-a\tau}$$

where $p(t) > 0$ satisfies

$$\frac{dp(t)}{dt} = -2ap(t) + \sigma^2.$$

We can solve explicitly for $p(t)$ since it is a (non-homogeneous) linear differential equation:

$$p(t) = e^{-2at}p(0) + (1 - e^{-2at})\frac{\sigma^2}{2a}.$$

Finally, making use of the fact that $Y = cX$ we have

$$r(t, t+\tau) = c^2(e^{-2at}p(0) + (1 - e^{-2at})\frac{\sigma^2}{2a})e^{-a\tau}.$$

In steady state, the correlation function for the output becomes

$$r(\tau) = \frac{c^2\sigma^2}{2a}e^{-a\tau}.$$

Note that the correlation function has the same form as the Ornstein-Uhlenbeck process in Example 4.6 (with $Q = c^2\sigma^2$). ∇

Random processes in the frequency domain

As in the case of deterministic linear systems, we can analyze a stochastic linear system either in the state space or the frequency domain. The frequency domain approach provides a very rich set of tools for modeling and analysis of interconnected systems, relying on the frequency response and transfer functions to represent the flow of signals around the system.

Given a random process $X(t)$, we can look at the frequency content of the properties of the response. In particular, if we let $\rho(\tau)$ be the correlation function for a (scalar) random process, then we define the *power spectral density function* as the Fourier transform of ρ:

$$S(\omega) = \int_{-\infty}^{\infty} \rho(\tau) e^{-j\omega\tau}\, d\tau, \qquad \rho(\tau) \quad = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) e^{j\omega\tau}\, d\tau.$$

The power spectral density provides an indication of how quickly the values of a random process can change through the frequency content: if there is high frequency content in the power spectral density, the values of the random variable can change quickly in time.

Example 4.8 (Ornstein-Uhlenbeck process)**.** To illustrate the use of these measures, consider the Ornstein-Uhlenbeck process whose correlation function we computed in Example 4.7:

$$\rho(\tau) = \frac{Q}{2\omega_0} e^{-\omega_0(\tau)}.$$

The power spectral density becomes

$$\begin{aligned} S(\omega) &= \int_{-\infty}^{\infty} \frac{Q}{2\omega_0} e^{-\omega|\tau|} e^{-j\omega\tau}\, d\tau \\ &= \int_{-\infty}^{0} \frac{Q}{2\omega_0} e^{(\omega - j\omega)\tau}\, d\tau + \int_{0}^{\infty} \frac{Q}{2\omega_0} e^{(-\omega - j\omega)\tau}\, d\tau = \frac{Q}{\omega^2 + \omega_0^2}. \end{aligned}$$

We see that the power spectral density is similar to a transfer function and we can plot $S(\omega)$ as a function of ω in a manner similar to a Bode plot, as shown in Figure 4.5. Note that although $S(\omega)$ has a form similar to a transfer function, it is a real-valued function and is not defined for complex ω. ∇

Using the power spectral density, we can give a more intuitive definition of "white noise" as a zero-mean, random process with power spectral density $S(\omega) =$ constant for all ω. If $X(t) \in \mathbb{R}^n$ (a random vector), then $S(\omega) \in \mathbb{R}^{n \times n}$. We see that a random process is white if all frequencies are equally represented in its power spectral density; this spectral property is the reason for the terminology "white."

Given a linear system

$$\dot{X} = AX + FW, \qquad Y = CX,$$

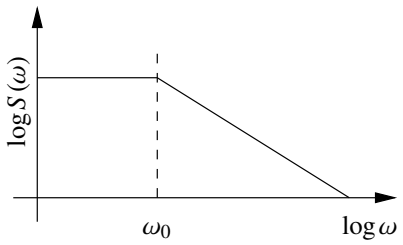

Figure 4.5: Spectral power density for a first-order Markov process.

with W given by white noise, we can compute the spectral density function corresponding to the output Y. Let $H(s) = C(sI - A)^{-1}B$ be the transfer function from W to Y. We start by computing the Fourier transform of the steady state correlation function (4.22):

$$\begin{aligned} S_Y(\omega) &= \int_{-\infty}^{\infty} \left[\int_0^{\infty} h(\xi) Q h(\xi + \tau) d\xi \right] e^{-j\omega\tau} d\tau \\ &= \int_0^{\infty} h(\xi) Q \left[\int_{-\infty}^{\infty} h(\xi + \tau) e^{-j\omega\tau} d\tau \right] d\xi \\ &= \int_0^{\infty} h(\xi) Q \left[\int_0^{\infty} h(\lambda) e^{-j\omega(\lambda - \xi)} d\lambda \right] d\xi \\ &= \int_0^{\infty} h(\xi) e^{j\omega\xi} d\xi \cdot QH(j\omega) = H(-j\omega) Q H(j\omega). \end{aligned}$$

This is then the (steady state) response of a linear system to white noise.

As with transfer functions, one of the advantages of computations in the frequency domain is that the composition of two linear systems can be represented by multiplication. In the case of the power spectral density, if we pass white noise through a system with transfer function $H_1(s)$ followed by transfer function $H_2(s)$, the resulting power spectral density of the output is given by

$$S_Y(\omega) = H_1(-j\omega) H_2(-j\omega) Q_u H_2(j\omega) H_1(j\omega).$$

Exercises

4.1 Consider a standard model of transcription and translation with probabilistic creation and degradation of discrete mRNA and protein molecules. The *propensity functions* for each reaction are as follows:

- Probability of transcribing 1 mRNA molecule: $0.2dt$
- Probability of degrading 1 mRNA molecule: $0.5dt$ and is proportional to the number of mRNA molecules.

- Probability of translating 1 protein: $5dt$ and is proportional to the number of mRNA molecules.
- Probability of degrading 1 protein molecule: $0.5dt$ and is proportional to the number of protein molecules.

In each case, dt will be the time step chosen for your simulation, which we take as $dt = 0.05$ sec.

(i) Simulate the stochastic system above until time $T = 100$. Plot the resulting number of mRNA and protein over time.

(ii) Now assume that the proteins are degraded much more slowly than mRNA and the propensity function of protein degradation is now $0.05dt$. To maintain similar protein levels, the translation probability is now $0.5dt$ (and still proportional to the number of mRNA molecules). Simulate this system as above. What difference do you see in protein level? Comment on the effect of protein degradation rates on noise.

4.2 Compare a simple model of negative autoregulation to one without autoregulation:

$$\frac{dX}{dt} = \beta_0 - \gamma X$$

and

$$\frac{dX}{dt} = \frac{\beta}{1 + X/K} - \gamma X.$$

(i) Assume that the basal transcription rates β and β_0 vary between cells, following a Gaussian distribution with $\sigma/\mu = 0.1$. Simulate time courses of both models for 100 different "cells" using the following parameters: $\beta = 2, \beta_0 = 1, \gamma = 1, K = 1$. Plot the nonregulated and autoregulated systems in two separate plots. Comment on the variation you see in the time courses.

(ii) Calculate the deterministic steady state for both models above. How does variation in the basal transcription rate β or β_0 enter into the steady state? Relate it to what you see in part (i).

4.3 Consider a simple model for gene expression with reactions

$$\phi \xrightarrow{\alpha} \mathrm{m}, \qquad \mathrm{m} \xrightarrow{\kappa} \mathrm{m} + \mathrm{P}, \qquad \mathrm{m} \xrightarrow{\delta} \phi, \qquad \mathrm{P} \xrightarrow{\gamma} \emptyset.$$

Let $\alpha = 1/2$, $\kappa = 20\log(2)/120$, $\delta = \log(2)/120$ and $\gamma = \log(2)/600$, and answer the following questions:

(i) Use the stochastic simulation algorithm (SSA) to obtain realizations of the stochastic process of gene expression and numerically compare with the deterministic ODE solution. Explore how the realizations become close to or apart from the ODE solution when the volume is changed. Determine the stationary probability distribution for the protein (you can do this numerically).

(ii) Now consider the additional binding reaction of protein P with downstream DNA binding sites D:

$$\mathrm{P} + \mathrm{D} \underset{d}{\overset{a}{\rightleftharpoons}} \mathrm{C}.$$

Note that the system is no longer linear due to the presence of a bimolecular reaction. Use the SSA algorithm to obtain sample realizations and numerically compute the probability distribution of the protein. Compare it to what you obtained in part (i). Explore how this probability distribution and the one of C change as the rate constants a and d become larger with respect to $\gamma, \alpha, \kappa, \delta$. Do you think we can use a QSS approximation similar to what we have done for ODE models?

(iii) Determine the Langevin equation for the system in part (ii) and obtain sample realizations. Explore numerically how good this approximation is when the volume decreases/increases.

4.4 Consider the bimolecular reaction

$$\mathrm{A} + \mathrm{B} \underset{d}{\overset{a}{\rightleftharpoons}} \mathrm{C},$$

in which A and B are in total amounts A_{tot} and B_{tot}, respectively. Compare the steady state value of C obtained from the deterministic model to the mean value of C obtained from the stochastic model as the volume is changed in the stochastic model. What do you observe? You can perform this investigation through numerical simulation.

4.5 Consider the simple birth and death process:

$$\mathrm{Z} \underset{k_1 G}{\overset{k_2 G}{\rightleftharpoons}} \emptyset,$$

in which G is a "gain." Assume that the reactions are catalyzed by enzymes and that the gain G can be tuned by changing the amounts of these enzymes. A deterministic ODE model for this system incorporating disturbances due to environmental perturbations is given by

$$\frac{dZ}{dt} = k_1 G - k_2 G Z + d(t).$$

Determine the Langevin equation for this birth and death process and compare its form to the deterministic one. Also, determine the frequency response of Z to noise for both the deterministic model and for the Langevin model. Does increasing the gain G have the same effect in both models? Explain.

4.6 Consider a second-order system with dynamics

$$\frac{d}{dt}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} -a & 0 \\ 0 & -b \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} w, \qquad Y = \begin{pmatrix} 1 & 1 \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

that is forced by Gaussian white noise w with zero mean and variance σ^2. Assume $a, b > 0$.

(i) Compute the correlation function $\rho(\tau)$ for the output of the system. Your answer should be an explicit formula in terms of a, b and σ.

(ii) Assuming that the input transients have died out, compute the mean and variance of the output.