The stochastic dynamics of biochemical reaction networks can be accurately described by discrete-state Markov processes where each chemical reaction corresponds to a state transition of the process. Due to the largeness problem of the state space, analysis techniques based on an exploration of the state space are often not feasible and the integration of the moments of the underlying probability distribution has become a very popular alternative. In this paper the focus is on a comparison of reconstructed distributions from their moments obtained by two different moment-based analysis methods, the method of moments (MM) and the method of conditional moments (MCM). We use the maximum entropy principle to derive a distribution that fits best to a given sequence of (conditional) moments. For the two gene regulatory networks that we consider we find that the MCM approach is more suitable to describe multimodal distributions and that the reconstruction is more accurate if conditional distributions are considered.

Keywords: Chemical Master Equation, Moment Closure, Method of Conditional Moments, Maximum Entropy.

PACS: 02.50.Ga, 87.18.Vf, 02.50.Cw

MSC: 60J22, 44A60, 37N25

1. INTRODUCTION

Discrete-state Markov processes are frequently used to describe the dynamics of reactions occurring in a single cell. The main advantage of such models over continuous deterministic approaches is that they are able to capture the inherent discreteness and randomness of molecular interactions. The states of the considered Markov models are population vectors that count the number of molecules of different types and chemical reactions are modeled as transitions between population vectors. The evolution of the underlying probability distribution is governed by the Chemical Master Equation (CME) which gives the change of probability over time for each state of the process. Due to the combinatorial nature of the population vectors, the number of states that have to be considered during the integration of the CME can be enormous which makes a direct numerical solution of the CME infeasible. Therefore, analysis methods based on the integration of the moments of the underlying probability distribution have become very popular. Instead of integrating one equation for each population vector only equations for the expectations, variances, covariances and possibly higher moments of the joint distribution are integrated. Thus, the complexity of the analysis method is no longer dependent on the molecular counts but only on the number of different chemical species. The most widely used approach is a moment closure whose moment equations are derived from the CME but truncated after a Taylor expansion about the expected population numbers \[3, 8\]. In the sequel we refer to it as the method of moments (MM). Recently, a more general moment-based approach was proposed that allows to keep for some of the chemical populations the original probability-based representation. This makes sense for those populations that are very small. For instance, a population may represent the binding of a transcription factor to a promoter region where only the values 0 (promoter free) and 1 (transcription factor bound to promoter) are possible. Similarly, in many systems one has to deal at the same time with very large and very small populations, the latter having, say, at most ten molecules with significant probability. Clearly, for such cases a moment-based description is not appropriate and it is usually better to integrate over time the probabilities of having 0, 1, \ldots, 10 molecules. In Hasenauer et al. \[13\] a hybrid integration scheme for the moments of large populations conditioned on the actual molecule numbers of small populations is derived from the CME. In the sequel, we refer to it as the method of conditional moments (MCM). In \[13\] the MM and MCM approach are compared in terms of accuracy but the comparison is limited to the moment values. Here we want to determine probabilities instead of moments, such as the likelihood of molecule levels measured in vitro (in order to calibrate parameters) or probabilities of certain stochastic events of interest. The focus is on a comparison of the underlying probability distributions that are reconstructed from the moments. In Fig.1 we illustrate the difference between a direct integration of the CME (lower arrow) and a moment-based approach (left, upper and right arrow). Here ME stands for the moment equations that are used to integrate the moments from time \(t_0\) until time \(t\).

In previous work \[4\] we proposed a reconstruction scheme based on the maximum entropy principle stating that from those distributions that fulfil the moment conditions (sequence of moments up to a certain order equals a given sequence of values) we choose the distribution that maximizes the entropy \[4\]. In \[4\] we applied this principle...
to the MM approach and found that, for all examples that we considered, the integration of the moment equations and the reconstruction of the distributions is several orders of magnitude faster than a direct numerical integration of the CME. By increasing the population sizes the speed up can be made arbitrarily large. However, the accuracy of the reconstructed distributions is not always satisfying and simply increasing the order of the considered moments does, from a certain order on, not improve the results, since the optimization procedure often becomes numerically unstable. In particular if the distributions have a complex shape (such as multimodal distributions) the reconstruction is in many parts not very accurate.

Here we propose to reconstruct the conditional probability distributions that result from the MCM approach and multiply with the probability of the condition (i.e. marginal distributions of small populations) to obtain the full distribution. This has several advantages compared to the MM reconstruction. First, in the MM approach the reconstruction of the probability distributions of small populations is usually very imprecise while in the MCM approach these distributions are directly available. Second, the conditional distribution in the MCM approach are often less complex (e.g. unimodal instead of bimodal) and thus easier to reconstruct. Finally, as the maximal order $m$ of considered moments increases, the MCM approach requires less variables (conditional moments and conditional probabilities) than in the MM approach. This is because the number of conditional probabilities is fixed while the number of considered (conditional) moments grows exponentially in $m$. Nevertheless, we find that the information provided by this smaller set of variables is as good or even better for reconstructing the original distributions. We consider two case studies and compare the reconstructed distributions for both, the MM and the MCM approach. We find that for most parameter combinations the reconstruction based on the MCM approach is more accurate. Thus, we argue that the MCM approach should be preferred when deriving the probability of certain events from moment values.

In the next two sections we shortly introduce the chemical master equation (Section 2) and the method of moments (Section 3). Section 4 then describes in detail the reconstruction based on the maximum entropy approach. The numerical results of two case studies are presented in Section 5 and with Section 6 we conclude the paper. A detailed derivation of the moment equations is given in the Appendix, Sections 1 (MM approach) and 2 (MCM approach).

2. THE CHEMICAL MASTER EQUATION

Given $n$ different chemical species $S_1, \ldots, S_n$ and a set of $m$ reactions

$$R_j : \ell^-_{j,1}S_1 + \ldots + \ell^-_{j,n}S_n \xrightarrow{c_j} \ell^+_{j,1}S_1 + \ldots + \ell^+_{j,n}S_n,$$

where $1 \leq j \leq m$, we construct a Markov process $(\tilde{X}(t), t \geq 0)$. The random vector $\tilde{X}(t) = (X_1(t), \ldots, X_n(t))$ is such that the $i$-th entry $X_i(t)$ is the number of molecules of type $i$ at time $t$ for $i \in \{1, \ldots, n\}$. If $\tilde{X}(t) = (x_1, \ldots, x_n) \in \mathbb{N}_0^n$ is the state of the process at time $t$ and $x_i \geq \ell^-_{i,j}$ for all $i$, then the $j$-th reaction corresponds to a possible transition from state $\tilde{x}$ to state $\tilde{x} + \tilde{v}_j$ where $\tilde{v}_j$ is the change vector with entries $(-\ell^-_{j,i} + \ell^+_{j,i}) \in \mathbb{Z}^n$. Assuming that the reaction volume is well-stirred and in thermal equilibrium, it is possible to physically justify that the transition rate is determined as $\alpha_j(x) := c_j \prod_{i=1}^n (x_i / \ell^-_{i,j})$ where $c_j$ depends on the temperature, the volume and the physical properties of the reactant species of the $j$-th reaction [10]. Here we make the usual assumption that temperature and volume are constant over time such that $c_j$ is a constant, often called stochastic reaction rate constant. The time evolution of $\tilde{X}$ is given by the Chemical Master Equation (CME)

$$\frac{dp(\tilde{x},t)}{dt} = \sum_{j=1}^m \left( \alpha_j(\tilde{x} - \tilde{v}_j)p(\tilde{x} - \tilde{v}_j,t) - \alpha_j(\tilde{x})p(\tilde{x},t) \right). $$ (1)

Here, the initial conditions are fixed and for a time point $t \geq 0$ we let $p(\tilde{x},t)$ denote the probability $P(\tilde{X}(t) = \tilde{x})$, where $\tilde{x}$ is an $n$-vector of molecule counts. We remark that the probabilities $p(\tilde{x},t)$ are uniquely determined if we consider the equations of all states that are reachable from the initial set of possible states $\{ \tilde{x} \mid p(\tilde{x},0) > 0 \}$. The main drawback is, however, that for most models the number of reachable states is extremely large or even infinite which renders an efficient numerical integration of Eq. (1) impossible.

Numerical approximations based on a static or dynamic truncation of the state space have been developed which allow for accurate approximations whenever the probability of having a large population is very small for all involved chemical species [17,19]. For our experimental results we make use of such methods in order to compare their results to the reconstructed distributions. Clearly, this is only feasible for systems where all chemical populations remain small. The error introduced by the truncation is then controlled by a small threshold $\delta > 0$ that determines in every integration step whether a state has a significant amount of probability ($p(\tilde{x},t) > \delta$) or not. In the latter case the probability of the state is approximated by zero and over time.
the probability mass that is ‘lost’ accumulates. The amount of probability loss equals the sum of all state-wise absolute approximation errors and therefore we can, by choosing δ very small, compute an accurate approximation of the true distribution. The following example is used as a running example throughout the rest of the paper.

**Example 1.** We consider a simple gene expression model that describes the formation of mRNA (R) and protein (P) molecules \( \mathbb{R} \). The production of R is controlled by the state of the DNA which can be either active \((\text{on})\) or inactive \((\text{off})\). We assume we have a single copy of the corresponding gene, i.e., it always holds that \( \text{on} + \text{off} = 1 \). If \( \text{on} = 1 \) then mRNA molecules are synthesized and can be further translated into proteins. The proteins induce the activation of the DNA forming a positive feedback mechanism. Moreover, mRNA and proteins can degrade. The chemical reactions are as follows:

\[
\begin{align*}
\text{on} \xrightarrow{\tau_{\text{off}}} & \text{off} \\
\text{on} \xrightarrow{k_{\text{off}}} & \text{off} + R \\
R \xrightarrow{k_{\text{p}}} & R + P \\
R \xrightarrow{\gamma_R} & P + \text{off}
\end{align*}
\]

A state of the associated Markov process is a vector \( \vec{x} = (x_{\text{on}}, x_{\text{off}}, x_R, x_P) \) where \( x_{\text{on}} \) and \( x_{\text{off}} \) are the \( \text{on} \) and \( \text{off} \) populations \((x_{\text{on}} + x_{\text{off}} = 1)\) and \( x_R, x_P \in \mathbb{N} \) are the number of mRNA and protein molecules, respectively. The left plot in Fig. 2 shows the two-dimensional marginal distributions of mRNA and proteins at time \( t = 10 \) (darker points correspond to larger probability values) where we chose rate constants \((\gamma_{\text{on}}, \tau_{\text{off}}, k_{\text{off}}, k_{\text{p}}, \gamma_R, \gamma_P, \tau_{\text{p}}) = (1, 1, 10, 1, 4, 1, 0.015)\) and \( x = (1, 0, 0, 0) \) for the initial state as in \( \mathbb{R} \). The other two plots in Fig. 2 show the distributions of mRNA and proteins after conditioning on the two states of the DNA. The distributions were computed by integrating the CME based on a fourth-fifth order Runge-Kutta method as described above. We chose \( \delta = 10^{-15} \) as truncation threshold yielding a total error of \( \epsilon = 3 \cdot 10^{-10} \) at time instant \( t = 10 \).

3. **METHOD OF MOMENTS**

For systems with large population sizes the CME gives an extremely large number of differential equations that must be integrated for a transient solution. Even if we truncate the state space and consider the sets \( \{\vec{x} \mid p(\vec{x}, t) > \delta\} \) for some small \( \delta \) and \( t \geq 0 \), the numerical integration of Eq. (1) is infeasible since the probability mass distributes on a large set of states. In such cases a very efficient way of simulating the system is the integration of the moments of the joint distribution over time. It is, for instance, straightforward to derive a differential equation for the time evolution of the first-order moments \( E[X(t)] \). However, this equation may involve moments of higher order. Thus, one has to derive equations for second or higher order moments as well which may again involve moments of even higher order. In order to get a finite number of differential equations we can, for instance, assume that all centered moments of order higher than \( M \) are zero. In this way we get equations to approximate the first \( M \) moments of the distribution. We discuss this in detail in the Supporting Information, Sect. [3] and refer to this in the following as the MM approach. For simple systems such as the gene expression in Example 1 we find that the approximation provided by the moment closure method is very accurate. In general, however, experimental results show that the approximation tends to become worse if systems exhibit complex behavior such as multistability or oscillations. Increasing the number of moments typically improves the accuracy [3] but sometimes the resulting equations become very stiff [8].

Grima has investigated the accuracy of the approximation for \( M = 2 \) and \( M = 3 \) by a comparison with the system size expansion of the master equation [12]. He found that for monostable systems with large volumes the approximation of the means \( E[X(t)] \) have a relative error that scale as \( \Omega^{-M} \) while the relative errors of the variances and covariances scale as \( \Omega^{-(M-1)}, M \in \{2,3\} \). For small volumes or systems with multiple modes, however, only experimental evaluations of the accuracy are available [3,8], where the approximated moments are compared to statistical estimates based on Monte Carlo simulations of the process.

In many chemical reaction networks we find a joint distribution for which a representation in terms of the moments is not ideal. For instance, many networks describing gene regulatory processes contain binding events where the number of binding sites is very low (often there is just a single binding site). Then the size of the populations of the corresponding chemical species is bounded by the number of binding sites. For instance, in the gene expression example (Example 1) we only have two possibilities - either the DNA is in the on or in the off state. Then the marginal probability distribution consists of a small number of discrete probabilities (e.g. the probability that the DNA is in the on or in the off state) and the joint distribution is typically divided into several modes that correspond to the different binding states. In such cases it is obvious that for the small popula-
tions a moment representation is not adequate compared to considering the discrete mode probabilities. This, however, implies that one has to consider conditional moments (conditioned on the mode) for the remaining (large) populations. In the work of Hasenauer et al. [13] equations have been derived for the integration of the mode probabilities over time (where the mode corresponds to the state of the small populations, e.g. state of binding sites, etc.) and equations to integrate the moments of the large populations, conditioned on the mode. In the Supporting Information, Sect. 2 we shortly describe how to derive these equations (mainly following the lines of Hasenauer et al. [13]). We refer to this as the MCM approach and in the experimental results presented in the sequel, we reconstruct the joint distribution based on both, the MM approach and the MCM approach to see whether the reconstructed joint distributions are more accurate if conditional moments are used.

4. MAXIMUM ENTROPY RECONSTRUCTION

In this section we focus on reconstructing a probability distribution when the moments of the distribution up to a certain order are given. Since the moments may correspond to a set of distributions, we apply the maximum entropy principle to choose a distribution from this set [13]. The idea is to choose among the distributions, that fulfill the moment equations, a distribution that maximizes the entropy. For instance, the exponential distribution maximizes the entropy among all continuous distributions on $[0,\infty)$ with the same mean. Similarly, the normal distribution is chosen among all continuous distributions if mean and variance are known. The maximum entropy principle assumes a minimum amount of prior information and avoids any other latent assumption about the distribution. It is successfully applied in thermodynamics and statistical mechanics [12], climate prediction [2], performance analysis [22] and many other areas.

Here, we use the moments of the $i$-th population up to order $M$ to obtain the one- and two-dimensional marginal probability distributions of a reaction network. The reconstruction of the distributions of higher dimension is more involved and hence advanced numerical techniques must be applied [11].

4.1. Maximum Entropy Approach

In the sequel we simply write $X$ (and $x$) for any random vector (and state) of at most $n$ molecular populations at some fixed time instant $t$. Given a sequence of $M$ non-central moments $E(X^k) = \mu_k, k = 0, 1, \ldots, M,$ the set $G$ of allowed (discrete) probability distributions consists of all non-negative functions $g$ for which the following conditions hold

$$\sum_x x^k g(x) = \mu_k, k = 0, 1, \ldots, M. \quad (2)$$

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 $q \in G$ that maximizes the entropy $H(g)$, i.e.,

$$q = \arg \max_{g \in G} H(g) = \arg \max_{g \in G} \left(-\sum_x g(x) \ln g(x)\right). \quad (3)$$

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

$$\mathcal{L}(g, \lambda) = H(g) - \sum_{k=1}^M \lambda_k \left(\sum_x x^k g(x) - \mu_k\right),$$

where $\lambda = (\lambda_0, \ldots, \lambda_M)$ are the corresponding Lagrangian multipliers. The maximum of the unconstrained Lagrangian $\mathcal{L}$ corresponds to the solution of the constrained maximum entropy problem (4). Note that setting the derivatives of $\mathcal{L}(g, \lambda)$ w.r.t. $\lambda_k$ to zero results in the moment constraints. The general form of the maximum is obtained by setting $\frac{\partial \mathcal{L}}{\partial q(x)}$ to zero which yields

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

where

$$Z = e^{1+\lambda_0} \sum_x \exp \left(-\sum_{k=1}^M \lambda_k x^k\right) \quad (4)$$

is a normalization constant. The last equality in Eq. (4) follows from the fact that $q$ is a distribution and thus $\lambda_0$ is uniquely determined by $\lambda_1, \ldots, \lambda_M$. 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 $\lambda_k$. This yields the dual function

$$\Psi(\lambda) = \ln Z + \sum_{k=1}^M \lambda_k \mu_k. \quad (5)$$

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. (5) (see [6]).

The constrained optimization problem (4) can be solved analytically for $M \leq 2$. When $M > 2$ moments are given, numerical methods have to be applied. Theoretical conditions for the existence of a solution are well-elaborated (for example in [15, 20, 22]) however they do not provide an algorithmic way to derive a reconstruction. Here
we solve the maximum entropy problem by minimizing the dual function [5] using the iterative Levenberg-Marquardt method [23]. Let us first consider the reconstruction of one-dimensional discrete marginal distributions on $[0, \infty)$. Given an approximation $\lambda^{(t)} = (\lambda_1^{(t)}, \ldots, \lambda_M^{(t)})$ of the vector $\lambda = (\lambda_1, \ldots, \lambda_M)$ in the $t$-th step of the iteration (where we omit $\lambda_0$ due to Eq. (4)), the elements of the gradient vector are computed as $\partial \Psi / \partial \lambda_i \approx \mu_i - \frac{1}{2} \bar{\mu}_i$. The approximation $\bar{\mu}_i$ of the $i$-th moment is given by

$$\bar{\mu}_i = \sum_x x^i \exp \left( - \sum_{k=1}^M \lambda_k^{(t)} x^k \right), \quad i = 1, \ldots, 2M. \quad (6)$$

We discuss the truncation of the infinite support below and assume for now that the above sum ranges over $x \in \mathbb{N}_0$. The entries of the Hessian $H$ are approximated by

$$H_{i,j} = \frac{\partial^2 \Psi}{\partial \lambda_i \partial \lambda_j} \approx \frac{Z \cdot \bar{\mu}_{i+j} - \bar{\mu}_i \bar{\mu}_1}{Z^2}, \quad i, j = 1, \ldots, M.$$

Note that the normalization constant $Z$ and the approximated moments $\bar{\mu}_i$ are updated in each step of the iteration, i.e. they depend on $\ell$ but we omit the superscript $\ell$ here. Now, we make use of the Levenberg-Marquardt formula to compute the next approximation

$$\lambda^{(t+1)} = \lambda^{(t)} - \left( H + \gamma^{(t)} \cdot \text{diag}(H) \right)^{-1} \partial \Psi / \partial \lambda, \quad (7)$$

where we apply the simple updating strategy suggested in [23] for the damping factor $\gamma$. As an initial starting point for the optimization, we use $\lambda^{(0)} = (0, \ldots, 0)$ and stop the iteration when the difference between the solutions becomes smaller than a certain threshold $\delta_\lambda$, i.e. when the condition $|\lambda^{(t+1)} - \lambda^{(t)}| < \delta_\lambda$ is satisfied. Since for the systems that we consider the dual function is convex [18, 24], there exists a unique minimum $\lambda^* = (\lambda_1^*, \ldots, \lambda_M^*)$ where all first derivatives are zero and where the Hessian is positive definite. Note that the dimensionality of the optimization problem is $M$ due to the fact that $\lambda_0^* = \ln Z - 1$. In order to approximate the moments in Eq. (6) we need to sum over all possible states (the corresponding set of states might be infinite). Instead, during the iterative procedure, we consider a subset $D^* = \{x_L, \ldots, x_R\} \subset \mathbb{N}_0$ that contains the main part of the probability mass [21]. In the Supporting Information, Sect. 3 we describe in detail how the distribution support is approximated. The final results $\lambda^*$ and $D^*$ of the iteration yield the distribution

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

which is an approximation of the marginal distribution of the process at time $t$ (where $\tilde{q}(x) = 0$ if $x \notin D^*$).

To find the minimum of the dual function we use the Levenberg-Marquardt method which might fail due to the numerical instabilities when the inverse of the Hessian is calculated. To gain better numerical stability other functions of random variables need to be considered (other than monomials that generate algebraic moments) such as Chebyshev polynomials [5] and Fup basis functions [11]. A faster convergence of the iteration procedure can be obtained by exploiting other optimization methods such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) procedure [7] and the combined approach [1], where a series of transformations is used to overcome numerical difficulties. Note that the above approach provides a reasonable approximation of the individual probabilities only in the region where the main part of the probability mass is located. In order to accurately approximate the tails of the distribution special methods have been developed [9].

5. CASE STUDIES

In this section we present numerical results of the maximum entropy reconstruction when it is applied to the moments of a reaction network. We estimate the accuracy of the reconstruction by comparing the obtained maximum entropy distributions to the distributions computed via a direct numerical solution of the CME (as explained in Section 2). Our main focus is on investigating whether the MCM approach is more suitable for this than the MM approach. We reconstruct one- and two-dimensional marginal distributions in three different ways, in the sequel referred to as weighted sum MCM, joint MCM and MM. The last case refers to a reconstruction based on the moments obtained from the MM approach. The former two refer to reconstructions based on the MCM approach where weighted sum means that we reconstruct conditional distributions and derive the full marginal by multiplying with the probability of the corresponding condition (e.g. gene is active or not) and summing up. In contrast, joint MCM means that we use the unconditional moments (approximated by the product of conditional moments and the probability of the condition) for the reconstruction. In addition we reconstruct the conditional distribution from the conditional moments and compare them to the conditional distributions obtained via a direct numerical integration of the CME. In the Supporting Information, Sect. 5 we describe in detail how the reconstructed distributions are derived based on the approximated moments. We also check whether approximation errors are already introduced during the integration of the moments or whether they are due to the reconstruction step.

We consider only systems where a direct numerical simulation is possible such that we are able to reason about the accuracy of the moment-based analysis and the reconstruction of the distribution. For more complex systems with high population sizes a direct numerical solution is not feasible. In contrast, a moment-based analysis is possible when the population sizes are large since the computational demands are independent of the size of the populations. In the sequel we use the relative $L_\infty$ metric given by
\[ ||\epsilon||^2_{\infty} = \max_{x \in \{\hat{S}, D^*\}} \left| \frac{\hat{g}(x) - p_s(x)}{\tilde{p}_s(x)} \right| \]

where \( p_s(x) \) refers to the marginal probability distribution obtained by solving the CME directly, \( \hat{S} \) is the set of significant states for truncation parameter \( \delta = 10^{-15} \), and \( D^* \) is the support of the reconstructed distribution \( q \). For the reconstruction we set \( \delta_{\lambda} = 10^{-5} \) and \( \delta_{\text{prob}} = 10^{-4} \).

We present the most important findings in the following paragraphs and list all details of our numerical results in the Supporting Information, Sect. 6.

a. Gene Expression Model. For the reaction network of Example 1, the moment-based analysis based on the MM and the MCM approach is very accurate but the MCM approach has lower relative errors for high moments (see Table IV). Thus, for a higher order closure we cannot expect the MM approach to be more accurate than the joint MCM approach.

We first consider the one-dimensional marginal distributions of \( R \) and \( P \) where the moments \( \mu_0, \ldots, \mu_{M+1} \) for \( M \in \{3, 5, 7\} \) are computed based on the MM and the MCM approach. As constraints in Eq. (2) we use the equations for \( \mu_0, \ldots, \mu_M \) (but not the one of \( \mu_{M+1} \)) since the optimization procedure is sensitive to the relative high approximation errors of the moment of highest order. We then compare the approximation errors of all three reconstruction methods (weighted sum MCM, joint MCM and MM) at time \( t = 10 \) (see Supporting Information, Table III and Table IV) using the relative \( L_\infty \) metric. We find that the reconstruction is accurate for all \( M \in \{3, 5, 7\} \) and the best result for the distribution of mRNA is obtained when the joint MCM method is applied with \( M = 7 \) yielding a relative error of 0.15%. Thus, for this distribution an accurate approximation of the unconditional moments yields the best reconstruction since here the conditional distributions are slightly harder to reconstruct than the (unconditional) marginal distributions.

The distribution of protein molecules is reconstructed most accurately when the weighted sum MCM is applied but with \( M = 3 \) while the results for \( M \in \{5, 7\} \) are worse, i.e., a reconstruction based on the information given by the first three moments of the conditional distributions (\( |||\epsilon|||_{\infty}^2 = 5.94\% \)) and the conditional probabilities is the most accurate one. However, the reconstructed conditional distributions are less accurate for \( M = 3 \) (approximation errors are 14.6% and 28.1%) than for \( M \in \{5, 7\} \). Further investigations yield that the reason why the fit for \( M = 3 \) performs best is that within the weighted sum under- and overestimations at the point \( x = 0 \) are summed up and thus the rather inaccurate conditional distributions yield better results for the unconditional distribution than the more accurate conditional distributions for the case \( M \in \{5, 7\} \).

In Fig. 3 we plot the corresponding reconstructions (shown as crosses) both for conditional (left and middle plots) and marginal (right plot) distributions of proteins, where we condition on the state of the DNA, e.g. \( P|D_{\text{on}} \) refers to the approximation of the protein distribution when the DNA is active (\( D_{\text{on}} = 1 \)). These reconstructions are compared to the results of a direct numerical integration of the CME (shown in green).

The reconstruction of the two-dimensional marginal distributions also requires to solve the maximum entropy problem but with larger dimensionality. In the Supporting Information, Section 5 we discuss the differences that arise in the two-dimensional case. We observe that the relative error of the reconstruction decreases when higher-order moments are used. The minimum value is obtained when \( M = 7 \) and the best reconstruction is provided by joint MCM with an error of \( |||\epsilon|||_{\infty}^2 = 24.7\% \). However, all three methods have similar relative errors but the MCM approach uses less moment equations. The best reconstructions are shown in Fig. 4 where the left and middle plots refer to the conditional distributions (where conditions are \( D_{\text{on}} = 1 \) and \( D_{\text{off}} = 1 \) correspondingly) and the marginal distribution is shown in the right plot.

b. Exclusive Switch Model. We consider a gene regulatory network called exclusive switch [10]. It describes the dynamics of two genes with an overlapping promoter region, and the corresponding proteins \( P_1 \) and \( P_2 \). Both \( P_1 \) and \( P_2 \) are produced if no transcription factor is bound to the promoter region. However if a molecule of type \( P_1 \) (\( P_2 \)) is bound to the promoter then it inhibits the expression of the other protein, i.e. molecules of type \( P_2 \) (\( P_1 \)) cannot be produced. Only one molecule can be bound to the promoter region at a time which gives three possibilities for the state of the promoter region (free, \( P_1 \) bound, \( P_2 \) bound). The model is infinite in two dimensions (\( P_1 \) and \( P_2 \)) and the
In Fig. 5, where the left plot corresponds to the distribution of the proteins we also reconstruct the (bi-stable) two-dimensional distribution for the distribution of the proteins. We observe that the accuracy increases when we use information about higher-order moments and the best results are obtained for \( M = 7\). The most accurate reconstruction for the distribution of \( P_1 \) is obtained when we use joint MCM (\( ||\epsilon||_\infty = 10.7\% \)) and weighted sum MCM provides the best reconstruction for the distribution of \( P_2 \) (\( ||\epsilon||_\infty = 9.3\% \)). We also reconstruct the (bi-stable) two-dimensional distribution of the proteins \( P_1 \) and \( P_2 \). The approximation error \( ||\epsilon||_\infty = 14.2\% \) is minimal when we use the weighted sum MCM approach. The reconstruction results are shown in Fig. 5 where the left plot corresponds to the distribution obtained via a direct numerical solution of the CME and the reconstruction obtained using weighted sum MCM is shown in the right plot.

We observe that in most cases the reconstructions obtained using MCM-based approaches are more accurate than those using the MM approach both for high and low amount of moment constraints. In particular when a small number of moments is used (\( M = 3 \)), MCM-based approaches provide much better reconstructions. In the case of a high number of moments, the MCM approach uses a much smaller number of moment equations but the reconstructed distributions are as good or better than those obtained from the MM approach. For the distributions that have a simple shape (such as one-dimensional distributions in the gene expression model) the difference between MCM and MM is small, but for more complex distributions such as bi-modal distributions it is more advantageous to first reconstruct the conditional distributions based on the conditional moments and then derive the (unconditional) distributions as a weighted sum.

### 6. Conclusions

We considered moment-based approaches for the analysis of models based on the theory of stochastic chemical kinetics. We described how the maximum entropy approach can be used to reconstruct the underlying probability distributions from the moments. The accuracy of this combined procedure is investigated by comparing the obtained distributions with those arising from a direct numerical solution of the chemical master equation. Our experimental results show that the proposed combination of moment-based analysis and maximum entropy reconstruction is both fast and reasonably accurate. Moreover, we found that conditional moments often provide more information about the distributions leading to more accurate reconstructions. We expect that the proposed approach will allow a fast and accurate approximation of likelihoods and other event probabilities in large systems for which a direct solution of the CME is not feasible. As future work, we plan to combine the MCM approach with an efficient maximum likelihood approach for parameter calibration and experiment design approaches based on maximising the Fisher information.
Drabold. Maximum entropy and the problem of moments: A stable algorithm. *Physical Review E*, 71(5):057701/1–057701/4, 2005.

[6] Adam L. Berger, Vincent J. Della Pietra, and Stephen A. Della Pietra. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–71, 1996. ISSN 0891-2017.

[7] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scient. Comp.*, 16(5):1190–1208, 1995.

[8] Stefan Engblom. Computing the moments of high dimensional solutions of the master equation. *Applied Mathematics and Computation*, 180(2):498–515, 2006.

[9] P. N. Gavrilidis and G. A. Athanassoulis. The truncated stieltjes moment problem solved by using kernel density functions. *Journal of Computational and Applied Mathematics*, 236(17):4193–4213, 2012. ISSN 0377-0427.

[10] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 71(25):2340–2361, 1977.

[11] H. Gotovac and B. Gotovac. Maximum entropy algorithm with inexact upper entropy bound based on fup basis functions with compact support. *Journal of Computational Physics*, 228(24):9079–9091, 2009.

[12] R. Grima. A study of the accuracy of moment-closure approximations for stochastic chemical kinetics. *The Journal of Chemical Physics*, 136(15):4105, 2012.

[13] J. Hasenauer, V. Wolf, A. Kazeroonian, and F.J. Theis. Method of conditional moments for the chemical master equation. *Journal of Mathematical Biology*, pages 1–49, 2013. ISSN 0303-6812.

[14] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630, 1957.

[15] C. Kleiber and J. Stoyanov. Multivariate distributions and the moment problem. *Journal of Multivariate Analysis*, 113:7–18, 2013. ISSN 0047259X.

[16] A. Loinger, A. Lipshat, N. Q. Balaban, and O. Biham. Stochastic simulations of genetic switch systems. *Physical Review E*, 75:021904, 2007.

[17] M. Mateescu, V. Wolf, F. Didier, and T. A. Henzinger. Fast adaptive uniformisation of the chemical master equation. *IET Systems Biology*, 4(6):441–452, 2010.

[18] L. R. Mead and N. Papanicolaou. Maximum entropy in the problem of moments. *Journal of Mathematical Physics*, 25:2404, 1984.

[19] B. Munsky and M. Khammash. The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.*, 124:044144, 2006.

[20] J. Stoyanov. Krein condition in probabilistic moment problems. *Bernoulli*, 6(5):pp. 939–949, 2000. ISSN 13507265.

[21] A. Tari, M. Telek, and P. Buchholz. A simplified moment-based estimation method for extreme probabilities, infinite and positive cases. *Formal Techniques for Computer Systems and Business Processes*, pages 79–93. Springer, 2005.

[22] Árpád Tari, Miklós Telek, and Peter Buchholz. A unified approach to the moments based distribution estimation–unbounded support. In *Formal Techniques for Computer Systems and Business Processes*, pages 79–93. Springer, 2005.

[23] Matt Transtrum and J. Sethna. Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization. *arXiv preprint:1201.5885*, 2012.

[24] Zhijun Wu, George N Phillips Jr, Richard Tapia, and Yin Zhang. A fast newton algorithm for entropy maximization in phase determination. *SIAM review*, 43(4):623–642, 2001.
Appendix A: Supporting Information

In the following sections we first describe in detail how the moment equations are obtained (Section 1 and 2) and how we approximate the support of the distribution (Section 3). In Section 4 we then discuss the differences that arise during the reconstruction of distributions with two instead of only one dimension and in Section 5 we discuss the details of the reconstruction for the distributions of the case studies of Section 5. More numerical results for the two case studies are then provided in Section 6.

1. Method of Moments

For the time derivative of the expectation of a function \( f : \mathbb{N}_0^n \to \mathbb{R}^n \) applied to the vector of species we directly get from Eq. (1):

\[
\frac{d}{dt} E\left(f(\bar{X}(t))\right) = \sum_x f(x) \frac{d}{dt} p(\bar{x}, t) = \sum_{j=1}^{m} E\left(\alpha_j(\bar{X}(t)) (f(\bar{X}(t) + \bar{v}_j) - f(\bar{X}(t)))\right). \tag{A1}
\]

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

\[
\frac{d}{dt} E\left(\bar{X}(t)\right) = \sum_{j=1}^{m} \bar{v}_j E\left(\alpha_j(\bar{X}(t))\right). \tag{A2}
\]

Note that the system of ODEs in Eq. (A2) is only closed if at most monomolecular reactions \( \sum_{i=1}^{n} \ell_{j,i} \leq 1 \) are involved. Otherwise \( E\left(\alpha_j(\bar{X}(t))\right) \) involves moments of second order. However, in this case we can approximate the unknown second order moments, say \( E(X_i(t) \cdot X_j(t)) \) if the reaction is of the form \( S_i + S_j \rightarrow \ldots, i \neq j \), either by assuming that the covariance is zero, which gives \( E(X_i(t) \cdot X_j(t)) = E(X_i(t)) \cdot E(X_j(t)) \) or by extending the system in (A2) with additional equations for the second moments. The general strategy is to replace \( \alpha_j(\bar{X}(t)) \) by a Taylor series about the mean \( E(\bar{X}(t)) \). Let us write \( \mu_i(t) \) for \( E(X_i(t)) \) and \( \bar{\mu}(t) \) for the vector with entries \( \mu_i(t), 1 \leq i \leq n \). Then

\[
E\left(\alpha_j(\bar{X})\right) = \alpha_j(\bar{\mu}) + \frac{1}{1!} \sum_{i=1}^{n} E\left(X_i - \mu_i\right) \frac{\partial}{\partial x_i} \alpha_j(\bar{\mu}) + \frac{1}{2!} \sum_{i=1}^{n} \sum_{k=1}^{n} E\left((X_i - \mu_i)(X_k - \mu_k)\right) \frac{\partial^2}{\partial x_i \partial x_k} \alpha_j(\bar{\mu}) + \ldots \tag{A3}
\]

where we omitted \( t \) in the equation to improve the readability. Note that \( E(X_i(t) - \mu_i) = 0 \) and since we restrict to reactions that are at most bimolecular, all 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

\[
E\left(\alpha_j(\bar{X})\right) = \alpha_j(\bar{\mu}) + \frac{1}{1!} \sum_{i=1}^{n} \sum_{k=1}^{n} C_{ik} \frac{\partial}{\partial x_i \partial x_k} \alpha_j(\bar{\mu}) \tag{A4}
\]

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

\[
\frac{d}{dt} C_{ik} = \frac{d}{dt} E(X_i X_k) - \frac{d}{dt} (\mu_i \mu_k) = \frac{d}{dt} E(X_i X_k) - (\frac{d}{dt} \mu_i) \mu_k - \mu_i (\frac{d}{dt} \mu_k) \tag{A5}
\]

and if we couple this equation with the equations for the means, the only unknown term that remains is the derivative \( \frac{d}{dt} E(X_i X_k) \) of the second moment. For this we can use the same strategy as before, i.e., from Eq. (A1) we get

\[
\frac{d}{dt} E(X_i X_k) = \sum_{j=1}^{m} \left( v_{j,i} v_{j,k} E\left(\alpha_j(\bar{X})\right) + v_{j,i} E\left(\alpha_j(\bar{X}) X_j\right) + v_{j,k} E\left(\alpha_j(\bar{X}) X_k\right) \right), \tag{A6}
\]

where \( v_{j,i} \) and \( v_{j,k} \) are the corresponding entries of the vector \( v_j \). Clearly, we can use Eq. (A4) for the term \( E(\alpha_j(\bar{X})) \) while the terms \( E(\alpha_j(\bar{X}) X_j) \) and \( E(\alpha_j(\bar{X}) X_k) \) have to be replaced by the corresponding Taylor series about the mean. Let \( f_j(\bar{x}) := \alpha_j(\bar{x}) \bar{x}_i \). Similarly to Eq. (A4) we get that \( E(\alpha_j(\bar{X}) X_j) \) equals

\[
\alpha_j(\bar{\mu}) \mu_i + \frac{1}{1!} \sum_{i=1}^{n} E\left(X_i - \mu_i\right) \frac{\partial}{\partial x_i} f_j(\bar{\mu}) + \frac{1}{2!} \sum_{i=1}^{n} \sum_{k=1}^{n} E\left((X_i - \mu_i)(X_k - \mu_k)\right) \frac{\partial^2}{\partial x_i \partial x_k} f_j(\bar{\mu}) + \ldots \tag{A7}
\]

Here, it is important to note that moments of order three come into play since derivatives of order three of \( f_j(\bar{x}) = \alpha_j(\bar{x}) \bar{x}_i \) may be nonzero. It is possible to take these terms into account by deriving additional equations for moments of order three and higher. Obviously, these equations will then include moments of even higher order such that theoretically we end up with an infinite system of equations. However, a popular strategy is to close the equations by assuming that all moments of order \( M \) that are centered around the mean are equal to zero. E.g., if we choose \( M = 2 \), then we can simply use the approximation

\[
E\left(\alpha_j(\bar{X}) X_i\right) \approx \alpha_j(\bar{\mu}) \mu_i + \frac{1}{1!} \sum_{i=1}^{n} \sum_{k=1}^{n} E\left((X_i - \mu_i)(X_k - \mu_k)\right) \frac{\partial^2}{\partial x_i \partial x_k} f_j(\bar{\mu}).
\]

This approximation is then inserted into Eq. (A6) and the result is used to replace the term \( \frac{d}{dt} E(X_i X_k) \) in Eq. (A5). Finally, we can integrate the time evolution of the means and that of the covariances and variances.
Example 2. We apply the standard moment closure technique to the gene expression system from Example 1. When we consider only the moments up to second order the corresponding equations for the average number of molecules are, for instance, given by

\[
\begin{align*}
d\mu_{D_{\text{off}}} &= \tau_{\text{off}} \mu_{D_{\text{on}}} - E(\tau_{\text{on}} X_{D_{\text{off}}} X_P) \\
d\mu_{D_{\text{on}}} &= \tau_{\text{on}} \mu_{D_{\text{off}}} + E(\tau_{\text{on}} X_{D_{\text{off}}} X_P) \\
d\mu_R &= k_r \mu_{D_{\text{off}}} - \gamma_r \mu_R \\
d\mu_P &= k_p \mu_R - \gamma_p \mu_P,
\end{align*}
\]

where \(\mu_{D_{\text{off}}}, \mu_{D_{\text{on}}}, \mu_R, \mu_P\) are the expected numbers of \(D_{\text{off}}\), \(D_{\text{on}}\), respectively, and \(\mu_R, \mu_P\) are the expected numbers of mRNA and proteins.

Next we compute the obtained moments with those computed via a direct numerical integration of the CME (Table 1). We consider the following three cases. The moment closure approximation is carried out using all moments up to order 1, 6, and 8. For each case we list the number of moment equations, the running time, and the maximum relative errors in the first four moments (columns 4–7).

Please notice that in the reconstruction procedure we do not use the moment of the highest order. For example, if we approximate moments up to order 6 then the highest order that is taken into account during the reconstruction is 5 (which corresponds to the case \(M = 5\), cf. Sect. 6) because of the high sensitivity of the numerical procedure even to the small absolute error in the moment approximation.

2. Method of Conditional Moments

We first decompose the chemical populations described by \(\tilde{X}(t)\) into small and large populations. Here we assume that this decomposition is static. However, it is obvious that during the integration over time, we can (after reconstructing the joint distribution) choose a different decomposition for the remaining time. The question from what size on a population is considered as small is typically dependent on the amount of main memory that is available and on the maximum order of the moments that we consider for the large populations. Note that considering conditional moments yields a smaller amount of equations if the order of the considered moments is high. The reason is that the number of equations for representing the dynamics of the small populations does not increase as the order of considered conditional moments increases. Also, for many systems the decomposition is obvious as the small populations are exactly those that have a maximal size of, say, less than 10 (because they represent binding sites) and the large populations count protein numbers which may become rather large.

Formally, we write the random vector \(\tilde{X}(t)\) at time \(t\) as \(\tilde{X}(t) = (\tilde{Y}(t), \tilde{Z}(t))\) where \(\tilde{Y}(t)\) corresponds to the small, \(\tilde{Z}(t)\) to the large populations. Similarly, we write \(\tilde{x} = (\tilde{y}, \tilde{z})\) for the states of the process. We also assume that the propensity functions \(\alpha_j(\tilde{x})\) can be decomposed such that \(\alpha_j(\tilde{x}) = \tilde{\alpha}(\tilde{y})\tilde{\alpha}(\tilde{z})\) for \(\tilde{x} = (\tilde{y}, \tilde{z})\) where \(\tilde{\alpha}\) and \(\tilde{\alpha}\) are nonnegative functions of the vectors of small and large populations. Similarly, we decompose the change vectors \(\tilde{v}_j\) such that \(\tilde{v}_j = (\tilde{v}_y, \tilde{v}_z)\). Again, the first component refers to the small and the second component to the large populations. Now, Eq. (1) becomes

\[
\frac{d p(\tilde{y}, \tilde{z})}{dt} = \sum_{j=1}^{m} (\tilde{\alpha}_j(\tilde{y} - \tilde{v}_y)\tilde{\alpha}_j(\tilde{z} - \tilde{v}_z)p(\tilde{y} - \tilde{v}_y, \tilde{z} - \tilde{v}_z) - \tilde{\alpha}_j(\tilde{y})\tilde{\alpha}_j(\tilde{z})p(\tilde{y}, \tilde{z}))
\]

where we omitted the time parameter \(t\) to improve readability. Next, we sum over all possible \(\tilde{z}\) to get the time evolution of the marginal distribution \(\tilde{p}(\tilde{y}) = \sum_{\tilde{z}} p(\tilde{y}, \tilde{z})\) of the small populations.

\[
\frac{d}{dt} \tilde{p}(\tilde{y}) = \sum_{\tilde{z}} \sum_{j=1}^{m} \tilde{\alpha}_j(\tilde{y} - \tilde{v}_y)\tilde{\alpha}_j(\tilde{z} - \tilde{v}_z)p(\tilde{y} - \tilde{v}_y, \tilde{z} - \tilde{v}_z) - \sum_{\tilde{z}} \sum_{j=1}^{m} \tilde{\alpha}_j(\tilde{y})\tilde{\alpha}_j(\tilde{z})p(\tilde{y}, \tilde{z})
\]

\[
\sum_{\tilde{y}} \tilde{\alpha}_j(\tilde{y} - \tilde{v}_y)\tilde{\alpha}_j(\tilde{z})p(\tilde{y} - \tilde{v}_y, \tilde{z})E[\tilde{\alpha}_j(\tilde{Z}) | Y = \tilde{y} - \tilde{v}_y]
\]

\[
- \sum_{\tilde{y}} \tilde{\alpha}_j(\tilde{y})\tilde{\alpha}_j(\tilde{Z})E[\tilde{\alpha}_j(\tilde{Z}) | Y = \tilde{y}]
\]

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 \(\tilde{y}\) 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 continuous part of the reaction rate. For the latter, we can use a Taylor expansion about the conditional population means. Similar to Eq. (A3) 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 probability \(\tilde{p}(\tilde{y})\) may become zero, we first derive an equation for the evolution

| moment closure order | time (sec) | error ord. 1 | error ord. 2 | error ord. 3 | error ord. 4 |
|---------------------|-----------|--------------|--------------|--------------|--------------|
| 4                   | 70        | 8 · 10^{-6}  | 8.3 · 10^{-3} | 9.6 · 10^{-3} | 8.24 · 10^{-3} |
| 6                   | 210       | 2 · 10^{-6}  | 2 · 10^{-6}  | 1 · 10^{-3}  | 3.6 · 10^{-3}  |
| 8                   | 495       | 1 · 10^{-6}  | 2 · 10^{-6}  | 2 · 10^{-6}  | 4 · 10^{-6}  |

TABLE 1: Moment closure approximation results for the gene expression system
of the partial means (conditional means multiplied by the probability of the condition).

\[
\frac{d}{dt} \left( E[\vec{Z} \mid \vec{y}] p(\vec{y}) \right) = \sum_{\vec{y}} \frac{d}{dt} p(\vec{y}, \vec{z}) = \sum_{\vec{y}} \hat{\alpha}_j(\vec{y} - \hat{\vec{v}}_j) E[(Z_i + \hat{v}_{ij})\alpha_j(\vec{Z}) \mid \vec{y} - \hat{\vec{v}}_j] p(\vec{y} - \hat{\vec{v}}_j)
\]

where in the second line we applied Eq. (A8) and simplified the result. Here \(v_{ij}\) corresponds to \(j\)-th element of the change vector for \(i\)-th reaction. The conditional expectations \(E[(Z_i + \hat{v}_{ij})\alpha_j(\vec{Z}) \mid \vec{y} - \hat{\vec{v}}_j]\) are then replaced by their Taylor expansion about the conditional means such that the equation involves only conditional means and higher centered conditional moments \([13]\). 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 \(p(\vec{y})\) 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.

**Example 3.** We apply the method of conditional moments to the gene expression system from Example 1. The modes of the system are then given by the state of the DNA. The equations for the mode probabilities (\(P_{\text{off}}, P_{\text{on}}\)) and the expected number of mRNA (\(\mu_{R,\text{off}}, \mu_{R,\text{on}}\)) and proteins (\(\mu_{P,\text{off}}, \mu_{P,\text{on}}\)) as are follows:

\[
\frac{d}{dt} P_{\text{off}} = \tau_{\text{on}} P_{\text{on}} - (\tau_{\text{off}} + \tau_{\text{on}} \mu_{P,\text{off}}) P_{\text{off}}
\]

\[
\frac{d}{dt} (\mu_{R,\text{off}} P_{\text{off}}) = -\gamma R \mu_{R,\text{off}} P_{\text{off}}
\]

\[
\frac{d}{dt} (\mu_{P,\text{off}} P_{\text{off}}) = (k_p \mu_{R,\text{on}} - \gamma_p \mu_{P,\text{off}}) P_{\text{off}}
\]

\[
\frac{d}{dt} P_{\text{on}} = (\tau_{\text{off}} + \tau_{\text{on}} \mu_{P,\text{off}}) P_{\text{off}} - \tau_{\text{on}} P_{\text{on}}
\]

\[
\frac{d}{dt} (\mu_{R,\text{on}} P_{\text{on}}) = (k_r - \gamma R \mu_{R,\text{on}}) P_{\text{on}}
\]

\[
\frac{d}{dt} (\mu_{P,\text{on}} P_{\text{on}}) = (k_p \mu_{R,\text{on}} - \gamma_p \mu_{P,\text{on}}) P_{\text{on}}
\]

We computed the conditional moments and conditional probabilities of the running example (cf. Ex. 1 and Ex. 2) over time by considering moments up to the order of 4, 6, and 8. For these three cases the number of equations, when compared to the method of moments (MM), are as follows:

| moment order | M   | 4   | 6   | 8   |
|--------------|-----|-----|-----|-----|
| # equations for MM | 70  | 210 | 495 |
| # equations for MCM | 30  | 50  | 90  |

The maximum relative errors of the results of the method of conditional moments (MCM) are given in Table III where we again compared to the results obtained via a direct numerical solution.

Our experiments show that the MCM performs much faster (due to the smaller number of equations) and still yields accurate approximation of the moments. Our observation was that the MCM tends to provide a better approximation for higher moments whereas the MM approach is more accurate for lower moments when the same number of moments is considered. For example, in case of 6 moments the maximum relative error for the first moments computed by the MM approach is 0.000002 compared to 0.000032 when computed using the MCM. At the same time, the maximum relative errors of the sixth moments are 0.000647 and 0.000203 for the MM and the MCM respectively. Note that the (unconditional) moments for the MCM are computed via multiplication of the conditional moments with the mode probabilities.

### 3. Approximation of the Support

During the iteration we approximate the moments using Eq. (6) where we do not sum over all states \(x \in \mathbb{N}_0\) but consider a subset \(D = \{x_L, \ldots, x_R\} \subset \mathbb{N}_0\). Note that we have to find appropriate values for \(x_L\) and \(x_R\) since the iteration might fail to converge if the chosen value of \(x_R\) is very large (and if \(x_L = 0\)) as the conditional number of the matrix \((H + \gamma(H) \cdot \text{diag}(H))\) is very large in this case.

Thus, we use the results of the in [21] to find a region that contains the main part of the probability mass. We consider the roots of the function

\[
\Delta^0(w) = \begin{vmatrix}
\mu_0 & \mu_1 & \cdots & \mu_k \\
\vdots & \vdots & & \vdots \\
\mu_{k-1} & \mu_k & \cdots & \mu_{2k-1} \\
1 & w & \cdots & w^k
\end{vmatrix},
\]

where \(k = \left\lfloor \frac{M}{2} \right\rfloor\), and \(M\) is even. Let \(W = \{w_1, \ldots, w_k\}\) be the set of the solutions of \(\Delta^0(w) = 0\), where \(w_1 < \ldots < w_k\) are real and simple roots. The set \(D^{(0)} = \{x_L^{(0)}, \ldots, x_R^{(0)}\}\) with \(x_L^{(0)} = |w_1|\) and \(x_R^{(0)} = |w_k|\) is used as an initial guess for the approximated support when we start the optimization procedure. In the \(i\)-th iteration we refine the approximation by checking if the probability of the
right-most state $x_R^{(i)}$ is reasonably small in comparison to the maximum value of $g(x)$ for $x \in D^{(i)}$, i.e.,
\[
g(x_R^{(i)}) < \delta_{\text{prob}} \cdot \max_{x \in D^{(i)}} g(x),
\]
where $\delta_{\text{prob}}$ is a small threshold. The support is extended until inequality (A11) is satisfied. The final results $\lambda^*$ and $D^*$ of the iteration yields the distribution
\[
\bar{q}(x) = \exp\left(-1 - \sum_{k=0}^M \lambda_k^* x^k\right),
\]
which is an approximation of the marginal distribution $p_*(x,t) = P(X(t) = x)$, i.e. $p_*(x,t) \approx \bar{q}(x)$ if $x \in D^*$ and $p_*(x,t) \approx 0$ if $x \notin D^*$.

We can also account for the case of an odd number of moments. In addition to the function $\Delta^0(w)$ defined in Eq. (A10), we also consider the function $\Delta^1(\eta)$
\[
\Delta^1(\eta) = \begin{bmatrix}
\mu_1 - w_1 \mu_0 & \cdots & \mu_z - w_1 \mu_{z-1} \\
\vdots & \ddots & \vdots \\
\mu_{z-1} - w_1 \mu_{z-2} & \cdots & \mu_2 - w_1 \mu_1 - w_2 \mu_0 \\
1 & \cdots & \eta^2 - 1
\end{bmatrix},
\]
where $z = \lfloor \frac{M}{2} \rfloor + 1$ and $w_1$ is the smallest root of the equation $\Delta^0(w) = 0$. Again, let $W = \{w_1, \ldots, w_k\}$ be the set of solutions of $\Delta^0(w) = 0$ and $H = \{\eta_1, \ldots, \eta_k\}$ be the set of solutions of $\Delta^1(\eta) = 0$, where all the elements of $W$ and $H$ are real and simple. The first approximation for the truncated support of the distribution is then given by the set $D^{(0)} = \{x_L^{(0)}, \ldots, x_R^{(0)}\}$ with $x_L^{(0)} = \min(w_1, \eta_1)$ and $x_R^{(0)} = \max(w_k, \eta_k)$. We extend the support until inequality (A11) is satisfied by adding a new state in each iteration
\[
(x_L^{(l+1)}, x_R^{(l+1)}) = \begin{cases}
(x^{(l)}, x_R^{(l)}) & \text{if } \text{even}, \\
(x_L^{(l)}, x_R^{(l)} + 1) & \text{if } \text{odd}.
\end{cases}
\]
The final results $\bar{\lambda}$ and $\bar{D}$ of the iteration yields the distribution $\bar{q}(x)$ that approximates the marginal distribution of interest.

4. Numerical Approach for the Two-dimensional Maximum Entropy Problem

In case of two-dimensional distributions, the maximum entropy problem is modified as follows. We consider a sequence of non-central moments $E(X^r X^l) = \mu_{r,l}$, $0 \leq r + l \leq M$, and the set $G^2$ of all two-dimensional discrete distributions that satisfy the following constraints
\[
\sum_{x,y} x^r y^l g(x, y) = \mu_{r,l}, \quad 0 \leq r + l \leq M.
\]
Here $X$ and $X_o$ correspond to the populations of two different species, i.e. to two distinct elements of the random vector $\hat{X}(t) = (X_1(t), \ldots, X_u(t))$ at some fixed time instant $t$. Similarly to the optimization problem (3), we seek for the distribution $q \in G^2$ that maximizes the entropy $H(g)$
\[
q = \arg \max_{g \in G^2} \sum_{x,y} g(x, y) \ln g(x, y)
\]
We proceed similar to the one-dimensional case. The general form of the solution for the maximum entropy problem is given by
\[
q(x, y) = \exp\left(-1 - \sum_{0 \leq r + l \leq M} \lambda_{r,l} x^r y^l\right) = \frac{1}{Z} \exp\left(- \sum_{0 \leq r + l \leq M} \lambda_{r,l} x^r y^l\right),
\]
where normalization constant $Z$ is calculated as
\[
Z = e^{1+\lambda_{0,0}} \sum_{x,y} \exp\left(- \sum_{0 \leq r + l \leq M} \lambda_{r,l} x^r y^l\right).
\]
To solve the maximum entropy optimization problem numerically, we apply the Levenberg-Marquardt method in Eq. (7) where $\lambda_{l} = (\lambda_{0,1}, \lambda_{1,0}, \ldots, \lambda_{0,L}, \lambda_{L,0})$ is an approximation of the vector $\lambda$ in Eq. (A15). The elements of the gradient vector are computed as $\frac{\partial g}{\partial \lambda_{l}} \approx \mu_{r,l} - \frac{1}{Z} \tilde{\mu}_{r,l}$, where $\tilde{\mu}_{r,l}$ is approximated by
\[
\tilde{\mu}_{r,l} = \sum_{x,y} x^r y^l \exp\left(- \sum_{0 \leq r + l \leq M} \lambda_{r,l} x^r y^l\right).
\]
Here $r, l \in \{0, \ldots, 2M\}$ and the sum is taken over all $(x,y) \in N^2_0$. Finally, the elements of the Hessian matrix are computed as
\[
H_{r+u,l+v} = \frac{\partial^2 \bar{g}}{\partial \lambda_{u,v}} \approx \frac{Z \tilde{\mu}_{r+u,l+v} - \tilde{\mu}_{r,l} \tilde{\mu}_{u,v}}{Z^2},
\]
where $0 \leq r + u \leq M, 0 \leq u + v \leq M$. Following the same procedure as in Section 4, the vector $\lambda^* = (\lambda_{0,1}, \lambda_{1,0}, \ldots, \lambda_{0,M}, \lambda_{M,0})$ is found. The dimensionality of the optimization problem is $0.5 (M^2 + 3M)$ and $\lambda_{0,0}$ can be calculated from (A16) as $\lambda_{0,0} = \ln Z - 1$. In comparison to the one-dimensional case, the range of the values of $\tilde{\mu}_{r,l}$ becomes wider due to the larger dimensionality, so that the conditional number of the matrix $(H + \gamma (\cdot) \cdot \text{diag}(H))$ is even higher and the iteration might fail. To approximate the moment values in (A17) we truncate the infinite support and consider the subset $D^2_{xy} = D_x^2 \times D_y^2$ instead. Again, we choose $D^2_{xy}$ such that the probability of the right-most state in each dimension is small enough in comparison to the maximum value of the corresponding one-dimensional marginal distribution, that is we check in the $i$-th iteration whether the following two conditions hold
\[
g(x_R^{(i)}), \quad g(x_R^{(i)}) < \delta_{\text{prob}} \cdot \max_{x \in D_{xy}^{(i)}} g(x, \cdot),
\]
\[
g(x_R^{(i)}, y_R^{(i)}) < \delta_{\text{prob}} \cdot \max_{y \in D_{xy}^{(i)}} g(\cdot, y).
\]
Here $g(x, \cdot)$ denotes the one-dimensional marginal distribution for $x \in D_k^{(1)}$ obtained as $g(x, \cdot) = \sum_{y \in D_k^{(2)}} g(x, y)$. The approximation $\tilde{q}(x, y)$ of the marginal distribution $p_{\cdot,\cdot}(x, y, t) = P(X_t(x) = x, X_0(t) = y)$ is then defined by the result $\lambda^*$ of iteration procedure such that $p_{\cdot,\cdot}(x, y, t) \approx \tilde{q}(x, y)$ if $(x, y) \in D_{xy}^*$ and $p_{\cdot,\cdot}(x, y, t) \approx 0$ if $(x, y) \notin D_{xy}^*$.

5. Reconstruction of Distributions from Approximated Moments

In the sequel we discuss the details of the reconstruction of marginal probability distributions based on solving the moment problem using the maximum entropy approach. We consider the three possibilities introduced in Section 4: weighted sum MCM, joint MCM and MM. We illustrate the details of all three approaches by means of an example.

Example 4. We consider the gene expression model (see Example 1) where we reconstruct the marginal distribution of protein molecules $P(X_P(x) = x) = p_X(x, t)$. The moments $\mu_k = E(X_k^p)$ and the corresponding conditional moments are obtained using the MCM and MM equations, for $k = 0, \ldots, M + 1$. In the case of joint MCM and MM we use the first $M + 1$ moments as values as constraints in Eq. (4) and solve the maximum entropy optimization problem in Eq. (3). In both cases, the solution is given by a pair $(\lambda^*, D^*)$ of parameter vector and truncated support. The corresponding reconstructed distribution is defined as

$$\tilde{q}(x) = \exp\left(-1 - \sum_{k=0}^{M} \lambda_k^* x^k\right), \quad x \in D^*.$$  

In order to apply the weighted sum MCM, we reconstruct the conditional distribution from the sequences $\mu_{P_{\text{off}},k}$ and $\mu_{P_{\text{on}},k}$ which approximate the conditional moments $E(X_k^p|D_{\text{off}} = 1)$ and $E(X_k^p|D_{\text{on}} = 1)$ for $k = 0, \ldots, M + 1$. Here, $X_p$ counts the number of proteins and the condition $D_{\text{off}} = 1$ ($D_{\text{on}} = 1$) refers to the case where the gene is inactive (active). These sequences are obtained using the MCM approach together with the approximation of the mode probabilities $p_{\text{off}}$ and $p_{\text{on}}$ (cf. Example 1). We solve the maximum entropy problem for each moment sequence and the reconstruction of marginal unconditional distribution is given by

$$\tilde{q}_{\text{wsMCM}}(x) = \begin{cases} p_{\text{off}} \tilde{q}_{\text{off}}(x), & x \in D_{\text{off}}^p \setminus D_{\text{on}}^p \\ p_{\text{on}} \tilde{q}_{\text{on}}(x), & x \in D_{\text{on}}^p \setminus D_{\text{off}}^p \\ p_{\text{off}} \tilde{q}_{\text{off}}(x), & x \in D_{\text{off}}^p \cap D_{\text{on}}^p \\ + p_{\text{on}} \tilde{q}_{\text{on}}(x), & x \in D_{\text{on}}^p \cap D_{\text{off}}^p, \end{cases}$$

where $\tilde{q}_{\text{off}}(x)$ and $\tilde{q}_{\text{on}}(x)$ are the reconstructions of the conditional distributions.

To reconstruct two-dimensional marginal distributions we numerically solve the two-dimensional maximum entropy problem as described in Section 4. We illustrate how two-dimensional distributions are reconstructed by means of an example and apply the weighted sum MCM approach.

Example 5. We consider the exclusive switch system described in Section 5. The goal here is to reconstruct the two-dimensional marginal distribution $P(X_P = x, X_P(t) = y)$ of proteins $P_1$ and $P_2$. We first approximate the mode probabilities $p_1 = P(DNA = 1)$, $p_2 = P(DNA.P1 = 1)$ and $p_3 = P(DNA.P2 = 1)$ (cf. Eq. A9). In addition, the conditional moments

$$\mu_{1:r,l} = E(X_{P_1}^r X_{P_2}^l | DNA = 1)$$

are approximated for $0 \leq r + l \leq M + 1$, where $DNA = 1$ refers to the case where the promoter is free and $DNA.P1 = 1$ ($DNA.P2 = 1$) to the case where a molecule of type $P_1$ (type $P_2$) is bound to the promoter. The constraints for the maximum entropy problem are given by the elements of these three sequences for $0 \leq r + l \leq M$ and the corresponding solutions of the optimization problem are given by the pairs $(\lambda_i^*, D_i^*)$, $i = \{1, 2, 3\}$. Then the reconstructed distribution is given by

$$\tilde{q}_{\text{wsMCM}}(x, y) = \begin{cases} p_1 \tilde{q}_1(x, y), & (x, y) \in D_1^* \setminus (D_2^* \cup D_3^*) \\ p_2 \tilde{q}_2(x, y), & (x, y) \in D_2^* \setminus (D_1^* \cup D_3^*) \\ p_3 \tilde{q}_3(x, y), & (x, y) \in D_3^* \setminus (D_1^* \cup D_2^*) \\ \sum_{i=1}^2 p_i \tilde{q}_i(x, y), & (x, y) \in (D_1^* \cap D_2^*) \setminus D_3^* \\ \sum_{i=1}^3 p_i \tilde{q}_i(x, y), & (x, y) \in (D_1^* \cap D_3^*) \setminus D_2^* \\ \sum_{i=1}^3 p_i \tilde{q}_i(x, y), & (x, y) \in (D_2^* \cap D_3^*) \setminus D_1^* \\ \sum_{i=1}^3 p_i \tilde{q}_i(x, y), & (x, y) \in D_1^* \cup D_2^* \cup D_3^*, \end{cases}$$

where $\tilde{q}_i(x, y) = \exp(-1 - \sum_{r+l \leq M} \lambda_i^{r,l} x^r y^l)$.

6. Case Studies

Here we present detailed results of the reconstruction of the marginal distributions that were discussed in Sect. 4 where we introduced only one metric $||\epsilon||_\infty$ to measure the relative approximation error in $L_\infty$. Here we use the following two additional metrics:

$$||\epsilon||_\infty = \max_{x \in (S_{\text{crit}} \cup D^*)} |\tilde{q}(x) - p(x)|$$

$$||\epsilon||_1 = \sum_{x \in (S_{\text{crit}} \cup D^*)} |\tilde{q}(x) - p(x)|$$

Again, we compute these error measures by comparing the reconstructed distributions with those given by the numerical integration of the CME.
TABLE III: Approximation errors of the mRNA distribution (gene expression).

| metric | M | $R|D_\text{off}$ | $R|D_\text{on}$ | $R|w\text{MCM}$ | $R|j\text{MCM}$ | $R|\text{MM}$ |
|--------|---|-----------------|----------------|-----------------|-----------------|----------------|
| $||\epsilon||_\infty$ | 3 | 0.0013          | 0.0028         | 0.0011          | 0.0028          | 0.0027         |
|        | 5 | 0.0006          | 0.0014         | 0.0001          | 0.0002          | 0.0004         |
|        | 7 | 0.0019          | 0.0006         | 0.0011          | 0.0001          | 0.0007         |
| $||\epsilon||_{\%}$ | 3 | 0.9554          | 19.9998        | 1.1784          | 3.9024          | 3.8281         |
|        | 5 | 0.4752          | 4.6059         | 1.1528          | 0.2085          | 0.489          |
|        | 7 | 1.3378          | 0.5088         | 1.3816          | 0.1493          | 0.862          |
| $||\epsilon||_1$ | 3 | 0.0075          | 0.0217         | 0.0072          | 0.0192          | 0.0192         |
|        | 5 | 0.0033          | 0.0118         | 0.0077          | 0.0012          | 0.0021         |
|        | 7 | 0.0086          | 0.0050         | 0.0063          | 0.0005          | 0.0053         |

TABLE IV: Approximation errors of the protein distribution (gene expression).

| metric | M | $P|D_\text{off}$ | $P|D_\text{on}$ | $P|w\text{MCM}$ | $P|j\text{MCM}$ | $P|\text{MM}$ |
|--------|---|-----------------|----------------|-----------------|-----------------|----------------|
| $||\epsilon||_\infty$ | 3 | 0.0032          | 0.0012         | 0.0007          | 0.0001          | 0.0011         |
|        | 5 | 0.0016          | 0.0006         | 0.0007          | 0.0001          | 0.0008         |
|        | 7 | 0.0019          | 0.0006         | 0.0011          | 0.0001          | 0.0007         |
| $||\epsilon||_{\%}$ | 3 | 14.6062         | 28.1174        | 5.8293          | 9.8885          | 9.5832         |
|        | 5 | 7.958           | 14.5507        | 5.9393          | 8.8261          | 7.006          |
|        | 7 | 13.4622         | 1.439          | 11.0841         | 8.6216          | 5.8925         |
| $||\epsilon||_1$ | 3 | 0.0173          | 0.0187         | 0.0104          | 0.0052          | 0.0052         |
|        | 5 | 0.0269          | 0.0124         | 0.0154          | 0.0055          | 0.0091         |
|        | 7 | 0.0141          | 0.012          | 0.0106          | 0.0059          | 0.0058         |

Gene Expression Model. We show the approximation errors $||\epsilon||_{\infty}$, $||\epsilon||_{\%}$ and $||\epsilon||_1$ for the reconstruction of both conditional and unconditional distributions for mRNA (protein) in Table III (Table IV). Here, the first two columns refer to the approximation error of the conditional distributions for mRNA (protein) denoted by $R|D_\text{off}$ ($P|D_\text{off}$) and $R|D_\text{on}$ ($P|D_\text{on}$). The last three columns refer to the reconstructions of the marginal distribution obtained using weighted sum MCM, joint MCM and MM, respectively. We observe that the reconstruction is most accurate for the distributions of mRNA when the joint MCM method is applied with $M = 7$ in terms of both maximum deviation $||\epsilon||_{\infty}$ and sum of absolute errors $||\epsilon||_1$. The distribution of protein molecules is reconstructed most accurately (with respect to $||\epsilon||_1$) when joint MCM or MM is applied but with $M = 3$. It shows that the main part of this distribution can already be nicely explained using the information that is contained in the first three moments. Please note that the large approximation errors of conditional distribution reconstructions may still provide an accurate reconstruction for the unconditional distribution because of the computation of weighted sum that can average out individual deviations from the true probability value.

The sensitivity of the optimization procedure can also influence the final result. The reconstruction that uses less degrees of freedom can provide an accurate solution since the distribution of the simple shape is able to explain the main behavior. At the same time, adding more moments into the consideration allows to capture more details, however it may change the reconstruction drastically due to the sensitivity and the corresponding approximation error can become larger. To the best of our knowledge, there exist no criteria that provides the number of moments that have to be considered such that adding more information do not change the maximum entropy reconstruction much.

The number of degrees of freedom We notice that the approximation error of the marginal distribution computed using joint MCM does not have to be given by the weighted sum of the approximation errors of conditional distribution recover. The joint MCM method performs best in terms of $||\epsilon||_{\infty}$. We notice that the reconstruction results are generally quite similar for the approaches that are based on an approximation of the unconditional moments. However, the MCM approach has the advantage that the distribution of species such as DNA is very accurate since they are directly available and are not reconstructed from the moments. A moment based approach such as MM needs a large number of moments for an accurate reconstruction [5].

We plot the best (with respect to $||\epsilon||_1$) reconstructions of the mRNA distributions in Fig. 6, where the left (middle) plot corresponds to the distribution of mRNA conditioned on the state of DNA, $D_\text{off} = 1$ ($D_\text{on} = 1$). The right plot in Fig. 6 shows the reconstruction of the marginal distribution of mRNA (using joint MCM and $M = 3$). All reconstructed distributions (shown as crosses) are compared to the distributions obtained via a direct numerical integration of the CME (shown in green).

We also notice that the approximation of conditional moments in MCM method is less accurate then the approximation of unconditional moments in MM method (cf. Tables I and II). It means that the main source of the approximation error is the optimization procedure of maximum entropy method.

An example of a two-dimensional distribution reconstruction was shown in Fig. 4. Here we present in addition the approximation errors for all three reconstruction methods in Tab. VII both for conditional and marginal two-dimensional distributions of mRNA and protein. We observe that both approximation errors $||\epsilon||_{\infty}$ and $||\epsilon||_1$ decrease when we increase the order of the moments. For the sake of readability we denote the reconstructed distribution by $\tilde{q}$ in the following tables. For instance, the approximation of the joint marginal distribution of $R$ and $P$ under the

FIG. 6: Approximation of conditional and marginal distribution of mRNA (gene expression).
TABLE V: Two-dimensional distribution reconstruction (gene expression).

| metric M | $q_{D_{off}}$ | $q_{D_{on}}$ | $\hat{q}_{ws\ MCM}$ | $\check{q}_{MM}$ |
|----------|--------------|--------------|------------------|----------------|
| 3        | 0.0123       | 8.4531 $\times 10^{-4}$ | 0.0059           | 0.0063 0.0062 |
| 5        | 0.0112       | 6.2558 $\times 10^{-4}$ | 0.0052           | 0.0059 0.0059 |
| 7        | 0.0059       | 5.0335 $\times 10^{-4}$ | 0.0028           | 0.0025 0.0029 |
| 3        | 59.1054      | 48.8522      | 58.1141          | 62.1209 61.9539 |
| 5        | 53.4683      | 36.1533      | 51.7948          | 58.1554 58.0910 |
| 7        | 28.0351      | 29.1008      | 28.1394          | 24.7305 29.135 |
| 3        | 0.2284       | 0.0581       | 0.1148           | 0.1292 0.1294 |
| 5        | 0.1237       | 0.0345       | 0.0631           | 0.0975 0.0963 |
| 7        | 0.0209       | 0.0243       | 0.0413           | 0.0468 0.0644 |

TABLE VI: Results for protein $P_1$ (exclusive switch).

| metric M | $q_{DNA}$ | $q_{DNA}|P_1$ | $q_{DNA}|P_2$ | $\check{q}_{MM}$ |
|----------|-----------|--------------|--------------|----------------|
| 3        | 0.0199    | 0.0067       | 0.0212       | 0.0116 0.0239 0.0253 |
| 5        | 0.0017    | 0.0023       | 0.0044       | 0.0002 0.0044 0.0041 |
| 7        | 0.0015    | 0.0018       | 0.0027       | 0.0014 0.0017 0.0078 |
| 3        | >100      | 20.3589      | >100         | >100 >100 >100 |
| 5        | 10.239    | 7.4295       | >100         | >100 >100 >100 |
| 7        | 9.1583    | 6.3346       | >100         | >100 >100 >100 |
| 3        | 0.6399    | 0.2592       | 0.3671       | 0.2656 0.5096 0.5886 |
| 5        | 0.0727    | 0.0801       | 0.0354       | 0.0562 0.0798 0.3068 |
| 7        | 0.0699    | 0.0593       | 0.0308       | 0.0421 0.0752 0.1236 |

condition $D_{off} = 1$ is denoted by $\hat{q}_{D_{off}}$.

Exclusive Switch Model. Next we address the accuracy of the reconstruction of conditional and marginal distributions of the exclusive switch model introduced in Sect. 5. In Tables VII and VIII, the approximation errors are listed for the conditional distributions of the proteins where we condition on the three possible states of the promoter, i.e., $DNA = 1$, $DNA.P_1 = 1$ or $DNA.P_2 = 1$.

We observe that the total deviation $||e||_1$ is minimal for both proteins $P_1$ and $P_2$ when the weighted sum MCM approach is applied for all $M \in \{3, 5, 7\}$. Thus, for the exclusive switch system it is advantageous to approximate the marginal distributions by first reconstructing the conditional distributions and computing the weighted sum. In almost all cases the error decreases when more information about the moments is used. Because of the complex bi-modal shape of the distributions it is beneficial to consider higher-order moments. It is important to note also that the large value of $||e||_\infty$ ($||e||_\infty > 100$) comes from the probabilities around the boundary points of the support $(x_L, x_R)$. In the remaining parts of the support $D^*$ the reconstruction is accurate. For example, in Fig. 7 we show the reconstructions of both marginal (right plot) and conditional (left and middle plots) distributions of $P_1$, where weighted sum MCM was used with $M = 5$. The maximum relative error for the reconstruction of the marginal distribution is large ($||e||_\infty > 100$) however the distance in $L_1$ does not differ much from the one obtained when $M = 7$ ($||e||_1 = 0.056$ and $||e||_1 = 0.042$).

We also consider the conditional and marginal two-dimensional distributions of proteins $P_1$ and $P_2$ in Tables VII and VIII. Again we condition on the state of the promoter region, e.g. $q_{DNA|P_1}$ corresponds to the joint distribution of proteins $P_1$ and $P_2$ when $DNA.P_1 = 1$.

We can see that both approximation error metrics in $L_\infty$ and in $L_1$ decrease when more moments are taken into consideration (for instance, in the case of $M = 5$ we have 21 constrains in Eq. (11)). The marginal distribution $P(X_{P_1} = x, X_{P_2} = y)$ is best approximated when the weighted sum MCM approach is applied in terms of both metrics $||e||_1$ and $||e||_\infty$ ($M = 7$). Generally, the MCM approach gives more accurate results, i.e., both weighted sum MCM and joint MCM perform better than MM. We show the reconstructions of three conditional distributions in Fig. 8 shown for the case when $M = 5$, where the plots refer to the conditions $DNA = 1$, $DNA.P_1 = 1$ and $DNA.P_2 = 1$. The reconstruction of the marginal distribu-

**FIG. 7:** Conditional and marginal distributions of $P_1$ (exclusive switch).

TABLE VIII: Two-dimensional conditional protein distributions (exclusive switch).

| metric M | $q_{DNA}$ | $q_{DNA|P_1}$ | $q_{DNA|P_2}$ |
|----------|-----------|--------------|--------------|
| 3        | 3.1408 $\times 10^{-4}$ | 3.0629 $\times 10^{-4}$ | 4.5354 $\times 10^{-4}$ |
| 5        | 3.1623 $\times 10^{-4}$ | 7.4072 $\times 10^{-5}$ | 4.3186 $\times 10^{-4}$ |
| 7        | 4.4949 $\times 10^{-5}$ | 1.1953 $\times 10^{-4}$ | 1.1110 $\times 10^{-4}$ |
| 3        | 69.089     | >100         | 49.3173      |
| 5        | 32.528     | >100         | 47.1483      |
| 7        | 19.7620    | >100         | 12.9727      |
| 3        | 0.8090     | 0.4763       | 0.6014       |
| 5        | 0.2451     | 0.0997       | 0.4606       |
| 7        | 0.1299     | 0.0908       | 0.1001       |

**TABLE VII: Results for protein $P_2$ (exclusive switch).**

| metric M | $q_{DNA}$ | $q_{DNA|P_1}$ | $q_{DNA|P_2}$ | $\check{q}_{MM}$ |
|----------|-----------|--------------|--------------|----------------|
| 3        | 0.0086    | 0.0095       | 0.0083       | 0.0055 0.0107 0.0117 |
| 5        | 0.0017    | 0.0014       | 0.0027       | 0.0014 0.0017 0.0017 |
| 7        | 0.0009    | 0.0008       | 0.0018       | 0.0009 0.0012 0.0013 |
| 3        | >100      | >100         | >100         | >100 >100 >100 |
| 5        | 16.2041   | 14.3546      | 16.3616      | 16.1399 >100 |
| 7        | 9.2318    | 5.6999       | 9.3179       | 9.321 12.1589 17.5795 |
| 3        | 0.7966    | 0.4038       | 0.5169       | 0.398 0.7462 0.7462 |
| 5        | 0.155     | 0.0433       | 0.1352       | 0.083 0.1676 0.166 |
| 7        | 0.1063    | 0.04171      | 0.0917       | 0.0719 0.1022 0.1242 |
TABLE IX: Two-dimensional marginal distributions (exclusive switch).

| metric $\tilde{M}$ | $\tilde{q}_{MCM}$ | $\tilde{q}_{q}$ | $\tilde{q}_{MM}$ |
|-------------------|----------------|----------------|----------------|
| $\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|s...