

Continuous time Markov chain models for chemical reaction networks

David F. Anderson ^{*}Departments of Mathematics
University of Wisconsin - Madison
480 Lincoln Drive
Madison, WI 53706-1388
<http://www.math.wisc.edu/~anderson>
anderson@math.wisc.edu

Thomas G. Kurtz, [†]Departments of Mathematics
University of Wisconsin - Madison
480 Lincoln Drive
Madison, WI 53706-1388
<http://www.math.wisc.edu/~kurtz>
kurtz@math.wisc.edu

July 27, 2010

Abstract

A reaction network is a chemical system involving multiple reactions and chemical species. The simplest stochastic models of such networks treat the system as a continuous time Markov chain with the state being the number of molecules of each species and with reactions modeled as possible transitions of the chain. This chapter is devoted to the mathematical study of such stochastic models. We begin by developing much of the mathematical machinery we need to describe the stochastic models we are most interested in. We show how one can represent counting processes of the type we need in terms of the Poisson process. This leads to a stochastic equation, usually termed the random time change representation, for continuous time Markov chain models. We include a discussion on the relationship between this stochastic equation and the corresponding martingale problem and Kolmogorov forward (master) equation. Next, we exploit the representation of the stochastic equation for chemical reaction networks and, under what is often called the classical scaling, show how to derive the deterministic law of mass action from the Markov chain model. We also review the diffusion,

^{*}Research supported in part by NSF grant DMS 05-53687

[†]Research supported in part by NSF grants DMS 05-53687 and DMS 08-05793

or Langevin, approximation, include a discussion of first order reaction networks, and present a large class of networks, those that are weakly reversible and have a deficiency of zero, that induce product-form stationary distributions. Finally, we discuss models in which the numbers of molecules and/or the reaction rate constants of the system vary over several orders of magnitude. We show that one consequence of this wide variation in scales is that different subsystems may evolve on different time scales and this time-scale variation can be exploited to identify reduced models that capture the behavior of parts of the system. We will discuss systematic ways of identifying the different time scales and deriving the reduced models.

MSC 2000 subject classifications: 60J27, 60J28, 60J80, 60F17, 80A30, 92C40

Keywords: Reaction network, Markov chain, law of mass action, law of large numbers, central limit theorem, diffusion approximation, Langevin approximation, stochastic equations, multiscale analysis, stationary distributions

1 Introduction

The idea of modeling chemical reactions as a stochastic process at the molecular level dates back at least to [12] with a rapid development beginning in the 1950s and 1960s. (See, for example, [6, 7, 39].) For the reaction



in which one molecule of A and one molecule of B are consumed to produce one molecule of C , the intuition for the model for the reaction is that the probability of the reaction occurring in a small time interval $(t, t + \Delta t]$ should be proportional to the product of the numbers of molecules of each of the reactants and to the length of the time interval. In other words, since for the reaction to occur a molecule of A and a molecule of B must be close to each other, the probability should be proportional to the number of pairs of molecules that can react. A more systematic approach to this conclusion might be to consider the following probability problem: Suppose k red balls (molecules of A) and l black balls (molecules of B) are placed uniformly at random in n boxes, where n is much larger than k and l . What is the probability that at least one red ball ends up in the same box as a black ball? We leave it to the reader to figure that out. For a more physically based argument, see [22].

Our more immediate concern is that the calculation, however justified, assumes that the numbers of molecules of the chemical species are known. That assumption means that what is to be computed is a *conditional probability*, that is, a computation that uses information that might not (or could not) have been known when the experiment was first set up.

Assuming that at time t there are $X_A(t)$ molecules of A and $X_B(t)$ molecules of B in our system, we express our assumption about the probability of the reaction occurring by

$$P\{\text{reaction occurs in } (t, t + \Delta t] | \mathcal{F}_t\} \approx \kappa X_A(t) X_B(t) \Delta t \quad (1.1)$$

where \mathcal{F}_t represents the information about the system that is available at time t and κ is a positive constant, the *reaction rate constant*. Since Kolmogorov's fundamental work [29], probabilists have modeled information as a σ -algebra (a collection of sets with particular

properties) of events (subsets of possible outcomes) in the sample space (the set of all possible outcomes). Consequently, mathematically, \mathcal{F}_t is a σ -algebra, but readers unfamiliar with this terminology should just keep the idea of information in mind when we write expressions like this, that is, \mathcal{F}_t just represents the information available at time t .

One of our first goals will be to show how to make the intuitive assumption in (1.1) into a precise mathematical model. Our model will be formulated in terms of X_A , X_B , and X_C which will be *stochastic processes*, that is, random functions of time. The triple $X(t) = (X_A(t), X_B(t), X_C(t))$ gives the *state* of the process at time t . Simple bookkeeping implies

$$X(t) = X(0) + R(t) \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}, \quad (1.2)$$

where $R(t)$ is the number of times the reaction has occurred by time t and $X(0)$ is the vector giving the numbers of molecules of each of the chemical species in the system at time zero. We will assume that two reactions cannot occur at exactly the same time, so R is a counting process, that is, $R(0) = 0$ and R is constant except for jumps of plus one.

Our first task, in Section 2, will be to show how one can represent counting processes of the type we need in terms of the most elementary counting process, namely, the Poisson process. Implicit in the fact that the right side of (1.1) depends only on the current values of X_A and X_B is the assumption that the model satisfies the Markov property, that is, the future of the process only depends on the current value, not on values at earlier times. The representation of counting processes in terms of Poisson processes then gives a stochastic equation for a general continuous-time Markov chain. There are, of course, other ways of specifying a continuous-time Markov chain model, and Section 2 includes a discussion of the relationship between the stochastic equation and the corresponding martingale problem and Kolmogorov forward (master) equation. We also include a brief description of the common methods of simulating the models.

Exploiting the representation as a solution of a stochastic equation, in Section 3 we discuss stochastic models for chemical reaction networks. Under what we will refer to as the classical scaling, we show how to derive the deterministic law of mass action from the Markov chain model and introduce the diffusion or Langevin approximation. We also discuss the simple class of networks in which all reactions are unary and indicate how the large literature on branching processes and queueing networks provides useful information about this class of networks. Many of these networks have what is known in the queueing literature as product form stationary distributions, which makes the stationary distributions easy to compute. The class of networks that have stationary distributions of this form is not restricted to unary networks, however. In particular, all networks that satisfy the conditions of the zero-deficiency theorem of Feinberg [16, 17], well-known in deterministic reaction network theory, have product-form stationary distributions. There is also a brief discussion of models of reaction networks with delays.

The biological systems that motivate the current discussion may involve reaction networks in which the numbers of molecules of the chemical species present in the system vary over several orders of magnitude. The reaction rates may also vary widely. One consequence of this wide variation in scales is that different subsystems may evolve on different time

scales and this time-scale variation can be exploited to identify reduced models that capture the behavior of parts of the system. Section 4 discusses systematic ways of identifying the different time scales and deriving the reduced models.

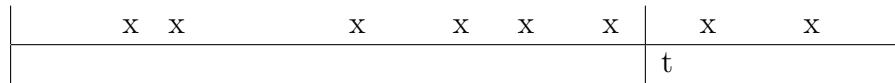
Although much of the discussion that follows is informal and is intended to motivate rather than rigorously demonstrate the ideas and methods we present, any lemma or theorem explicitly identified as such is rigorously justifiable, or at least we intend that to be the case. Our intention is to prepare an extended version of this paper that includes detailed proofs of most or all of the theorems included.

2 Counting processes and continuous time Markov chains

The simplest counting process is a Poisson process, and Poisson processes will be the basic building blocks that we use to obtain more complex models.

2.1 Poisson processes

A Poisson process is a model for a series of random observations occurring in time.



Let $Y(t)$ denote the number of observations by time t . In the figure above, $Y(t) = 6$. Note that for $t < s$, $Y(s) - Y(t)$ is the number of observations in the time interval $(t, s]$. We make the following assumptions about the model.

- 1) Observations occur one at a time.
- 2) Numbers of observations in disjoint time intervals are independent random variables, i.e., if $t_0 < t_1 < \dots < t_m$, then $Y(t_k) - Y(t_{k-1})$, $k = 1, \dots, m$ are independent random variables.
- 3) The distribution of $Y(t + a) - Y(t)$ does not depend on t .

The following result can be found in many elementary books on probability and stochastic processes. See, for example, Ross [41].

Theorem 2.1 *Under assumptions 1), 2), and 3), there is a constant $\lambda > 0$ such that, for $t < s$, $Y(s) - Y(t)$ is Poisson distributed with parameter $\lambda(s - t)$, that is,*

$$P\{Y(s) - Y(t) = k\} = \frac{(\lambda(s - t))^k}{k!} e^{-\lambda(s-t)}. \quad (2.1)$$

If $\lambda = 1$, then Y is a *unit* (or rate one) Poisson process. If Y is a unit Poisson process and $Y_\lambda(t) \equiv Y(\lambda t)$, then Y_λ is a Poisson process with parameter λ . Suppose $Y_\lambda(t) = Y(\lambda t)$ and \mathcal{F}_t represents the information obtained by observing $Y_\lambda(s)$, for $s \leq t$. Then by the independence assumption and (2.1)

$$P\{Y_\lambda(t + \Delta t) - Y_\lambda(t) > 0 | \mathcal{F}_t\} = P\{Y_\lambda(t + \Delta t) - Y_\lambda(t) > 0\} = 1 - e^{-\lambda \Delta t} \approx \lambda \Delta t. \quad (2.2)$$

The following facts about Poisson processes play a significant role in our analysis of the models we will discuss.

Theorem 2.2 *If Y is a unit Poisson process, then for each $u_0 > 0$,*

$$\lim_{n \rightarrow \infty} \sup_{u \leq u_0} \left| \frac{Y(nu)}{n} - u \right| = 0 \quad a.s.$$

Proof. For fixed u , by the independent increments assumption, the result is just the ordinary law of large numbers. The uniformity follows by monotonicity. \square

The classical central limit theorem implies

$$\lim_{n \rightarrow \infty} P\left\{ \frac{Y(nu) - nu}{\sqrt{n}} \leq x \right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = P\{W(u) \leq x\},$$

where W is a *standard Brownian motion*. In fact, the approximation is uniform on bounded time intervals in much the same sense that the limit in Theorem 2.2 is uniform. This result is essentially Donsker's functional central limit theorem [13]. It suggests that for large n

$$\frac{Y(nu) - nu}{\sqrt{n}} \approx W(u), \quad \frac{Y(nu)}{n} \approx u + \frac{1}{\sqrt{n}} W(u)$$

where the approximation is uniform on bounded time intervals. One way to make this approximation precise is through the strong approximation theorem of Komlós, Major, and Tusnády [30, 31], which implies the following.

Lemma 2.3 *A unit Poisson process Y and a standard Brownian motion W can be constructed so that*

$$\Gamma \equiv \sup_{t \geq 0} \frac{|Y(t) - t - W(t)|}{\log(2 \vee t)} < \infty \quad a.s.$$

and there exists $c > 0$ such that $E[e^{c\Gamma}] < \infty$.

Proof. See Corollary 7.5.5 of [15]. \square

Note that

$$\left| \frac{Y(nt) - nt}{\sqrt{n}} - \frac{1}{\sqrt{n}} W(nt) \right| \leq \frac{\log(nt \vee 2)}{\sqrt{n}} \Gamma, \quad (2.3)$$

and that $\frac{1}{\sqrt{n}} W(nt)$ is a standard Brownian motion.

2.2 Continuous time Markov chains

The calculation in (2.2) and the time-change representation $Y_\lambda(t) = Y(\lambda t)$ suggest the possibility of writing R in (1.2) as

$$R(t) = Y\left(\int_0^t \kappa X_A(s)X_B(s)ds\right)$$

and hence

$$\begin{pmatrix} X_A(t) \\ X_B(t) \\ X_C(t) \end{pmatrix} \equiv X(t) = X(0) + \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} Y\left(\int_0^t \kappa X_A(s)X_B(s)ds\right). \quad (2.4)$$

Given Y and the initial state $X(0)$ (which we assume is independent of Y), (2.4) is an equation that uniquely determines X for all $t > 0$. To see that this assertion is correct, let τ_k be the k th jump time of Y . Then letting

$$\zeta = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix},$$

(2.4) implies $X(t) = X(0)$ for $0 \leq t < \tau_1$, $X(t) = X(0) + \zeta$ for $\tau_1 \leq t < \tau_2$, and so forth. To see that the solution of this equation has the properties suggested by (1.1), let $\lambda(X(t)) = \kappa X_A(t)X_B(t)$ and observe that occurrence of the reaction in $(t, t + \Delta t]$ is equivalent to $R(t + \Delta t) > R(t)$, so the left side of (1.1) becomes

$$\begin{aligned} & P\{R(t + \Delta t) > R(t) | \mathcal{F}_t\} \\ &= 1 - P\{R(t + \Delta t) = R(t) | \mathcal{F}_t\} \\ &= 1 - P\left\{Y\left(\int_0^t \lambda(X(s))ds + \lambda(X(t))\Delta t\right) = Y\left(\int_0^t \lambda(X(s))ds\right) | \mathcal{F}_t\right\} \\ &= 1 - e^{-\lambda(X(t))\Delta t} \approx \lambda(X(t))\Delta t, \end{aligned}$$

where the third equality follows from the fact that $Y(\int_0^t \lambda(X(s))ds)$ and $X(t)$ are part of the information in \mathcal{F}_t (are \mathcal{F}_t -measurable in the mathematical terminology) and the independence properties of Y .

More generally, a continuous time Markov chain X taking values in \mathbb{Z}^d is specified by giving its *transition intensities* (*propensities* in much of the chemical physics literature) λ_l that determine

$$P\{X(t + \Delta t) - X(t) = \zeta_l | \mathcal{F}_t^X\} \approx \lambda_l(X(t))\Delta t, \quad (2.5)$$

for the different possible jumps $\zeta_l \in \mathbb{Z}^d$, where \mathcal{F}_t^X is the σ -algebra generated by X (all the information available from the observation of the process up to time t). If we write

$$X(t) = X(0) + \sum_l \zeta_l R_l(t)$$

where $R_l(t)$ is the number of jumps of ζ_l at or before time t , then (2.5) implies

$$P\{R_l(t + \Delta t) - R_l(t) = 1 | \mathcal{F}_t^X\} \approx \lambda_l(X(t))\Delta t, \quad l \in \mathbb{Z}^d.$$

R_l is a *counting process* with intensity $\lambda_l(X(t))$ and by analogy with (2.4) we write

$$X(t) = X(0) + \sum \zeta_l Y_l \left(\int_0^t \lambda_l(X(s)) ds \right), \quad (2.6)$$

where the Y_l are independent unit Poisson processes. This equation has a unique solution by the same jump by jump argument used above provided $\sum_l \lambda_l(x) < \infty$ for all x . Unless we add additional assumptions, we cannot rule out the possibility that the solution only exists up to some finite time. For example, if $d = 1$ and $\lambda_1(k) = (1 + k)^2$, the solution of

$$X(t) = Y_1 \left(\int_0^t (1 + X(s))^2 ds \right)$$

hits infinity in finite time. To see why this is the case, compare the above equation to the ordinary differential equation

$$\dot{x}(t) = (1 + x(t))^2, \quad x(0) = 0.$$

2.3 Equivalence of stochastic equations and martingale problems

There are many ways of relating the intensities λ_l to the stochastic process X , and we will review some of these in later sections, but the stochastic equation (2.6) has the advantage of being intuitive (λ_l has a natural interpretation as a “rate”) and easily generalized to take into account such properties as external noise, in which (2.6) becomes

$$X(t) = X(0) + \sum \zeta_l Y_l \left(\int_0^t \lambda_l(X(s), Z(s)) ds \right)$$

where Z is a stochastic process independent of $X(0)$ and the Y_l , or delays, in which (2.6) becomes

$$X(t) = X(0) + \sum \zeta_l Y_l \left(\int_0^t \lambda_l(X(s), X(s - \delta)) ds \right),$$

or perhaps the λ_l become even more complicated functions of the past of X . We will also see that these stochastic equations let us exploit well-known properties of the Poisson processes Y_l to study the properties of X .

The basic building blocks of our models remain the counting processes R_l and their intensities expressed as functions of the past of the R_l and possibly some additional stochastic input independent of the Y_l (for example, the initial condition $X(0)$ or the environmental noise Z).

For the moment, we focus on a finite system of counting processes $R = (R_1, \dots, R_m)$ given as the solution of a system of equations

$$R_l(t) = Y_l \left(\int_0^t \gamma_l(s, R) ds \right), \quad (2.7)$$

where the γ_l are *nonanticipating* in the sense that

$$\gamma_l(t, R) = \gamma_l(t, R(\cdot \wedge t)), \quad t \geq 0,$$

that is, at time t , $\gamma_l(t, R)$ depends only on the past of R up to time t , and the Y_l are independent, unit Poisson processes. The independence of the Y_l ensures that only one of the R_l jumps at a time. Let τ_k be the k th jump time of R . Then any system of this form has the property that for all l and k ,

$$M_l^k(t) \equiv R_l(t \wedge \tau_k) - \int_0^{t \wedge \tau_k} \gamma_l(s, R) ds$$

is a *martingale*, that is, there exists a filtration $\{\mathcal{F}_t\}$ such that

$$E[M_l^k(t+s)|\mathcal{F}_t] = M_l^k(t), \quad t, s \geq 0.$$

Note that

$$\lim_{k \rightarrow \infty} E[R_l(t \wedge \tau_k)] = \lim_{k \rightarrow \infty} E\left[\int_0^{t \wedge \tau_k} \gamma_l(s, R) ds\right],$$

allowing $\infty = \infty$, and if the limit is finite for all l and t , then $\tau_\infty = \infty$ and for each l ,

$$M_l(t) = R_l(t) - \int_0^t \gamma_l(s, R) ds$$

is a martingale.

There is a converse to these assertions. If (R_1, \dots, R_m) are counting processes adapted to a filtration $\{\mathcal{F}_t\}$ and $(\lambda_1, \dots, \lambda_m)$ are nonnegative stochastic processes adapted to $\{\mathcal{F}_t\}$ such that for each k and l ,

$$R_l(t \wedge \tau_k) - \int_0^{t \wedge \tau_k} \lambda_l(s) ds$$

is a $\{\mathcal{F}_t\}$ -martingale, we say that λ_l is the $\{\mathcal{F}_t\}$ -intensity for R_l .

Lemma 2.4 *Assume that $R = (R_1, \dots, R_m)$ is a system of counting processes with no common jumps and λ_l is the $\{\mathcal{F}_t\}$ -intensity for R_l . Then there exist independent unit Poisson processes Y_1, \dots, Y_m (perhaps on an enlarged sample space) such that*

$$R_l(t) = Y_l\left(\int_0^t \lambda_l(s) ds\right).$$

Proof. See Meyer [40] and Kurtz [36]. □

This lemma suggests the following alternative approach to relating the intensity of a counting process to the corresponding counting process. Again, given nonnegative, nonanticipating functions γ_l , the intuitive problem is to find counting processes R_l such that

$$P\{R_l(t + \Delta t) > R_l(t) | \mathcal{F}_t\} \approx \gamma_l(t, R) \Delta t,$$

which we now translate into the following *martingale problem*. In the following definition $\mathbb{J}_m[0, \infty)$ denotes the set of m -dimensional cadlag counting paths.

Definition 2.5 Let γ_l , $l = 1, \dots, m$, be nonnegative, nonanticipating functions defined on $\mathbb{J}_m[0, \infty)$. Then a family of counting processes $R = (R_1, \dots, R_m)$ is a solution of the martingale problem for $(\gamma_1, \dots, \gamma_m)$ if the R_l have no simultaneous jumps and there exists a filtration $\{\mathcal{F}_t\}$ such that R is adapted to $\{\mathcal{F}_t\}$ and for each l and k ,

$$R_l(t \wedge \tau_k) - \int_0^{t \wedge \tau_k} \gamma_l(s, R) ds$$

is a $\{\mathcal{F}_t\}$ -martingale.

Of course, the solution of (2.7) is a solution of the martingale problem and Lemma 2.4 implies that every solution of the martingale problem can be written as a solution of the stochastic equation. Consequently, the stochastic equation and the martingale problem are equivalent ways of specifying the system of counting processes that corresponds to the γ_l . The fact that the martingale problem uniquely characterizes the system of counting processes is a special case of a theorem of Jacod [24].

2.4 Thinning of counting processes

Consider a single counting process R_0 with $\{\mathcal{F}_t\}$ -intensity λ_0 , and let $p(t, R_0)$ be a cadlag (right continuous with left limits at each $t > 0$), nonanticipating function with values in $[0, 1]$. For simplicity, assume

$$E[R_0(t)] = E\left[\int_0^t \lambda_0(s) ds\right] < \infty.$$

We want to construct a new counting process R_1 such that at each jump of R_0 , R_1 jumps with probability $p(t-, R_0)$. Perhaps the simplest construction is to let ξ_0, ξ_1, \dots be independent, uniform $[0, 1]$ random variables that are independent of R_0 and to define

$$R_1(t) = \int_0^t \mathbf{1}_{[0, p(s-, R_0)]}(\xi_{R_0(s-)}) dR_0(s).$$

Since with probability one,

$$R_1(t) = \lim_{n \rightarrow \infty} \sum_{k=0}^{\lfloor nt \rfloor} \mathbf{1}_{[0, p(\frac{k}{n}, R_0)]}(\xi_{R_0(\frac{k}{n})}) (R_0(\frac{k+1}{n}) - R_0(\frac{k}{n})),$$

where $\lfloor z \rfloor$ is the integer part of z , setting $\tilde{R}_0(t) = R_0(t) - \int_0^t \lambda_0(s) ds$, we see that

$$\begin{aligned} R_1(t) - \int_0^t \lambda_0(s) p(s, R_0) ds &= \int_0^t (\mathbf{1}_{[0, p(s-, R_0)]}(\xi_{R_0(s-)}) - p(s-, R_0)) dR_0(s) \\ &\quad + \int_0^t p(s-, R_0) d\tilde{R}_0(s) \end{aligned}$$

is a martingale (because both terms on the right are martingales). Hence, R_1 is a counting process with intensity $\lambda_0(t)p(t, R_0)$. We could also define

$$R_2(t) = \int_0^t \mathbf{1}_{(p(s-, R_0), 1]}(\xi_{R_0(s-)}) dR_0(s),$$

so that R_1 and R_2 would be counting processes without simultaneous jumps having intensities $\lambda_0(t)p(t, R_0)$ and $\lambda_0(t)(1 - p(t, R_0))$.

Note that we could let p be a nonanticipating function of both R_0 and R_1 , or equivalently, R_1 and R_2 . With that observation in mind, let $\gamma_0(t, R)$ be a nonnegative, nonanticipating function of $R = (R_1, \dots, R_m)$, and let $p_l(t, R)$, $l = 1, \dots, m$, be cadlag nonnegative, nonanticipating functions satisfying $\sum_{l=1}^m p_l(t, R) \equiv 1$. Let Y be a unit Poisson process and ξ_0, ξ_1, \dots be independent, uniform $[0, 1]$ random variables that are independent of Y , and set $q_0 = 0$ and for $1 \leq l \leq m$ set $q_l(t, R) = \sum_{i=1}^l p_i(t, R)$. Now consider the system

$$R_0(t) = Y\left(\int_0^t \gamma_0(s, R) ds\right) \quad (2.8)$$

$$R_l(t) = \int_0^t \mathbf{1}_{(q_{l-1}(s-, R), q_l(s-, R)]}(\xi_{R_0(s-)}) dR_0(s). \quad (2.9)$$

Then $R = (R_1, \dots, R_m)$ is a system of counting processes with intensities $\lambda_l(t) = \gamma_0(t, R)p_l(t, R)$.

If, as in the time-change equation (2.7) and the equivalent martingale problem described in Definition 2.5, we start with intensities $\gamma_1, \dots, \gamma_m$, we can define

$$\gamma_0(t, R) = \sum_{l=1}^m \gamma_l(t, R), \quad p_l(t, R) = \frac{\gamma_l(t, R)}{\gamma_0(t, R)},$$

and the solution of the system (2.8) and (2.9) will give a system of counting processes with the same distribution as the solution of the time-change equation or the martingale problem. Specializing to continuous-time Markov chains and defining

$$\lambda_0(x) = \sum_l \lambda_l(x), \quad q_l(x) = \sum_{i=1}^l \lambda_i(x) / \lambda_0(x),$$

the equations become

$$\begin{aligned} R_0(t) &= Y\left(\int_0^t \lambda_0(X(s)) ds\right) \\ X(t) &= X(0) + \sum_l \zeta_l \int_0^t \mathbf{1}_{(q_{l-1}(X(s-)), q_l(X(s-))]}(\xi_{R_0(s-)}) dR_0(s). \end{aligned} \quad (2.10)$$

2.5 The martingale problem and forward equation for Markov chains

Let X satisfy (2.6), and for simplicity, assume that $\tau_\infty = \infty$, that only finitely many of the λ_l are not identically zero, and that

$$E[R_l(t)] = E\left[\int_0^t \lambda_l(X(s)) ds\right] < \infty, \quad l = 1, \dots, m.$$

Then for f a bounded function on \mathbb{Z}^d ,

$$f(X(t)) = f(X(0)) + \sum_l \int_0^t (f(X(s-) + \zeta_l) - f(X(s))) dR_l(t)$$

and defining

$$\tilde{R}_l(t) = R_l(t) - \int_0^t \lambda_l(X(s))ds,$$

we see that

$$\begin{aligned} f(X(t)) - f(X(0)) - \int_0^t \sum_l \lambda_l(X(s))(f(X(s) + \zeta_l) - f(X(s)))ds \\ = \sum_l \int_0^t (f(X(s-) + \zeta_l) - f(X(s)))d\tilde{R}_l(t) \end{aligned}$$

is a martingale.

Define

$$Af(x) = \sum_l \lambda_l(x)(f(x + \zeta_l) - f(x)).$$

Allowing $\tau_\infty < \infty$, define $X(t) = \infty$ for $t \geq \tau_\infty$. If $\tau_\infty < \infty$,

$$\lim_{k \rightarrow \infty} |X(\tau_k)| = \infty,$$

and this definition gives a “continuous” extension of X to the time interval $[0, \infty)$. Let f satisfy $f(x) = 0$ for $|x|$ sufficiently large, and define $f(\infty) = 0$. Then for any solution of (2.6),

$$f(X(t)) - f(X(0)) - \int_0^t Af(X(s))ds \tag{2.11}$$

is a martingale.

Definition 2.6 *A right continuous, $\mathbb{Z}^d \cup \{\infty\}$ -valued stochastic process X is a solution of the martingale problem for A if there exists a filtration $\{\mathcal{F}_t\}$ such that for each f satisfying $f(x) = 0$ for $|x|$ sufficiently large, (2.11) is a $\{\mathcal{F}_t\}$ -martingale. X is a minimal solution, if in addition, $X(t) = \infty$ for $t \geq \tau_\infty$.*

The following lemma follows from Lemma 2.4.

Lemma 2.7 *If X is a minimal solution of the martingale problem for A , then there exist independent unit Poisson processes Y_l (perhaps on an enlarged sample space) such that*

$$R_l(t) = Y_l\left(\int_0^t \lambda_l(X(s))ds\right).$$

The martingale property implies

$$E[f(X(t))] = E[f(X(0))] + \int_0^t E[Af(X(s))]ds$$

and taking $f(x) = \mathbf{1}_{\{y\}}(x)$, we have

$$P\{X(t) = y\} = P\{X(0) = y\} + \int_0^t \left(\sum_l \lambda_l(y - \zeta_l) P\{X(s) = y - \zeta_l\} - \sum_l \lambda_l(y) P\{X(s) = y\} \right) ds$$

giving the Kolmogorov forward or *master equation* for the distribution of X . In particular, defining $p_y(t) = P\{X(t) = y\}$ and $\nu_y = P\{X(0) = y\}$, $\{p_y\}$ satisfies the system of differential equations

$$\dot{p}_y(t) = \sum_l \lambda_l(y - \zeta_l) p_{y-\zeta_l}(t) - \left(\sum_l \lambda_l(y) \right) p_y(t), \quad (2.12)$$

with initial condition $p_y(0) = \nu_y$.

Lemma 2.8 *Let $\{\nu_y\}$ be a probability distribution on \mathbb{Z}^d , and let $X(0)$ satisfy $P\{X(0) = y\} = \nu_y$. The system of differential equations (2.12) has a unique solution satisfying $p_y(0) = \nu_y$ and $\sum_y p_y(t) \equiv 1$ if and only if the solution of (2.6) satisfies $\tau_\infty = \infty$.*

2.6 Simulation

The stochastic equations (2.6) and (2.10) suggest methods of simulating continuous-time Markov chains, and these methods are, in fact, well known. Equation (2.6) corresponds to the *next reaction* (next jump) method as defined by Gibson and Bruck [19].

The algorithm obtained by simulating (2.10) is known variously as the *embedded chain method* or Gillespie's [20, 21] *direct method* or the *stochastic simulation algorithm* (SSA).

If we define an Euler-type approximation for (2.6), that is, for $0 = \tau_0 < \tau_1 < \dots$, recursively define

$$\widehat{X}(\tau_n) = X(0) + \sum_l \zeta_l Y_l \left(\sum_{k=0}^{n-1} \lambda_l(\widehat{X}(\tau_k)) (\tau_{k+1} - \tau_k) \right),$$

we obtain Gillespie's [23] τ -leap method.

2.7 Stationary distributions

We restrict our attention to continuous-time Markov chains for which $\tau_\infty = \infty$ for all initial values and hence, given $X(0)$, the process is uniquely determined as a solution of (2.6), (2.10), or the martingale problem given by Definition 2.6, and the one-dimensional distributions are uniquely determined by (2.12). A probability distribution π is called a stationary distribution for the Markov chain if $X(0)$ having distribution π implies X is a stationary process, that is, for each choice of $0 \leq t_1 < \dots < t_k$, the joint distribution of

$$(X(t + t_1), \dots, X(t + t_k))$$

does not depend on t .

If $X(0)$ has distribution π , then since $E[f(X(0))] = E[f(X(t))] = \sum_x f(x)\pi(x)$, the martingale property for (2.11) implies

$$0 = E\left[\int_0^t Af(X(s)) ds \right] = t \sum_x Af(x)\pi(x),$$

and as in the derivation of (2.12),

$$\sum_l \lambda_l(y - \zeta_l)\pi(y - \zeta_l) - \left(\sum_l \lambda_l(y) \right)\pi(y) = 0.$$

3 Reaction networks

We consider a *network* of r_0 chemical reactions involving s_0 chemical species, S_1, \dots, S_{s_0} ,



where the ν_{ik} and ν'_{ik} are nonnegative integers. Let the components of $X(t)$ give the numbers of molecules of each species in the system at time t . Let ν_k be the vector whose i th component is ν_{ik} , the number of molecules of the i th chemical species consumed in the k th reaction, and let ν'_k be the vector whose i th component is ν'_{ik} , the number of molecules of the i th species produced by the k th reaction. Let $\lambda_k(x)$ be the rate at which the k th reaction occurs, that is, it gives the propensity/intensity of the k th reaction as a function of the numbers of molecules of the chemical species.

If the k th reaction occurs at time t , the new state becomes

$$X(t) = X(t-) + \nu'_k - \nu_k.$$

The number of times that the k th reaction occurs by time t is given by the counting process satisfying

$$R_k(t) = Y_k\left(\int_0^t \lambda_k(X(s)) ds\right),$$

where the Y_k are independent unit Poisson processes. The state of the system then satisfies

$$\begin{aligned} X(t) &= X(0) + \sum_k R_k(t)(\nu'_k - \nu_k) \\ &= X(0) + \sum_k Y_k\left(\int_0^t \lambda_k(X(s)) ds\right)(\nu'_k - \nu_k). \end{aligned}$$

To simplify notation, we will write

$$\zeta_k = \nu'_k - \nu_k.$$

3.1 Rates for the law of mass action

The stochastic form of the law of mass action says that the rate at which a reaction occurs should be proportional to the number of distinct subsets of the molecules present that can form the inputs for the reaction. Intuitively, the mass action assumption reflects the idea that the system is well-stirred in the sense that all molecules are equally likely to be at any location at any time. For example, for a binary reaction $S_1 + S_2 \rightarrow S_3$ or $S_1 + S_2 \rightarrow S_3 + S_4$,

$$\lambda_k(x) = \kappa_k x_1 x_2,$$

where κ_k is a rate constant. For a unary reaction $S_1 \rightarrow S_2$ or $S_1 \rightarrow S_2 + S_3$, $\lambda_k(x) = \kappa_k x_1$. For $2S_1 \rightarrow S_2$, $\lambda_k(x) = \kappa_k x_1(x_1 - 1)$.

For a binary reaction $S_1 + S_2 \rightarrow S_3$, the rate should vary inversely with volume, so it would be better to write

$$\lambda_k^N(x) = \kappa_k N^{-1} x_1 x_2 = N \kappa_k z_1 z_2,$$

where classically, N is taken to be the volume of the system times Avogadro's number and $z_i = N^{-1}x_i$ is the concentration in moles per unit volume. For $2S_1 \rightarrow S_2$, since N is very large,

$$\frac{1}{N}\kappa_k x_1(x_1 - 1) = N\kappa_k z_1(z_1 - \frac{1}{N}) \approx N\kappa_k z_1^2.$$

Note that unary reaction rates also satisfy

$$\lambda_k(x) = \kappa_k x_i = N\kappa_k z_i.$$

Although, reactions of order higher than binary may not be physical, if they were, the analogous form for the intensity would be

$$\lambda_k^N(x) = \kappa_k \frac{\prod_i \nu_{ik}!}{N^{|\nu_k|-1}} \prod_i \binom{x_i}{\nu_{ik}} = N\kappa_k \frac{\prod_i \nu_{ik}!}{N^{|\nu_k|}} \prod \binom{x_i}{\nu_{ik}},$$

where $|\nu_k| = \sum_i \nu_{ik}$. Again $z = N^{-1}x$ gives the concentrations in moles per unit volume, and

$$\lambda_k^N(x) \approx N\kappa_k \prod_i z_i^{\nu_{ik}} \equiv N\tilde{\lambda}_k(z), \quad (3.1)$$

where $\tilde{\lambda}_k$ is the usual deterministic form of mass action kinetics.

3.2 General form for the classical scaling

Setting $C^N(t) = N^{-1}X(t)$ and using (3.1)

$$\begin{aligned} C^N(t) &= C^N(0) + \sum_k N^{-1}Y_k \left(\int_0^t \lambda_k^N(X(s)) ds \right) \zeta_k \\ &\approx C^N(0) + \sum_k N^{-1}Y_k \left(N \int_0^t \tilde{\lambda}_k(C^N(s)) ds \right) \zeta_k \\ &= C^N(0) + \sum_k N^{-1}\tilde{Y}_k \left(N \int_0^t \tilde{\lambda}_k(C^N(s)) ds \right) \zeta_k + \int_0^t F(C^N(s)) ds, \end{aligned}$$

where $\tilde{Y}_k(u) = Y_k(u) - u$ is the centered process and

$$F(z) \equiv \sum_k \kappa_k \prod_i z_i^{\nu_{ik}} \zeta_k.$$

The law of large numbers for the Poisson process, Lemma 2.2, implies $N^{-1}\tilde{Y}(Nu) \approx 0$, so

$$C^N(t) \approx C^N(0) + \sum_k \int_0^t \kappa_k \prod_i C_i^N(s)^{\nu_{ik}} \zeta_k ds = C^N(0) + \int_0^t F(C^N(s)) ds,$$

which in the limit as $N \rightarrow \infty$ gives the classical deterministic law of mass action

$$\dot{C}(t) = \sum_k \kappa_k \prod_i C_i(t)^{\nu_{ik}} \zeta_k = F(C(t)). \quad (3.2)$$

(See [32, 34, 35].)

Since by (2.3),

$$\frac{1}{\sqrt{N}}\tilde{Y}_k(Nu) = \frac{Y_k(Nu) - Nu}{\sqrt{N}}$$

is approximately a Brownian motion,

$$\begin{aligned} V^N(t) &\equiv \sqrt{N}(C^N(t) - C(t)) \\ &\approx V^N(0) + \sqrt{N}\left(\sum_k \frac{1}{N}Y_k(N) \int_0^t \tilde{\lambda}_k(C^N(s))ds\right)\zeta_k - \int_0^t F(C(s))ds \\ &= V^N(0) + \sum_k \frac{1}{\sqrt{N}}\tilde{Y}_k(N) \int_0^t \tilde{\lambda}_k(C^N(s))ds\zeta_k + \int_0^t \sqrt{N}(F(C^N(s)) - F(C(s)))ds \\ &\approx V^N(0) + \sum_k W_k\left(\int_0^t \tilde{\lambda}_k(C(s))ds\right)\zeta_k + \int_0^t \nabla F(C(s))V^N(s)ds, \end{aligned}$$

where the second approximation follows from (3.1), and the limit as N goes to infinity gives $V^N \Rightarrow V$ where

$$V(t) = V(0) + \sum_k W_k\left(\int_0^t \tilde{\lambda}_k(C(s))ds\right)\zeta_k + \int_0^t \nabla F(C(s))V(s)ds. \quad (3.3)$$

(See [33, 35, 42] and Chapter 11 of [15].) This limit suggests the approximation

$$C^N(t) \approx \hat{C}_N(t) \equiv C(t) + \frac{1}{\sqrt{N}}V(t). \quad (3.4)$$

Since (3.3) is a linear equation driven by a Gaussian process, V is Gaussian as is \hat{C}_N .

3.3 Diffusion/Langevin approximations

The first steps in the argument in the previous section suggest simply replacing the rescaled centered Poisson processes $\frac{1}{\sqrt{N}}\tilde{Y}_k(N\cdot)$ by independent Brownian motions and considering a solution of

$$D^N(t) = D^N(0) + \sum_k \frac{1}{\sqrt{N}}W_k\left(\int_0^t \tilde{\lambda}_k(D^N(s))ds\right)\zeta_k + \int_0^t F(D^N(s))ds \quad (3.5)$$

as a possible approximation for C^N . Unfortunately, even though only ordinary integrals appear in this equation, the theory of the equation is not quite as simple as it looks. Unlike (2.6) where uniqueness of solutions is immediate, no general uniqueness theorem is known for (3.5) without an additional requirement on the solution. In particular, setting

$$\tau_k^N(t) = \int_0^t \tilde{\lambda}_k(D^N(s))ds,$$

we must require that the solution D^N is *compatible* with the Brownian motions W_k in the sense that $W_k(\tau_k^N(t) + u) - W_k(\tau_k^N(t))$ is independent of $\mathcal{F}_t^{D^N}$ for all $k, t \geq 0$, and $u \geq 0$.

This requirement is intuitively natural and is analogous to the requirement that a solution of an Itô equation be *nonanticipating*. In fact, we have the following relationship between (3.5) and a corresponding Itô equation.

Lemma 3.1 *If D^N is a compatible solution of (3.5), then there exist independent standard Brownian motions B_k (perhaps on an enlarged sample space) such that D^N is a solution of the Itô equation*

$$D^N(t) = D^N(0) + \sum_k \frac{1}{\sqrt{N}} \int_0^t \sqrt{\tilde{\lambda}(D^N(s))} dB_k(s) \zeta_k + \int_0^t F(D^N(s)) ds. \quad (3.6)$$

Proof. See [35, 36] and Chapter 11 of [15]. For a general discussion of compatibility, see [37], in particular, Example 3.20. \square

In the chemical physics literature, D^N is known as the Langevin approximation for the continuous-time Markov chain model determined by the master equation. Just as there are alternative ways of determining the continuous-time Markov chain model, there are alternative approaches to deriving the Langevin approximation. For example, C^N is a solution of the martingale problem corresponding to

$$A_N f(x) = \sum_k N \lambda_k(x) (f(x + N^{-1} \zeta_k) - f(x)),$$

and if f is three times continuously differentiable with compact support,

$$A_N f(x) = L_N f(x) + O(N^{-2}),$$

where

$$L_N f(x) = \frac{1}{2N} \sum_k \zeta_k^\top \partial^2 f(x) \zeta_k^\top + F(x) \cdot \nabla f(x),$$

and any compatible solution of (3.5) is a solution of the martingale problem for L_N , that is, there is a filtration $\{\mathcal{F}_t^N\}$ such that

$$f(D^N(t)) - f(D^N(0)) - \int_0^t L_N f(D^N(s)) ds$$

is a $\{\mathcal{F}_t^N\}$ -martingale for each twice continuously differentiable function having compact support. The converse also holds, that is, any solution of the martingale problem for L_N that does not hit infinity in finite time can be obtained as a compatible solution of (3.5) or equivalently, as a solution of (3.6).

Finally, the Langevin approximation can be derived starting with the master equation. First rewrite (2.12) as

$$\dot{p}^N(y, t) = \sum_l N \lambda_l(y - N^{-1} \zeta_l) p^N(y - N^{-1} \zeta_l, t) - \left(\sum_l N \lambda_l(y) \right) p^N(y, t), \quad (3.7)$$

where now

$$p^N(y, t) = P\{C^N(t) = y\}.$$

Expanding $\lambda_l(y - N^{-1}\zeta_l)p^N(y - N^{-1}\zeta_l)$ in a Taylor series (the Kramers-Moyal expansion, or in this context, the system-size expansion of van Kampen; see [42]) and discarding higher order terms gives

$$\dot{p}^N(y, t) \approx \frac{1}{2N} \sum_l \zeta_l^\top \partial^2 (\lambda_l(y) p^N(y, t)) \zeta_k - \sum_l \zeta_l \cdot \nabla (\lambda_l(y) p^N(y, t)).$$

Replacing \approx by $=$ gives the Fokker-Planck equation

$$\dot{q}^N(y, t) = \frac{1}{2N} \sum_l \zeta_l^\top \partial^2 (\lambda_l(y) q^N(y, t)) \zeta_k - \sum_l \zeta_l \cdot \nabla (\lambda_l(y) q^N(y, t))$$

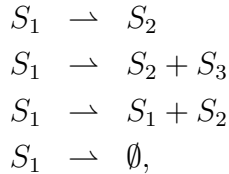
corresponding to (3.6). These three derivations are equivalent in the sense that any solution of the Fokker-Planck equation for which $q^N(\cdot, t)$ is a probability density for all t gives the one-dimensional distributions of a solution of the martingale problem for L_N , and as noted before, any solution of the martingale problem that does not hit infinity in finite time can be obtained as a solution of (3.6) or (3.5). See [38] for a more detailed discussion.

The approximation (3.4) is justified by the convergence of V^N to V , but the justification for taking D^N as an approximation of C^N is less clear. One can, however, apply the strong approximation result, Lemma 2.3, to construct D^N and C^N in such a way that in a precise sense, for each $T > 0$,

$$\sup_{t \leq T} |D^N(t) - C^N(t)| = \mathcal{O}\left(\frac{\log N}{N}\right).$$

3.4 First order reaction networks

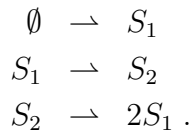
If all reactions in the network are unary, for example,



then the resulting process is a multitype branching process, and if reactions of the form



are included, the process is a branching process with immigration. Networks that only include the above reaction types are termed first order reaction networks. For simplicity, first consider the system



The stochastic equation for the model becomes

$$X(t) = X(0) + Y_1(\kappa_1 t) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + Y_2(\kappa_2 \int_0^t X_1(s) ds) \begin{pmatrix} -1 \\ 1 \end{pmatrix} + Y_3(\kappa_3 \int_0^t X_2(s) ds) \begin{pmatrix} 2 \\ -1 \end{pmatrix},$$

for some choice of $\kappa_1, \kappa_2, \kappa_3 > 0$. Using the fact that $E[Y_k(\int_0^t \lambda_k(s) ds)] = E[\int_0^t \lambda_k(s) ds]$, we have

$$\begin{aligned} E[X(t)] &= E[X(0)] + \begin{pmatrix} \kappa_1 \\ 0 \end{pmatrix} t + \int_0^t \kappa_2 E[X_1(s)] ds \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \kappa_3 \int_0^t E[X_2(s)] ds \begin{pmatrix} 2 \\ -1 \end{pmatrix} \\ &= E[X(0)] + \begin{pmatrix} \kappa_1 \\ 0 \end{pmatrix} t + \int_0^t \begin{pmatrix} -\kappa_2 & 2\kappa_3 \\ \kappa_2 & -\kappa_3 \end{pmatrix} E[X(s)] ds \end{aligned}$$

giving a simple linear system for the first moments, $E[X(t)]$. For the second moments, note that

$$X(t)X(t)^\top = X(0)X(0)^\top + \int_0^t X(s-)dX(s)^\top + \int_0^t dX(s)X(s-)^\top + [X]_t,$$

where $[X]_t$ is the matrix of quadratic variations which in this case is simply

$$[X]_t = Y_1(\kappa_1 t) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + Y_2(\kappa_2 \int_0^t X_1(s) ds) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + Y_3(\kappa_3 \int_0^t X_2(s) ds) \begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix}.$$

Since

$$X(t) - X(0) - \kappa_1 t \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \kappa_2 \int_0^t X_1(s) ds \begin{pmatrix} -1 \\ 1 \end{pmatrix} - \kappa_3 \int_0^t X_2(s) ds \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

is a martingale,

$$\begin{aligned} &E[X(t)X(t)^\top] \\ &= E[X(0)X(0)^\top] + \int_0^t E \left[X(s) \left(\begin{pmatrix} \kappa_1 & 0 \end{pmatrix} + X(s)^\top \begin{pmatrix} -\kappa_2 & 2\kappa_3 \\ \kappa_2 & -\kappa_3 \end{pmatrix} \right)^\top \right] ds \\ &\quad + \int_0^t E \left[\left(\begin{pmatrix} \kappa_1 \\ 0 \end{pmatrix} + \begin{pmatrix} -\kappa_2 & 2\kappa_3 \\ \kappa_2 & -\kappa_3 \end{pmatrix} X(s) \right) X(s)^\top \right] ds \\ &\quad + \begin{pmatrix} \kappa_1 & 0 \\ 0 & 0 \end{pmatrix} t + \int_0^t \left(\kappa_2 E[X_1(s)] \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + \kappa_3 E[X_2(s)] \begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix} \right) ds \\ &= E[X(0)X(0)^\top] + \int_0^t \left(E[X(s)] \begin{pmatrix} \kappa_1 & 0 \end{pmatrix} + \begin{pmatrix} \kappa_1 \\ 0 \end{pmatrix} E[X(s)^\top] \right) ds \\ &\quad + \int_0^t \left(E[X(s)X(s)^\top] \begin{pmatrix} -\kappa_2 & 2\kappa_3 \\ \kappa_2 & -\kappa_3 \end{pmatrix}^\top + \begin{pmatrix} -\kappa_2 & 2\kappa_3 \\ \kappa_2 & -\kappa_3 \end{pmatrix} E[X(s)X(s)^\top] \right) ds \\ &\quad + \begin{pmatrix} \kappa_1 & 0 \\ 0 & 0 \end{pmatrix} t + \int_0^t \left(\kappa_2 E[X_1(s)] \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + \kappa_3 E[X_2(s)] \begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix} \right) ds. \end{aligned}$$

In general, the stochastic equation for first order networks will be of the form

$$X(t) = X(0) + \sum_k Y_k^0(\alpha_k^0 t) \zeta_k^0 + \sum_{l=1}^{s_0} \sum_k Y_k^l(\alpha_k^l \int_0^t X_l(s) ds) \zeta_k^l,$$

where all components of ζ_k^0 are nonnegative and all components of ζ_k^l are nonnegative except for the possibility that the l th component of ζ_k^l may be -1 . The martingale properties of the Y_k^l imply that the expectation of X satisfies

$$E[X(t)] = E[X(0)] + at + \int_0^t AE[X(s)]ds \quad (3.8)$$

where $a = \sum_k \alpha_k^0 \zeta_k^0$ and A is the matrix whose l th column is $A_l = \sum_k \alpha_k^l \zeta_k^l$. Note that the solution of (3.8) is given by

$$E[X(t)] = e^{At}E[X(0)] + \int_0^t e^{A(t-s)}a ds,$$

and if A is invertible

$$E[X(t)] = e^{At}E[X(0)] + A^{-1}(e^{At} - I)a, \quad (3.9)$$

where I is the identity matrix.

Similarly to before, the matrix of second moments satisfies

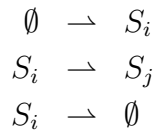
$$\begin{aligned} E[X(t)X(t)^\top] &= E[X(0)X(0)^\top] + \int_0^t (E[X(s)]a^\top + aE[X(s)]^\top)ds \\ &\quad + \int_0^t (AE[X(s)X(s)^\top] + E[X(s)X(s)^\top]A^\top)ds \\ &\quad + B_0t + \sum_l \int_0^t E[X_l(s)]B_l ds, \end{aligned}$$

where

$$B_0 = \sum_k \alpha_k^0 \zeta_k^0 \zeta_k^{0\top}, \quad B_l = \sum_k \alpha_k^l \zeta_k^l \zeta_k^{l\top}.$$

See [3], Section V.7.

A system that only includes reactions of the form



can be interpreted as an infinite server queueing network, with $\emptyset \rightarrow S_i$ corresponding to an ‘‘arrival’’, $S_i \rightarrow \emptyset$, a ‘‘departure’’, and $S_i \rightarrow S_j$ the movement of a ‘‘customer’’ from station i to station j . Customers (molecules) that start in or enter the system move (change type) independently until they leave the system. This independence implies that if $\{X_i(0)\}$ are independent Poisson distributed random variables, then $\{X_i(t)\}$ are independent Poisson distributed random variables for all $t \geq 0$. Since the Poisson distribution is determined by its expectation, under the assumption of an independent Poisson initial distribution, the distribution of $X(t)$ is determined by $E[X(t)]$, that is, by the solution of (3.8).

Suppose that for each pair of species S_i and S_j , it is possible for a molecule of S_i to be converted, perhaps through a sequence of intermediate steps, to a molecule of S_j . In

addition, assume that the system is open in the sense that there is at least one reaction of the form $\emptyset \rightarrow S_i$ and one reaction of the form $S_j \rightarrow \emptyset$. Then A is invertible, so $E[X(t)]$ is given by (3.9), and as $t \rightarrow \infty$, $e^{At} \rightarrow 0$ so $E[X(t)] \rightarrow -A^{-1}a$. It follows that the stationary distribution for X is given by a vector \bar{X} of independent Poisson distributed random variables with $E[\bar{X}] = -A^{-1}a$.

If the system is closed so that the only reactions are of the form $S_i \rightarrow S_j$ and the initial distribution is multinomial with parameters $(n, p_1(0), \dots, p_{s_0}(0))$, that is, for $k = (k_1, \dots, k_{s_0})$ with $\sum_i k_i = n$,

$$P\{X(0) = k\} = \binom{n}{k_1, \dots, k_{s_0}} \prod p_i(0)^{k_i},$$

then $X(t)$ has a multinomial distribution with parameters $(n, p_1(t), \dots, p_{s_0}(t))$, where $p(t) = (p_1(t), \dots, p_{s_0}(t))$ is given by

$$p(t) = e^{At}p(0).$$

Note that if the intensity for the reaction $S_i \rightarrow S_j$ is $\kappa_{ij}X_i(t)$, then the model is equivalent to n independent continuous-time Markov chains with state space $\{1, \dots, s_0\}$ and transition intensities given by the κ_{ij} . Consequently, if the independent chains have the same initial distribution, $p(0) = (p_1(0), \dots, p_{s_0}(0))$, then they have the same distribution at time t , namely $p(t)$. The multinomial distribution with parameters (n, \bar{p}) with $\bar{p} = \lim_{t \rightarrow \infty} p(t)$ will be a stationary distribution, but \bar{p} is not unique unless the assumption that every chemical species S_i can be converted into every other chemical species S_j holds.

See [18] for additional material on first order networks.

3.5 Product form stationary distributions

The Poisson and multinomial stationary distributions discussed above for unary systems are special cases of what are known as *product form stationary distributions* in the queueing literature. As noted in Chapter 8 of [28] and discussed in detail in [2], a much larger class of reaction networks also has product form stationary distributions. In fact, stochastic models of reaction networks that satisfy the conditions of the zero deficiency theorem of Feinberg [16] from deterministic reaction network theory have this property.

Let $\mathcal{S} = \{S_i : i = 1, \dots, s_0\}$ denote the collection of chemical species, $\mathcal{C} = \{\nu_k, \nu'_k : k = 1, \dots, r_0\}$ the collection of *complexes*, that is, the vectors that give either the inputs or the outputs of a reaction, and $\mathcal{R} = \{\nu_k \rightarrow \nu'_k : k = 1, \dots, r_0\}$ the collection of reactions. The triple, $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ determines the reaction network.

Definition 3.2 *A chemical reaction network, $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$, is called weakly reversible if for any reaction $\nu_k \rightarrow \nu'_k$, there is a sequence of directed reactions beginning with ν'_k as a source complex and ending with ν_k as a product complex. That is, there exist complexes ν_1, \dots, ν_r such that $\nu'_k \rightarrow \nu_1, \nu_1 \rightarrow \nu_2, \dots, \nu_r \rightarrow \nu_k \in \mathcal{R}$. A network is called reversible if $\nu'_k \rightarrow \nu_k \in \mathcal{R}$ whenever $\nu_k \rightarrow \nu'_k \in \mathcal{R}$.*

Let \mathcal{G} be the directed graph with nodes given by the complexes \mathcal{C} and directed edges given by the reactions $\mathcal{R} = \{\nu_k \rightarrow \nu'_k\}$, and let $\mathcal{G}_1, \dots, \mathcal{G}_\ell$ denote the connected components of \mathcal{G} . $\{\mathcal{G}_j\}$ are the *linkage classes* of the reaction network. Note that a reaction network is weakly reversible if and only if the linkage classes are strongly connected.

Definition 3.3 $S = \text{span}_{\{\nu_k \rightarrow \nu'_k \in \mathcal{R}\}} \{\nu'_k - \nu_k\}$ is the stoichiometric subspace of the network. For $c \in \mathbb{R}^{s_0}$ we say $c + S$ and $(c + S) \cap \mathbb{R}_{>0}^{s_0}$ are the stoichiometric compatibility classes and positive stoichiometric compatibility classes of the network, respectively. Denote $\dim(S) = s$.

Definition 3.4 The deficiency of a chemical reaction network, $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$, is $\delta = |\mathcal{C}| - \ell - s$, where $|\mathcal{C}|$ is the number of complexes, ℓ is the number of linkage classes, and s is the dimension of the stoichiometric subspace.

For $x, c \in \mathbb{Z}_{\geq 0}^{s_0}$, we define $c^x \equiv \prod_{i=1}^{s_0} c_i^{x_i}$, where we interpret $0^0 = 1$, and $x! \equiv \prod_{i=1}^{s_0} x_i!$. If for each complex $\eta \in \mathcal{C}$, $c \in \mathbb{R}_{>0}^{s_0}$ satisfies

$$\sum_{k:\nu_k=\eta} \kappa_k c^{\nu_k} = \sum_{k:\nu'_k=\eta} \kappa_k c^{\nu'_k}, \quad (3.10)$$

where the sum on the left is over reactions for which η is the source complex and the sum on the right is over those for which η is the product complex, then c is a special type of equilibrium of the system (you can see this by summing each side of (3.10) over the complexes), and the network is called *complex balanced*. The following is the Deficiency Zero Theorem of Feinberg [16].

Theorem 3.5 Let $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ be a weakly reversible, deficiency zero chemical reaction network governed by deterministic mass action kinetics, (3.2). Then, for any choice of rate constants κ_k , within each positive stoichiometric compatibility class there is precisely one equilibrium value c , that is $\sum_k \kappa_k c^{\nu_k} (\nu'_k - \nu_k) = 0$, and that equilibrium value is locally asymptotically stable relative to its compatibility class. Moreover, for each $\eta \in \mathcal{C}$,

$$\sum_{k:\nu_k=\eta} \kappa_k c^{\nu_k} = \sum_{k:\nu'_k=\eta} \kappa_k c^{\nu'_k}. \quad (3.11)$$

For stochastically modeled systems we have the following theorem.

Theorem 3.6 Let $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ be a chemical reaction network with rate constants κ_k . Suppose that the deterministically modeled system is complex balanced with equilibrium $\bar{c} \in \mathbb{R}_{>0}^m$. Then, for any irreducible communicating equivalence class, Γ , the stochastic system has a product form stationary measure

$$\pi(x) = M \frac{\bar{c}^x}{x!}, \quad x \in \Gamma, \quad (3.12)$$

where M is a normalizing constant.

Theorem 3.5 then shows that the conclusion of Theorem 3.6 holds, regardless of the choice of rate constants, for all stochastically modeled systems with a reaction network that is weakly reversible and has a deficiency of zero.

3.6 Models with delay

Modeling chemical reaction networks as continuous-time Markov chains is intuitively appealing and, as noted, consistent with the classical deterministic law of mass action. Cellular reaction networks, however, include reactions for which the exponential timing of the simple Markov chain model is almost certainly wrong. These networks typically involve assembly processes (transcription or translation), referred to as elongation, in which an enzyme or ribosome follows a DNA or RNA template to create a new DNA, RNA, or protein molecule. The exponential holding times in the Markov chain model reflect an assumption that once the molecules come together in the right configuration, the time it takes to complete the reaction is negligible. That is not, in general, the case for elongation. While each step of the assembly process might reasonably be assumed to take an exponentially distributed time, the total time is a sum of such steps with the number of summands equal to the number of nucleotides or amino acids. Since this number is large and essentially fixed, if the individual steps have small expectations, the total time that the reaction takes once the assembly is initiated may be closer to deterministic than exponential. See [5, 8] for examples of stochastic models of cellular reaction networks with delays.

One reasonable (though by no means only) way to incorporate delays into the models is to assume that for a reaction with deterministic delay ξ_k that initiates at time t^* the input molecules are lost at time t^* and the product molecules are produced at time $t^* + \xi_k$. Noting that the number of initiations of a reaction by time t can still be modeled by the counting process $Y_k(\int_0^t \lambda_k(X(s))ds)$, we may let Γ_1 denote those reactions with no delay and Γ_2 those with a delay, and conclude that the system should satisfy the equation

$$\begin{aligned}
 X(t) = & X(0) + \sum_{k \in \Gamma_1} Y_{k,1}(\int_0^t \lambda_k(X(s))ds)(\nu'_k - \nu_k) \\
 & - \sum_{k \in \Gamma_2} Y_{k,2}(\int_0^t \lambda_k(X(s))ds)\nu_k + \sum_{k \in \Gamma_2} Y_{k,2}(\int_0^{t-\xi_k} \lambda_k(X(s))ds)\nu'_k,
 \end{aligned}$$

where we take $X(s) \equiv 0$, and hence $\lambda_k(X(s)) \equiv 0$, for $s < 0$. Existence and uniqueness of solutions to this equation follow by the same jump by jump argument used in Section 2.2.

Simulation of reaction networks modeled with delay is no more difficult than simulating those without delay. For example, the above equation suggests a simulation strategy equivalent to the next reaction method [1, 19]. There are also analogues of the stochastic simulation algorithm, or Gillespie's algorithm [8].

4 Multiple scales

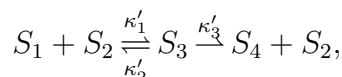
The classical scaling that leads to the deterministic law of mass action assumes that all chemical species are present in numbers of the same order of magnitude. For reaction networks in biological cells, this assumption is usually clearly violated. Consequently, models derived by the classical scaling may not be appropriate. For these networks some species are present in such small numbers that they should be modeled by discrete variables while others are present in large enough numbers to reasonably be modeled by continuous variables.

These large numbers may still differ by several orders of magnitude, so normalizing all “large” quantities in the same way may still be inappropriate. Consequently, methods are developed in [4], [26], and [27] for deriving simplified models in which different species numbers are normalized in different ways appropriate to their numbers in the system.

4.1 Derivation of the Michaelis-Menten equation

Perhaps the best known examples of reaction networks in which multiple scales play a role are models that lead to the Michaelis-Menten equation. Darden [9, 10] gave a derivation starting from a stochastic model, and we prove his result using our methodology.

Consider the reaction system



where S_1 is the substrate, S_2 the enzyme, S_3 the enzyme-substrate complex, and S_4 the product. Assume that the parameters scale so that

$$\begin{aligned} Z_1^N(t) &= Z_1^N(0) - N^{-1}Y_1(N \int_0^t \kappa_1 Z_1^N(s) Z_2^N(s) ds) + N^{-1}Y_2(N \int_0^t \kappa_2 Z_3^N(s) ds) \\ Z_2^N(t) &= Z_2^N(0) - Y_1(N \int_0^t \kappa_1 Z_1^N(s) Z_2^N(s) ds) + Y_2(N \int_0^t \kappa_2 Z_3^N(s) ds) + Y_3(N \int_0^t \kappa_3 Z_3^N(s) ds) \\ Z_3^N(t) &= Z_2^N(0) + Y_1(N \int_0^t \kappa_1 Z_1^N(s) Z_2^N(s) ds) - Y_2(N \int_0^t \kappa_2 Z_3^N(s) ds) - Y_3(N \int_0^t \kappa_3 Z_3^N(s) ds) \\ Z_4^N(t) &= N^{-1}Y_3(N \int_0^t \kappa_3 Z_3^N(s) ds), \end{aligned}$$

where $\kappa_1, \kappa_2, \kappa_3$ do not depend upon N . Note that we scale the numbers of molecules of the substrate and the product as in the previous section, but we leave the enzyme and enzyme-substrate variables discrete. Note that $M = Z_3^N(t) + Z_2^N(t)$ is constant, and define

$$\widehat{Z}_2^N(t) = \int_0^t Z_2^N(s) ds = Mt - \int_0^t Z_3^N(s) ds.$$

Theorem 4.1 *Assume that $Z_1^N(0) \rightarrow Z_1(0)$. Then (Z_1^N, \widehat{Z}_2^N) converges to $(Z_1(t), \widehat{Z}_2(t))$ satisfying*

$$\begin{aligned} Z_1(t) &= Z_1(0) - \int_0^t \kappa_1 Z_1(s) \dot{\widehat{Z}}_2(s) ds + \int_0^t \kappa_2 (M - \widehat{Z}_2(s)) ds \\ 0 &= - \int_0^t \kappa_1 Z_1(s) \dot{\widehat{Z}}_2(s) ds + \int_0^t (\kappa_2 + \kappa_3) (M - \widehat{Z}_2(s)) ds, \end{aligned} \tag{4.1}$$

and hence $\dot{\widehat{Z}}_2(s) = \frac{(\kappa_2 + \kappa_3)M}{\kappa_2 + \kappa_3 + \kappa_1 Z_1(s)}$ and

$$\dot{Z}_1(t) = - \frac{M \kappa_1 \kappa_3 Z_1(t)}{\kappa_2 + \kappa_3 + \kappa_1 Z_1(s)}. \tag{4.2}$$

Proof. Relative compactness of the sequence (Z_1^N, \widehat{Z}_2^N) is straightforward, that is, at least along a subsequence, we can assume that (Z_1^N, \widehat{Z}_2^N) converges in distribution to a continuous process (Z_1, \widehat{Z}_2) (which turns out to be deterministic). Dividing the second equation by N and passing to the limit, we see (Z_1, \widehat{Z}_2) must satisfy

$$0 = - \int_0^t \kappa_1 Z_1(s) d\widehat{Z}_2(s) + (\kappa_2 + \kappa_3)Mt - \int_0^t (\kappa_2 + \kappa_3) d\widehat{Z}_2(s). \quad (4.3)$$

Since \widehat{Z}_2 is Lipschitz, it is absolutely continuous, and rewriting (4.3) in terms of the derivative gives the second equation in (4.1). The first equation follows by a similar argument. \square

Of course, (4.2) is the Michaelis-Menten equation.

4.2 Scaling species numbers and rate constants

Assume that we are given a model of the form

$$X(t) = X(0) + \sum_k Y_k \left(\int_0^t \lambda'_k(X(s)) ds \right) (\nu'_k - \nu_k)$$

where the λ'_k are of the form

$$\lambda'_k(x) = \kappa'_k \prod_i \nu_{ik}! \prod_i \binom{x_i}{\nu_{ik}}.$$

Let $N_0 \gg 1$. For each species i , define the *normalized abundance* (or simply, the abundance) by

$$Z_i(t) = N_0^{-\alpha_i} X_i(t),$$

where $\alpha_i \geq 0$ should be selected so that $Z_i = O(1)$. Note that the abundance may be the species number ($\alpha_i = 0$) or the species concentration or something else.

Since the rate constants may also vary over several orders of magnitude, we write $\kappa'_k = \kappa_k N_0^{\beta_k}$ where the β_k are selected so that $\kappa_k = O(1)$. For a binary reaction

$$\kappa'_k x_i x_j = N_0^{\beta_k + \alpha_i + \alpha_j} \kappa_k z_i z_j,$$

and we can write

$$\beta_k + \alpha_i + \alpha_j = \beta_k + \nu_k \cdot \alpha.$$

We also have,

$$\kappa'_k x_i = N_0^{\beta_k + \nu_k \cdot \alpha} z_i, \quad \kappa'_k x_i (x_i - 1) = N_0^{\beta_k + \nu_k \cdot \alpha} z_i (z_i - N_0^{-\alpha_i}),$$

with similar expressions for intensities involving higher order reactions.

We replace N_0 by N in the above expressions and consider a family of models,

$$Z_i^N(t) = Z_i^N(0) + \sum_k N^{-\alpha_i} Y_k \left(\int_0^t N^{\beta_k + \nu_k \cdot \alpha} \lambda_k(Z^N(s)) ds \right) (\nu'_{ik} - \nu_{ik}),$$

where the original model is $Z = Z^{N_0}$. Note that for reactions of the form $2S_i \rightarrow *$, where $*$ represents an arbitrary linear combination of the species, the rate is $N^{\beta_k+2\alpha_i} Z_i^N(t)(Z_i^N(t) - N^{-\alpha_i})$, so if $\alpha_i > 0$, we should write λ_k^N instead of λ_k , but to simplify notation, we will simply write λ_k .

We have a family of models indexed by N for which $N = N_0$ gives the “correct” or original model. Other values of N and any limits as $N \rightarrow \infty$ (perhaps with a change of time-scale) give approximate models. The challenge is to select the α_i and the β_k in a reasonable way, but once that is done, the initial condition for index N is given by

$$Z_i^N(0) = N^{-\alpha_i} \left\lfloor N^{\alpha_i} \frac{X_i(0)}{N_0^{\alpha_i}} \right\rfloor,$$

where $\lfloor z \rfloor$ is the integer part of z and the $X_i(0)$ are the initial species numbers in the original model.

Allowing a change of time-scale, where t is replaced by tN^γ , suppose $\lim_{N \rightarrow \infty} Z_i^N(\cdot N^\gamma) = Z_i^\infty$. Then we should have

$$X_i(t) \approx N_0^{\alpha_i} Z_i^\infty(tN_0^{-\gamma}).$$

4.3 Determining the scaling exponents

There are, of course, many ways of selecting the α_i and β_k , but we want to make this selection so that there are limiting models that give reasonable approximations for the original model. Consequently, we look for natural constraints on the α_i and β_k .

For example, suppose that the rate constants satisfy

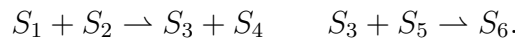
$$\kappa'_1 \geq \kappa'_2 \geq \dots \geq \kappa'_{r_0}.$$

Then it seems natural to select

$$\beta_1 \geq \dots \geq \beta_{r_0},$$

although it may be reasonable to separate the binary reactions from the unary reaction.

To get a sense of the issues involved in selecting exponents that lead to reasonable limits, consider a reaction network in which the reactions involving S_3 are



Then

$$\begin{aligned} Z_3^N(t) = & Z_3^N(0) + N^{-\alpha_3} Y_1(N^{\beta_1+\alpha_1+\alpha_2} \int_0^t \kappa_1 Z_1^N(s) Z_2^N(s) ds) \\ & - N^{-\alpha_3} Y_2(N^{\beta_2+\alpha_3+\alpha_5} \int_0^t \kappa_2 Z_3^N(s) Z_5^N(s) ds), \end{aligned}$$

or scaling time

$$\begin{aligned} Z_3^N(tN^\gamma) = & Z_3^N(0) + N^{-\alpha_3} Y_1(N^{\beta_1+\alpha_1+\alpha_2+\gamma} \int_0^t \kappa_1 Z_1^N(sN^\gamma) Z_2^N(sN^\gamma) ds) \\ & - N^{-\alpha_3} Y_2(N^{\beta_2+\alpha_3+\alpha_5+\gamma} \int_0^t \kappa_2 Z_3^N(sN^\gamma) Z_5^N(sN^\gamma) ds). \end{aligned}$$

Assuming that for the other species in the system $Z_i^N = O(1)$, we see that $Z_3^N = O(1)$ if

$$(\beta_1 + \alpha_1 + \alpha_2 + \gamma) \vee (\beta_2 + \alpha_3 + \alpha_5 + \gamma) \leq \alpha_3$$

or if

$$\beta_1 + \alpha_1 + \alpha_2 = \beta_2 + \alpha_3 + \alpha_5 > \alpha_3.$$

Note that in the latter case, we would expect $Z_3^N(t) \approx \frac{\kappa_1 Z_1^N(t) Z_2^N(t)}{\kappa_2 Z_5^N(t)}$. If these conditions both fail, then either Z_3^N will blow up as $N \rightarrow \infty$ or will be driven to zero.

With this example in mind, define $Z_i^{N,\gamma}(t) = Z_i^N(tN^\gamma)$ so

$$Z_i^{N,\gamma}(t) = Z_i^N(0) + \sum_k N^{-\alpha_i} Y_k \left(\int_0^t N^{\gamma+\beta_k+\nu_k \cdot \alpha} \lambda_k(Z^{N,\gamma}(s)) ds \right) (\nu'_{ik} - \nu_{ik}).$$

Recalling that $\zeta_k = \nu'_k - \nu_k$, for $\theta_i \geq 0$, consider

$$\sum_i \theta_i N^{\alpha_i} Z_i^{N,\gamma}(t) = \sum_i \theta_i N^{\alpha_i} Z_i^N(0) + \sum_k Y_k \left(\int_0^t N^{\gamma+\beta_k+\nu_k \cdot \alpha} \lambda_k(Z^{N,\gamma}(s)) ds \right) \langle \theta, \zeta_k \rangle,$$

where $\langle \theta, \zeta_k \rangle = \sum_i \theta_i \zeta_{ik}$, and define $\alpha_\theta = \max\{\alpha_i : \theta_i > 0\}$. If all $Z_i^{N,\gamma} = O(1)$, then the left side is $O(N^{\alpha_\theta})$, and as in the single species example above, we must have

$$\gamma + \max\{\beta_k + \nu_k \cdot \alpha : \langle \theta, \zeta_k \rangle \neq 0\} \leq \alpha_\theta. \quad (4.4)$$

or

$$\max\{\beta_k + \nu_k \cdot \alpha : \langle \theta, \zeta_k \rangle > 0\} = \max\{\beta_k + \nu_k \cdot \alpha : \langle \theta, \zeta_k \rangle < 0\}. \quad (4.5)$$

Note that (4.4) is really a constraint on the time-scale determined by γ saying that if (4.5) fails for some θ , then γ must satisfy

$$\gamma \leq \alpha_\theta - \max\{\beta_k + \nu_k \cdot \alpha : \langle \theta, \zeta_k \rangle \neq 0\}.$$

The value of γ given by

$$\gamma_i = \alpha_i - \max\{\beta_k + \nu_k \cdot \alpha : \zeta_{ik} \neq 0\}$$

gives the natural time-scale for S_i in the sense that $Z_i^{N,\gamma}$ is neither asymptotically constant nor too rapidly oscillating to have a limit. The γ_i are values of γ for which interesting limits may hold. Linear combinations $\langle \theta, Z^{N,\gamma} \rangle$ may have time-scales

$$\gamma_\theta = \alpha_\theta - \max\{\beta_k + \nu_k \cdot \alpha : \langle \theta, \zeta_k \rangle \neq 0\}$$

that are different from all of the species time-scales and may give *auxiliary variables* (see, for example, [14]) whose limits capture interesting properties of the system.

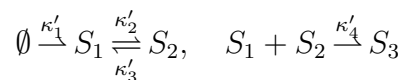
The equation (4.5) is called the *balance equation*, and together, the alternative (4.5) and (4.4) is referred to as the *balance condition*. Note that it is not necessary to solve the balance equations for every choice of θ . The equations that fail simply place restrictions on the time-scales γ that can be used without something blowing up. The goal is to find α_i and β_k that

give useful limiting models, and solving some subset of the balance equations can be a useful first step. Natural choices of θ in selecting the subset of balance equations to solve include those for which $\langle \theta, \zeta_k \rangle = 0$ for one or more of the ζ_k . See Section 3.4 of [26] for a more detailed discussion.

In the next subsection, we apply the balance conditions to identify exponents useful in deriving a reduced model for a simple reaction network. For an application to a much more complex model of the heat shock response in *E. coli*, see [25].

4.4 An application of the balance conditions

Consider the simple example



Assume $\kappa'_k = \kappa_k N_0^{\beta_k}$. Then a useful subset of the balance equations is

$$\begin{array}{ll} S_2 & \beta_2 + \alpha_1 = (\beta_3 + \alpha_2) \vee (\beta_4 + \alpha_1 + \alpha_2) \\ S_1 & \beta_1 \vee (\beta_3 + \alpha_2) = (\beta_2 + \alpha_1) \vee (\beta_4 + \alpha_1 + \alpha_2) \\ S_3 & \beta_4 + \alpha_1 + \alpha_2 = -\infty \\ S_1 + S_2 & \beta_1 = \beta_4 + \alpha_1 + \alpha_2 \end{array}$$

where we take the maximum of the empty set to be $-\infty$. Of course, it is not possible to select parameters satisfying the balance equation for S_3 , so we must restrict γ by

$$\gamma \leq \alpha_3 - (\beta_4 + \alpha_1 + \alpha_2). \quad (4.6)$$

Let $\alpha_1 = 0$ and $\beta_1 = \beta_2 > \beta_3 = \beta_4$, so balance for S_1 , S_2 , and $S_1 + S_2$ is satisfied if $\alpha_2 = \beta_2 - \beta_3$, which we assume. Taking $\alpha_3 = \alpha_2$, (4.6) becomes

$$\gamma \leq -\beta_4 = -\beta_3.$$

The system of equations becomes

$$\begin{aligned} Z_1^N(t) &= Z_1^N(0) + Y_1(\kappa_1 N^{\beta_1} t) - Y_2(\kappa_2 N^{\beta_2} \int_0^t Z_1^N(s) ds) \\ &\quad + Y_3(\kappa_3 N^{\beta_3 + \alpha_2} \int_0^t Z_2^N(s) ds) - Y_4(\kappa_4 N^{\beta_4 + \alpha_2} \int_0^t Z_1^N(s) Z_2^N(s) ds) \\ Z_2^N(t) &= Z_2^N(0) + N^{-\alpha_2} Y_2(\kappa_2 N^{\beta_2} \int_0^t Z_1^N(s) ds) \\ &\quad - N^{-\alpha_2} Y_3(\kappa_3 N^{\beta_3 + \alpha_2} \int_0^t Z_2^N(s) ds) - N^{-\alpha_2} Y_4(\kappa_4 N^{\beta_4 + \alpha_2} \int_0^t Z_1^N(s) Z_2^N(s) ds) \\ Z_3^N(t) &= Z_3^N(0) + N^{-\alpha_3} Y_4(\kappa_4 N^{\beta_4 + \alpha_2} \int_0^t Z_1^N(s) Z_2^N(s) ds) \end{aligned}$$

There are two time-scales of interest in this model, $\gamma = -\beta_1$, the time-scale of S_1 , and $\gamma = -\beta_3$, the time-scale of S_2 and S_3 . For $\gamma = -\beta_1$, and recalling that $\alpha_2 + \beta_3 = \alpha_2 + \beta_4 = \beta_1 = \beta_2$,

$$\begin{aligned} Z_1^{N,-\beta_1}(t) &= Z_1^N(0) + Y_1(\kappa_1 t) - Y_2(\kappa_2 \int_0^t Z_1^{N,-\beta_1}(s) ds) \\ &\quad + Y_3(\kappa_3 \int_0^t Z_2^{N,-\beta_1}(s) ds) - Y_4(\kappa_4 \int_0^t Z_1^{N,-\beta_1}(s) Z_2^{N,-\beta_1}(s) ds) \\ Z_2^{N,-\beta_1}(t) &= Z_2^N(0) + N^{-\alpha_2} Y_2(\kappa_2 \int_0^t Z_1^{N,-\beta_1}(s) ds) \\ &\quad - N^{-\alpha_2} Y_3(\kappa_3 \int_0^t Z_2^{N,-\beta_1}(s) ds) - N^{-\alpha_2} Y_4(\kappa_4 \int_0^t Z_1^{N,-\beta_1}(s) Z_2^{N,-\beta_1}(s) ds) \\ Z_3^{N,-\beta_1}(t) &= Z_3^N(0) + N^{-\alpha_3} Y_4(\kappa_4 \int_0^t Z_1^{N,-\beta_1}(s) Z_2^{N,-\beta_1}(s) ds), \end{aligned}$$

and the limit of $Z^{N,-\beta_1}$ satisfies

$$\begin{aligned} Z_1(t) &= Z_1(0) + Y_1(\kappa_1 t) - Y_2(\kappa_2 \int_0^t Z_1(s) ds) + Y_3(\kappa_3 \int_0^t Z_2(s) ds) - Y_4(\kappa_4 \int_0^t Z_1(s) Z_2(s) ds) \\ Z_2(t) &= Z_2(0) \\ Z_3(t) &= Z_3(0) \end{aligned}$$

Note that the stationary distribution for Z_1 is Poisson with $E[Z_1] = \frac{\kappa_1 + \kappa_3 Z_2(0)}{\kappa_2 + \kappa_4 Z_2(0)}$.

For $\gamma = -\beta_3$,

$$\begin{aligned} Z_1^{N,-\beta_3}(t) &= Z_1^N(0) + Y_1(\kappa_1 N^{\beta_1 - \beta_3} t) - Y_2(\kappa_2 N^{\beta_2 - \beta_3} \int_0^t Z_1^{N,-\beta_3}(s) ds) \\ &\quad + Y_3(\kappa_3 N^{\alpha_2} \int_0^t Z_2^{N,-\beta_3}(s) ds) - Y_4(\kappa_4 N^{\alpha_2} \int_0^t Z_1^{N,-\beta_3}(s) Z_2^{N,-\beta_3}(s) ds) \\ Z_2^{N,-\beta_3}(t) &= Z_2^N(0) + N^{-\alpha_2} Y_2(\kappa_2 N^{\beta_2 - \beta_3} \int_0^t Z_1^{N,-\beta_3}(s) ds) \\ &\quad - N^{-\alpha_2} Y_3(\kappa_3 N^{\alpha_2} \int_0^t Z_2^{N,-\beta_3}(s) ds) - N^{-\alpha_2} Y_4(\kappa_4 N^{\alpha_2} \int_0^t Z_1^{N,-\beta_3}(s) Z_2^{N,-\beta_3}(s) ds) \\ Z_3^{N,-\beta_3}(t) &= Z_3^N(0) + N^{-\alpha_3} Y_4(\kappa_4 N^{\alpha_2} \int_0^t Z_1^{N,-\beta_3}(s) Z_2^{N,-\beta_3}(s) ds), \end{aligned}$$

and dividing the first equation by $N^{\beta_1 - \beta_3} = N^{\beta_2 - \beta_3} = N^{\alpha_2}$, we see that

$$\int_0^t Z_1^{N,-\beta_3}(s) (\kappa_2 + \kappa_4 Z_2^{N,-\beta_3}(s)) ds - \int_0^t (\kappa_1 + \kappa_3 Z_2^{N,-\beta_3}(s)) ds \rightarrow 0.$$

Since $Z_2^{N,-\beta_3}$ is well-behaved, this limit can be shown to imply

$$\int_0^t Z_1^{N,-\beta_3}(s) ds - \int_0^t \frac{\kappa_1 + \kappa_3 Z_2^{N,-\beta_3}(s)}{\kappa_2 + \kappa_4 Z_2^{N,-\beta_3}(s)} ds \rightarrow 0.$$

We emphasize that $Z_1^{N,-\beta_3}$ is not converging, but it is oscillating rapidly and averages locally so that this limit holds. It follows that $(Z_2^{N,-\beta_3}, Z_3^{N,-\beta_3})$ converges to the solution of

$$\begin{aligned} Z_2(t) &= Z_2(0) + \int_0^t \left((\kappa_2 - \kappa_4 Z_2(s)) \frac{\kappa_1 + \kappa_3 Z_2(s)}{\kappa_2 + \kappa_4 Z_2(s)} - \kappa_3 Z_2(s) \right) ds. \\ &= Z_2(0) + \int_0^t \left(\kappa_1 - \frac{2\kappa_4 Z_2(s)(\kappa_1 + \kappa_3 Z_2(s))}{\kappa_2 + \kappa_4 Z_2(s)} \right) ds. \\ Z_3(t) &= Z_3(0) + \int_0^t \kappa_4 Z_2(s) \frac{\kappa_1 + \kappa_3 Z_2(s)}{\kappa_2 + \kappa_4 Z_2(s)} ds. \end{aligned}$$

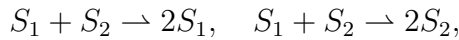
4.5 Hybrid limits

If $Z_i^\gamma = \lim_{N \rightarrow \infty} Z_i^{N,\gamma}$ exists and is a well-behaved process for some choice of γ , then if $\alpha_i = 0$, $Z_i^{\infty,\gamma}$ will be an integer-valued, pure-jump process and if $\alpha_i > 0$, Z_i^γ will have continuous sample paths. In fact, if $\alpha_i > 0$, typically Z_i^γ will satisfy an equation of the form

$$Z_i^\gamma(t) = Z_i(0) + \int_0^t F_i(Z^\gamma(s)) ds.$$

Consequently, the natural class of limits will be *hybrid* or *piecewise deterministic* (in the sense of Davis [11]) *models* in which some components are discrete and some are absolutely continuous. See Section 3 of [4] and Section 6.3 of [26] for examples.

It is possible to obtain diffusion processes as limits, but these are not typical for reaction networks. (Note that the diffusion approximations discussed in Section 3.3 do not arise as *limits* of a sequence of processes.) One example that is more naturally interpreted as a model in population genetics (a Moran model) but can be interpreted as a reaction network would be



where both reactions have the same rate constant. Suppose the normalized system has the form

$$\begin{aligned} Z_1^N(t) &= Z_1^N(0) + N^{-1/2} Y_1(\kappa N \int_0^t Z_1^N(s) Z_2^N(s) ds) - N^{-1/2} Y_2(\kappa N \int_0^t Z_1^N(s) Z_2^N(s) ds) \\ Z_2^N(t) &= Z_2^N(0) + N^{-1/2} Y_2(\kappa N \int_0^t Z_1^N(s) Z_2^N(s) ds) - N^{-1/2} Y_1(\kappa N \int_0^t Z_1^N(s) Z_2^N(s) ds). \end{aligned}$$

If we center Y_1 and Y_2 , the centerings cancel, and assuming $(Z_1^N(0), Z_2^N(0)) \Rightarrow (Z_1^\infty(0), Z_2^\infty(0))$, (Z_1^N, Z_2^N) converges to a solution of

$$\begin{aligned} Z_1(t) &= Z_1(0) + W_1(\kappa \int_0^t Z_1(s) Z_2(s) ds) - W_2(\kappa \int_0^t Z_1(s) Z_2(s) ds) \\ Z_2(t) &= Z_2(0) + W_2(\kappa \int_0^t Z_1(s) Z_2(s) ds) - W_1(\kappa \int_0^t Z_1(s) Z_2(s) ds). \end{aligned}$$

4.6 Central limit theorems and diffusion approximations

Note that in Section 4.1, Z_2^N and Z_3^N do not converge, but $\int_0^t Z_2^N(s)ds$ and $\int_0^t Z_3^N(s)ds$ do, that is, the rapid fluctuations in Z_2^N and Z_3^N average out. Similarly, to obtain (4.7), we used the fact that for $\gamma = -\beta_3$, the rapid fluctuations in $Z_1^{N,\gamma} = Z_1^N(\cdot N^\gamma)$ average to something well-behaved.

Both of these examples have deterministic limits, and it is natural to seek the same kind of central limit theorem that holds under the classical scaling. For fluctuations around (4.7), we have

$$\begin{aligned}
V^N(t) &= N^{\alpha_2/2}(Z_2^{N,\gamma}(t) - Z_2(t)) & (4.7) \\
&= V^N(0) + N^{-\alpha_2/2}\tilde{Y}_2(\kappa_2 N^{\alpha_2} \int_0^t Z_1^{N,\gamma}(s)ds) \\
&\quad - N^{-\alpha_2/2}\tilde{Y}_3(\kappa_3 N^{\alpha_2} \int_0^t Z_2^{N,\gamma}(s)ds) - N^{-\alpha_2/2}\tilde{Y}_4(\kappa_4 N^{\alpha_2} \int_0^t Z_1^{N,\gamma}(s)Z_2^{N,\gamma}(s)ds) \\
&\quad + N^{\alpha_2/2} \int_0^t \left(Z_1^{N,\gamma}(s) \left(\kappa_2 - \kappa_4 Z_2^{N,\gamma}(s) \right) - \frac{\kappa_1 + \kappa_3 Z_2(s)}{\kappa_2 + \kappa_4 Z_2(s)} (\kappa_2 - \kappa_4 Z_2(s)) \right) \\
&\quad - \kappa_3 \int_0^t V^N(s)ds.
\end{aligned}$$

Assuming $V^N(0)$ converges, the convergence of $Z_2^{N,\gamma}$ and $\int Z_1^{N,\gamma}ds$ and the functional central limit theorem for the renormalized Poisson processes imply the convergence of the first four terms on the right and we would have a central limit theorem similar to that described in Section 3.2 if it were not for the fifth term on the right.

To treat the fifth term, we exploit the martingale properties discussed in Section 2.5. In particular, if

$$f_N(z_1, z_2) = N^{-\alpha_2/2} z_1 \frac{\kappa_2 - \kappa_4 z_2}{\kappa_2 + \kappa_4 z_2}$$

and we define

$$F(z_2) = \frac{\kappa_1 + \kappa_3 z_2}{\kappa_2 + \kappa_4 z_2} (\kappa_2 - \kappa_4 z_2)$$

as in (2.11),

$$\begin{aligned}
M_N(t) &= f_N(Z_1^{N,\gamma}(t), Z_2^{N,\gamma}(t)) - f_N(Z_1^{N,\gamma}(0), Z_2^{N,\gamma}(0)) - \int_0^t A_N f_N(Z_1^{N,\gamma}(s), Z_2^{N,\gamma}(s))ds \\
&= -N^{\alpha_2/2} \int_0^t \left(Z_1^{N,\gamma}(s) (\kappa_2 - \kappa_4 Z_2^{N,\gamma}(s)) - F(Z_2^{N,\gamma}(s)) \right) ds + O(N^{-\alpha_2/2})
\end{aligned}$$

is a martingale, and (4.7) becomes

$$\begin{aligned}
V^N(t) &= V^N(0) + N^{-\alpha_2/2} \tilde{Y}_2(\kappa_2 N^{\alpha_2} \int_0^t Z_1^{N,\gamma}(s) ds) \\
&\quad - N^{-\alpha_2/2} \tilde{Y}_3(\kappa_3 N^{\alpha_2} \int_0^t Z_2^{N,\gamma}(s) ds) - N^{-\alpha_2/2} \tilde{Y}_4(\kappa_4 N^{\alpha_2} \int_0^t Z_1^{N,\gamma}(s) Z_2^{N,\gamma}(s) ds) \\
&\quad - M_N(t) + N^{\alpha_2/2} \int_0^t \left(F(Z_2^{N,\gamma}(s)) - F(Z_2(s)) \right) ds \\
&\quad - \kappa_3 \int_0^t V^N(s) ds + O(N^{-\alpha_2/2}) \\
&= V^N(0) + \widehat{M}_N(t) + N^{\alpha_2/2} \int_0^t \left(F(Z_2^{N,\gamma}(s)) - F(Z_2(s)) \right) ds \\
&\quad - \kappa_3 \int_0^t V^N(s) ds + O(N^{-\alpha_2/2}),
\end{aligned}$$

where \widehat{M}_N is defined by the above equality. For M_N above, the quadratic variation of M_N is

$$\begin{aligned}
[M_N]_t &= \int_0^t (f_N(Z_1^{N,\gamma}(s-)) + 1, Z_2^{N,\gamma}(s-)) - f_N(Z_1^{N,\gamma}(s-), Z_2^{N,\gamma}(s-))^2 dR_1^N(s) \\
&\quad + \int_0^t (f_N(Z_1^{N,\gamma}(s-)) - 1, Z_2^{N,\gamma}(s-) + N^{-\alpha_2}) - f_N(Z_1^{N,\gamma}(s-), Z_2^{N,\gamma}(s-))^2 dR_2^N(s) \\
&\quad + \int_0^t (f_N(Z_1^{N,\gamma}(s-)) + 1, Z_2^{N,\gamma}(s-) - N^{-\alpha_2}) - f_N(Z_1^{N,\gamma}(s-), Z_2^{N,\gamma}(s-))^2 dR_3^N(s) \\
&\quad + \int_0^t (f_N(Z_1^{N,\gamma}(s-)) - 1, Z_2^{N,\gamma}(s-) - N^{-\alpha_2}) - f_N(Z_1^{N,\gamma}(s-), Z_2^{N,\gamma}(s-))^2 dR_4^N(s).
\end{aligned}$$

Observing that each of the integrands is asymptotically

$$N^{-\alpha_2} \left(\frac{\kappa_2 - \kappa_4 Z_2^{N,\gamma}(s)}{\kappa_2 + \kappa_4 Z_2^{N,\gamma}(s)} \right)^2$$

and that, for example,

$$N^{-\alpha_2} R_2^N(t) \rightarrow \int_0^t \kappa_2 \frac{\kappa_1 + \kappa_3 Z_2(s)}{\kappa_2 + \kappa_4 Z_2(s)} ds,$$

we have $[M_N]_t \rightarrow C(t)$ where

$$C(t) = \int_0^t \left(\frac{\kappa_2 - \kappa_4 Z_2(s)}{\kappa_2 + \kappa_4 Z_2(s)} \right)^2 \left(\kappa_1 + \kappa_2 \frac{\kappa_1 + \kappa_3 Z_2(s)}{\kappa_2 + \kappa_4 Z_2(s)} + \kappa_3 Z_2(s) + \kappa_4 Z_2(s) \frac{\kappa_1 + \kappa_3 Z_2(s)}{\kappa_2 + \kappa_4 Z_2(s)} \right) ds,$$

which, by the martingale central limit theorem (see, for example, Theorem 7.1.4 of [15]), implies $M_N \Rightarrow M$ where M can be written as the time change of a Brownian motion, that is, $M(t) = W(C(t))$.

Unfortunately, M is not independent of the limits of the three renormalized Poisson processes, so rather than applying the martingale central limit theorem to M_N , we need to apply it to \widehat{M}_N . The quadratic variation for \widehat{M}_N is

$$\begin{aligned} [\widehat{M}_N]_t &= \int_0^t (f_N(Z_1^{N,\gamma}(s-) + 1, Z_2^{N,\gamma}(s-)) - f_N(Z_1^{N,\gamma}(s-), Z_2^{N,\gamma}(s-)))^2 dR_1^N(s) \\ &+ \int_0^t (N^{-\alpha_2/2} - f_N(Z_1^{N,\gamma}(s-) - 1, Z_2^{N,\gamma}(s-) + N^{-\alpha_2}) + f_N(Z_1^{N,\gamma}(s-), Z_2^{N,\gamma}(s-)))^2 dR_2^N(s) \\ &+ \int_0^t (-N^{-\alpha_2/2} - f_N(Z_1^{N,\gamma}(s-) + 1, Z_2^{N,\gamma}(s-) - N^{-\alpha_2}) + f_N(Z_1^{N,\gamma}(s-), Z_2^{N,\gamma}(s-)))^2 dR_3^N(s) \\ &+ \int_0^t (-N^{-\alpha_2/2} - f_N(Z_1^{N,\gamma}(s-) - 1, Z_2^{N,\gamma}(s-) - N^{-\alpha_2}) + f_N(Z_1^{N,\gamma}(s-), Z_2^{N,\gamma}(s-)))^2 dR_4^N(s), \end{aligned}$$

and $[\widehat{M}_N]_t$ converges to

$$\begin{aligned} \widehat{C}(t) &= \int_0^t \left(\kappa_1 \left(\frac{\kappa_2 - \kappa_4 Z_2(s)}{\kappa_2 + \kappa_4 Z_2(s)} \right)^2 + \left(1 + \frac{\kappa_2 - \kappa_4 Z_2(s)}{\kappa_2 + \kappa_4 Z_2(s)} \right)^2 \left(\kappa_2 \frac{\kappa_1 + \kappa_3 Z_2(s)}{\kappa_2 + \kappa_4 Z_2(s)} + \kappa_3 Z_2(s) \right) \right. \\ &\quad \left. + \left(\frac{\kappa_2 - \kappa_4 Z_2(s)}{\kappa_2 + \kappa_4 Z_2(s)} - 1 \right)^2 \kappa_4 Z_2(s) \frac{\kappa_1 + \kappa_3 Z_2(s)}{\kappa_2 + \kappa_4 Z_2(s)} \right) ds. \end{aligned}$$

Consequently, $\widehat{M}_N \Rightarrow W(\widehat{C}(t))$ and $V^N \Rightarrow V$ satisfying

$$V(t) = V(0) + W(\widehat{C}(t)) + \int_0^t (F'(Z_2(s)) - \kappa_3)V(s)ds,$$

which, as in (3.3) is a Gaussian process.

Let

$$\begin{aligned} G(z_2) &= \left(\kappa_1 \left(\frac{\kappa_2 - \kappa_4 z_2}{\kappa_2 + \kappa_4 z_2} \right)^2 + \left(1 + \frac{\kappa_2 - \kappa_4 z_2}{\kappa_2 + \kappa_4 z_2} \right)^2 \left(\kappa_2 \frac{\kappa_1 + \kappa_3 z_2}{\kappa_2 + \kappa_4 z_2} + \kappa_3 z_2 \right) \right. \\ &\quad \left. + \left(\frac{\kappa_2 - \kappa_4 z_2}{\kappa_2 + \kappa_4 z_2} - 1 \right)^2 \kappa_4 z_2 \frac{\kappa_1 + \kappa_3 z_2}{\kappa_2 + \kappa_4 z_2} \right). \end{aligned}$$

Then the analysis above suggests the following diffusion or Langevin approximation for $Z_2^{N,\gamma}$:

$$D^N(t) = D^N(0) + N^{-\alpha_2/2} W \left(\int_0^t G(D^N(s)) ds \right) + \int_0^t (F(D^N(s)) - \kappa_3 D^N(s)) ds.$$

See [27] for a detailed discussion of the central limit theorem and diffusion approximations for multiscaled models. In particular, that paper contains a systematic discussion of the treatment of integral terms with rapidly oscillating integrands.

References

- [1] David F. Anderson. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. *J. Chem. Phys.*, 127(21):214107, 2007.
- [2] David F. Anderson, Gheorge Craciun, and Thomas G. Kurtz. Product-form stationary distributions for deficiency zero chemical reaction networks. *Bull. Math. Biol.*, 2010. doi: 10.1007/s11538-010-9517-4.
- [3] Krishna B. Athreya and Peter E. Ney. *Branching processes*. Springer-Verlag, New York, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196.
- [4] Karen Ball, Thomas G. Kurtz, Lea Popovic, and Greg Rempala. Asymptotic analysis of multiscale approximations to reaction networks. *Ann. Appl. Probab.*, 16(4):1925–1961, 2006. ISSN 1050-5164.
- [5] Manuel Barrio, Kevin Burrage, André Leier, and Tianhai Tian. Oscillatory regulation of Hes1: Discrete stochastic delay modelling and simulation. *PLoS Comp. Biol.*, 2: 1017–1030, 2006.
- [6] Anthony F. Bartholomay. Stochastic models for chemical reactions. I. Theory of the unimolecular reaction process. *Bull. Math. Biophys.*, 20:175–190, 1958.
- [7] Anthony F. Bartholomay. Stochastic models for chemical reactions. II. The unimolecular rate constant. *Bull. Math. Biophys.*, 21:363–373, 1959.
- [8] Dmitri Bratsun, Dmitri Volfson, Lev S. Tsimring, and Jeff Hasty. Delay-induced stochastic oscillations in gene regulation. *PNAS*, 102:14593 – 14598, 2005.
- [9] Thomas Darden. A pseudo-steady state approximation for stochastic chemical kinetics. *Rocky Mountain J. Math.*, 9(1):51–71, 1979. ISSN 0035-7596. Conference on Deterministic Differential Equations and Stochastic Processes Models for Biological Systems (San Cristobal, N.M., 1977).
- [10] Thomas A. Darden. Enzyme kinetics: stochastic vs. deterministic models. In *Instabilities, bifurcations, and fluctuations in chemical systems (Austin, Tex., 1980)*, pages 248–272. Univ. Texas Press, Austin, TX, 1982.
- [11] M. H. A. Davis. *Markov models and optimization*, volume 49 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1993. ISBN 0-412-31410-X.
- [12] Max Delbrück. Statistical fluctuations in autocatalytic reactions. *The Journal of Chemical Physics*, 8(1):120–124, 1940. doi: 10.1063/1.1750549. URL <http://link.aip.org/link/?JCP/8/120/1>.
- [13] Monroe D. Donsker. An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.*, 1951(6):12, 1951. ISSN 0065-9266.

- [14] Weinan E, Di Liu, and Eric Vanden-Eijnden. Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates. *The Journal of Chemical Physics*, 123(19):194107, 2005. URL <http://link.aip.org/link/?JCP/123/194107/1>.
- [15] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. ISBN 0-471-08186-8. Characterization and convergence.
- [16] Martin Feinberg. Chemical reaction network structure and the stability of complex isothermal reactors i. the deficiency zero and deficiency one theorems. *Chem. Engr. Sci.*, 42(10):2229–2268, 1987.
- [17] Martin Feinberg. Chemical reaction network structure and the stability of complex isothermal reactors ii. multiple steady states for networks of deficiency one. *Chem. Engr. Sci.*, 43(1):1–25, 1988.
- [18] Chetan Gadgil, Chang Hyeong Lee, and Hans G. Othmer. A stochastic analysis of first-order reaction networks. *Bull. Math. Biol.*, 67(5):901–946, 2005. ISSN 0092-8240. doi: 10.1016/j.bulm.2004.09.009. URL <http://dx.doi.org/10.1016/j.bulm.2004.09.009>.
- [19] M. A. Gibson and Bruck J. Efficient exact simulation of chemical systems with many species and many channels. *J. Phys. Chem. A*, 104(9):1876–1889, 2000.
- [20] Daniel T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Computational Phys.*, 22(4):403–434, 1976.
- [21] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340–61, 1977.
- [22] Daniel T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A*, 188:404–425, 1992.
- [23] Daniel T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001. doi: 10.1063/1.1378322. URL <http://link.aip.org/link/?JCP/115/1716/1>.
- [24] Jean Jacod. Multivariate point processes: predictable projection, Radon-Nikodým derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31:235–253, 1974/75.
- [25] Hye-Won Kang. The multiple scaling approximation in the heat shock model of e. coli. In Preparation, 2009.
- [26] Hye-Won Kang and Thomas G. Kurtz. Separation of time-scales and model reduction for stochastic reaction networks. In preparation, 2010.
- [27] Hye-Won Kang, Thomas G. Kurtz, and Lea Popovic. Diffusion approximations for multiscale chemical reaction models. in preparation, 2010.

- [28] Frank P. Kelly. *Reversibility and stochastic networks*. John Wiley & Sons Ltd., Chichester, 1979. ISBN 0-471-27601-4. Wiley Series in Probability and Mathematical Statistics.
- [29] A. N. Kolmogorov. *Foundations of the theory of probability*. Chelsea Publishing Co., New York, 1956. Translation edited by Nathan Morrison, with an added bibliography by A. T. Bharucha-Reid.
- [30] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 32:111–131, 1975.
- [31] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 34(1):33–58, 1976.
- [32] Thomas G. Kurtz. Solutions of ordinary differential equations as limits of pure jump Markov processes. *J. Appl. Probability*, 7:49–58, 1970. ISSN 0021-9002.
- [33] Thomas G. Kurtz. Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Probability*, 8:344–356, 1971. ISSN 0021-9002.
- [34] Thomas G. Kurtz. The relationship between stochastic and deterministic models for chemical reactions. *J. Chem. Phys.*, 57(7):2976–2978, 1972.
- [35] Thomas G. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes Appl.*, 6(3):223–240, 1977/78.
- [36] Thomas G. Kurtz. Representations of Markov processes as multiparameter time changes. *Ann. Probab.*, 8(4):682–715, 1980. ISSN 0091-1798. URL [http://links.jstor.org/sici?sici=0091-1798\(198008\)8:4<682:ROMPAM>2.0.CO;2-W&origin=MSN](http://links.jstor.org/sici?sici=0091-1798(198008)8:4<682:ROMPAM>2.0.CO;2-W&origin=MSN).
- [37] Thomas G. Kurtz. The Yamada-Watanabe-Engelbert theorem for general stochastic equations and inequalities. *Electron. J. Probab.*, 12:951–965, 2007. ISSN 1083-6489.
- [38] Thomas G. Kurtz. Equivalence of stochastic equations and martingale problems. In *Stochastic Analysis 2010*. Springer, 2010.
- [39] Donald A. McQuarrie. Stochastic approach to chemical kinetics. *J. Appl. Probability*, 4:413–478, 1967.
- [40] P. A. Meyer. Démonstration simplifiée d'un théorème de Knight. In *Séminaire de Probabilités, V (Univ. Strasbourg, année universitaire 1969–1970)*, pages 191–195. Lecture Notes in Math., Vol. 191. Springer, Berlin, 1971.
- [41] Sheldon Ross. *A first course in probability*. Macmillan Co., New York, second edition, 1984. ISBN 0-02-403910-1.

- [42] N. G. van Kampen. A power series expansion of the master equation. *Canad. J. Phys.*, 39:551–567, 1961. ISSN 0008-4204.