Modeling stochasticity in biochemical reaction networks

This content has been downloaded from IOPscience. Please scroll down to see the full text.
(http://iopscience.iop.org/0022-3727/49/9/093001)

View the table of contents for this issue, or go to the journal homepage for more.

Download details:

IP Address: 134.84.165.204
This content was downloaded on 04/02/2016 at 20:52

Please note that terms and conditions apply.
1. Introduction

The molecular revolution in biology is now more than 60 years old, with the canonical start date of this revolution in April 1953, when Watson and Crick published ‘A Structure of Deoxyribose Nucleic Acid’ [1]. This revolution launched a vast reductionist research program in the biological sciences. The approach is founded on the worldview that biological phenomena, like physical and chemical phenomena, are the result of nothing more than matter in motion. Following the scientific discovery paradigm of the physical sciences, theories in the biological sciences must be grounded in physics and chemistry and are widely accepted when they have direct evidential support, possess explanatory generality and demonstrate predictive precision [2].

Explanatory generality and predictive precision can adequately be assessed by the development of mathematical models that explain and predict direct evidence. For example, in physics, it was the development of differential calculus models by Newton and Leibniz that launched the scientific revolution, which in turn ushered modern times.

In biology, mathematical modeling efforts are increasingly important in discoveries. From high-resolution, atomistic models that capture the structure of protein or DNA molecules [3, 4], to metabolic networks [5], to whole-organ simulations [6, 7], and to ecological-scale systems [8], there is now an astonishing wealth of modeling formalisms widely accepted in the biological sciences.

However, we progress slowly in explaining cellular phenotypes in terms of networks of molecular interactions. The reasons are: (1) the number of distinct molecular components in any cell is irreducibly large. And although we now have better maps of molecular species inside cells, thanks to genomic,
proteomic and metabolomic technologies, these maps are rarely close to being complete; (2) the interactions between these components are now better understood than ever before. But, again, we do not know with certainty all the biologically relevant interactions between proteins, nucleic acids, metabolites, etc; (3) biological systems are not closed ones. They exchange mass and energy with their environment, often in complicated, difficult to decipher transport or reaction mechanisms. These mechanisms render context-dependent modeling formalisms with additional degrees of freedom; (4) biological systems are often far from the thermodynamic limit, with only a few molecules in cells for numerous molecular species. Stochasticity then impacts the observed cellular phenotypes, requiring modeling formalisms that take into account the inherent variability of biomolecular system behaviors.

1.1. Models for systems and synthetic biology

Systems biology and synthetic biology share a common goal: they are both aimed at understanding and engineering complexity in biomolecular systems. Systems and synthetic biology may be considered as the two sides of the same coin: systems biology is reverse engineering, whereas synthetic biology is forward engineering. Systems biology generates information on components and interactions that comprise biological complexity. Synthetic biology uses these components and interactions in new, non-naturally occurring contexts.

In systems biology, advances in genomic, transcriptomic, proteomic and metabolomic technologies have provided strong foundation for modeling the dynamic behavior of biomolecular systems [9–12].

We have argued elsewhere that synthetic biology can further catalyze the development of mathematical models that are well grounded in physicochemical laws and which explain complex phenotypes in terms of interacting biomolecular ensembles synthetic biological systems [13, 14]. For over a decade now, DNA synthesis technologies have revolutionized the construction of DNA sequences encoding for non-naturally occurring biomolecular systems, such as bistable switches, oscillators, and logic gates [15–18].

These artificial biological systems may be amenable to high-resolution molecular models that explain dynamic phenotypes in terms of cascades of biomolecular interactions because they are small, user-defined systems with few, rather well-understood context dependencies [13, 19–23].

In any event, for either systems or synthetic biology modeling efforts, the challenge of incorporating intrinsic, thermal noise in the models remains. Therefore, the important topic of modeling stochasticity in biochemical reaction networks will be discussed, as well as how this theoretical framework gives rise to revolutionary computational tools.

2. Stochastic chemical kinetics

The canonical approach to describe chemically reacting systems involves the use of deterministic rate laws [24]. This viewpoint considers the reactions to take place as a continuum. Given an initial state, the system will always evolve to the same final state predicted by the mathematical model. However, this approach fails to take into account the fact that molecular populations are finite and countable. Rigorously, these molecular populations must be described by integer variables that can only change stochastically and by discrete amounts. When dealing with large systems, at the thermodynamic limit, the classical approach is well suited and correctly describes their physical behavior, because molecular fluctuations are small and can be neglected [25]. Nevertheless, for small systems, which includes most biological chemical reaction networks, another mathematical approach must be used.

In the 1960s, McQuarrie and Oppenheim [26–28] were the first to propose a probabilistic approach to the problem and they introduced some of the theoretical concepts of what became famously known as the chemical master equation (CME). A meticulous derivation of the CME, however, came only 25 years later through the work of Gillespie [29], using statistical thermodynamics and probability theory.

It is important to emphasize that two different types of noise can affect biochemical networks: intrinsic and extrinsic noise. In processes like gene expression and regulatory control during cell cycle, for example, the number of proteins and nucleic acids in the network is considerably low, which leads to important thermal fluctuations (intrinsic noise) [30]. Extrinsic noise with longer time-scales can also cause variability. Over the course of the cell cycle, especially during cell division, the number of certain enzymes, polymerases and ribosomes can change. For example, according to the Koch–Schaechter model, unequal cell divisions explain much of the variability of properties of the cell cycle [31, 32]. Although some authors suggest that extrinsic fluctuations dominate cellular variation [33–35], others will argue that intrinsic and extrinsic noises are independent in their effects, with intrinsic fluctuations being pronounced in the division process [31]. It is not the goal of the present work to settle this debate. However, it is commonly agreed upon that if the system is far from the thermodynamic limit (i.e. very low number of molecules) intrinsic variability will be important [30]. The discussion in this review paper focuses solely on intrinsic stochasticity.

2.1. Chemical master equation

Consider, for example, an isothermal chemical reaction network at constant volume involving N chemical species connected through M reactions. Let $X_i(t)$ be the number of molecules of species $i$ at a given time $t$, so that the vector $\mathbf{X}(t)$ describes the state of the system at time $t$. Assume the system is well stirred and the majority of the molecular collisions are nonreactive, i.e. they are just elastic collisions. In this case, it is not necessary to worry about molecular positions and velocities [29]. This means that the system can be fully described by vector $\mathbf{X}(t)$ at time $t$. If each reaction in the network is supposed elementary, a state-change matrix $\nu$ of size $N \times M$ can be defined where the elements $\nu_{ij}$ represent the stoichiometric coefficients of species $i$ in reaction $j$. Then, the chemical master equation is formally written as:
expresses the probability that one reaction will occur in the next infinitesimal time interval \([t, t + dt]\). The form of the propensity function is given by the order of the reaction, the number of possible combinations among the reacting molecules and the reaction rate constant according to molecular physics [25].

To make the above more clear, we will provide an example with the Michaelis–Menten reaction network:

\[
\begin{align*}
S + E & \underset{k_1}{\overset{k_2}{\rightleftharpoons}} S : E & S : E & \underset{k_3}{\overset{k_4}{\rightleftharpoons}} S + E \\
S : E & \overset{k_5}{\rightarrow} P + E & P & \overset{k_6}{\rightarrow} S
\end{align*}
\]

For this system the CME becomes:

\[
\frac{\partial P(X, t)}{\partial t} = k_1(X_E + 1)(X_S + 1)P(X - \nu_1, t) + k_2(X_{S:E} + 1)P(X - \nu_2, t) + k_3(X_{S:E} + 1)P(X - \nu_3, t) + k_4(X_P + 1)P(X - \nu_4, t) - [k_2X_EX_S + (k_2 + k_3)X_{S:E}] P(X, t) - k_3XP(X, t)
\]

where \(X = [X_E, X_S, X_{S:E}, X_P]^T\) and

\[
\nu = \begin{bmatrix} -1 & 1 & 1 & 0 \\ -1 & 1 & 0 & 1 \\ 1 & -1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}.
\]

The chemical master equation governs how the probability distribution of the state space (molecular populations) changes over time. It can be simply viewed as a balance equation based on the law of conservation of probability for each possible state of the system. The two terms on the right hand side of equation (1) account for the rate of probability of arriving into state \(X(t)\) minus the rate of probability of leaving that state.

At a first glance, it might be hard to recognize the difficulty of solving the CME. The model is incredibly simple and extremely powerful. Nonetheless, as pointed out by Gillespie [25], for each possible combination of reactant molecules there will be one extra equation, which makes the CME a system of coupled differential equations that can only be solved analytically for very simple problems.

One of the classical approximation methods commonly used to solve the CME analytically is the Linear Noise Approximation (LNA). Van Kampen [28, 36] was the first to propose a system-size expansion by using infinite series of powers of the inverse volume of the system to represent the moment equations of the probability distribution, which shall be discussed in depth later. Keeping only the leading-order terms of this perturbation analysis, the expected value for concentrations are the same as those obtained by deterministic kinetic models, while fluctuations (variances) are inversely proportional to the volume [28]. For reaction networks involving only zero and first order kinetics the LNA is exact up to second-order moments. Furthermore, recent theoretical developments show that even for some specific classes of networks with second order kinetics the LNA is exact for all volumes and parameters of the system [37].

As the reaction network becomes larger and the nonlinearity of the problem increases, numerical solutions based on LNA become challenging to obtain, if not impossible.

\[2.2. \text{The stochastic simulation algorithm (SSA) and its variations} \]

In 1976, Gillespie [38] proposed that instead of solving the CME for the probability distribution, it would be simpler to simulate individual reaction trajectories of \(X(t)\) over time. Evidently, this is not the same as suggesting a numerical solution to the master equation. Yet, by making use of the ergodic hypothesis, if one is able to generate a large ensemble of trajectories and sample the state space, it is possible to reconstruct the probability distribution as an exact solution by statistical analysis.

For that reason, Gillespie [39] developed the stochastic simulation algorithm (SSA), which employs a kinetic Monte Carlo procedure to generate time trajectories. The main idea is to select which reaction will take place (integer random variable) and the time for the next reaction (exponential random variable) by drawing two uniformly distributed random numbers and applying consistent mathematical criteria. Essentially, given an initial condition, the system transitions from one state to another through jump Markov processes, which have a memoryless property, and thus evolves in time stochastically. The major issue with this approach, however, is the computational cost. Because every reaction event is simulated and many trajectories are needed in order to obtain a good solution, the SSA becomes considerably slow for most systems of interest. This is especially true if the system to be dealt with involves multiple time-scale phenomena (stiff systems).

Many other suggestions have been made by the scientific community with the purpose of improving the SSA and its efficiency. Gibson and Bruck [40], for instance, proposed the Next Reaction method, which creates a priority queue for reactions to happen and is able to use only one uniform random number per event, thus saving computational time. Other similar works include that of Cao and Petzold [41], Lok and Brent [42] and McCollum [43]. However, as long as these methods are still simulating every single reaction event step-by-step, the process will remain too slow, even though it may be efficient. It becomes, then, necessary to make use of approximate solutions in order to gain speed in computation.

Suppose, for example, that some reactions can be considered fast, i.e. they occur much more frequently. Gillespie [44] observed that, in this case, there will be a small interval of time \(\tau\) in which all propensity functions are nearly constant.
(tau-leaping method). Therefore, the representation of such reactions becomes a continuous Markov process. The amount of times that a certain reaction happens in that time interval (number of arrivals in mathematics jargon of renewal processes) follows a Poisson distribution. If the interval \( \tau \) is, however, also large enough for many reaction events to take place, the Poisson random variable is well approximated by a normal distribution.

Using this rationale, a general chemical Langevin equation (CLE) can be derived [45] to describe each fast reaction as a continuous and stochastic process. In effect, the CLEs are stochastic differential equations, having an additive noise term. In the simplest of approximations this stochastic, or diffusion, term can be a Gaussian white-noise, independent of the system’s state, with a mean of zero and a variance dictated by the temperature. In less approximate formalisms, the CLE stochastic term can be derived rigorously from the CME and be dependent on the state of the system. The noise term is then called multiple, multiplicative noise.

In CLEs the deterministic drift is equivalent to the deterministic reaction rate equations, while the diffusion term is a source of stochastic perturbation to the system [45]. In the thermodynamic limit, the fluctuations caused by the diffusion term disappear and the continuous deterministic rate laws obtained from the classical approach are recovered.

Another method for numerically solving CME with promising results is Finite State Projection method. The method was introduced by Munsky and Khammash [46] and is based on the idea of partitioning the system. The method truncates the system state space and then solves the truncated system of ordinary differential equations generated from CME. The algorithm is more computationally efficient than SSA or the tau-leaping method [46], but it has a dependence on the method of truncation and more importantly on the size of the state space.

2.3. Multi-scale hybrid models for simulations

Many chemical reaction networks span multiple time scales and different population sizes for each species. The reaction events can be rare or frequent, while the species populations can be large or small. In the recent past, much attention has been given to the combination of modeling formalisms through the partition of the problem space into more homogeneous regions. Virtually, at least five different regimes for multi-scale simulations of chemical reactions can be identified according to time and population size: (I) slow-discrete; (II) slow-continuous; (III) fast-discrete; (IV) fast-continuous stochastic; and (V) fast-continuous deterministic [13].

In the slow-discrete region, both the population size is very small and the reaction events are rare, which implies that this subset must be described with the highest accuracy. Hence, for region (I) it is extremely important to obtain exact solutions of the CME using the SSA. In region (II), because the population size is large, it is possible to skip the simulation of some reaction events and not affect the accuracy of the solution. If, however, one wants to be more rigorous, a Poisson distribution can be used to approximate the number of reaction events in region (II).

In the fast-discrete region the populations are small, but reactions happen very frequently. This means that the species populations reach steady-state rather quickly and one is able to approximate the solution by skipping those reaction steps and sampling directly from the stationary probability distribution. In contrast, in region (IV) the reactions are fast and the populations are large. Hence, the continuous time Markov chain approximation with chemical Langevin equations is well suited, as described at the end of the last section. As the size of the system increases and the thermodynamic limit is approached one enters region (V), which is classically modeled through ordinary differential equations [13].

The essential idea of hybrid methods is to use approximate solutions to describe the dynamics of fast reactions, while applying exact solutions for slow ones. For instance, one of the simplest strategies was developed by Rao and Arkin [47] assuming quasi-steady state for fast reactions and applying the regular SSA for slow events. Another approach by Puchalka and Kierzek [48] used a next reaction variant for the slow regions and tau-leaping with Poisson distribution for fast regions. Haseltine and Rawlings [49] modeled a discrete jump Markov chain for the slow reactions and approximated the fast region by a continuous Markov process.

A stochastic algorithm for slow reactions spanning discrete and continuous systems (regions I and II) was first developed by Salis and Kaznessis [50], followed by a technique spanning regions (I) and (III) using the probabilistic steady-state approximation [51]. In order to couple regions (II) and (IV), an adaptive method employing time-stepping stochastic differential equations was proposed by Sotiropoulos and Kaznessis [52] followed by a decomposition scheme for stiff chemical Langevin equations [53].

The general approach behind those methods is to model fast reactions through continuous Markov chains and the CLE, while slow reactions are described by a slight modification of the Next Reaction method proposed by Gibson and Bruck [40]. The next jump of slow reactions is determined by solving a system of differential equations that depend on the CLE for fast reactions. This intrinsic coupling of diverse regions requires a simultaneous numerical integration of both mathematical models, which is carried out by the Euler–Maruyama scheme with efficiency and accuracy [50]. This methodology has been tested in the development of design principles for genetically engineered networks, including bistable switches and oscillators [51, 54].

In order to make such algorithms more accessible and to enable the study and design of large well-mixed biological systems, an avalanche of tools that model gene regulatory networks with stochastic simulations has been developed and is widely used [55–62].

It is worth mentioning that it is in the interest of synthetic and systems biology efforts to not only model a biochemical system but also to predict the response of the system to small perturbations. For example, knowing the solution for a specific set of kinetic constants for a biochemical reaction network, it is important to observe the changes in the solution with a small change in one or more of the kinetic constants. Stochastic sensitivity analysis attempts to address such problems. Some notable ways to achieve such goal include the
representation of Fisher informational matrix for the ordinary differential equations linear noise approximation [63] or separation of different time scales of a system implemented by Gupta and Khammash [64].

3. Moment equations of stochastic reaction networks

Even though Monte Carlo simulations are reliable and accurate methods, they are computationally expensive. There still exists the need to develop more efficient ways to solve stochastic systems. One popular alternative among researchers to explain such systems is the use of time derivatives of probability moments.

Moments are expected values which are used to describe a probability distribution [65]. The first moment is related to the mean of the distribution, the second moment is related to the variance, the third to the skewness, the fourth to the kurtosis etc [66]. There are many different types of moments, like the central [67], polynomial [68], jump [66], factorial [65] etc. In what follows, we demonstrate that the chemical master equation can be correlated to a set of ODEs called moment equations involving their time derivatives:

\[ \frac{\partial \mu}{\partial t} = A \cdot \mu + A' \cdot \mu' \]

where \( \mu \) and \( \mu' \) are vectors containing lower order moments (moments up to a specific order) and higher order moments, respectively, while \( A \) and \( A' \) are matrices. In nonlinear elementary reaction networks the lower order moments depend on the higher order ones \( A' \) is non zero), which creates difficulties in solving the system. All the different types of moments are valid for describing a probability distribution [65]. As a result, the moment equations may have different forms of expression.

In 1949 [69], Moyal was one of the first who used time derivatives of expected values in order to express stochastic systems. However, the first publication suggesting the use of probability moments for solving a stochastic chemical reaction network was from McQuarrie, Jachimowski and Russell in 1964 [70]. In their publication, they worked with nonlinear reaction networks, however they limited their discussion to up to two moments. A common assumption among researchers was that up to two moments would have been enough to describe a distribution [66]. Nowadays, we know that a system can require many more moments up to any specific order; this can be even eight or ten moments [66, 71]. Since then, there have been numerous attempts to create and solve moments equations [65, 66, 68, 72]. Most of the scientific efforts are focused in two areas: (1) how to generate moment equations more efficiently and (2) how to express the higher order moments in terms of the lower order ones.

3.1. Derivation of moment equations

Sotiropoulos and Kaznessis [66] suggested the use of jump moments for the derivation of moment equations. Jump moments are given by the following expression:

\[ d^{j,m_1,...,m_k}(X) = \sum_{X'} [(X'_i - X_i)^{m_i}(X'_j - X_j)^{m_j}...] T(X'|X) \]

where the above moment is of order \( m = m_1 + m_2 + ... \) and \( T(X'|X) \) is the transition rate from state \( X \) to state \( X' \). By using jump moments the time derivative of an \( m \) order moment is given by:

\[ \frac{d(X_1^{m_1}X_2^{m_2}...X_N^{m_N})}{dt} = \sum_{j_1=0}^{m_1} \sum_{j_2=0}^{m_2} ... \sum_{j_N=0}^{m_N} \left[ m_1 \right] \left[ m_2 \right] ... \left[ m_N \right] \times (X_1^{j_1}X_2^{j_2}...X_N^{j_N} - a_{extra}) \]

where \( a_{extra} = \sum_{i=1}^{M} \left[ \prod_{i=1}^{N} j_i^{\nu_i} \right] a_i(X) \). For further proofs and derivations the reader is referred to [66]. The difficulty with this formulation is that it is not computationally friendly. The summation in equation (5) depends on the number of molecules of the components, which usually is unknown at the beginning. The propensity term \( a_i(X) \) does not have a computationally favorable structure. Besides, the analytical form is not recursive, which means that the computational cost increases disproportionately with an increase of number of moments and components (combinatorial explosion) [65].

Another notable work is that of Gillespie in 2009 [68]. Using a univariate moment generating function and the polynomial moments, Gillespie managed to produce an efficient way of reconstructing the moment equation. The paper also suggests the use of probability cumulants instead of probability moments in order to solve the set of equations. Probability cumulants are formed via the natural logarithm of the moment generating function, and represent another mathematical form of a moment basis set.

Even though the use of polynomial moments gave good results for Gillespie [68], at least for small systems, Smadbeck and Kaznessis [65] proposed a more computationally efficient approach. They integrated the use of factorial moments for the generation of moment matrices and moment equations. Through the use of factorial moments, the matrices produced have a banded form, which requires less memory to store them. Besides, factorial moment matrices can be produced faster than polynomial ones [65]. Factorial moments are given by the expression:

\[ \frac{x_1!}{(x_1 - m_1)! (x_2 - m_2)! ... (x_N - m_N)!} P(X,t) \]

where \( x_i \) represents the possible values of \( X_i \), \( i = 1,...,N \). One of the reasons for using factorial moments is that they are the space derivatives of \( Z \)-transform of the probability distribution, which is simply a spatial transformation having the following form:

\[ G(S,t) = \sum_{x_0=0}^{\infty} \sum_{x_1=0}^{\infty} ... \sum_{x_N=0}^{\infty} s_1^{x_1} s_2^{x_2} ... s_N^{x_N} P(X,t) \]
with respect to time, we get:

$$\frac{\partial G(S,t)}{\partial t} \bigg|_{k=1} = \{x_i\} \quad (8a)$$

$$\frac{\partial^2 G(S,t)}{\partial t^2} \bigg|_{k=1} = \{x_i^2\} \quad (8b)$$

Besides, it is also true that: $G(S,t) \big|_{k=1} = 1$. Thus, it is easy to generate factorial moments through the $Z$-transformation of the probability distribution. Differentiating the function $G(S,t)$ with respect to time, we get:

$$\frac{\partial G(S,t)}{\partial t} = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \ldots \sum_{y=0}^{\infty} s_i^1 s_j^2 \ldots s_y^N \frac{\partial P(X,t)}{\partial t} \quad (9)$$

Applying it to the CME (1) and performing some mathematical manipulations, we get:

$$\frac{\partial G(S,t)}{\partial t} = \sum_{\mu=1}^{N} \prod_{i=1}^{N} \nu_{i,\mu}^{+} \prod_{i=1}^{N} \nu_{i,\mu}^{-} \frac{\partial G(S,t)}{\partial \mu} \quad (10)$$

where the $\nu_{i,\mu}^{+}$ and $\nu_{i,\mu}^{-}$ refers to the absolute values of the stoichiometries of the reactants and products of reaction $\mu$, respectively [65]. Equation (10) is the $Z$-transformation of the chemical master equation (Z-CME). Compared to the CME, that depends on the discrete space variable $X$, the Z-CME depends on a continuous space variables.

Combining equations (8a), (8b) and (10), we can construct the first two moment equations for any chemical reaction network. Hence, by differentiating equation (10) with respect to the appropriate space variable we can produce matrices $A_f$ and $A'_f$ from equation (3) (the subscript $f$ refers to the factorial moments).

It is possible to prove that the factorial moment vector $\mu^f$ can be mapped into the polynomial moment vector $\mu^p$: $\mu^f = T \cdot \mu^p$ [65]. So the moment equations for polynomial moments become:

$$\frac{\partial \mu^p}{\partial t} = A_p \mu^p \quad (11a)$$

$$T \frac{\partial \mu^f}{\partial t} = A_p T \mu^f \quad (11b)$$

$$\frac{\partial \mu^f}{\partial t} = T^{-1} A_p T \mu^f \quad (11c)$$

Comparing to equation (3) we observe that: $A_f = T^{-1} A_p T$. Therefore, for the same system the factorial moment matrix $(A_f)$ is related to the polynomial one $(A_p)$ through a similarity transformation [65]. This result implies that the two matrices have the same eigenvalues, they are just different basis set [65, 73].

Returning to the Michaelis–Menten reaction network example presented in section 2.1, equation (10) becomes:

$$\frac{\partial G(S,t)}{\partial t} = k_d(s_{SE} - s_{SE}) \frac{\partial^2 G(S,t)}{\partial s_{SE}^2} + k_d(s_{SE} - s_{SE}) \frac{\partial G(S,t)}{\partial s_{SE}}$$

$$+ k_d(s_{SE} - s_{SE}) \frac{\partial G(S,t)}{\partial s_{SE}} + k_d(s_{SE} - s_{SE}) \frac{\partial G(S,t)}{\partial s_{SE}} \quad (12)$$

Each term of the above equation represents a chemical reaction step. And then from the Z-CME we can generate the moment matrices $A_f$ and $A'_f$. For example, calculating the equation of the second order moment $\{S \cdot E\}$:

$$\frac{\partial \{S \cdot E\}}{\partial t} = E_T k_2 \{1\} + E_T k_3 \{3\} + [k_3(E_T - 2)$$

$$+ k_d(1 - E_T + S_T) \{E\} - (k_2 + k_3 + k_4)\{S \cdot E\}$$

$$+ (k_3 - k_2)\{E^2\} - k_3\{S^2 \cdot E\} - k_3\{S \cdot E^2\} \quad (13)$$

where $S_T$ and $E_T$ are the total amount of molecules of components $S$ and $E$, respectively.

So, the moment equations can be obtained through the $Z$-transformation of CME. Even though, the necessary number of extra higher order moments may be determined, it is still an open research interest to find what is the minimum number of lower order moments needed to describe a given reaction network [65].

3.2. Closure-scheme approximations

At equation (13), the second order moment $\{S \cdot E\}$ depends on the third order moments $\{S^2 \cdot E\}$ and $\{S \cdot E^2\}$. In general, in a system with nonlinear reactions the lower order moments depend on higher order ones [65]. Nevertheless, how to define and express higher-order moments is of ongoing interest.

Most of the moment closures techniques in the literature, i.e. techniques for connecting the higher order moments with lower order ones, are based around the idea of assuming a specific form of the probability distribution. By doing so, the higher order moments can be expressed as functions of the lower order moments only. Thus, the moment equations are transformed into a set of closed coupled equations and they can be solved with an ordinary differential equation solver. In matrix form equation (3) is transformed into:

$$\frac{\partial \mu}{\partial t} = A \cdot \mu + A' \cdot f(\mu) \quad (14)$$

where $f_i$ denoted the higher order moment $i$ as a function of the lower order ones.

Some of the most common assumed probability distribution forms are normal, lognormal, bionomial and poison [72, 74–76]. For instance, if a normal distribution is assumed, then it has been proven that the third order moment $(\mu^3)$ is related to the first $(\mu_1)$ and second $(\mu_2)$ as follows:

$$\mu^3 = f_3(\mu_1, \mu_2) = 3 \mu_2 \mu_3 - 2 \mu_1^3 \quad (72, 74)$$

or for the lognormal distribution it holds: $\mu^3 = \left( \frac{\mu_3}{\mu_1} \right)^3$ [72, 75].

Of course, if there is no a priori knowledge of the distribution of a system, there can be a hesitation in selecting a specific probability distribution. Singh and Hespanha [72] compared in more detail some of the common approaches in literature.
4. The zero-information entropy method

In 1948, Shannon published a simple and revolutionary paper in The Bell System Technical Journal [77], where he introduced the concept of information entropy (represented as $H$) as a measure of uncertainty or unexpectedness. The idea was that the more unexpected an event was, the more new information it carried within itself. If, on the other hand, an outcome of a system can be inferred from previous knowledge, then the new event is redundant, as it provides little new information. It has been said that when Shannon asked for advice from the famous mathematician Newmann about what to call his new measure, he replied: ‘you call it entropy, because most people do not know what entropy really is, and if you use the word ‘entropy’ in an argument, you will win every time!’ [78]

It has also been suggested that what Shannon did for information theory was equivalent to ‘what Gibbs did for physical chemistry’ [79].

However, it was not until 1957 that Jaynes first proposed the maximum entropy principle (MEP), applying it in the context of statistical mechanics and drawing an analogy with the second law of thermodynamics [80]. If some information about a probability distribution is known a priori, then the MEP states that one should look for the distribution with maximum uncertainty, while satisfying the given information (unbiased estimate).

Based on those principles, Smadbeck and Kaznessis [71] developed the zero-information (ZI) closure scheme for the system of moment equations. They assumed that all information necessary to build the exact probability distribution described by the CME is contained within a finite number of lower-order moments. In that case, all higher-order moments add no information to the problem and can, therefore, be obtained from the maximum entropy probability distribution.

Using Shannon’s definition, for a discrete multivariate system of $N$ components, the entropy is given by:

$$ H = -\sum_{\Omega} P(X_1,\ldots,X_N) \ln P(X_1,\ldots,X_N) $$

(15)

where $\Omega$ corresponds to the $N$-dimensional state space of all possible values for $(X_1,\ldots,X_N)$, used here to simplify the notation of multiple summation signs. Again, $P(X_1,\ldots,X_N)$ represents the discrete multivariate probability distribution.

Assuming that the first $\Psi$ lower-order moments are known, the entropy defined in equation (15) must be maximized with respect to these constrains. This can be accomplished through the method of Lagrange multipliers:

$$ \Lambda = H - \sum_{i=0}^{\Psi} \lambda_i \left[ \mu_i - \sum_{\Omega} f_{\mu_i}(X_1,\ldots,X_N) P(X_1,\ldots,X_N) \right] $$

(16)

where $\Lambda$ is the Lagrangian and $\lambda_i$ is the Lagrange multiplier associated with the lower-order moment $\mu_i$. The function $f_{\mu_i}(X_1,\ldots,X_N)$ defines the $i$-th lower moment. For example, if the first moment is the expected value of $X_1$, then $f_{\mu_1}(X_1,\ldots,X_N) = X_1$. Knowing the lower moments, their Lagrange multipliers might be computed using a root-finder numerical method such as Newton–Raphson. Hence, the maximum information entropy distribution is written as follows [81]:

$$ P_\Omega(X_1,\ldots,X_N) = \exp \left[ - \sum_{i=0}^{\Psi} \lambda_i f_{\mu_i}(X_1,\ldots,X_N) \right] $$

(17)

Now with the most unbiased estimate for the distribution being known, the higher-order moments can be calculated as:

$$ \mu'_H = \sum_{\Omega} f_{\mu'_i}(X_1,\ldots,X_N) P_\Omega(X_1,\ldots,X_N) $$

(18)

where $f_{\mu'_i}(X_1,\ldots,X_N)$ now represents the function for the $i$-th higher-order moment. The subscript $H$ emphasizes that the moment was obtained from the maximum entropy distribution. Replacing the originally unknown higher moments with the ones calculated assuming maximum uncertainty and zero information contribution, the system of differential equations for lower moments is now closed and ready to be integrated:

$$ \frac{\partial \mu}{\partial t} = A \cdot \mu + A' \cdot \mu'_H $$

(19)

Given an initial condition, the dynamical evolution of the lower moments can be found and one is able to reconstruct the probability distribution for each time step. Even more important than that, for many biological applications one is interested in evaluating the stationary distribution, which can be readily obtained without necessarily carrying out the numerical integration. Notice that if both lower and higher moments are expressed in terms of the maximum entropy distribution and the lhs of equation (3) is set to zero, then it becomes a system of $\psi$ nonlinear algebraic equations with $\psi$ unknown Lagrange multipliers:

$$ 0 = A \cdot \mu_H + A' \cdot \mu'_H $$

(20)

Again, using a multivariable root-finder method such as Newton–Raphson, determining the solution to this system will specify the stationary probability distribution. As an example, figure 1 shows the solutions from kinetic Monte Carlo simulations (SSA) and from ZI-closure scheme for one of the simplest biological systems already introduced earlier in this paper, the Michaelis–Menten enzymatic kinetics. By doing a mass balance, the original set of four components can be reduced to only two independent variables: substrate (S) and enzyme (E) concentrations. It is observed that the ZI-closure solution (continuous curve) matches the discrete points obtained through simulation very closely. In order to collect those points from SSA, 1000 000 reaction trajectories were simulated and the computational time was around 108 CPU hours. However, the Zero Information method using only up to fourth order moments reached convergence in a matter of 1.5 CPU seconds. The main reason for this difference is that, since only the stationary probability distribution was of interest, in ZI-closure the dynamical steps were not calculated. Smadbeck and Kaznessis [71, 85] have extensively reported more results on the Michaelis–Menten reaction network with ZI-closure scheme.
5. Towards a theory of stability

When analyzing the dynamics of nonlinear systems, one of the major concerns is whether the steady state solutions are stable. More than 120 years ago, the work of Lyapunov [82] established the classical theory of stability for deterministic systems. His ideas set the foundations of the fields of systems engineering and process control.

Up until the 1950s, virtually not a single person would conceive the idea of an oscillatory reaction, for example. However, after the Belousov–Zhabotinsky [83] reaction was discovered, the scientific community had to give more attention to the theory of stability governing the nonlinear dynamics of reaction networks. Lyapunov’s framework was also successfully employed to explain the dynamical behavior of those oscillatory reactions. Likewise in chemical engineering, Warden, Aris and Amundson [84] were able to apply Lyapunov’s theory to describe the possibility of multiple steady states and their stabilities in non-isothermal continuously stirred chemical reactors (CSTRs).

Although, to this day, little work has been done to develop a universal framework for the stability of stochastic chemical reaction networks, the ability to cast the chemical master equation into a deterministic system of moment equations increases our hope of attaining such goal. Smadbeck and Kaznessis [85] proposed the linearization of the dynamical system near its stationary solution through the computation of its steady state Jacobian ($J_{SS}$). Hence, by simply carrying out a Taylor expansion of the right side of equation (19) around the steady state, we can use the following approximation:

$$\frac{\partial \mu}{\partial t} = J_{SS} \mu$$  \hspace{1cm} (21)

Because of this approximation, we are able to evaluate the local stability of the steady state solution through the analysis of the eigenvalues and eigenvectors of the Jacobian matrix, which is formally obtained from equation (3) as:

$$J_{SS} = \left[ \frac{\partial \mu'_H}{\partial \mu} \right]_{SS} = A + \mathcal{N} \left( \frac{\partial \mu'_H}{\partial \lambda} \right)_{SS}$$  \hspace{1cm} (22)

where the subscript $SS$ emphasizes that the matrix is calculated at steady state. Now, if the ZI closure scheme is applied, both higher and lower-order moments are related to each other through the Lagrange multipliers of the maximum entropy distribution. In this case, the chain rule for partial derivatives and some differential manipulation can be used to expand the last term in equation (22):

$$\frac{\partial \mu'_H}{\partial \lambda} \bigg|_{SS} = \frac{\partial \mu'_H}{\partial \mu} \bigg|_{SS} \frac{\partial \lambda}{\partial \mu} \bigg|_{SS} = \frac{\partial \mu'_H}{\partial \lambda} \bigg|_{SS}$$  \hspace{1cm} (23)

Combining equations (22) and (23), we obtain a simple final expression for the steady state Jacobian matrix according to the ZI closure approach:

$$J_{SS} = A + \mathcal{N} \left( \frac{\partial \mu'_H}{\partial \lambda} \right)_{SS}$$  \hspace{1cm} (24)

Based on this definition for the Jacobian matrix, Smadbeck and Kaznessis were able to derive analytical expressions for time correlation and response functions involving its eigenvalues and eigenvectors [85]. Surprisingly, for many systems they observed that fluctuations around the steady state were congruent to perturbation responses, i.e. to the dynamical relaxation from one steady state to another due to external forces and/or parameter variations.

Similar to the fluctuation-dissipation theorem for linear regimes, autocorrelation functions were found to be congruent to response functions. Those results are in agreement with some rigorous mathematical derivations, which suggest that for ergodic continuous-time Markov chains the Lyapunov...
exponential stability is closely related to the stability under parametric perturbation [86, 87]. However, further investigation should be carried out in order to establish whether both phenomena are determined by a common factor.

6. Conclusions

Synthetic biologists may use mathematical models as a helpful tool for designing and engineering new biological phenotypes. And even though modeling of such systems can be a demanding task, there are numerous notable efforts to standardize the languages and formalisms used [22, 88–91]. Synthetic biology models can be employed in computational simulations and thereby be used to generate hypotheses that may be tested experimentally.

The goal of this review paper was to present numerous approaches to modeling the intrinsically stochastic phenomena that occur in biochemical reactions away from the thermodynamic limit. All those methods attempt to address the challenge of solving the chemical master equation and they are divided into two main classes: stochastic simulations involving kinetic Monte Carlo sampling of the probability distribution; and numerical solutions to the deterministic system of moment equations with closure schemes, which enable the reconstruction of the probability distribution.

While the primary challenge with stochastic simulations is their computational cost, the difficulties with the latter approach fall into two categories: efficient generation of moment equations and accurate proposals of closure schemes. It has been suggested in this paper that the use of factorial moments and the zero-information entropy closure scheme are promising strategies. Although thus far only small chemical reaction networks have been numerically explored (e.g. Michaelis–Menten system), the theoretical basis for the method is well founded. Furthermore, the capability of deriving analytical expressions for the stability analysis of nonlinear stochastic systems is intriguing. We speculate that such analysis may prove helpful in guiding the design of synthetic biological systems. After over 50 years from its introduction, the chemical master equation may now be solved for nontrivial systems and the expectation of its utility in synthetic biology efforts is becoming less unrealistic.

Acknowledgments

This work was supported by a grant from the National Institutes of Health (GM111358) and a grant from the National Science Foundation (CBET-1412283). This work utilized the high-performance computational resources of the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575. Computational support from the Minnesota Supercomputing Institute (MSI) is gratefully acknowledged. Support from the University of Minnesota Digital Technology Center, the University of Minnesota Biotechnology Institute and CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil is also acknowledged.

References

[37] Grimà R 2015 Linear-noise approximation and the chemical master equation agree up to second-order moments for a class of chemical systems *Phys. Rev. E* **92** 042124
[40] Gibson M A and Bruck J 2000 Exact stochastic simulation of chemical systems with many species and many channels *J. Phys. Chem.* **105** 1876–89
[68] Gillespie C S 2009 Moment-closure approximations for mass-action models *JET Syst. Biol.* **3** 52–8


