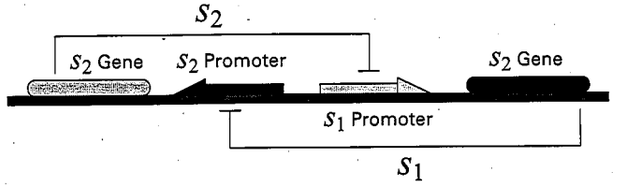
2 Modeling and Analysis of Stochastic Biochemical Networks

Mustafa Khammash

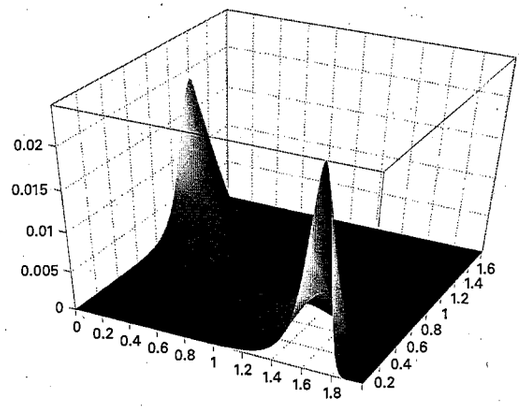
The cellular environment is replete with noisy processes. A key source of this "intrinsic" noise is the randomness that characterizes the motion of cellular constituents at the molecular level. Cellular noise results not only in random fluctuations (over time) within individual cells, but also in phenotypic variability among clonal cellular populations. In some instances, fluctuations are suppressed downstream through intricate dynamical networks that act as noise filters. Yet, in other important instances, noise-induced fluctuations are exploited to the cell's advantage. Researchers are just now beginning to understand that the richness of stochastic phenomena in biology depends directly on the interactions of dynamics and noise and on the mechanisms through which these interactions occur. This chapter outlines some of the key approaches for the modeling and analysis of cellular noise and the resulting fluctuations in the copy numbers of cellular constituents.

2.1 Noise in Biological Networks: Origins and Implications

Events in biological networks follow from biochemical reactions at the molecular level. The random nature of such reactions can be traced back to the random collisions among reactant molecules whose trajectories are driven by thermal motion. Such randomness leads to fluctuations in the molecular copy numbers of reactants both among similar cells and within a single cell over time. These fluctuations (commonly referred to as noise) can propagate downstream and impact events and processes in accordance to the dynamics of the network interconnections. Cellular noise has been measured experimentally and classified according to its source (Elowitz et al., 2002; Swain et al., 2002): *intrinsic noise* refers to noise originating within the boundaries of the process under consideration and is due to the inherent stochastic nature of chemical reactions, whereas *extrinsic noise* has origins that are more global and affects all processes in the cell under consideration in a similar way (for example, regulatory protein copy numbers, RNAP numbers). Noise, both intrinsic and extrinsic, can play a critical role in biological processes.



a



b

Figure 2.1
(a) Gardner-Cantor-Collins synthetic genetic toggle switch (Gardner et al., 2000). Protein s_1 suppresses the expression of the s_2 gene; protein s_2 suppresses the expression of the s_1 gene. (b) Bimodal nature of the distribution of proteins s_1 and s_2 that can arise for some parameter values of the toggle switch. The distribution shown was computed using the finite state projection method described in section 2.3.4.

McAdams and Arkin (1997, 1999) proposed that lysis-lysogeny fate decisions for phage λ are determined by a noise-driven stochastic switch, implying that the fate of a given cell can be determined only in a probabilistic sense. Another stochastic switch, which governs the piliation of *Escherichia coli*, has been modeled by Munsky et al. (2005). Aside from endogenous switches, bistable genetic switches have been constructed and tested (Gardner et al., 2000; Hasty et al., 2002). Depending on their parameter values, such switches are driven by stochastic noise to exhibit states with a bimodal probability distribution (figure 2.1). Elowitz and Leibler (2000) reported on the first synthetic oscillator, called the "repressilator," a novel circuit of three genes, each expressing a product that represses the next gene, thereby creating a feedback loop of three genes. A deterministic model and a discrete stochastic model that captures the effect of noise guided their design. The role of noise in the operation of the repressilator was recently studied by Yoda et al. (2007). Noise also appears to play an important role in the noise-enhanced robustness of oscillations in relaxation oscillators, for example, in the circadian rhythm. This effect, which is sometimes referred

to as coherence resonance, has been studied by Vilar et al. (2002) and El-Samad and Khammash (2006a). Yet another curious effect of noise can be seen in the fluctuation-enhanced sensitivity of intracellular regulation referred to as "stochastic focusing" (Paulsson et al., 2000). In gene expression, noise-induced fluctuations in gene products have been the subject of considerable interest (El-Samad and Khammash, 2006b; Isaacs et al., 2003; Paulsson, 2004; Rosenfeld et al., 2002; Swain, 2004; Thattai and van Oudenaarden, 2001). Many of these studies look at the propagation of noise in gene networks, as well as the impact and limitations of various types of feedback in suppressing such fluctuations.

2.1.1 Deterministic versus Stochastic Modeling

One approach to modeling reactions in biochemical networks, discussed in chapter 1, uses the law of mass action, which results in a set of differential equations that describe the evolution of concentrations of species adopted by the network over time. As an example, consider the reaction

$$A + B \xrightarrow{k} C.$$

A deterministic formulation of chemical kinetics yields the following description:

$$\frac{\mathrm{d}[C]}{\mathrm{d}t} = k[A] \times [B],$$

where $[\cdot]$ denotes the concentration, which is considered to be a continuous variable. In contrast, a discrete stochastic formulation of the same reaction describes the *probability* that the numbers of molecules of species A and B take certain integer values at a given time t. In this way, populations of the species within the network of interest are treated as random variables. In this description, reactions take place randomly according to certain probabilities determined by several factors including reaction rates and species populations. For example, given certain integer populations of A and B, say N_A and N_B, at time t, the probability that the above reaction takes place within the interval $[t, t + \mathrm{d}t)$ is proportional to $(N_A \times N_B/V)\,\mathrm{d}t$, where V is the volume of the space containing the molecules of A and B and $\mathrm{d}t$ is a small time increment.

Thus, in this mesoscopic stochastic formulation of chemical kinetics, molecular species are characterized by their probability density function, which quantifies the amount of fluctuations around a certain mean value. When molecule numbers get large (while maintaining the same initial concentration), fluctuations become negligible, and the mesoscopic description converges to the macroscopic description. In typical cellular environments where small volumes and molecule copy numbers are the norm, mesoscopic stochastic descriptions offer a more accurate representation of

chemical reactions and their accompanying fluctuations. Because they can generate distinct phenomena that simply cannot be captured by deterministic descriptions, such fluctuations need to be accounted for. The next section gives a more detailed description of the stochastic framework for modeling chemical reactions and provides a connection between the stochastic and deterministic descriptions.

2.2 Stochastic Chemical Kinetics

Consider a chemically reacting system of volume Ω containing N molecular species $S_1, \ldots, S_N$ that react through M allowable reactions $R_1, \ldots, R_M$. We assume that the system is well stirred and is in thermal equilibrium, thus the reaction volume is at a constant temperature T and the molecules move due to the thermal energy. Let $\boldsymbol{X}(t) = [X_1(t) \ldots X_N(t)]^T$ be the state vector, where $X_i(t)$ is a random variable that describes the number of molecules of species S_i in the system at time t. Elementary reactions may be either monomolecular: $S_i \rightarrow$ Products, or bimolecular: $S_i + S_j \rightarrow$ Products. Each reaction channel R_k defines a transition from some state $\boldsymbol{X} = \boldsymbol{x}_i$ to some other state $\boldsymbol{X} = \boldsymbol{x}_i + \boldsymbol{s}_k$, which reflects the change in the state after the reaction has taken place. The variable $\boldsymbol{s}_k$ is termed the *stoichiometric vector*, and the set of all M reactions define the *stoichiometry matrix*:

$$\boldsymbol{S} = [\boldsymbol{s}_1 \ldots \boldsymbol{s}_M].$$

Associated with each reaction R_k is a *propensity function*, $w_k(\boldsymbol{x})$, which captures the rate of the reaction k. Specifically, $w_k(\boldsymbol{x})\,\mathrm{d}t$ is the probability that, given the system is in state $\boldsymbol{x}$ at time t, the kth reaction will take place exactly once in the time interval $[t, t + \mathrm{d}t)$. The propensity function for various reaction types is given in table 2.1.

2.2.1 The Chemical Master Equation

Also known as the forward Kolmogorov equation, the *chemical master equation* (CME) describes the time evolution of the probability that the system is in a given

Table 2.1
Propensity functions for the various elementary reaction types

Reaction type		Propensity function
$S_i \rightarrow$ Products		$c\boldsymbol{x}_i$
$S_i + S_j \rightarrow$ Products	$(i \neq j)$	$c'\boldsymbol{x}_i\boldsymbol{x}_j$
$S_i + S_i \rightarrow$ Products		$c''\boldsymbol{x}_i(\boldsymbol{x}_i - 1)/2$

Note: If we denote by k, k', and k'' the reaction rate constants from deterministic mass action kinetics for the first, second, and third reaction types shown in the table, it can be shown that $c = k$, $c' = k'/\Omega$, and $c'' = 2k''/\Omega$.

state $\boldsymbol{x}$. The CME can be derived based on the Markov property of chemical reactions. Suppose the system is in state $\boldsymbol{x}$ at time t. Within an error of order $\mathcal{O}(\mathrm{d}t^2)$, the following statements apply:

- The probability that an R_k reaction fires exactly once in the time interval $[t, t + \mathrm{d}t]$ is given by $w_k(\boldsymbol{x})\,\mathrm{d}t$.
- The probability that no reactions fire in the time interval $[t, t + \mathrm{d}t]$ is given by $1 - \sum_k w_k(\boldsymbol{x})\,\mathrm{d}t$.
- The probability that more than one reaction fires in the time interval $[t, t + \mathrm{d}t]$ is zero.

Let $P(\boldsymbol{x}, t)$ denote the probability that the system is in state $\boldsymbol{x}$ at time t. We can express $P(\boldsymbol{x}, t + \mathrm{d}t)$ as follows:

$$P(\boldsymbol{x}, t + \mathrm{d}t) = P(\boldsymbol{x}, t)\left(1 - \sum_{k=1}^{M} w_k(\boldsymbol{x})\,\mathrm{d}t\right) + \sum_{k=1}^{M} P(\boldsymbol{x} - \boldsymbol{s}_k, t) w_k(\boldsymbol{x} - \boldsymbol{s}_k)\,\mathrm{d}t + \mathcal{O}(\mathrm{d}t^2).$$

On the right-hand side, the first term is the probability that the system is already in state $\boldsymbol{x}$ at time t and no reactions occur in the next $\mathrm{d}t$. In the second term, the kth term in the summation is the probability that the system at time t is an R_k reaction away from being at state $\boldsymbol{x}$, and that an R_k reaction takes place in the next $\mathrm{d}t$.

Moving $P(\boldsymbol{x}, t)$ to the left-hand side, dividing by $\mathrm{d}t$, and taking the limit as $\mathrm{d}t$ goes to zero yields the chemical master equation:

$$\frac{\mathrm{d}P(\boldsymbol{x}, t)}{\mathrm{d}t} = \sum_{k=1}^{M} (w_k(\boldsymbol{x} - \boldsymbol{s}_k) P(\boldsymbol{x} - \boldsymbol{s}_k, t) - w_k(\boldsymbol{x}) P(\boldsymbol{x}, t)). \tag{2.1}$$

2.2.2 Deterministic and Stochastic Models: A Connection

We now establish a connection between the stochastic process $\boldsymbol{X}(t)$ and the solution of the deterministic reaction rate equations arising from conventional mass-action kinetics. The latter corresponds to the trajectories of the concentrations of species $S_1, \ldots, S_N$. Let these concentrations be denoted by $\Phi(t) = [\Phi_1(t) \ldots \Phi_N(t)]^T$. Accordingly, $\Phi(\cdot)$ satisfies the mass-action ordinary differential equation:

$$\frac{\mathrm{d}\Phi}{\mathrm{d}t} = \boldsymbol{S} f(\Phi(t)), \qquad \Phi(0) = \Phi_0.$$

For a meaningful comparison with the stochastic solution, we shall compare the function $\Phi(t)$ with the volume-normalized stochastic process $\boldsymbol{X}^{\Omega}(t) = \boldsymbol{X}(t)/\Omega$. We now ask, how does $\boldsymbol{X}^{\Omega}(t)$ relate to $\Phi(t)$? The answer is given by the following fact, which is a consequence of the law of large numbers (Ethier and Kurtz, 1986):

Fact 2.1 Let $\Phi(t)$ be the deterministic solution to the reaction rate equations

$$\frac{d\Phi}{dt} = Sf(\Phi(t)), \qquad \Phi(0) = \Phi_0.$$

Let $X^\Omega(t)$ be the *stochastic* representation of the same chemical system with $X^\Omega(0) = \Phi_0$. Then, for every $t \geq 0$,

$$\lim_{\Omega\to\infty} \sup_{s\leq t} |X^\Omega(s) - \Phi(s)| = 0, \quad \text{almost surely.} \qquad \blacksquare$$

Put into words, this fact states that over any finite time interval, the stochastic description *converges* to the deterministic one *in the thermodynamic limit*. Although this result is reassuring, in practice, the large volume assumption cannot be justified: the cell volume is fixed. Indeed, a stochastic description could differ appreciably from its large volume limit—a fact that makes stochastic models necessary.

2.3 Stochastic Analysis Tools

Stochastic analysis tools may be broadly divided into four categories: (1) Monte Carlo methods, which compute sample paths whose statistics are used to extract information about the system; (2) methods that approximate the stochastic process $X(t)$ by solutions of certain stochastic differential equations; (3) methods that compute the trajectories of various moments of $X(t)$; and (4) methods that compute the evolution of probability densities of the stochastic process $X(t)$.

2.3.1 Monte Carlo Simulations

Because the chemical master equation is typically infinite dimensional with no obvious analytical solution, most analyses at the mesoscopic scale have been conducted using Monte Carlo algorithms. The most widely used of these algorithms is Gillespie's stochastic simulation algorithm (SSA; Gillespie, 1976) and its variants. These are described next.

The Gillespie Algorithm

Each step of the stochastic simulation algorithm begins at a time t and at a state $X(t) = x$ and comprises three substeps:

1. Generate the time until the next reaction;
2. Determine which reaction occurs at that time; and
3. Update the time and state to reflect the previous two choices.

The SSA approach is exact in the sense that it results in a random variable with a probability distribution exactly equal to the solution of the corresponding chemical master equation. However, each run of the SSA provides only a single trajectory. Thus numerous trajectories need to be generated that are then used to compute statistics of interest.

We now describe these substeps in more detail.

To each of the reactions $\{R_1, \ldots, R_M\}$, we associate a random variable $\mathcal{T}_i$ that describes the time for the next firing of reaction R_i. A key fact is that $\mathcal{T}_i$ is exponentially distributed with parameter w_i. From these, we can define two additional random variables, one continuous and the other discrete:

$$\mathcal{T} = \min_i\{\mathcal{T}_i\} \quad \text{(time to the next reaction)},$$

$$\mathcal{R} = \arg\min_i\{\mathcal{T}_i\} \quad \text{(index of the next reaction)}.$$

It can be shown that

1. $\mathcal{T}$ is exponentially distributed with parameter $\sum_i w_i$.
2. $\mathcal{R}$ has the discrete distribution $P(\mathcal{R} = k) = \frac{w_k}{\sum_i w_i}$.

Gillespie's simulation algorithm relies on samples of the random variables $\mathcal{T}$ and $\mathcal{R}$ to advance through each sample path.

Gillespie's Stochastic Simulation Algorithm:

- *Step 0* Initialize time t and state population x.
- *Step 1* Draw a sample τ from the distribution of $\mathcal{T}$ (figure 2.2).
- *Step 2* Draw a sample μ from the distribution of $\mathcal{R}$ (figure 2.2).
- *Step 3* Update time $t \leftarrow t + \tau$. Update the state $x \leftarrow x + s_\mu$.

Leaping Methods

The biggest drawback of the stochastic simulation algorithm is that it must step through one reaction at a time and thus is often prohibitively slow. One approximate accelerated simulation strategy is known as *tau leaping* (Gillespie, 2001), which advances the system by a *preselected* time τ that encompasses more than one reaction event. Roughly speaking, tau leaping requires that τ be chosen small enough so that the following *leap condition* is satisfied: the expected state change induced by the leap must be sufficiently small that no propensity function changes "appreciably" during that time. Tau leaping has been shown to speed up the simulation of certain systems significantly (Gillespie, 2001; Gillespie and Petzold, 2003; Rathinam et al., 2003, 2005), although it is not as foolproof as the SSA. If one takes leaps that are too large, the tau leap assumptions may be violated and the results may be inaccurate

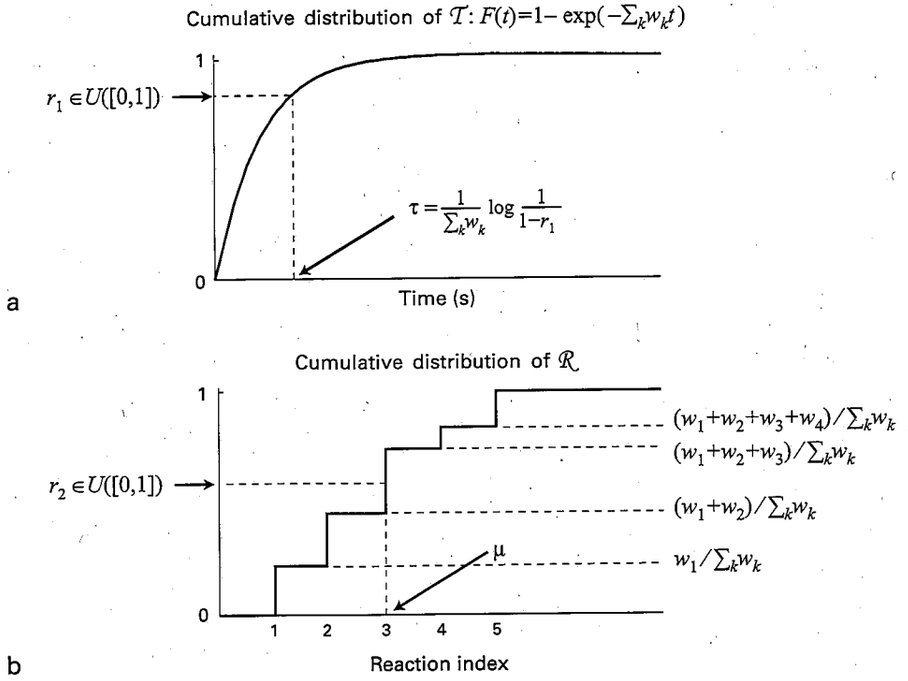


Figure 2.2
Cumulative distribution of the two random variables $\mathcal{T}$ and $\mathcal{R}$. (a) A sample of $\mathcal{T}$ is drawn by first drawing a uniformly distributed random number r_1 and then finding its inverse image under F, the cumulative distribution of $\mathcal{T}$. (b) A sample from the distribution of $\mathcal{R}$ is drawn using a similar procedure.

or even nonsensical. For example, some species populations might be driven negative. Moreover, if the system is stiff, meaning that it has widely varying time scales with the fastest mode being stable, the leap condition will generally limit the size of τ to the time scale of the fastest mode, with the result that large leaps cannot be taken.

2.3.2 Stochastic Differential Equation Approximations

The Chemical Langevin Equation: Diffusion Approximation

One approximation to the chemical master equation, the *chemical Langevin equation*, can be obtained through the tau-leaping multireaction update formula: with $\boldsymbol{X}(t) = \boldsymbol{x}$, suppose that the leap time τ can be taken small enough to satisfy the leap condition, but large enough that $w_k(\boldsymbol{x})\tau \gg 1$ for every reaction. If the leap condition holds, the number of reactions in the interval τ is a Poisson random variable with mean and variance equal to $w_k(\boldsymbol{x})\tau$. When this quantity is much larger than one, this

random variable is well approximated by a normal random variable with the same mean and variance. This leads to the *Langevin leaping formula*:

$$\boldsymbol{X}(t+\tau) \approx \boldsymbol{x} + \sum_{k=1}^{M} \boldsymbol{s}_k w_k(\boldsymbol{x})\tau + \sum_{k=1}^{M} \boldsymbol{s}_k \sqrt{w_k(\boldsymbol{x})}\mathcal{N}_k(0,1)\sqrt{\tau}, \tag{2.2}$$

which expresses the state increment $\boldsymbol{X}(t+\tau) - \boldsymbol{x}$ as the sum of two terms: a deterministic "drift" term proportional to τ and a fluctuating "diffusion" term proportional to $\sqrt{\tau}$, where the $\mathcal{N}_k(0,1)$ denote independent standard normal random variables (Gillespie, 2000, 2002). From equation (2.2), one can approximate the stochastic process $\boldsymbol{X}$ by another stochastic process $\boldsymbol{V}$, which is described by the following nonlinear stochastic differential equation:

$$d\boldsymbol{V}(t) = \sum_{k=1}^{M} \boldsymbol{s}_k w_k(\boldsymbol{V}(t))\,\mathrm{d}t + \sum_{k=1}^{M} \boldsymbol{s}_k \sqrt{w_k(\boldsymbol{V}(t))}\,\mathrm{d}B_k(t),$$

where the $B_k(t)$ are independent standard Brownian motion processes. This approximation is sometimes called the *chemical Langevin equation approximation* or the *diffusion approximation* (Gillespie, 2000, 2002; Kurtz, 1978).

van Kampen's Linear Noise Approximation

Another approximation that leads to a stochastic differential equation is van Kampen's linear noise approximation (LNA; Elf and Ehrenberg, 2003; Khammash and El-Samad, 2005; Tomioka et al., 2004; van Kampen, 1981). It is essentially an approximation to the process $\boldsymbol{X}(t)$ that takes advantage of the fact that in the large volume limit ($\Omega \to \infty$), the process $\boldsymbol{X}^{\Omega}(t) = \boldsymbol{X}(t)/\Omega$ converges to the solution $\Phi(t)$ of the deterministic reaction rate equation:

$$\frac{\mathrm{d}\Phi}{\mathrm{d}t} = \boldsymbol{S}f(\Phi(t)), \qquad \Phi(t_0) = \Phi_0.$$

Defining a scaled "error" process,

$$\boldsymbol{V}^{\Omega}(t) = \sqrt{\Omega}(\boldsymbol{X}^{\Omega}(t) - \Phi(t)),$$

and using the central limit theorem, it can be shown (Ethier and Kurtz, 1986) that $\boldsymbol{V}^{\Omega}(t)$ converges in distribution to the solution $\boldsymbol{V}(t)$ of the linear stochastic differential equation:

$$\mathrm{d}\boldsymbol{V}(t) = \boldsymbol{J}_{\boldsymbol{F}}(\Phi(t))\boldsymbol{V}(t)\,\mathrm{d}t + \sum_{k=1}^{M} \boldsymbol{s}_k \sqrt{w_k(\Phi(t))}\,\mathrm{d}B_k(t),$$

where J_F denotes the Jacobian of $F(\cdot) = Sf(\cdot)$. Hence the linear noise approximation results in a process $X(t) \approx \Omega\Phi(t) + \sqrt{\Omega}V(t)$, which can be viewed as the sum of a deterministic term given by the solution to the deterministic reaction rate equation, and a zero mean stochastic term given by the solution to a linear stochastic differential equation. Though it is reasonable for systems with sufficiently large numbers of molecules (and volume), examples show that the LNA can yield poor results when this assumption is violated. This includes scenarios where the mean of a stochastic representation does not coincide with the solution of the corresponding reaction rate equation (see Paulsson et al., 2000 on *stochastic focusing*).

2.3.3 Moment Computations

When studying stochastic fluctuations that arise in biochemical networks, one is often interested in computing moments and variances of biochemical species. The moment dynamics can be described using the chemical master equation. To compute the first moment $E[X_i]$, we multiply the CME by x_i and then sum over all $(x_1, \ldots, x_N) \in \mathbf{N}^N$ to get

$$\frac{dE[X_i]}{dt} = \sum_{k=1}^{M} s_{ik} E[w_k(\boldsymbol{X})].$$

Similarly, to get the second moments $E[X_iX_j]$, we multiply the CME by x_ix_j and sum over all $(x_1, \ldots, x_N) \in \mathbf{N}^N$, which gives

$$\frac{dE[X_iX_j]}{dt} = \sum_{k=1}^{M} s_{ik}E[X_jw_k(\boldsymbol{X})] + E[X_iw_k(\boldsymbol{X})]s_{jk} + s_{ik}E[w_k(\boldsymbol{X})]s_{jk}.$$

These last two equations can be expressed more compactly in matrix form. Defining $w(x) = [w_1(x) \ldots w_M(x)]^T$, the moment dynamics become

$$\frac{dE[\boldsymbol{X}]}{dt} = \boldsymbol{S}E[\boldsymbol{w}(\boldsymbol{X})],$$

$$\frac{dE[\boldsymbol{X}\boldsymbol{X}^T]}{dt} = \boldsymbol{S}E[\boldsymbol{w}(\boldsymbol{X})\boldsymbol{X}^T] + E[\boldsymbol{w}(\boldsymbol{X})\boldsymbol{X}^T]^T\boldsymbol{S}^T + \boldsymbol{S}(\text{diag } E[\boldsymbol{w}(\boldsymbol{X})])\boldsymbol{S}^T.$$

In general, this set of moment equations cannot be solved explicitly because they will not always be closed: depending on the form of the propensity vector $\boldsymbol{w}(\cdot)$, the dynamics of the first moment $E(\boldsymbol{X})$ may depend on the second moments $E(\boldsymbol{X}\boldsymbol{X}^T)$, the second-moment dynamics may in turn depend on the third moments, and so on, resulting in an infinite system of ordinary differential equations. The following subsections will elaborate on the important special case when moment equations *are* closed, and then discuss approaches to deal with scenarios when they are not.

Special Case: Affine Propensities

If the propensity function is affine:

$$\boldsymbol{w}(\boldsymbol{x}) = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{w}_0 \quad (\boldsymbol{W} \text{ is an } N \times N \text{ matrix; } \boldsymbol{w}_0 \text{ is } N \times 1),$$

then

$$E[\boldsymbol{w}(\boldsymbol{X})] = \boldsymbol{W}E[\boldsymbol{X}] + \boldsymbol{w}_0,$$

and

$$E[\boldsymbol{w}(\boldsymbol{X})\boldsymbol{X}^T] = \boldsymbol{W}E[\boldsymbol{X}\boldsymbol{X}^T] + \boldsymbol{w}_0E[\boldsymbol{X}^T].$$

This gives us the following moment equations:

$$\frac{d}{dt}E[\boldsymbol{X}] = \boldsymbol{S}\boldsymbol{W}E[\boldsymbol{X}] + \boldsymbol{S}\boldsymbol{w}_0,$$

$$\frac{d}{dt}E[\boldsymbol{X}\boldsymbol{X}^T] = \boldsymbol{S}\boldsymbol{W}E[\boldsymbol{X}\boldsymbol{X}^T] + E[\boldsymbol{X}\boldsymbol{X}^T]\boldsymbol{W}^T\boldsymbol{S}^T + \boldsymbol{S}\,\text{diag}(\boldsymbol{W}E[\boldsymbol{X}] + \boldsymbol{w}_0)\boldsymbol{S}^T$$
$$+ \boldsymbol{S}\boldsymbol{w}_0E[\boldsymbol{X}^T] + E[\boldsymbol{X}]\boldsymbol{w}_0^T\boldsymbol{S}^T.$$

Clearly, this is a closed system of linear ODEs that can be solved directly for the first and second moments.

Defining the covariance matrix

$$\boldsymbol{\Sigma} = E[(\boldsymbol{X} - E[\boldsymbol{X}])(\boldsymbol{X} - E[\boldsymbol{X}])^T],$$

we can also compute covariance equations:

$$\frac{d}{dt}\boldsymbol{\Sigma} = \boldsymbol{S}\boldsymbol{W}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\boldsymbol{W}^T\boldsymbol{S}^T + \boldsymbol{S}\,\text{diag}(\boldsymbol{W}E[\boldsymbol{X}] + \boldsymbol{w}_0)\boldsymbol{S}^T.$$

The steady-state moments and covariances can be obtained by solving linear algebraic equations. Let

$$\overline{\boldsymbol{X}} = \lim_{t\to\infty} E[\boldsymbol{X}(t)]$$

and

$$\overline{\boldsymbol{\Sigma}} = \lim_{t\to\infty} \boldsymbol{\Sigma}(t).$$

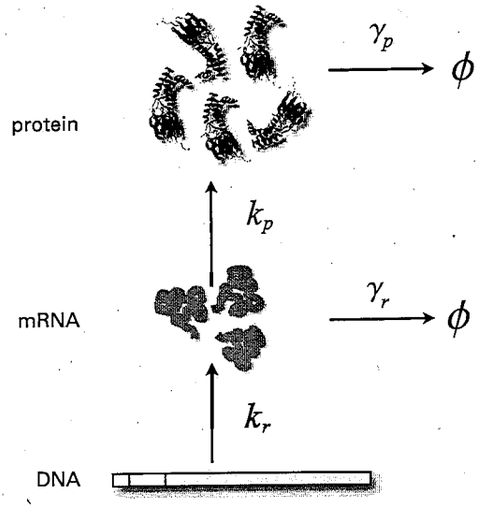


Figure 2.3
Simple model for gene expression. mRNA is transcribed at a rate k_r and is degraded at a rate γ_r. Protein is translated from mRNA at a rate k_p and is degraded at a rate γ_p.

Then

$$\boldsymbol{S}\boldsymbol{W}\overline{\boldsymbol{X}} = -\boldsymbol{S}\boldsymbol{w}_0,$$

and

$$\boldsymbol{S}\boldsymbol{W}\overline{\boldsymbol{\Sigma}} + \overline{\boldsymbol{\Sigma}}\boldsymbol{W}^T\boldsymbol{S}^T + \boldsymbol{S}\,\mathrm{diag}(\boldsymbol{W}\overline{\boldsymbol{X}} + \boldsymbol{w}_0)\boldsymbol{S}^T = \boldsymbol{0}.$$

The latter is an algebraic Lyapunov equation. Such equations arise in many applications in control theory, and efficient numerical techniques exist for their solution.

Example: Application to Gene Expression The above stationary covariance equations can be applied to compute exact expressions for the mean and coefficient of variation of a simple gene expression circuit. Consider the gene expression model shown in figure 2.3.

Let $X_1(t)$ and $X_2(t)$ be random variables describing the number of mRNA and protein molecules, respectively. The stoichiometry matrix for the gene expression reactions is

$$\boldsymbol{S} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix},$$

while the propensity vector is given by

$$\boldsymbol{w}(\boldsymbol{X}) = \begin{bmatrix} k_r \\ \gamma_r X_1 \\ k_p X_1 \\ \gamma_p X_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \gamma_r & 0 \\ k_p & 0 \\ 0 & \gamma_p \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} k_r \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$= \boldsymbol{W}\boldsymbol{X} + \boldsymbol{w}_0.$$

Defining $\boldsymbol{A} = \boldsymbol{S}\boldsymbol{W}$, we can compute the vector of stationary means:

$$\overline{\boldsymbol{X}} = -\boldsymbol{A}^{-1}\boldsymbol{S}\boldsymbol{w}_0 = \begin{bmatrix} \dfrac{k_r}{\gamma_r} \\ \dfrac{k_p k_r}{\gamma_p \gamma_r} \end{bmatrix}.$$

Defining

$$\boldsymbol{B}\boldsymbol{B}^T = \boldsymbol{S}\,\mathrm{diag}(\boldsymbol{W}\overline{\boldsymbol{X}} + \boldsymbol{w}_0)\boldsymbol{S}^T,$$

we see that the stationary covariance matrix is the solution to the Lyapunov equation:

$$\boldsymbol{A}\overline{\boldsymbol{\Sigma}} + \overline{\boldsymbol{\Sigma}}\boldsymbol{A}^T + \boldsymbol{B}\boldsymbol{B}^T = \boldsymbol{0},$$

which is given by

$$\overline{\boldsymbol{\Sigma}} = \begin{bmatrix} \dfrac{k_r}{\gamma_r} & \dfrac{k_p k_r}{\gamma_r(\gamma_r + \gamma_p)} \\ \dfrac{k_p k_r}{\gamma_r(\gamma_r + \gamma_p)} & \dfrac{k_p k_r}{\gamma_p \gamma_r}\left(1 + \dfrac{k_p}{\gamma_r + \gamma_p}\right) \end{bmatrix}.$$

A common measure of the degree of variability is the *coefficient of variation*, C_v, defined as the standard deviation divided by the mean. The coefficient of variation for the mRNA and protein can now be computed easily:

$$C_{vr} = \left(\frac{1}{k_r/\gamma_r}\right)^{1/2}, \quad \text{and} \quad C_{vp} = \left(\frac{\gamma_r \gamma_p}{k_r k_p}\right)^{1/2}\left(1 + \frac{k_p}{\gamma_r + \gamma_p}\right)^{1/2}.$$

In the case of mRNA, we have just shown the well-known result that the coefficient of variation equals $1/\sqrt{\text{mean}}$. ■

Moment Closure

An important property of the Markov processes that describe chemical reactions is that when one constructs a vector μ with all the first- and second-order statistical uncentered moments of the process state X, this vector evolves according to a *linear* equation of the form

$$\frac{d\mu}{dt} = A\mu + B\bar{\mu}. \tag{2.3}$$

Unfortunately, as pointed out earlier, equation (2.3) is not in general a closed system because the vector $\bar{\mu}$ may contain moments of order larger than two, whose evolution is not provided by equation (2.3). In fact, this will always be the case when bimolecular reactions are involved. To overcome this difficulty, one can approximate the *open linear system* (2.3) by the following *closed nonlinear system*:

$$\frac{dv}{dt} = Av + B\varphi(v), \tag{2.4}$$

where v is an approximation to the solution μ to (2.3) and $\varphi(\cdot)$ is a *moment closure function* that attempts to approximate the moments in $\bar{\mu}$ based on the values of the moments in μ. The construction of $\varphi(\cdot)$ often relies on postulating a given type for the distribution of X and then expressing the higher-order moments in $\bar{\mu}$ by a nonlinear function $\varphi(\mu)$ of the first- and second-order moments in μ. Postulating a normal distribution is quite popular (Gomez-Uribe and Verghese, 2007; Nasell, 2003b; Whittle, 1957), but when the population standard deviations are not much smaller than the means, choosing $\varphi(\cdot)$ based on a normal distribution assumption often leads to bad approximations. Other authors construct moment closure functions $\varphi(\cdot)$ based on different assumed distributions for X, which include lognormal (Keeling, 2000), Poisson, and binomial (Nasell, 2003a) distributions.

Hespanha (2005) proposed a new technique for moment closure that does not require a priori assumptions on the shape of the distribution for X. Instead, the moment closure $\varphi(\cdot)$ is computed by trying to match all (or a large number of) the time derivatives of the exact solution to equation (2.3) with the corresponding time derivatives of the approximate solution to equation (2.4), for a given set of initial conditions. With this approach, one can indeed match derivatives between equation (2.3) and equation (2.4) with small error (Singh and Hespanha, 2006, 2007a,b). Moreover, this can be done with moment closure functions $\varphi(\cdot)$ that do not depend on the (most often poorly known) parameters of the chemical reactions. This leads to an automated methodology to construct the approximate closed systems (2.4). A set of Matlab scripts that constructs truncated moment dynamics given a set of chemical reactions may be found in Hespanha (2006).

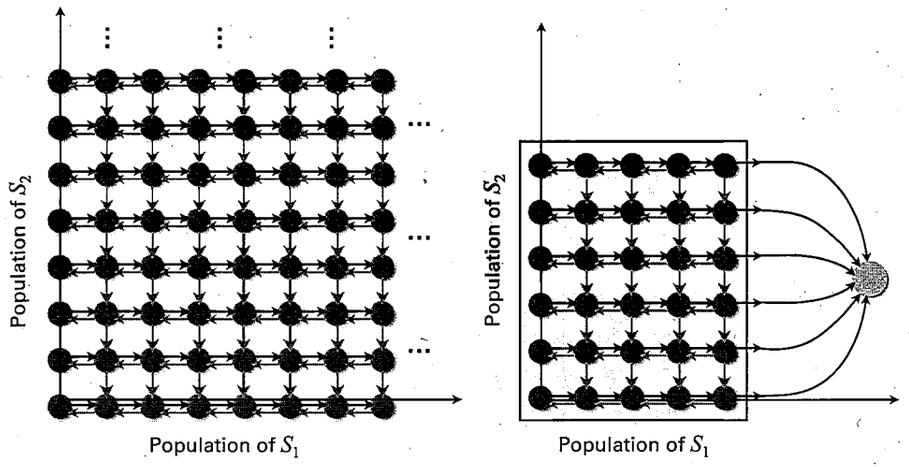


Figure 2.4
Finite-state projection. (Left) State space for a system with two species. The corresponding process is a continuous-time, discrete-state Markov process whose state space is typically quite large or infinite. Arrows indicate possible transitions within states. (Right) Projected system for a specific projection region (*gray box*). The projected system is obtained as follows. Transitions within the projection region are kept unchanged. Transitions that emanate from states within the region and end at states outside (in the original system) are routed to a single absorbing state in the projected system. Transitions into the projection region are deleted. As a result, the projected system is a finite-state Markov process, and the probability of each state can be computed exactly.

2.3.4 Density Computations

Another approach used to analyze models described by the chemical master equation aims to compute the probability density functions for the random variable X. This is achieved by approximate solutions of the CME, using a new analytical approach called the finite-state projection (FSP) (Munsky and Khammash, 2006; Peles et al., 2006). The FSP approach relies on a projection that preserves an important subset of the state space (for example, that supporting the bulk of the probability distribution), while projecting the remaining large or infinite states onto a single "absorbing" state (figure 2.4).

Probabilities for the resulting finite-state Markov chain can be computed exactly, and can be shown to give a lower bound for the corresponding probability for the original full system. The finite-state projection algorithm provides a means of systematically choosing a projection of the chemical master equation that satisfies any prespecified accuracy requirement. The basic idea of the FSP is as follows. In matrix form, the CME may be written as

$$\frac{dP(t)}{dt} = AP(t), \tag{2.5}$$

where $\boldsymbol{P}(t)$ is the (infinite) vector of probabilities corresponding to each possible state in the configuration space. The generator matrix $\boldsymbol{A}$ embodies the propensity functions for transitions from one configuration to another and is defined by the reactions and the enumeration of the configuration space. A projection can now be made to achieve an arbitrarily accurate approximation as outlined next. Given an index set of the form $J = \{j_1, j_2, j_3, \ldots\}$ and a vector $\boldsymbol{v}$, let $\boldsymbol{v}_J$ denote the subvector of $\boldsymbol{v}$ chosen according to J, and for any matrix $\boldsymbol{A}$, let $\boldsymbol{A}_J$ denote the submatrix of $\boldsymbol{A}$ whose rows and columns have been chosen according to J. With this notation, we can restate the result from Munsky and Khammash (2006):

Fact 2.2: Finite-State Projection Consider any distribution that evolves according to equation (2.5). Let $\boldsymbol{A}_J$ be a principal submatrix of $\boldsymbol{A}$ and $\boldsymbol{P}_J$ be a subvector of $\boldsymbol{P}$, both corresponding to the indexes in J. If, for a given $\varepsilon > 0$ and $t_f \geq 0$, we have

$$\mathbf{1}^T \exp(\boldsymbol{A}_J t_f)\boldsymbol{P}_J(0) \geq 1 - \varepsilon,$$

then

$$\|\exp(\boldsymbol{A}_J t_f)\boldsymbol{P}_J(0) - \boldsymbol{P}_J(t_f)\|_1 \leq \varepsilon. \tag{2.6}$$

■

Inequality (2.6) provides a bound on the error between the exact solution $\boldsymbol{P}_J$ to the (infinite) chemical master equation and the matrix exponential of the (finite) reduced system with generator $\boldsymbol{A}_J$. This result is the basis for an algorithm to compute the probability density function with guaranteed accuracy. The FSP approach and various improvements on the main algorithm are described by Munsky and Khammash (2008).

2.4 Conclusions and Future Directions

This chapter has presented several approaches for the stochastic analysis of chemical reactions arising in gene networks. Although, in isolation, none of these techniques provides a solution that is both computationally effective and accurate, they are clearly complementary in scope. The density computation methods (for example, finite-state projection) are especially suitable for very low molecule counts, moment closure methods for medium to large counts, and the linear noise approximation for very large counts. Combining these different types of approximations in a unified framework to devise models that are both accurate and computationally efficient remains an important research goal. Essentially, one would like to follow the limiting volume argument for chemical species that have a large number of molecules (resulting in linear models for the covariances), but use Monte Carlo or finite-projection techniques for chemical species that do not.