Distribution Approximations for the Chemical Master Equation: Comparison of the Method of Moments and the System Size Expansion

Alexander Andreychenko, Luca Bortolussi, Ramon Grima, Philipp Thomas and Verena Wolf

Abstract The stochastic nature of chemical reactions has resulted in an increasing research interest in discrete-state stochastic models and their analysis. A widely used approach is the description of the temporal evolution of such systems in terms of a chemical master equation (CME). In this paper we study two approaches for approximating the underlying probability distributions of the CME. The first approach is based on an integration of the statistical moments and the reconstruction of the distribution based on the maximum entropy principle. The second approach relies on an analytical approximation of the probability distribution of the CME using the system size expansion, considering higher order terms than the linear noise approximation. We consider gene expression networks with unimodal and multimodal protein distributions to compare the accuracy of the two approaches. We find that both methods provide accurate approximations to the distributions of the CME while having different benefits and limitations in applications.

A. Andreychenko · L. Bortolussi · V. Wolf (✉)
Modelling and Simulation Group, Saarland University, Saarbrücken, Germany
e-mail: wolf@cs.uni-saarland.de

A. Andreychenko
e-mail: makedon@cs.uni-saarland.de

L. Bortolussi
Department of Mathematics and Geosciences, University of Trieste, Trieste, Italy
e-mail: luca@dmi.units.it

R. Grima
School of Biological Sciences, University of Edinburgh, Edinburgh, UK
e-mail: ramon.grima@ed.ac.uk

P. Thomas
Department of Mathematics, Imperial College London, London, UK
e-mail: p.thomas@imperial.ac.uk

© Springer International Publishing Switzerland 2017

F. Graw et al. (eds.), *Modeling Cellular Systems*, Contributions in Mathematical and Computational Sciences 11, DOI 10.1007/978-3-319-45833-5_2

1 Introduction

It is widely recognized that noise plays a crucial role in shaping the behavior of biological systems [1–4]. Part of such noise can be explained by intrinsic fluctuations of molecular concentrations inside a living cell, caused by the randomness of biochemical reactions, and fostered by the low numbers of certain molecular species [2]. As a consequence of this insight, stochastic modeling has rapidly become very popular [5], dominated by Markov models based on the chemical master equation (CME) [5, 6].

The CME represents a system of differential equations that specifies the time evolution of a discrete-state stochastic model that explicitly accounts for the discreteness and randomness of molecular interactions. It has therefore been widely used to model gene regulatory networks, signaling cascades and other intracellular processes which are significantly affected by the stochasticity inherent in reactions involving low number of molecules [7].

A solution of the CME yields the probability distribution over population vectors that count the number of molecules of each chemical species. While a numerical solution of the CME is rather straightforward, i.e. via a truncation of the state space [8, 9], for most networks the combinatorial complexity of the underlying state space renders efficient numerical integration infeasible. Therefore, the stochastic simulation algorithm (SSA), a Monte Carlo technique, is commonly used to derive statistical estimates of the corresponding state probabilities [10].

An alternative to stochastic simulation is to rely on approximation methods that can provide fast and accurate estimates of some aspects of stochastic models. Typically, most approximation methods focus on the estimation of moments of the distributions [11–16]. However, two promising approaches for the approximate computation of the distribution underlying the CME have recently been developed, whose complexity is independent of the molecular population sizes.

The first method discussed here is based on the inverse problem, i.e., reconstructing the probability distribution from its moments. To this end, a closure on the moment equations is employed which yields an approximation of the evolution of all moments up to order K of the joint distribution [14–16]. Thus, instead of solving one equation per population vector we solve $\sum_{k=1}^{K} \binom{N_S+k-1}{k}$ equations if N_S is the number of different chemical species. Given the (approximate) moments at the final time instant, it is possible to reconstruct the corresponding marginal probability distributions using the maximum entropy principle [17, 18]. The reconstruction requires the solution of a nonlinear constrained optimization problem. Nevertheless, the integration of the moment equations and the reconstruction of the underlying distribution can for most systems be carried out very efficiently and thus allows a fast approximation of the CME.

The second method, which is based on van Kampen's system size expansion [19], does not resort to moments but instead represents a direct analytical approximation of the probability distribution of the CME. Unlike the method of moments, the technique assumes that the distribution can be expanded about its deterministic limit

rate equations using a small parameter called the system size. For biochemical systems, the system size coincides with the volume to which the reactants are confined. The leading order term of this expansion is given by the linear noise approximation which predicts that the fluctuations about the rate equation solution are approximately Gaussian distributed [19] in agreement with the central limit theorem that is valid for large number of molecules. For low molecule numbers, the non-Gaussian corrections to this law can be investigated systematically using the higher order terms in the system size expansion. A solution to this expansion has recently been given in closed form as a series of the probability distributions in the inverse system size [20]. Although in general the positivity of this distribution approximation cannot be guaranteed, it often provides simple and accurate analytical approximations to the non-Gaussian distributions underlying the CME.

Since many reaction networks involve very small molecular populations, it is often questionable whether the system size expansion and moment based approaches can appropriately capture their underlying discreteness. For example, the state of a gene that can either be 'on' or 'off' while the number of mRNA molecules is of the order of only a few tens on average. In these situations, hybrid approaches are more appropriate, and supported by convergence results in a hybrid sense in the thermodynamic limit [21]. A hybrid moment approach for the solution of the CME integrates a small master equation for the small populations while a system of conditional moment equations is integrated for the large populations which is coupled with the equation for the small populations. Equations of unlikely states of the small populations can be truncated and then, with a similar computational effort, a more accurate approximation of the CME is obtained compared to the standard method of moments. Similarly, a conditional system size expansion can be constructed that tracks the probabilities of the small populations and applies the system size expansion to the large populations conditionally. Presently, such an approach has employed the linear noise approximation for gene regulatory networks with slow promoters invoking timescale separation [22]. The validity of a conditional system size expansion including higher than linear noise approximation terms is, however, still under question.

Given these recent developments, it remains to be clarified how these two competing approximation methods perform in practice. Here, we carry out a comparative study between numerical results obtained using the method of moments and analytical results obtained from the system size expansion for two common gene expression networks. The outline of the manuscript is the following: In Sect. 2 we will briefly review the CME formulation. Section 3 outlines the methods of (conditional) moments and the reconstruction of the probability distributions using the maximum entropy principle. In Sect. 4 the approximate solution of the CME using the SSE is reviewed. We then carry out two detailed case studies in Sect. 5. In the first case study, we investigate a model of a constitutively expressed gene leading to a unimodal protein distribution. In a second example we study the efficacy of the described hybrid approximations using the method of conditional moments and the conditional system size expansion for the prediction of multimodal protein distributions from the expression of a self-activating gene. These two examples are typical

scenarios encountered in more complex models, and as such provide ideal benchmarks for a qualitative and quantitative comparison of the two methods. We conclude with a discussion in Sect. 6.

2 Stochastic Chemical Kinetics

A biochemical reaction network is specified by a set of N_S different chemical species $S_1, \ldots, S_{N_S}$ and by a set of R reactions of the form

$$\ell^-_{r,1} S_1 + \cdots + \ell^-_{r,N_S} S_{N_S} \xrightarrow{k_r} \ell^+_{r,1} S_1 + \cdots + \ell^+_{r,N_S} S_{N_S}, \quad 1 \le r \le R.$$

Given a reaction network, we define a continuous-time Markov chain $\{\mathbf{X}(t), t \ge 0\}$, where $\mathbf{X}(t) = (X_1(t), \ldots, X_{N_S}(t))$ is a random vector whose i-th entry $X_i(t)$ is the number of molecules of type S_i. If $\mathbf{X}(t) = \mathbf{x} = (x_1, \ldots, x_{N_S}) \in \mathbb{N}_0^{N_S}$ is the state of the process at time t and $x_i \ge \ell^-_{r,i}$ for all i, then the r-th reaction corresponds to a possible transition from state $\mathbf{x}$ to state $\mathbf{x} + \mathbf{v}_r$ where $\mathbf{v}_r$ is the change vector with entries $v_{r,i} = \ell^+_{r,i} - \ell^-_{r,i} \in \mathbb{Z}^{N_S}$. The rate of the reaction is given by the propensity function $\gamma_r(\mathbf{x})$, with $\gamma_r(\mathbf{x})\mathrm{d}t$ being the probability of a reaction of index r occurring in a time instant $\mathrm{d}t$, assuming that the reaction volume is well-stirred. The most common form of propensity function follows from the principle of mass action and depends on the volume Ω to which the reactants are confined, $\gamma_r(\mathbf{x}) := \Omega k_r \prod_{i=1}^{N_S} \Omega^{-\ell^-_{r,i}} \binom{x_i}{\ell^-_{r,i}}$, as it can be derived on the basis of physical kinetics [6, 23].

We now fix the initial condition $\mathbf{x}_0$ of the Markov process to be deterministic and let $\Pi(\mathbf{x}, t) = \mathrm{Prob}(\mathbf{X}(t) = \mathbf{x} \,|\, \mathbf{X}(0) = \mathbf{x}_0)$ for $t \ge 0$. The time evolution of $\Pi(\mathbf{x}, t)$ is governed by the Chemical Master Equation (CME) as

$$\frac{\mathrm{d}\Pi(\mathbf{x}, t)}{\mathrm{d}t} = \sum_{r=1}^{R} \left(\gamma_r(\mathbf{x} - \mathbf{v}_r)\Pi(\mathbf{x} - \mathbf{v}_r, t) - \gamma_r(\mathbf{x})\Pi(\mathbf{x}, t)\right). \tag{1}$$

We remark that the probabilities $\Pi(\mathbf{x}, t)$ are uniquely determined when we consider the equations of all states that are reachable from the initial conditions.

Example 1 In the sequel we will use the following system as a running example. It describes the expression of a gene with enzymatic degradation [24].

$$\varnothing \xrightarrow{k_0} M \xrightarrow{k_1} \varnothing, \quad M \xrightarrow{k_2} M + P, \quad P \xrightarrow{f(x_P)} \varnothing,$$

where $f(x_P)$ is a propensity function that depends on the number of molecules of type P (see also Sect. 5.1 for more details). A state of the associated Markov process is a vector $\mathbf{x} = (x_M, x_P)$ where $x_M, x_P \in \mathbb{N}_0$ are the number of mRNA (M) and protein (P) molecules in the system.

3 Method of Moments

For most realistic examples the number of reachable states is extremely large or even infinite, which renders an efficient numerical integration of Eq. (1) impossible. An approximation of the moments of the distribution over time can be obtained by considering the corresponding moment equations that describe the dynamics of the moments up to order $K < \infty$. We refer to this approach as the method of moments (MM). In this section we will briefly discuss the derivation of the corresponding moment equations following the lines of Ale et al. [16].

Let $\mathbf{f} : \mathbb{N}_0^{N_S} \to \mathbb{R}^n$ be a function that is independent of t where $n \in \mathbb{N}$. In the sequel we will exploit the following relationship,

$$\begin{aligned} \tfrac{d}{dt} E(\mathbf{f}(\mathbf{X}(t))) &= \sum_{\mathbf{x}} \mathbf{f}(\mathbf{x}) \cdot \tfrac{d}{dt} \Pi(\mathbf{X}(t) = \mathbf{x}) \\ &= \sum_{r=1}^{R} E\big[\gamma_r(\mathbf{X}(t)) \cdot (\mathbf{f}(\mathbf{X}(t) + \mathbf{v}_r) - \mathbf{f}(\mathbf{X}(t)))\big]. \end{aligned} \tag{2}$$

For $\mathbf{f}(\mathbf{x}) = \mathbf{x}$ this yields a system of equations for the population means

$$\tfrac{d}{dt} E(\mathbf{X}(t)) = \sum_{r=1}^{R} \mathbf{v}_r E(\gamma_r(\mathbf{X}(t))). \tag{3}$$

Note that the system of ODEs in Eq. (3) is only closed if at most monomolecular mass action reactions ($\sum_{i=1}^{N_S} \ell_{r,i}^- \leq 1$) are involved. For most networks the latter condition is not true and higher order moments appear on the right side.

Let us write $\mu_i(t)$ for $E(X_i(t))$ and $\boldsymbol{\mu}(t)$ for the vector with entries $\mu_i(t)$, $1 \leq i \leq N_S$. Then a Taylor expansion of the function $\gamma_r(\mathbf{X}(t))$ about the mean $E(\mathbf{X}(t))$ yields

$$\begin{aligned} E(\gamma_r(\mathbf{X})) = \gamma_r(\boldsymbol{\mu}) &+ \tfrac{1}{1!} \sum_{i=1}^{N_S} E(X_i - \mu_i) \tfrac{\partial}{\partial \mu_i} \gamma_r(\boldsymbol{\mu}) \\ &+ \tfrac{1}{2!} \sum_{i=1}^{N_S} \sum_{k=1}^{N_S} E((X_i - \mu_i)(X_k - \mu_k)) \tfrac{\partial^2}{\partial \mu_i \partial \mu_k} \gamma_r(\boldsymbol{\mu}) + \ldots, \end{aligned} \tag{4}$$

where we omitted t to improve the readability. Note that $E(X_i(t) - \mu_i) = 0$ and assuming that all propensities follow the law of mass action and all reactions are at most bimolecular, the terms of order three and more disappear. By letting C_{ik} be the covariance $E[(X_i(t) - \mu_i)(X_k(t) - \mu_k)]$ we get

$$E(\gamma_r(\mathbf{X})) = \gamma_r(\boldsymbol{\mu}) + \tfrac{1}{2} \textstyle\sum_{i=1}^{N_S} \sum_{k=1}^{N_S} C_{ik} \tfrac{\partial^2}{\partial \mu_i \partial \mu_k} \gamma_r(\boldsymbol{\mu}). \tag{5}$$

Next, we derive an equation for the covariances by first exploiting the relationship

$$\frac{d}{dt} C_{ik} = \frac{d}{dt} E(X_i X_k) - \frac{d}{dt}(\mu_i \mu_k) = \frac{d}{dt} E(X_i X_k) - \left(\frac{d}{dt}\mu_i\right)\mu_k - \mu_i \left(\frac{d}{dt}\mu_k\right), \tag{6}$$

and if we couple this equation with the equations for the means, the only unknown term that remains is the derivative $\frac{d}{dt}E(X_iX_k)$ of the second moment. For this, we can apply the same strategy as before by using Eq. (2) for the test function $f(\mathbf{x}) := x_i x_k$. This equation will contain the expectations $E(\gamma_r(\mathbf{X}))$ and $E(\gamma_r(\mathbf{X})X_i)$ for which we again consider the Taylor expansion about the mean. For the expansion of $E(\gamma_r(\mathbf{X})X_i)$ moments of order three come into play since derivatives of order three of $\gamma_r(\mathbf{x})x_i$ may be nonzero. It is possible to take these terms into account by deriving additional equations for moments of order three and higher. These equations will then include moments of even higher order such that theoretically we end up with an infinite system of equations. Different strategies to close the equations have been proposed in the literature [25–29]. Here we consider a low dispersion closure and assume that all moments of order $>K$ that are centered around the mean are equal to zero. E.g., if we choose $K = 2$, then we can obtain a closed system of equations that does not include higher order terms. Then we can integrate the time evolution of the means and that of the covariances and variances.

Example 2 To illustrate the method we consider Example 1. Assuming that all central moments of order three and higher are equal to zero, we get the following equations for the means of the species.

$$\begin{aligned} \frac{d}{dt}\mu_M &= \Omega k_0 - k_1\mu_M \\ \frac{d}{dt}\mu_P &= k_1\mu_M - f(\mu_P) - \frac{1}{2}\frac{\partial^2 f(\mu_P)}{\partial {\mu_P}^2}C_{P,P} \end{aligned} \tag{7}$$

where the expectations of the species are given by μ_M and μ_P and the variance of P is given by $C_{P,P}$. Equations for covariances are omitted.

In many situations, the approximation provided by the MM approach is very accurate even if only the means and covariances are considered. In general, however, numerical results show that the approximation tends to become worse if systems exhibit complex behavior such as multistability or oscillations and the resulting equations may become very stiff [14, 30]. For some systems increasing the number of moments improves the accuracy [16]. Generally, however such a convergence is not seen [31], except in the limit of large volumes [32].

3.1 Equations for Conditional Moments

For many reactions networks a hybrid moment approach, called method of conditional moments (MCM), can be more advantageous, in which we decompose $\mathbf{X}(t)$ into small and large populations. The reason is that small populations (often describing the activation state of a gene) have distributions that assign a significant mass of probability only to a comparatively small number of states. In this case we can

integrate the probabilities directly to get a more accurate approximation of the CME compared to an integration of the moments.

Formally, we consider $\mathbf{X}(t) = (\mathbf{Y}(t), \mathbf{Z}(t))$, where $\mathbf{Y}(t)$ corresponds to the small, and $\mathbf{Z}(t)$ to the large populations. Similarly, we write $\mathbf{x} = (\mathbf{y}, \mathbf{z})$ for the states of the process and $\mathbf{v}_r = (\hat{\mathbf{v}}_r, \tilde{\mathbf{v}}_r)$ for the change vectors, $r \in \{1, \ldots, R\}$. Then, we condition on the state of the small populations and apply the MM to the conditional moments but not to the distribution of $\mathbf{Y}(t)$. The probabilities for $\mathbf{Y}(t) = \mathbf{y}$ are called mode probabilities. Now, Eq. (1) becomes

$$\frac{d\Pi(\mathbf{y}, \mathbf{z})}{dt} = \sum_{r=1}^{R} [\gamma_r(\mathbf{y} - \hat{\mathbf{v}}_r, \mathbf{z} - \tilde{\mathbf{v}}_r)\Pi(\mathbf{y} - \hat{\mathbf{v}}_r, \mathbf{z} - \tilde{\mathbf{v}}_r) - \gamma_r(\mathbf{y}, \mathbf{z})\Pi(\mathbf{y}, \mathbf{z})]. \quad (8)$$

Next, we sum over all possible $\mathbf{z}$ to get the time evolution of the mode probabilities $\Pi(\mathbf{y}) = \sum_{\mathbf{z}} \Pi(\mathbf{y}, \mathbf{z})$.

$$\begin{aligned} \tfrac{d}{dt}\Pi(\mathbf{y}) &= \sum_{\mathbf{z}} \sum_{r=1}^{R} \gamma_r(\mathbf{y} - \hat{\mathbf{v}}_r, \mathbf{z} - \tilde{\mathbf{v}}_r)\Pi(\mathbf{y} - \hat{\mathbf{v}}_r, \mathbf{z} - \tilde{\mathbf{v}}_r) - \sum_{\mathbf{z}} \sum_{r=1}^{R} \gamma_r(\mathbf{y}, \mathbf{z})\Pi(\mathbf{y}, \mathbf{z}) \\ &= \sum_{r=1}^{R} \Pi(\mathbf{y} - \hat{\mathbf{v}}_r) E[\gamma_r(\mathbf{y} - \hat{\mathbf{v}}_r, \mathbf{Z}) \,|\, \mathbf{Y} = \mathbf{y} - \hat{\mathbf{v}}_r] - \sum_{r=1}^{R} \Pi(\mathbf{y}) E[\gamma_r(\mathbf{y}, \mathbf{Z}) \,|\, \mathbf{Y} = \mathbf{y}]. \end{aligned} \quad (9)$$

Note that in this small master equation that describes the change of the mode probabilities over time, the sum runs only over those reactions that modify $\mathbf{y}$, since for all other reactions the terms cancel out. Moreover, on the right side we have only mode probabilities of neighboring modes and conditional expectations of the $\mathbf{z}$-part of the reaction rate. For the latter, we can use a Taylor expansion about the conditional population means. Similar to Eq. (5), this yields an equation that involves the conditional means and centered conditional moments of second order (variances and covariances). Thus, in order to close the system of equations, we need to derive equations for the time evolution of the conditional means and centered conditional moments of higher order. Since the mode probabilities may become zero, we first derive an equation for the evolution of the partial means (conditional means multiplied by the probability of the condition)

$$\begin{aligned} \tfrac{d}{dt}\left(E[\mathbf{Z} \mid \mathbf{y}]\, \Pi(\mathbf{y})\right) &= \sum_{\mathbf{z}} \mathbf{z} \tfrac{d}{dt}\Pi(\mathbf{y}, \mathbf{z}) \\ &= \sum_{r=1}^{R} E[(\mathbf{Z} + \tilde{\mathbf{v}}_r)\gamma_r(\mathbf{y} - \hat{\mathbf{v}}_r, \mathbf{Z}) \mid \mathbf{y} - \hat{\mathbf{v}}_r]\, \Pi(\mathbf{y} - \tilde{\mathbf{v}}_r) \\ &\quad - \sum_{r=1}^{R} E[\mathbf{Z}\gamma_r(\mathbf{y}, \mathbf{Z}) \mid \mathbf{y}]\, \Pi(\mathbf{y}), \end{aligned} \quad (10)$$

where in the second line we applied Eq. (8) and simplified the result. The conditional expectations $E[(\mathbf{Z} + \tilde{\mathbf{v}}_r)\gamma_r(\mathbf{y} - \hat{\mathbf{v}}_r, \mathbf{Z}) \mid \mathbf{y} - \hat{\mathbf{v}}_r]$ and $E[\mathbf{Z}\gamma_r(\mathbf{y}, \mathbf{Z}) \mid \mathbf{y}]$ are then replaced by their Taylor expansion about the conditional means such that the

equation involves only conditional means and higher centered conditional moments [33]. For higher centered conditional moments, similar equations can be derived. If all centered conditional moments of order higher than K are assumed to be zero, the result is a (closed) system of differential algebraic equations (algebraic equations are obtained whenever a mode probability $\Pi(\mathbf{y})$ is equal to zero). However, it is possible to transform the system of differential algebraic equations into a system of (ordinary) differential equations after truncating modes with insignificant probabilities. Then we can get an accurate approximation of the solution after applying standard numerical integration methods. For the case study in Sect. 5 we derived the system of ODEs using the tool SHAVE [34] which implements a truncation based approach and solves the ODEs using an explicit Runge–Kutta method.

Example 3 We apply the MCM approach to Example 1. The modes of the system correspond to the number of molecules of type M. Here we provide equations only for the first two modes, $M = 0$ and $M = 1$. The equations for the corresponding mode probabilities (p_0, p_1) and the expected number of proteins ($\mu_{P,0}$, $\mu_{P,1}$) conditioned on the modes are given below. This equation system has to be closed not only by setting to zero all the central moments of order higher than 2, but also by truncating the infinite population of M. This can be done by assuming that $p_n = 0$ from a certain number n on.

$$
\begin{aligned}
&\frac{d}{dt} p_0 = -\Omega k_0 p_0 + k_1 p_1 \\
&\frac{d}{dt} p_1 = \Omega k_0 p_0 - k_1 p_1 - \Omega k_0 p_1 \\
&\frac{d}{dt} \left(\mu_{P,0}\, p_0\right) = -\Omega k_0 \mu_{P,0} p_0 + k_1 \mu_{P,1} p_1 - f(\mu_{P,0}) p_0 - \frac{1}{2} \frac{\partial^2 f(\mu_{P,0})}{\partial {\mu_{P,0}}^2} C_{P,P,0} p_0 \\
&\frac{d}{dt} \left(\mu_{P,1}\, p_1\right) = \Omega k_0 \mu_{P,0} p_0 - \Omega k_0 \mu_{P,1} p_1 - k_1 \mu_{P,1} p_1 - f(\mu_{P,1}) p_1 \\
&\qquad\qquad - \frac{1}{2} \frac{\partial^2 f(\mu_{P,1})}{\partial {\mu_{P,1}}^2} C_{P,P,1} p_1 + k_2 p_1
\end{aligned}
\tag{11}
$$

3.2 Maximum Entropy Distribution Reconstruction

Given the (approximated) moments of a distribution up to order K it is possible to reconstruct the corresponding probability distribution. Since a finite number of moments defines a set of distributions, we apply the *maximum entropy principle* where we choose among the distributions, that fulfill the moment equations, the one that maximizes the entropy. For instance, the normal distribution is chosen among all continuous distributions with equal mean and variance.

In the sequel we describe how to obtain the one-dimensional marginal probability distributions of a reaction network when we use the moments up to order K. We mostly follow Andreychenko et al. [17] and simply write X for the corresponding molecular count at some fixed time instant t. Given a sequence of $K + 1$ noncentral

moments[1] $E\left(X^k\right) = \mu^{(k)}, k = 0, 1, \ldots, K$, the set $\mathcal{G}$ of allowed (discrete) probability distributions consists of all non-negative functions g for which the following conditions hold:

$$\sum_x x^k g(x) = \mu^{(k)}, k = 0, 1, \ldots, K. \tag{12}$$

Here x ranges over possible arguments (usually $x \in \mathbb{N}_0$) with positive probability. Note that we have included the constraint $\mu_0 = 1$ in order to guarantee that g is a probability distribution. According to the maximum entropy principle, we choose the distribution $q \in \mathcal{G}$ that maximizes the entropy $H(g)$, i.e.,

$$q = \arg\max_{q \in \mathcal{G}} H(g) = \arg\max_{g \in \mathcal{G}} \left(-\sum_x g(x) \ln g(x) \right). \tag{13}$$

The problem of finding the maximum entropy distribution is a nonlinear constrained optimization problem that can be addressed by considering the Lagrangian functional

$$\mathcal{L}(g, \lambda) = H(g) - \sum_{k=0}^{K} \lambda_k \left(\sum_x x^k g(x) - \mu^{(k)} \right),$$

where $\lambda = (\lambda_0, \ldots, \lambda_K)$ are the corresponding Lagrangian multipliers. The maximum of the unconstrained Lagrangian $\mathcal{L}$ corresponds to the solution of the constrained maximum entropy problem (13). Note that setting the derivatives of $\mathcal{L}(g, \lambda)$ w.r.t. λ_k, to zero results in the moment constraints. The general form of the maximum is obtained by setting $\frac{\partial \mathcal{L}}{\partial g(x)}$ to zero which yields

$$q(x) = \exp\left(-1 - \sum_{k=0}^{K} \lambda_k x^k \right) = \tfrac{1}{Z} \exp\left(-\sum_{k=1}^{K} \lambda_k x^k \right),$$

where

$$Z = e^{1+\lambda_0} = \sum_x \exp\left(-\sum_{k=1}^{M} \lambda_k x^k \right) \tag{14}$$

is a normalization constant. The last equality in Eq. (14) follows from the fact that q is a distribution and thus λ_0 is uniquely determined by $\lambda_1, \ldots, \lambda_K$. Next we insert the above general form into the Lagrangian, thus transforming the problem into an unconstrained convex minimization problem of the dual function w.r.t. the variables λ_k. This yields the dual function

[1] Noncentral moments can be easily obtained from central ones. For instance, the second noncentral moment $\mu^{(2)}$ is obtained from the variance σ^2 and the mean μ as $\mu^{(2)} = \sigma^2 + \mu^2$.

$$\Psi(\lambda) = \ln Z + \sum_{k=1}^{K} \lambda_k \mu^{(k)}. \tag{15}$$

According to the Kuhn–Tucker theorem, the solution $\lambda^* = \arg\min \Psi(\lambda)$ of the minimization problem determines the solution q of the original constrained optimization problem in Eq. (13) (see [35]). We solve this unconstrained optimization problem using the Newton method from the MATLAB's numerical minimization package `minFunc`, where we choose $\lambda^{(0)} = (0, \ldots, 0)$ as an initial starting point and use the approximated gradient and Hessian matrix. Since for the systems that we consider, the dual function is convex [36–38], there exists a unique minimum $\lambda^* = \left(\lambda_1^*, \ldots, \lambda_K^*\right)$ where all first derivatives are zero and where the Hessian is positive definite. The final results λ^* of the iteration yields the distribution

$$\tilde{q}(x) = \exp\left(-1 - \sum_{k=0}^{K} \lambda_k^* x^k\right),$$

which is an approximation of the marginal distribution of the process at time t.

The sequence of moments $\mu^{(k)}, k = 0, \ldots, K$ obtained using MM or MCM serves as an input to the maximum entropy reconstruction procedure. Due to the high sensitivity with respect to the accuracy of the highest order moment $E\left(X_P^K\right)$, we compute all moments up to $E\left(X_P^{K+1}\right)$ to get a better estimation but use moments only up to $E\left(X_P^K\right)$ in the entropy maximization.

The maximum entropy approach provides a useful extension of moment-based integration methods of the CME. In order to reconstruct a probability distribution, we require that moments of the reconstructed distribution match with those obtained by the method of moments. We apply the maximum entropy principle to the infinitely many distributions satisfying these constraints, i.e., we pick the distribution that maximizes the uncertainty. The advantage of this choice are that we can efficiently learn the maximum entropy distribution in one or two dimensions. Furthermore, it seems to produce reasonably accurate results, as validated a posteriori [17, 18].

The reconstruction of the distributions of higher dimension is more involved due to numerical instabilities arising when using the ill-conditioned Hessian matrix [36, 39] in the optimization procedure. As mentioned in the numerical results that we present in the sequel, problems may arise if the support of the distribution (which serves as an input argument to the optimization procedure) is not chosen adequately. The true support of the distribution is often infinite and a reasonable truncation has to be used [40]. A possible solution is addressed in [18] where we introduce an iterative heuristic-based procedure of the support approximation. However, generally this approach gives very accurate results relative to the information about the distribution given by the moment constrains.

4 System Size Expansion of the Probability Distribution

We here describe the use of the system size expansion [19] to obtain approximate but simple expressions for the probability distributions of the CME. For simplicity, we will focus on the case of a single species and follow the approach developed by *Thomas and Grima* [20]. The system size expansion makes use of the macroscopic limit of the CME which is attained for large reaction volumes. When concentrations are held constant, large volumes imply large number of molecules, and hence the macroscopic concentration $[X]$ follows the deterministic rate equation

$$\frac{\mathrm{d}[X]}{\mathrm{d}t} = \sum_{r=1}^{R} \nu_r f_r^{(0)}([X]). \tag{16}$$

Note that here $\nu_r = \nu_{r,1}$ because we only consider a single species. A prerequisite for Eq. (16) to be the deterministic limit of the CME is that the rate functions satisfy the scaling

$$\gamma_j(\Omega[X], \Omega) = \Omega\left[f_r^{(0)}([X]) + \Omega^{-1} f_r^{(1)}([X]) + \ldots\right], \tag{17}$$

which holds for instance in the case of mass action kinetics [20]. The first term in this series, $f_r^{(0)}([X]) = \lim_{\Omega\to\infty} \frac{\gamma_r(\Omega[X])}{\Omega}$, is the macroscopic rate function. Note that for an unimolecular reaction only the first term in Eq. (17) is present while for a bimolecular one the first two terms are nonzero.

The expansion allows to characterize the deviations from this deterministic behavior by successively taking into account higher order terms. Specifically, van Kampen proposed separating the dynamics of the molecular concentration into the deterministic part $[X]$ and a fluctuating component ε that reads

$$\frac{x}{\Omega} = [X] + \Omega^{-1/2}\varepsilon. \tag{18}$$

This ansatz can be used to expand the CME in powers of the inverse square root of the volume by changing variables from x to ε. Assuming that this transformation is continuous, i.e., $\Pi(\varepsilon, t) = \Pi(\Omega[X] + \Omega^{1/2}\varepsilon, t)\Omega^{1/2}$, the CME becomes

$$\frac{\partial}{\partial t}\Pi(\varepsilon, t) = \mathcal{L}_0\Pi(\varepsilon, t) + \sum_{k=1}^{N} \Omega^{-k/2}\mathcal{L}_k\Pi(\varepsilon, t) + O(\Omega^{-(N+1)/2}). \tag{19}$$

When the above equation is truncated after the N^{th} term, it approximates the CME by a partial differential equation. It is shown in Ref. [20] that the differential operators $\mathcal{L}_k$ can be written down explicitly and are given by

$$\mathcal{L}_k = \sum_{s=0}^{\lceil k/2 \rceil} \sum_{p=1}^{k-2(s-1)} \frac{\mathcal{D}_{p,s}^{k-p-2(s-1)}}{p!(k-p-2(s-1))!} (-\partial_\varepsilon)^p \varepsilon^{k-p-2(s-1)}, \tag{20}$$

where the coefficients

$$\mathcal{D}_{p,s}^{q} = \sum_{r=1}^{R} (\nu_r)^p \frac{\partial^q f_r^{(s)}([X])}{\partial [X]^q}, \tag{21}$$

depend only on the solution of the rate equation. We will solve Eq. (19) perturbatively by expanding the probability density as

$$\Pi(\varepsilon, t) = \sum_{j=0}^{N} \Omega^{-j/2} \pi_j(\varepsilon, t) + O(\Omega^{-(N+1)/2}). \tag{22}$$

Using the above series in Eq. (19) and equating order Ω^0 terms, one finds

$$\left(\frac{\partial}{\partial t} - \mathcal{L}_0 \right) \pi_0 = 0, \tag{23a}$$

while equating terms to order $\Omega^{-j/2}$, one finds

$$\left(\frac{\partial}{\partial t} - \mathcal{L}_0 \right) \pi_j(\varepsilon) = \mathcal{L}_1 \pi_{j-1} + \cdots + \mathcal{L}_j \pi_0, \tag{23b}$$

for $j > 0$. The first equation, Eq. (23a), is called the linear noise approximation [19] and its solution is a Gaussian distribution

$$\pi_0(\varepsilon, t) = \frac{1}{\sqrt{2\pi\sigma^2(t)}} \exp\left(-\frac{\varepsilon^2}{2\sigma^2(t)} \right), \tag{24}$$

with zero mean meaning that the rate equation are valid on average. Its variance $\sigma^2(t)$ follows the equation

$$\frac{\partial \sigma^2}{\partial t} = 2\mathcal{J}(t)\sigma^2 + \mathcal{D}_2^0(t), \tag{25}$$

where we have denoted by $\mathcal{J}(t) = \mathcal{D}_1^1(t)$ the Jacobian of the rate Eq. (16).

The solution of the partial differential equations (23b) can then be obtained using the eigenfunctions of $\mathcal{L}_0$. In particular, we can write

$$\pi_j(\varepsilon, t) = \sum_{m=1}^{3j} a_m^{(j)}(t) \psi_m(\varepsilon, t) \pi_0(\varepsilon, t)$$

where $\psi_m(\varepsilon,t) = \pi_0^{-1}(-\partial_\varepsilon)^m\pi_0 = \frac{1}{\sigma^m(t)}H_m\left(\frac{\varepsilon}{\sigma(t)}\right)$ and H_m are the Hermite polynomials. The solution of the system size expansion is therefore

$$\Pi(\varepsilon,t) = \pi_0(\varepsilon,t)\left(1+\sum_{j=1}^{N}\Omega^{-j/2}\sum_{m=1}^{3j}a_m^{(j)}(t)\psi_m(\varepsilon,t)\right) + O(\Omega^{-(N+1)/2}). \quad (26a)$$

Mathematically speaking the above equation represents an asymptotic series solution to the CME. Equations for the coefficients can be obtained using the orthogonality of the Hermite polynomials, and are given the ordinary differential equations

$$\left(\frac{\partial}{\partial t} - n\mathcal{J}\right)a_n^{(j)} = \sum_{k=1}^{j}\sum_{m=0}^{3(j-k)}a_m^{(j-k)}\sum_{s=0}^{\lceil k/2\rceil}\sum_{p=1}^{k-2(s-1)}\mathcal{D}_{p,s}^{k-p-2(s-1)}\mathcal{I}_{mn}^{p,k-p-2(s-1)}, \quad (26b)$$

with

$$\mathcal{I}_{mn}^{\alpha\beta} = \frac{\sigma^{\beta-\alpha+n-m}}{\alpha!}\sum_{s=0}^{\min(n-\alpha,m)}\binom{m}{s}\frac{(\beta+\alpha+2s-(m+n)-1)!!}{(\beta+\alpha+2s-(m+n))!(n-\alpha-s)!}, \quad (26c)$$

and zero for odd $(\alpha+\beta)-(m+n)$. Note that $a_n^{(j)} = 0$ when $n+j$ is odd. Explicit expressions for the probability density can now be evaluated to any desired order. Relatively simple and explicit solutions are obtained in steady state conditions by letting the time derivative on the left-hand side of Eq. (26b) go to zero. It follows that the first term in Eq. (26a), the linear noise approximation π_0, describes the distribution in the infinite volume limit. For finite volumes, implying finite number of molecules, higher order terms given by Eq. (26) have to be taken into account.

It is, however, the case that this approximation can become inaccurate for processes whose mean behavior differs significantly from the rate equation. This is because in van Kampen's ansatz, Eq. (18), the average concentration is approximated by the solution of the rate equation $[X]$. For biochemical networks involving bimolecular reactions, the rate equation provides only an approximation to the true average predicted by the CME because its propensities depend nonlinearly on the concentrations [41]. In applications it is important to account for these deviations from the rate equation solution and the linear noise approximation. We therefore calculate the true concentration mean and variance using the system size expansion a priori and then perform the expansion about the true mean. A posteriori, this leads to an expansion about the mean

$$\frac{x}{\Omega} = \underbrace{\left\langle\frac{x}{\Omega}\right\rangle}_{\text{mean}} + \underbrace{\Omega^{-1/2}\bar{\varepsilon}}_{\text{fluctuations}}, \quad (27)$$

which is different than the one proposed by van Kampen, Eq. (18), who expands the CME about the solution of the rate equation. Here, the averages are calculated from $\left\langle \frac{x}{\Omega} \right\rangle = [X] + \Omega^{-1/2}\langle \varepsilon \rangle$ such that $\bar{\varepsilon} = \varepsilon - \langle \varepsilon \rangle$ is a centered random variable quantifying the fluctuations about the true average. The result is the following expansion

$$\bar{\pi}(\bar{\varepsilon}, t) = \bar{\pi}_0(\bar{\varepsilon}, t) + \sum_{j=1}^{N} \Omega^{-j/2} \sum_{m=1}^{3j} \bar{a}_m^{(j)} \psi(\bar{\varepsilon}, t) \bar{\pi}_0(\bar{\varepsilon}, t) + O(\Omega^{-(N+1)/2}), \quad (28a)$$

where $\bar{\pi}_0(\bar{\varepsilon})$ is a Gaussian different from the linear noise approximation because it is centered about the true mean (instead of the rate equation). The coefficients in the above equation can be calculated from

$$\bar{a}_m^{(j)} = \sum_{k=0}^{j} \sum_{n=0}^{3k} a_n^{(k)} \kappa_{m-n}^{(j-k)}, \quad (28b)$$

and

$$\kappa_j^{(n)} = \frac{1}{n!} \sum_{m=0}^{\lfloor j/2 \rfloor} (-1)^{(j+m)} \sum_{k=j-2m}^{n-m} \binom{n}{k} B_{k,j-2m}\left(\left\{\zeta! a_1^{(\zeta)}\right\}\right) B_{n-k,m}\left(\left\{\frac{\zeta!}{2} \bar{\sigma}_{(\zeta)}^2\right\}\right). \quad (28c)$$

Here $a_1^{(j)}$ and $\bar{\sigma}_{(j)}^2 = 2\left[a_2^{(j)} - B_{j,2}(\{\zeta! a_1^{(\zeta)}\})/j!\right]$ denote the coefficients in the expansions of mean and variance

$$\langle \varepsilon \rangle = \sum_{j=1}^{N} \Omega^{-j/2} a_1^{(j)} + O(\Omega^{-(N+1)/2}), \quad (29a)$$

$$\bar{\sigma}^2 = \sigma^2 + \sum_{j=1}^{N} \Omega^{-j/2} \sigma_{(j)}^2 + O(\Omega^{-(N+1)/2}), \quad (29b)$$

respectively, and $B_{k,n}(\{y_\zeta\})$ denotes the partial Bell polynomials [42] defined as

$$B_{n,k}(\{y_\zeta\}_{\zeta=1}^{n-k+1}) = \sum{}' \frac{n!}{j_1! \dots j_{n-k+1}!} \left(\frac{y_1}{1!}\right)^{j_1} \cdots \left(\frac{y_{n-k+1}}{(n-k+1)!}\right)^{j_{n-k+1}}. \quad (30)$$

Note that $\sum'$ denotes the summation over all sequences $j_1, \dots, j_{n-k+1}$ of non-negative integers such that $j_1 + \cdots + j_{n-k+1} = k$ and $j_1 + 2j_2 + \cdots + (n-k+1)$ $j_{n-k+1} = n$. Note that the expansion about the mean has generally less nonzero coefficients because $\bar{a}_1^{(j)} = \bar{a}_2^{(j)} = 0$ for all j. Note also that for systems whose propensities depend at most linearly on the concentrations, mean and variance are exact to order Ω^0 (linear noise approximation), and hence for this case expansion (26a) coincides

with Eq. (28a). In Sect. 5 we show that typically a few terms in this expansion (28a) are sufficient to capture the underlying distributions of the CME.

A particularly relevant case, namely the stationary solution of the CME, turns out to be obtained quite straightforwardly. For example, truncating after Ω^{-1}-terms, from Eq. (29) it follows that $\langle \varepsilon \rangle = \Omega^{-1/2} a_1^{(1)} + O(\Omega^{-3/2})$ and $\bar{\sigma}^2 = \sigma^2 + \Omega^{-1}(2a_2^{(2)} - (a_1^{(1)})^2) + O(\Omega^{-3/2})$. Letting now the left hand side of Eq. (26b) go to zero, solving for the coefficients $a_n^{(j)}$ and using the solution in Eq. (28b) one finds that to order $\Omega^{-1/2}$ the only nonzero coefficient is

$$\bar{a}_3^{(1)} = -\frac{\sigma^4 \mathcal{D}_1^2}{6\mathcal{J}} + \frac{\sigma^2 \mathcal{D}_2^1}{6\mathcal{J}} + \frac{\mathcal{D}_3^0}{18\mathcal{J}}, \tag{31a}$$

while to order Ω^{-1}, one finds

$$\begin{aligned} \bar{a}_4^{(2)} &= -\frac{\mathcal{D}_4^0}{96\mathcal{J}} - \frac{\sigma^2 \mathcal{D}_3^1}{24\mathcal{J}} - \frac{\sigma^4 \mathcal{D}_2^2}{16\mathcal{J}} - \frac{\sigma^6 \mathcal{D}_1^3}{24\mathcal{J}} - \bar{a}_3^{(1)} \left(\frac{3\mathcal{D}_2^1}{8\mathcal{J}} + \frac{3\sigma^2 \mathcal{D}_1^2}{4\mathcal{J}} \right), \\ \bar{a}_6^{(2)} &= \frac{1}{2} (\bar{a}_3^{(1)})^2. \end{aligned} \tag{31b}$$

The above equations determine the series solution of the CME, Eq. (28a), in stationary conditions to order Ω^{-1}.

5 Results

In this section we will compare the (hybrid) method of moments (MM and MCM approach) and the (conditional) system size expansion (SSE approach) by considering the quality of the resulting distribution approximations. In the former case we use the maximum entropy approach outlined in Sect. 3.2 and the approximate solution obtained from the SSE in the latter case. We base our comparison on two simple but challenging examples. The first one describes the bursty expression of a protein which degrades via an enzyme-catalyzed reaction. The second example describes the expression of a protein activating its own expression resulting in a multimodal protein distribution.

5.1 Example 1: Bursty Protein Production

In order to compare the performance of the two methods, we will employ a simple model of gene expression with enzymatic degradation described in Ref. [24]. The system is described by the following set of biochemical reactions:

$$\varnothing \xrightarrow{k_0} M \xrightarrow{k_1} \varnothing, \quad M \xrightarrow{k_2} M + P,$$
$$P + E \underset{k_4}{\overset{k_3}{\rightleftharpoons}} C \xrightarrow{k_5} E. \tag{32}$$

Recently, such active protein degradation has been recognized to be an important factor in skewing nuclear protein distributions in mammalian cells toward high molecule numbers [43]. In our model we explicitly account for mRNA which is important even when the mRNA half-life is shorter than the one of the corresponding protein [44]. Assuming binding and unbinding of P and E are fast compared to protein degradation, the degradation kinetics can be simplified as in Refs. [43, 45], resulting in a Michaelis–Menten like kinetic rate:

$$P \xrightarrow{f(x_P)} \varnothing, \tag{33}$$

with $f(x_P) = \frac{\Omega\, v_M x_P}{\Omega K_M + x_P}$, where x_P is the number of proteins and $K_M = \frac{k_4+k_5}{k_3}$ is the Michaelis–Menten constant. Less obviously for this reduction to hold in the stochastic case, one also requires $k_4 \gg k_5$ as shown in Refs. [45, 46]. Here rates are set to $k_0 = 8, k_1 = 10, k_2 = 100, v_M = 100, K_M = 20, \Omega = 1$. In particular, the reaction rates involving mRNA are chosen such that proteins are produced in bursts of size $b = k_2/k_1$, i.e., on average 10 protein molecules are synthesized from a single transcript.

5.1.1 Method of Moments and Maximum Entropy Reconstruction

We compute an approximation of the moments of species M and P up to order 4 ($K + 1 = 4$) and 6 ($K + 1 = 6$) using the MM as explained in Sect. 3. The moments of P are then used to reconstruct the marginal probability distribution of P. For instance, given the moments $E\left(X_P^0\right), E\left(X_P^1\right), E\left(X_P^2\right), E\left(X_P^3\right), E\left(X_P^4\right)$ we approximate the distribution of P with $\tilde{q}(x) \approx \Pi(X_P = x)$ (see Sect. 3.2). Due to the high sensitivity of the maximum entropy method even to small approximation errors of the moments of highest order considered (in this case the approximation of $E\left(X_P^4\right)$ and $E\left(X_P^6\right)$, respectively), we use moments only up to $E\left(X_P^3\right)$ for the reconstruction. The same holds for all results presented below. In Fig. 1 we plot the two reconstructed distributions.

While for most parts of the support (including the tail) the distribution is accurately reconstructed even with $K = 3$, the method is less precise when considering the probability of small copy numbers of P. To improve the reconstruction, we may use more moments, for instance $K = 7$. However, in this case the moment equations become so stiff that the numerical integration fails completely. This happens due to a combination of highly nonlinear derivatives of the rate function $f(x_P)$ with large values of the higher order moments.

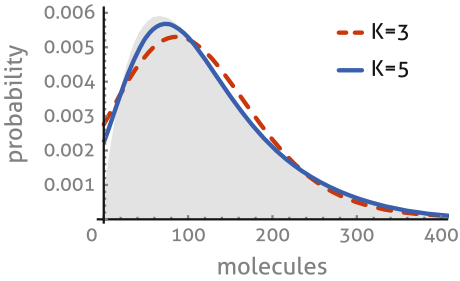

Fig. 1 Bursty protein production: reconstruction based on MM. The reconstructed distribution (cf. Eq. (3.2)) is compared to the distribution estimated with stochastic simulations (*shaded gray area*), where we use $K = 3$ (*dashed red*) and $K = 5$ (*solid blue line*) for the moments in the maximum entropy method. We find that taking into account moments of higher order increases the accuracy significantly, in particular in the region of 0–200 proteins

5.1.2 Solution Using the System Size Expansion

The approximate solution to the distribution functions using the system size expansion as outlined in Sect. 4 is available for networks of a single species only. We therefore concentrate on the case where the mRNA dynamics is much faster than the one of the protein, called the burst approximation [44]. It can be shown that the reaction scheme (32) follows the reduced CME

$$\frac{\mathrm{d}}{\mathrm{d}t}\Pi(x) = \Omega h_0 \sum_{z=0}^{\infty}(E_x^{-z} - 1)\varphi(z)\Pi(x) + \Omega(E_x^{+1} - 1)\frac{v_M x}{\Omega K_M + x}\Pi(x), \quad (34)$$

where E_x^{-z} is the step operator defined by $E_x^{-z} f(x) = f(x - z)$ for any function f. Note that protein synthesis occurs in random bursts z following a geometric distribution $\varphi(z) = \frac{1}{1+b}\left(\frac{b}{1+b}\right)^z$ with average b. The relation between the parameters in scheme (32) and Eq. (34) are $h_0 = k_0 k_2/k_1$ and $b = k_2/k_1$. Within this description protein synthesis involves many reactions: one for each value of z with probability $\varphi(z)$. The coefficients in the expansion of the CME follow from Eq. (21), and are given by

$$\mathcal{D}_n^m = \delta_{m,0} h_0 \langle z^n \rangle_\varphi + (-1)^n v_M \frac{\partial^m}{\partial [P]^m} \frac{[P]}{K_M + [P]}, \quad (35)$$

and $\mathcal{D}_{n,s}^m = 0$ for $s > 0$, where $[P]$ denotes the protein concentration according to the rate equation solution and $\langle z^n \rangle_\varphi = \sum_{z=0}^{\infty} z^n \varphi(z) = \frac{1}{1+b}\,\mathrm{Li}_{-n}(\frac{b}{1+b})$ denotes the average over the geometric distribution in terms of the polylogarithm function $\mathrm{Li}_{-n}(x) = \sum_{k=1}^{\infty} k^n x^k$. The rate Eq. (16) can be solved together with the linear noise approximation, Eq. (25), in steady state conditions. For $v_M < b h_0$ the solution is

$$[P] = K_M \left(1 - \frac{v_M}{bh_0}\right)^{-1}, \quad \sigma^2 = K_M(b+1)\varsigma(\varsigma+1), \tag{36a}$$

where we have defined by $\varsigma = \frac{[P]}{K_M}$ the reduced substrate concentration. In Fig. 2a we show that the expansion performed about the rate equation solution leads to large undulations, we therefore focus on the expansion about the mean. To this end, we have to take into account higher order corrections to the first two moments, we find $\langle\varepsilon\rangle = \Omega^{-1/2}(1+b)\varsigma$ and $\bar{\sigma}^2 = \sigma^2 + \Omega^{-1}(b+1)\varsigma(b(\varsigma+2)+\varsigma+1)$. The nonzero coefficients to order Ω^{-1} given by Eq. (31) then evaluate to

$$\begin{aligned}
\bar{a}_3^{(1)} &= \frac{\sigma^2}{6}(2b(\varsigma+1)+2\varsigma+1),\\
\bar{a}_4^{(2)} &= \frac{\sigma^2}{4}\left[(b+1)^2\varsigma^2 + (b+1)(2b+1)\varsigma + \frac{1}{6}(6b(b+1)+1)\right],\\
\bar{a}_6^{(2)} &= \frac{1}{2}\left[\bar{a}_3^{(1)}\right]^2.
\end{aligned} \tag{36b}$$

The coefficients to order $\Omega^{-3/2}$ can be obtained from Eq. (28) and read

$$\begin{aligned}
\bar{a}_3^{(3)} &= \frac{1}{6}(b+1)\varsigma\left[2(b+1)^2\varsigma^2 + 3b(2b+3)\varsigma + 6b(b+1) + 3\varsigma + 1\right],\\
\bar{a}_5^{(3)} &= \frac{\bar{a}_3^{(1)}}{20}\left[1 + 12b(b+1) + 12(b+1)^2\varsigma^2 + 12(b+1)(2b+1)\varsigma\right],\\
\bar{a}_7^{(3)} &= \bar{a}_3^{(1)}\bar{a}_4^{(2)}, \quad \bar{a}_9^{(3)} = \frac{1}{6}\left[\bar{a}_3^{(1)}\right]^3.
\end{aligned} \tag{36c}$$

The analytical form of these coefficients represents a particularly simple way of solving the CME. The approximation resulting from using these in Eq. (28a) is shown in Fig. 2b (dot-dashed blue line). This approximation captures much better the true distribution obtained from exact stochastic simulation using the SSA (shaded gray area) than the linear noise approximation (dashed red line). We found that including higher order terms in Eq. (28) helped to improve the agreement. However, the resulting expressions turn out to be more elaborate and are hence omitted. A quantitative assessment of this agreement will be given in Sect. 5.3.

5.2 *Example 2: Cooperative Self-Activation of Gene Expression*

As a second application of our methods we consider the regulatory dynamics of a single gene inducing its own, leaky expression. We therefore consider the case where gene activation occurs by binding of its own protein P to two independent sites

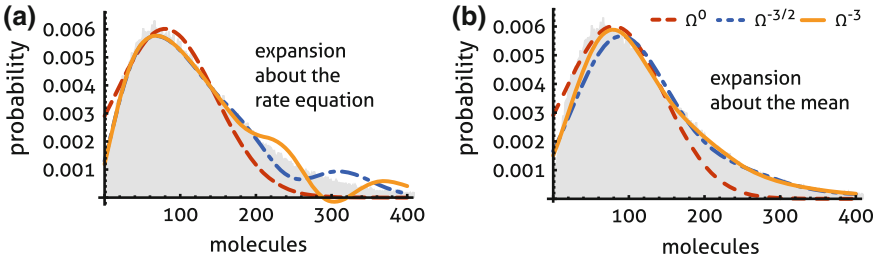

Fig. 2 Bursty protein production: system size expansion solution. **a** We compare the solution obtained using the system size expansion about the rate equation solution, Eq. (26a), that is truncated after Ω^0 (*dashed red*), $\Omega^{-3/2}$ (*dot-dashed blue*) and Ω^{-3}-terms (*solid yellow line*) to stochastic simulations (*shaded gray area*). We observe that the series yields large undulations and negative probabilities. **b** The system size expansion about the mean is shown when the series in Eq. (28a) truncated after Ω^0 (*dashed red*), $\Omega^{-3/2}$ (*dot-dashed blue*) and Ω^{-3}-terms (*solid yellow line*). We find that these approximations avoid undulations and converge rapidly with increasing truncation order to the distributions obtained from simulations using the SSA

$$\begin{aligned} G + P &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} G^*, \\ G^* + P &\underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} G^{**}, \end{aligned} \tag{37}$$

which is modeled explicitly using mass action kinetics. In effect, there are three gene states with increasing transcriptional activity, G, G^* and G^{**}, corresponding to zero, one or two activators bound and leading to a cooperative form of activation. Translation of a transcript denoted by M therefore must occur via one of the following reactions

$$\begin{aligned} G &\xrightarrow{k_G} G + M, \\ G^* &\xrightarrow{k_{G^*}} G^* + M, \\ G^{**} &\xrightarrow{k_{G^{**}}} G^{**} + M, \end{aligned} \tag{38}$$

where k_G denotes the basal transcription rate, k_{G^*} the transcription rate of the semi-induced state G^*, and $k_{G^{**}}$ the rate of the fully induced gene. Finally, by the standard model of translation and neglecting active degradation we have

$$\begin{aligned} M &\xrightarrow{k_3} M + P, \\ M &\xrightarrow{k_4} \varnothing, \quad P \xrightarrow{k_5} \varnothing, \end{aligned} \tag{39}$$

In the following two parameter sets listed in Table 1 leading to moderate and low protein numbers are considered. As we shall see the protein distributions are multi-modal in both cases representing ideal test cases for the distribution reconstruction using conditional moment closures and the conditional system size expansion.

Table 1 The two parameter sets used to study the multimodal expression of a self-activating gene described by the reactions (37)–(39). Set (A) leads moderate protein levels while set (B) yields low protein levels. Note that we have set $\Omega = 1$

Parameter	k_1	k_{-1}	k_2	k_{-2}	k_G	k_{G^*}	$k_{G^{**}}$	k_3	k_4	k_5
Set (A)	5×10^{-4}	3×10^{-3}	5×10^{-4}	2.5×10^{-2}	4	12	24	1200	300	1
Set (B)	5×10^{-4}	2×10^{-4}	5×10^{-4}	2×10^{-3}	4	60	160	30	300	1

5.2.1 Method of Conditional Moments and Maximum Entropy Reconstruction

We compare the distribution reconstruction using an approximation of the first 3, 5, and 7 moments of all the species obtained by the MCM (see Sect. 3). As for the previous case study, the values of the moments of P are used to reconstruct the corresponding marginal distribution. However, here we use the conditional moments $E\left(X^k|G=1\right)$, $E\left(X^k|G^*=1\right)$, $E\left(X^k|G^{**}=1\right)$ instead. We construct the function $\tilde{q}(x)$ that approximates the distribution of P by applying the law of total probability

$$\tilde{q}(x) = \sum_{S\in\{G,G^*,G^{**}\}} P\,(S=1)\,\tilde{q}(x|S=1). \tag{40}$$

The results of the approximation are plotted in Fig. 3. As we can see, the multimodality of the distribution is captured pretty well, and the quality of the reconstructed distribution is quite good in particular when using 7 moments.

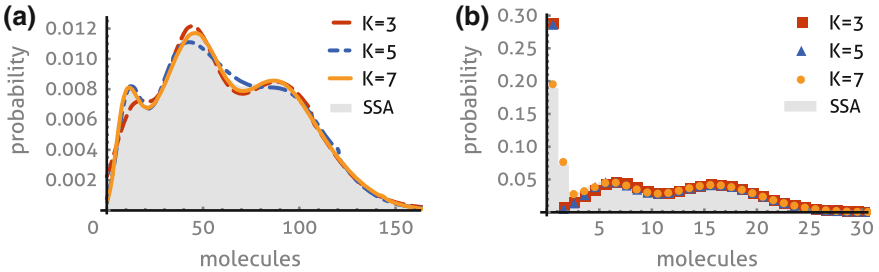

Fig. 3 Self-activating gene: reconstruction based on MCM. The reconstructed distribution (cf. Eq. (40)) for the case of moderate (**a**) and slow (**b**) protein production is compared to the distribution estimated with stochastic simulations (*shaded gray area*), where we use $K = 3$, $K = 5$, and $K = 7$ for the conditional moments in the maximum entropy method. Again we find a significant improvement if moments of higher order are taken into account, in particular in those regions where the shape of the distribution is complex such as the region for 0–5 proteins in case of (**b**)

5.2.2 Conditional System Size Expansion

An alternative technique to approximate distributions for gene regulatory networks with multiple gene states has been given by Thomas et al. [22]. The method makes use of timescale separation by grouping reactions into reactions that change the gene state and reactions that affect only the protein distributions. Based on a conditional variant of the linear noise approximation it was shown that when the gene transitions are slow, protein distributions are well approximated by Gaussian mixture distributions. Implicit in this approach was, of course, that the protein numbers are sufficiently large to justify an application of an linear noise approximation. We will here extend this framework considering higher order terms in the system size expansion to account for low number of protein molecules.

To this end, we describe by the vector $\mathbf{G}$ one of the three gene states $G = (1, 0, 0)$, $G^* = (0, 1, 0)$, and $G^{**} = (0, 0, 1)$ and by x the number of proteins. We will assume that gene transitions between these states evolve slowly on a timescale $1/\mu$. Rescaling time via $\tau = t/\mu$, the CME on the slow timescale reads

$$\frac{\mathrm{d}\Pi(\mathbf{G}, x, \tau)}{\mathrm{d}\tau} = \mu \mathcal{W}_0(\mathbf{G})\Pi(\mathbf{G}, x, \tau) + \mathcal{W}_1\Pi(\mathbf{G}, x, \tau), \tag{41}$$

where $\mathcal{W}_0(\mathbf{G})$ describes the reactions (38) and (39) in the burst approximation

$$\mathcal{W}_0(\mathbf{G}) = \sum_{z=0}^{\infty}(E_x^{-z} - 1)\,\mathbf{k}\cdot\mathbf{G}\,\varphi(z) + (E_x^{+1} - 1)k_4 x, \tag{42}$$

where $\mathbf{k} = (k_G, k_{G^*}, k_{G^{**}})$ and $\mathcal{W}_1$ denotes the transition matrix of the slow gene binding kinetics given by the reactions (37). Note that as before $\varphi(z)$ is the geometric distribution with mean $b = k_3/k_4$. Using the definition of conditional probability, we can write $\Pi(\mathbf{G}, x, \tau) = \Pi(x\,|\mathbf{G}, \tau)\Pi(\mathbf{G}, \tau)$ which transforms Eq. (41) into

$$\begin{aligned}&\frac{\mathrm{d}\Pi(x|\mathbf{G}, \tau)}{\mathrm{d}\tau}\Pi(\mathbf{G}, \tau) + \Pi(x|\mathbf{G}, \tau)\frac{\mathrm{d}\Pi(\mathbf{G}, \tau)}{\mathrm{d}\tau}\\ &\quad= \mu\,\Pi(\mathbf{G}, \tau)\,\mathcal{W}_0(\mathbf{G})\Pi(x|\mathbf{G}, \tau) + \mathcal{W}_1\Pi(x|\mathbf{G}, \tau)\Pi(\mathbf{G}, \tau).\end{aligned} \tag{43}$$

Marginalizing the above equation we find

$$\frac{\mathrm{d}\Pi(\mathbf{G}, \tau)}{\mathrm{d}\tau} = \left(\sum_{x=0}^{\infty}\mathcal{W}_1\,\Pi(x|\mathbf{G})\right)\Pi(\mathbf{G}, \tau), \tag{44}$$

where the term in brackets is a conditional average of the slow dynamics over the protein concentrations. In steady state conditions the above is equal to the equations $0 = \Pi(G)[P|G] - \Pi(G^*)K_1$, $0 = \Pi(G^*)[P|G^*] - K_2\Pi(G^{**})$ and $\Pi(G^{**}) = 1 - \Pi(G) - \Pi(G^*)$ by conservation of probability. Here $[P|\mathbf{G}]$ denotes the conditional protein concentration that remains to be obtained from $\Pi\,(x|\mathbf{G})$. The steady

state solution can be found analytically

$$\begin{aligned}\Pi(G) &= \frac{K_1 K_2}{K_1 K_2 + K_2[P|G] + [P|G][P|G^*]},\\ \Pi(G^*) &= \frac{K_2[P|G]}{K_1 K_2 + K_2[P|G] + [P|G][P|G^*]},\\ \Pi(G^{**}) &= \frac{[P|G][P|G^*]}{K_1 K_2 + K_2[P|G] + [P|G][P|G^*]},\end{aligned} \tag{45}$$

where $K_1 = \frac{k_{-1}}{k_1}$ and $K_2 = \frac{k_{-2}}{k_2}$ are the respective association constants of the DNA-protein binding. The protein distribution is then given by a weighted sum of the probability that a product is found given a particular gene state, times the probability of the gene being in that state:

$$\Pi(x) = \sum_{\mathbf{G}} \Pi(x\,|\mathbf{G})\Pi(\mathbf{G}). \tag{46}$$

It is however difficult to obtain $\Pi(x\,|\mathbf{G})$ analytically, we will therefore employ the limit of slow gene transitions ($\mu \to \infty$) in Eq. (43) to obtain

$$\mathcal{W}_0(\mathbf{G})\,\Pi_\infty\,(x|\mathbf{G}) = 0, \tag{47}$$

where $\Pi_\infty\,(x|\mathbf{G}) = \lim_{\mu\to\infty} \Pi\,(x|\mathbf{G}, \tau)$. We can now perform the system size expansion for the conditional distribution $\Pi_\infty(x\,|\mathbf{G})$ that is determined by Eq. (47) using the ansatz

$$\frac{x}{\Omega}\Big|\,\mathbf{G} = [P|\mathbf{G}] + \Omega^{-1/2}\varepsilon|\mathbf{G}, \tag{48}$$

for the conditional random variables describing the protein fluctuations. The coefficients in the expansion of the CME (47) are

$$\mathcal{D}_n^m = \delta_{m,0}\left(\mathbf{k}'\cdot\mathbf{G}\,\langle z^n\rangle_\varphi + (-1)^n k_1[P|\mathbf{G}]\right) + \delta_{m,1}(-1)^n k_1, \tag{49}$$

with $\mathbf{k}' = \mathbf{k}/\Omega$ and $\mathcal{D}_{n,s}^m = 0$ for $s > 0$. The solution of the rate equation and the conditional linear noise approximation are given by

$$[P|\mathbf{G}] = \mathbf{k}'\cdot\mathbf{G}\,\frac{b}{k_5},\quad \sigma^2_{P|\mathbf{G}} = [P|\mathbf{G}](1+b), \tag{50}$$

respectively. Note that there are no further corrections in Ω to this conditional linear noise approximation because the conditional CME (47) depends only linearly on the protein variables. The conditional distribution can now be obtained using the result (26a). Associating with $\pi_0(\varepsilon\,|\mathbf{G})$ a centered Gaussian of variance as given in Eq. (50), we find to order Ω^{-1},

$$\Pi_\infty(\varepsilon|\mathbf{G}) = \pi_0(\varepsilon|\mathbf{G})\Big[1 + \Omega^{-1/2}a_3^{(1)}(\mathbf{G})\psi_{3,\mathbf{G}}(\varepsilon) + \Omega^{-1}a_4^{(2)}(\mathbf{G})\psi_{4,\mathbf{G}}(\varepsilon) + \Omega^{-1}a_6^{(2)}(\mathbf{G})\psi_{6,\mathbf{G}}(\varepsilon)\Big] + O(\Omega^{-3/2}). \tag{51a}$$

By the definition given before Eq. (26a) the polynomials $\psi_{m,\mathbf{G}}(\varepsilon)$ depend on the gene state via the conditional variance Eq. (50). The coefficients obtained from Eq. (31) take the particularly simple form

$$a_3^{(1)}(\mathbf{G}) = \frac{1}{6}\left(2b^2 + 3b + 1\right)[P|\mathbf{G}], \tag{51b}$$

$$a_4^{(2)}(\mathbf{G}) = \frac{1}{24}(b+1)\left(6b^2 + 6b + 1\right)[P|\mathbf{G}], \qquad a_6^{(2)}(\mathbf{G}) = \frac{1}{2}\left[a_3^{(1)}(\mathbf{G})\right]^2. \tag{51c}$$

Finally, we remark that $\Pi(x\,|\mathbf{G})$ and $\Pi(\varepsilon\,|\mathbf{G})$ are related by $\Pi(x\,|\mathbf{G}) = \Omega^{-1/2}\Pi(\varepsilon = \Omega^{-1/2}x - \Omega^{1/2}[P|\mathbf{G}]\,|\mathbf{G})$.

In Fig. 4a, this conditional system size expansion is compared to stochastic simulations using the SSA. We find that the conditional linear noise approximation, Eq. (51), truncated after Ω^0 captures well the multimodal character of the distributions but misses to predict the precise location of its modes. In contrast, the conditional approximation of Eq. (51) taking into account up to Ω^{-1} terms accurately describes the location of these distribution maxima. We note however that a continuous approximation such as Eq. (51a) may fail in situations when the effective support of the conditional distributions represents the range of a few molecules. Such case is depicted in Fig. 4b. In this case the distributions are captured better by an approximation with discrete support as has been given in Ref. [20], Eqs. (35) and (36) therein. The resulting approximation, (blue filled circles) using the analytical form of the coefficients (51) is in excellent agreement with stochastic simulations performed using the SSA (shaded gray area). These findings highlight clearly the need to go beyond the conditional Gaussian approximation for the two cases studied here.

5.3 *Comparison of Numerical Results*

In order to compare both methods quantitatively, we calculated absolute $|\Pi(x) - \Pi^*(x)|$ and relative errors $|\Pi(x) - \Pi^*(x)|/\Pi(x)$ between the exact distribution $\Pi(x)$, which was estimated from simulations via the SSA, and the distribution approximation obtained either via the MM or via the SSE denoted by $\Pi^*(x)$. The results for the first case study are shown in Fig. 5. We observe that the maximum absolute and relative errors occur close to the boundary of zero molecules for both approximation methods. However, the SSE is more accurate than MM here. In this

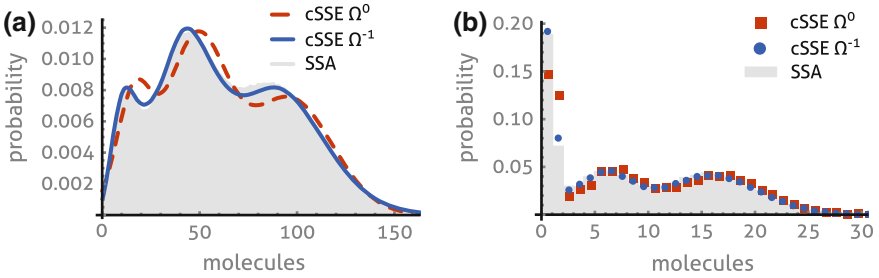

Fig. 4 Self-activating gene: conditional system size expansion. The conditional system size expansion (cSSE), Eqs. (45), (46) and (51), for the moderate (**a**) and slow (**b**) protein production is compared to stochastic simulations (SSA). While the conditional linear noise approximation (Ω^0) captures the multimodal character of the distribution, it misses the precise location of its modes. We find that the Ω^{-1}-estimate of the cSSE given by Eq. (51) agrees much better with the SSA. **b** The discrete approximation of Ref. [20], see main text for details, is shown when truncated after Ω^0 and Ω^{-1} terms. The analytical form of the coefficients in Eq. (51) has been used with parameter set B in Table 1

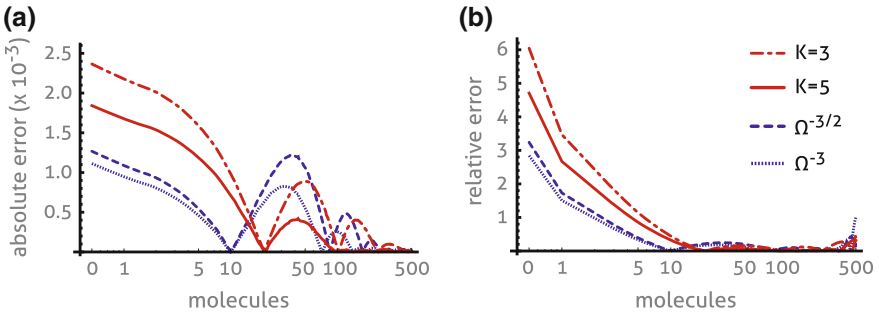

Fig. 5 Bursty protein production: absolute and relative error. We plot the absolute (**a**) and relative (**b**) errors of the MM (for moments up to order $K = 3$ and $K = 5$, Fig. 1) and the SSE (truncated after $\Omega^{-3/2}$ and Ω^{-3} terms, Fig. 2). The SSE for this example yields the percentage error $||\varepsilon||_V$ to be 5.1% ($\Omega^{-3/2}$) and 2.8% (Ω^{-3}) while the moment based approach yields 5.6% ($K = 3$) and 2.0% ($K = 5$). Both approaches become more accurate as more moments or higher order terms in the SSE are taken into account. For both methods, the maximum errors occurs at zero molecules

region a direct representation of the probabilities (as in the hybrid approaches) may be more appropriate. For measuring the overall agreement of the distributions we computed the percentage statistical distance

$$||\varepsilon||_V = \frac{100\%}{2} \sum_{x=0}^{\infty} |\Pi(x) - \Pi^*(x)|. \tag{52}$$

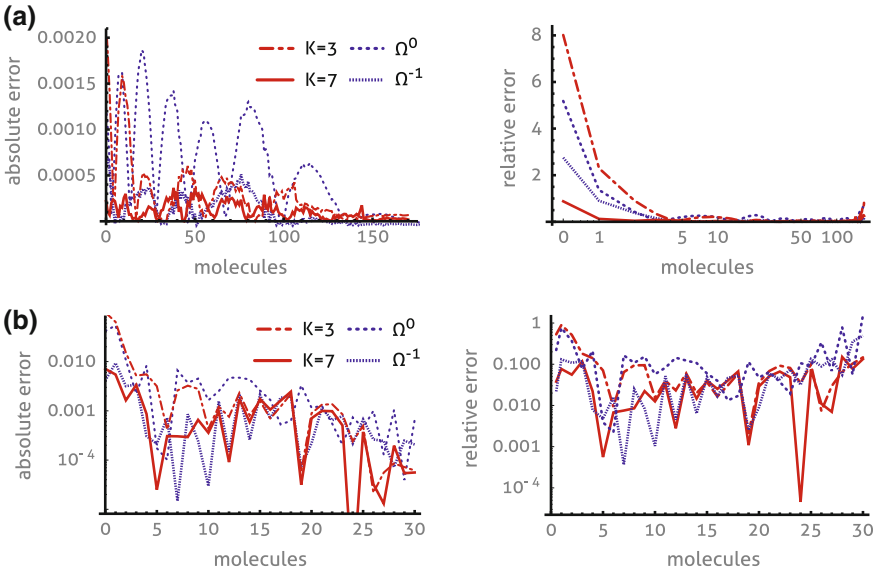

Fig. 6 Self-activating gene: absolute and relative error. We plot the absolute (*left*) and relative (*right*) errors of the MCM (for moments up to order $K = 3$ and $K = 7$) and the cSSE (truncated after Ω^0 and Ω^{-1} terms). For moderate protein production (**a**), corresponding to the distributions shown in Figs. 3a and 4a, the MCM yields a percentage error $||\varepsilon||_V$ of 2.4% ($K = 3$) and 1.0% ($K = 7$) while the cSSE yields 5.4% (Ω^0) and 1.4% (Ω^{-1}), respectively. For slow protein production (**b**), corresponding to the distributions shown in Figs. 3b and 4b, we find $||\varepsilon||_V = 10.9\%$ ($K = 3$) and 1.7% ($K = 7$) for the MCM as well as $||\varepsilon||_V = 8.3\%$ (Ω^0) and 1.9% (Ω^{-1}) for the cSSE, respectively

This distance can also be interpreted as the maximum percentage difference between the probabilities of all possible events assigned by the two distributions [47, 48] and achieves its maximum (100% error) when the distributions do not overlap. The numerical values given in the caption of Fig. 5 reveal that the estimation errors of the MM and the SSE decrease as more moments or higher order terms in the SSE are taken into account. The respective error estimates are of the same order of magnitude. However, the analytical solution obtained using the SSE, given in Sect. 5.1.2 including only low order terms is slightly more accurate than the MM with only few moments, while the MM with a larger number of moments becomes more accurate than the SSE including up to Ω^{-3}-terms.

For the second case study, we find that the absolute and relative estimation errors of the method of moments and the SSE are of the same order of magnitude, compare Fig. 6. However, we found that the method of moments including three conditional moments ($K = 3$) is overall more accurate than the conditional linear noise approximation (Ω^0). The approximations become comparable as higher order conditional moments and higher orders in conditional SSE are employed. However, the method

of moments including 7 conditional moments turned out to be slightly more accurate than the analytical SSE solution to order Ω^{-1} given in Sect. 5.2.2.

6 Discussion

We have here studied the accuracy of two recently proposed approximations for the probability distribution of the CME. The method of moments utilizes moment closure to approximate the first few moments of the CME from which the distribution is reconstructed via the maximum entropy principle. In contrast, the SSE method does not rely on a moment approximation but instead the probability distribution is obtained analytically as a series in a large parameter that corresponds roughly to the number of molecules involved in the reactions. Interestingly, our comparative study revealed that both methods yield comparable results. While generally both methods provide highly accurate approximations for the distribution tails and capture well the overall shape of the distributions, we found that for both methods the largest errors occur close to the boundary of zero molecules. We observed a similar behavior when conditional moments or the conditional SSE were used. These discrepancies could be resolved by taking into account higher order moment closure schemes or, equivalently, by taking into account higher order terms in the expansion of the probability distribution.

From a computational point of view, the method based on moment closure is limited by the number of moments that can be numerically integrated due to the stiffness of the resulting high-dimensional ODE system. Our investigation showed that such difficulties are encountered when one closes the moment equations beyond the 8^{th} moment. In contrast, the analytical solution provided by the SSE technique does not suffer from these issues because it is provided in closed form. We note however that the SSE solution given here is limited to a single species while the method of moments has no such limitation. Moreover, the conditional SSE solution relies on timescale separation which the method of conditional moments does not assume.

The computational cost of the analytical approximation provided by the SSE is generally less than that of the moment based approach because it avoids numerical integration and optimization. This fact may be particularly advantageous when wide ranges of parameters are to be studied, as for instance in parameter estimation from experimental data. We note, however, that the moment-based approach is still much faster than the one of the SSA because it avoids large amounts of ensemble averaging. Therefore the moment based approach may be preferable in situations where the SSE cannot be applied as we have mentioned above. We hence conclude that both approximation schemes provide complementary strategies for the analysis of stochastic behavior of biological systems. Most importantly, they both provide computationally more efficient strategies compared with simulation and numerical integration of the CME while preserving an high degree of accuracy. The methods thus have an high potential for the analysis of large-scale models.

Acknowledgements PT acknowledges support from the Royal Commission for the Exhibition of 1851 in form of a Research Fellowship.

References

1. H. H. McAdams and A. Arkin.: "Stochastic mechanisms in gene expression". Proc Natl Acad Sci 94.3 (1997), pp. 814–819.
2. P. S. Swain, M. B. Elowitz, and E. D. Siggia.: "Intrinsic and extrinsic contributions to stochasticity in gene expression". Proc Natl Acad Sci 99.20 (2002), pp. 12795–12800.
3. T. J. Perkins and P. S. Swain.: "Strategies for cellular decision-making". Mol Syst Biol 5 (2009).
4. M. B. Elowitz et al.: "Stochastic gene expression in a single cell". Sci Signal 297.5584 (2002), p. 1183.
5. D. Wilkinson.: Stochastic Modelling for Systems Biology. Chapman & Hall, 2006.
6. D. T. Gillespie.: "Exact stochastic simulation of coupled chemical reactions". J Phys Chem 81.25 (1977), pp. 2340–2361.
7. N. Maheshri and E. K. O'Shea.: "Living with noisy genes: how cells function reliably with inherent variability in gene expression". Annu Rev Biophys Biomol Struct 36 (2007), pp. 413–434.
8. B. Munsky and M. Khammash.: "The finite state projection algorithm for the solution of the chemical master equation". J Chem Phys 124.4 (2006), p. 044104.
9. M. Mateescu et al.: "Fast Adaptive Uniformisation of the Chemical Master Equation". IET Syst Biol 4.6 (2010), pp. 441–452.
10. D. T. Gillespie.: "Stochastic simulation of chemical kinetics". Annu Rev Phys Chem 58 (2007), pp. 35–55.
11. J. Elf and M. Ehrenberg.: "Fast evaluation of fluctuations in biochemical networks with the linear noise approximation". Genome Res 13.11 (2003), pp. 2475–2484.
12. R Grima. "An effective rate equation approach to reaction kinetics in small volumes: Theory and application to biochemical reactions in nonequilibrium steady-state conditions". J Chem Phys 133.3 (2010), p. 035101.
13. P. Thomas, H. Matuschek, and R. Grima.: "How reliable is the linear noise approximation of gene regulatory networks?" BMC Genomics 14. Suppl 4 (2013), S5.
14. S. Engblom.: "Computing the moments of high dimensional solutions of the master equation". Appl Math Comput 180 (2 2006), pp. 498 -515.
15. C. Gillespie. "Moment-closure approximations for mass-action models". IET Syst Biol 3.1 (2009), pp. 52–58.
16. A. Ale, P. Kirk, and M. P. H. Stumpf.: "A general moment expansion method for stochastic kinetic models". J Chem Phys 138.17 (2013), p. 174101.
17. A. Andreychenko, L. Mikeev, and V.Wolf.: "Model Reconstruction for Moment-Based Stochastic Chemical Kinetics". ACM Trans Model Comput Simul 25.2 (2015), 12:1–12:19.
18. A. Andreychenko, L. Mikeev, and V. Wolf.: "Reconstruction of Multimodal Distributions for Hybrid Moment-based Chemical Kinetics". To appear in Journal of Coupled Systems and Multiscale Dynamics (2015).
19. N. G. Van Kampen.: Stochastic Processes in Physics and Chemistry. Third. Amsterdam: Elsevier, Amsterdam, 1997.
20. P. Thomas and R. Grima. : "Approximate probability distributions of the master equation". Phys Rev E 92.1 (2015), p. 012120.
21. L. Bortolussi.: "Hybrid Behaviour of Markov Population Models". Information and Computation (2015 (accepted)).
22. P. Thomas, N. Popović, and R. Grima.: "Phenotypic switching in gene regulatory networks". Proc Natl Acad Sci 111.19 (2014), pp. 6994–6999.

23. D. T. Gillespie: “A diffusional bimolecular propensity function”. J Chem Phys 131.16 (2009), p. 164109.
24. P. Thomas, H. Matuschek, and R. Grima.: “Computation of biochemical pathway fluctuations beyond the linear noise approximation using iNA”. Bioinformatics and Biomedicine (BIBM), 2012 IEEE International Conference on. IEEE. 2012, pp. 1–5.
25. P. Whittle.: “On the Use of the Normal Approximation in the Treatment of Stochastic Processes”. J R Stat Soc Series B Stat Methodol 19.2 (1957), pp. 268–281.
26. J. H. Matis and T. R. Kiffe.: “On interacting bee/mite populations: a stochastic model with analysis using cumulant truncation”. Environ Ecol Stat 9.3 (2002), pp. 237–258.
27. I. Krishnarajah et al.: “Novel moment closure approximations in stochastic epidemics”. Bull Math Biol 67.4 (2005), pp. 855–873.
28. A. Singh and J. P. Hespanha.: “Lognormal moment closures for biochemical reactions”. Decision and Control, 2006 45th IEEE Conference on. IEEE. 2006, pp. 2063–2068.
29. A. Singh and J. P. Hespanha.: “Approximate moment dynamics for chemically reacting systems”. Automatic Control, IEEE Transactions on 56.2 (2011), pp. 414–418.
30. D. Schnoerr, G. Sanguinetti, and R. Grima.: “Comparison of different momentclosure approximations for stochastic chemical kinetics”. J Chem Phys 143.18 (2015), p. 185101.
31. D. Schnoerr, G. Sanguinetti, and R. Grima.: “Validity conditions for moment closure approximations in stochastic chemical kinetics”. J Chem Phys 141.8 (2014), p. 084103.
32. R. Grima.: “A study of the accuracy of moment-closure approximations for stochastic chemical kinetics”. J Chem Phys 136.15 (2012), p. 154105.
33. J. Hasenauer et al.: “Method of conditional moments for the Chemical Master Equation”. J Math Biol (2013), pp. 1–49.
34. M. Lapin, L. Mikeev, and V. Wolf.: “SHAVE - Stochastic Hybrid Analysis of Markov Population Models”. Proceedings of the 14th International Conference on Hybrid Systems: Computation and Control (HSCC’11). ACM International Conference Proceeding Series. 2011.
35. A.L. Berger, V.J.D. Pietra, S.A.D. Pietra, A Maximum Entropy Approach to Natural Language Processing. Comput Ling **22**(1), 39–71 (1996)
36. R. Abramov.: “The multidimensional maximum entropy moment problem: a review of numerical methods”. Commun Math Sci 8.2 (2010), pp. 377–392.
37. Z. Wu et al.: “A fast Newton algorithm for entropy maximization in phase determination”. SIAM Rev 43.4 (2001), pp. 623–642.
38. L. R. Mead and N. Papanicolaou: “Maximum entropy in the problem of moments”. J Math Phys 25 (1984), p. 2404.
39. G. W. Alldredge et al.: “Adaptive change of basis in entropy-based moment closures for linear kinetic equations”. J Comput Phys 258 (2014), pp. 489–508.
40. Á . Tari, M. Telek, and P. Buchholz.: “A unified approach to the moments based distribution estimation-unbounded support”. Formal Techniques for Computer Systems and Business Processes. Springer, 2005, pp. 79–93.
41. J. Elf et al.: “Mesoscopic kinetics and its applications in protein synthesis”. Systems Biology. Springer, 2005, pp. 95–18.
42. L. Comtet.: Advanced Combinatorics: The art of finite and infinite expansions. Springer Science & Business Media, 1974.
43. E. Giampieri et al.: “Active Degradation Explains the Distribution of Nuclear Proteins during Cellular Senescence”. PloS one 10.6 (2015), e0118442.
44. V. Shahrezaei and P. S. Swain.: “Analytical distributions for stochastic gene expression”. Proc Natl Acad Sci 105.45 (2008), pp. 17256–17261.
45. P. Thomas, A. V. Straube, and R. Grima.: “Communication: Limitations of the stochastic quasi-steady-state approximation in open biochemical reaction networks”. J Chem Phys **135**(18), 181103 (2011)
46. K. R. Sanft, D. T. Gillespie, and L. R. Petzold.: “Legitimacy of the stochastic Michaelis-Menten approximation”. Syst Biol, IET 5.1 (2011), pp. 58–69.
47. D. A. Levin, Y. Peres, and E. L. Wilmer.: Markov chains and mixing times. American Mathematical Soc., 2009.
48. T. M. Cover and J. A. Thomas.: Elements of information theory. John Wiley & Sons, 2012.

http://www.springer.com/978-3-319-45831-1

Modeling Cellular Systems

Graw, F.; Matthäus, F.; Pahle, J. (Eds.)

2017, XI, 161 p. 35 illus., 29 illus. in color., Hardcover
ISBN: 978-3-319-45831-1