Moment-Based Methods for Parameter Inference and Experiment Design for Stochastic Biochemical Reaction Networks

JAKOB RUESS and JOHN LYGEROS, ETH Zurich

Continuous-time Markov chains are commonly used in practice for modeling biochemical reaction networks in which the inherent randomness of the molecular interactions cannot be ignored. This has motivated recent research effort into methods for parameter inference and experiment design for such models. The major difficulty is that such methods usually require one to iteratively solve the chemical master equation that governs the time evolution of the probability distribution of the system. This, however, is rarely possible, and even approximation techniques remain limited to relatively small and simple systems. An alternative explored in this article is to base methods on only some low-order moments of the entire probability distribution. We summarize the theory behind such moment-based methods for parameter inference and experiment design and provide new case studies where we investigate their performance.

Categories and Subject Descriptors: G.3 [**Probability and Statistics**]: Markov Processes

General Terms: Algorithms, Theory

Additional Key Words and Phrases: Continuous-time Markov chains, experiment design, Fisher information, moment equations, parameter inference

ACM Reference Format:
Jakob Ruess and John Lygeros. 2015. Moment-based methods for parameter inference and experiment design for stochastic biochemical reaction networks. ACM Trans. Model. Comput. Simul. 25, 2, Article 8 (February 2015), 25 pages.
DOI: http://dx.doi.org/10.1145/2688906

1. INTRODUCTION

Numerous studies have shown that stochasticity is an inherent feature of biochemical reaction networks, especially if the molecule counts of some of the reacting chemical species are low [McAdams and Arkin 1997; Hasty et al. 2000; Samoilov and Arkin 2006]. In some cases, this variability may play an important role; for instance, it may ensure that a certain fraction of the cells commits to one fate and the rest to some other fate [Balazsi et al. 2011]. Further examples where the stochasticity fundamentally shapes the dynamics of the system are stochastic switches [Acar et al. 2008] and systems with noise-induced oscillations [Ko et al. 2010]. In other cases, variability in the system output may be undesirable and is suppressed as much as possible, for instance, through feedback loops [Raser and O'Shea 2005]. Irrespectively of whether the variability shapes the dynamics, has to be suppressed, or is just a by-product, it is generated by the system and carries a signature of the underlying molecular mechanisms. We can thus hope that by investigating the variability of single-cell measurements, we

This work was supported by the European Commission under the Network of Excellence HYCON2.
Authors' addresses: J. Ruess, IST Austria, Am Campus 1, 3400 Klosterneuburg, Austria; email: jruess@ist.ac.at; J. Lygeros, Automatic Control Laboratoy, ETH Zurich, Physikstrasse 3, 8092 Zurich, Switzerland; email: lygeros@control.ee.ethz.ch.

DOI: http://dx.doi.org/10.1145/2688906

can gain knowledge about the system. For instance, we can try to estimate the parameters of continuous-time Markov chain (CTMC) models from the measured distribution of the amount of molecules in a cell population. This is the focus of this article, and in the following sections, we discuss how parameter inference can be performed in this setting.

An example where the measured cell-to-cell variability adds the crucial piece of information to identify all the model parameters is the simplest model of gene expression, which consists only of messenger RNA and protein. If only the average amount of protein in the cell population is measured, it is fundamentally impossible to identify all the reaction rates. If, on the other hand, the variance of the amount of protein is also measured, this nonidentifiability is resolved [Munsky et al. 2009]. For this simple model, this is intuitive and relatively easy to see. For larger models, however, which contain more species and more unknown parameters, it is not as straightforward to determine what is required for the identification of the parameters. This may depend on which species are measured and how the measurement time points are chosen. Moreover, as was shown in Ruess et al. [2013], it is also possible that information about the model parameters can be gained through planned perturbations of the system, which can, for instance, be implemented using a microfluidic device [Uhlendorf et al. 2012]. Consequently, it may be essential to carefully design the experiments that are to be performed to ensure that the resulting measurements contain as much information as possible.

Both parameter inference and experiment design for CTMC models are computationally very difficult [Poovathingal and Gunawan 2010; Lillacci and Khammash 2013]. The main reason is that methods for these tasks usually require one to iteratively solve the chemical master equation (CME), which governs the time evolution of the probability distribution of the system [Gillespie 1992]. Solving the CME analytically is rarely possible, and approximate methods [Munsky and Khammash 2006; Wolf et al. 2010; Mateescu et al. 2010] are often computationally too expensive. An alternative is to resort to hybrid stochastic–deterministic descriptions of the system or to approximate only some low-order moments of the stochastic model [Henzinger et al. 2010; Hasenauer et al. 2014; Hespanha 2008]. Recently, first attempts have been made to use such reduced descriptions of the system for parameter inference [Mikeev and Wolf 2012; Kügler 2012; Zechner et al. 2012] and experiment design [Ruess et al. 2013].

In this article, we give a comprehensive overview of the theory behind moment-based methods for parameter inference and experiment design for single-cell distribution measurements. We start by introducing biochemical reaction networks in Section 2. In Section 3, we show how equations for the moments of the system can be derived in the presence of variability in the parameters. These equations form the basis for computing the likelihood in Section 4 and the Fisher information in Section 5, which are at the heart of parameter inference and experiment design, respectively. We conclude the article by studying three biochemical reaction networks in which the moment dynamics are nonclosed. To the best of our knowledge, moment-based inference for systems with nonclosed moment dynamics has barely been addressed in the literature [Zechner et al. 2012; Gonzalez et al. 2013], while moment-based experiment design has so far only been applied to systems where the moment dynamics are closed [Ruess et al. 2013].

2. BIOCHEMICAL REACTION NETWORKS

A biochemical reaction network consists of a set of m distinct chemical species that can undergo K reactions. Assuming that the system is well mixed, the time evolution of the number of molecules of the species can be modeled by an m-dimensional CTMC $X(t) = [X^1(t) \cdots X^m(t)]^T$, which can take states $x = [x^1 \cdots x^m]^T$ in $\mathbb{N}^m$. Note that we use uppercase letters $X^j(t)$, $j = 1, \ldots, m$ to index the different chemical species in order to reserve lowercase letters for indexing cells later. This should not be confused with exponentiation. Denote by $a_k(x, \theta) \in \mathbb{R}_0^+$, $k = 1, \ldots, K$ the propensities of the K

reactions; that is, $a_k(x,\theta)dt$ is the probability that reaction k takes place in the infinitesimal time interval $[t, t+dt]$ given that $X(t) = x$. We assume that the propensities are given by mass action kinetics with at most bimolecular reactions, which means that $a_k(x,\theta) = \theta_k g_k(x),\ k = 1,\dots,K$, where $\theta_k \in \mathbb{R}_0^+,\ k = 1,\dots,K$ are reaction rate constants and the functions $g_k(x),\ k = 1,\dots,K$ are at most quadratic in x and determined by the model structure. Consequently, the propensities (and also the system dynamics) are parameterized by $\theta = [\theta_1 \cdots \theta_K]^T$. For ease of notation, we will sometimes omit explicitly stating this dependency on θ and just write $a_k(x)$ instead of $a_k(x,\theta)$.

The probability $p(x,t) := P(X(t) = x)$ that $x^i,\ i = 1,\dots,m$ molecules of the m species are present at a certain time point t evolves according to the chemical master equation

$$\dot{p}(x,t) = -p(x,t)\sum_{k=1}^{K} a_k(x) + \sum_{k=1}^{K} p(x-\nu_k,t)a_k(x-\nu_k), \tag{1}$$

where $\nu_k \in \mathbb{Z}^m,\ k = 1,\dots,K$ are the stoichiometric transition vectors of the K reactions (if reaction k takes place, the state is updated by ν_k). According to Equation (1), the time evolution of the entire probability distribution can be computed by solving a system of coupled linear ordinary differential equations with one equation for each state x that can be reached by the CTMC. At a first glance, this may appear to be a simple task. However, for most systems, the number of states that can be reached by the process is very large or even infinite, which often makes computing the time evolution of the probability distribution intractable. In such cases, approximate methods can sometimes be used (see, e.g., Munsky and Khammash [2006], Wolf et al. [2010], and Hjartarson et al. [2013]). The idea of these methods is that the number of states that are likely to be reached by the process may be much smaller than the number of states that can theoretically be reached and thus the state space can be truncated. However, even if the state space is truncated, the number of differential equations still grows exponentially in the number of species, which limits the applicability of such approximate methods to systems that contain only a few species. Throughout this article, we will refer to such methods as finite state projection (FSP)-based methods.

Another way to approximate the time evolution of the probability distribution is by using Gillespie's stochastic simulation algorithm (SSA) [Gillespie 1976]. SSA allows one to draw sample paths from the CTMC. The probability distribution can then be estimated by extracting the statistics of sufficiently many sample paths. The main limitation of this approach is that a large number of sample paths might be needed to obtain accurate approximations of the probability distribution. This often makes SSA computationally too expensive, at least for iterative use in parameter inference and experiment design.

If computing the time evolution of the entire probability distribution of the CTMC is computationally too expensive, an alternative may be to focus only on some low-order moments. Let $\mu_l(t),\ l = 1, 2, \dots$ denote vectors containing the uncentered moments of $X(t)$ of order l, that is,

$$\begin{aligned}
\mu_1(t) &= [\mathbb{E}[X^1(t)] \cdots \mathbb{E}[X^m(t)]],\\
\mu_2(t) &= [\mathbb{E}[X^1(t)X^1(t)] \cdots \mathbb{E}[X^1(t)X^m(t)]\ \ \mathbb{E}[X^2(t)X^2(t)] \cdots \mathbb{E}[X^m(t)X^m(t)]],\\
\mu_3(t) &= [\mathbb{E}[X^1(t)X^1(t)X^1(t)]\ \ \mathbb{E}[X^1(t)X^1(t)X^2(t)] \cdots \mathbb{E}[X^m(t)X^m(t)X^m(t)]],\\
&\ \ \vdots \qquad\qquad\qquad\qquad \vdots \qquad\qquad\qquad\qquad \vdots
\end{aligned}$$

A system that describes the time evolution of the moments can be derived from the CME [Engblom 2006] and written as

$$\dot{\mu}_\infty(t) = \widetilde{A}(\theta)\mu_\infty(t), \tag{2}$$

where $\mu_\infty(t) = [\mu_1(t)\ \mu_2(t)\ldots]^T$ is an infinite vector containing all uncentered moments of the joint probability distribution of all the species and $\widetilde{A}(\theta)$ is an infinite matrix whose entries are (linear) functions of θ that are determined by the model structure. From the system of Equation (2), we can extract the equations for the moments up to some order L that evolve according to

$$\dot{\mu}(t) = A(\theta)\mu(t) + B(\theta)\bar{\mu}(t), \tag{3}$$

where $\mu(t) = [\mu_1(t)\ \mu_2(t)\cdots\mu_L(t)]^T$ is a vector containing the moments of order up to L of the joint probability distribution of all the species and $\bar{\mu}(t)$ is a vector containing moments of higher order. $A(\theta)$ and $B(\theta)$ are submatrices of $\widetilde{A}(\theta)$. $A(\theta)$ contains all rows and columns corresponding to moments of order up to L, whereas $B(\theta)$ contains the terms that determine how the moments of order up to L depend on higher-order moments. Under the assumption that the propensities $a_k(x),\ k = 1, \ldots, K$ are at most quadratic in x, $\bar{\mu}(t)$ contains only a number of moments of order $L+1$ and $\bar{\mu}(t)$ is a subvector of $\mu_{L+1}(t)$. The time evolution of $\mu(t)$ depends on $\bar{\mu}(t)$ and the system of moment equations is nonclosed and cannot be solved, unless $B(\theta)$ consists only of zeros, which is the case if all the propensities are affine in x. To obtain a closed and solvable system, so-called moment closure methods can be used, where an approximate system is obtained by replacing the higher-order moments by some function of the lower-order moments, that is, $\bar{\mu}(t) = f(\mu(t))$ [Singh and Hespanha 2011]. In most moment closure methods, the function f is chosen in accordance with some assumed distribution (for instance, Gaussian or log-normal) for which the higher-order moments are uniquely determined by the lower-order moments [Whittle 1957; Singh and Hespanha 2006], but there also exist other possibilities such as the method presented in Ruess et al. [2011], where $\bar{\mu}(t)$ is estimated from a small number of SSA runs. The major advantage of approximating the moments is that the number of moment equations is only polynomial in the number of species of the reaction network, and thus solving the moment equations may be possible in cases where approximating the entire probability distribution is intractable. The disadvantage of moment closure methods is that it is a priori often hard to say which method will give good approximations for a given system, and for some systems none of the existing methods leads to good approximations. Further background on CTMC models in biology and on different methods for their analysis can be found in the review paper of Goutsias and Jenkinson [2013].

3. EXTRINSIC VARIABILITY AND POPULATION MOMENT DYNAMICS

If the chemical master equation in the form (1) is used to describe the dynamics of a population of cells, it is implicitly assumed that the dynamics in all of the cells are governed by exactly the same CTMC. In reality, such an assumption cannot be expected to hold since even genetically identical cells may differ from each other in terms of cell size, expression capacity, local growth conditions, and so forth [Elowitz et al. 2002; Colman-Lerner et al. 2005; Volfson et al. 2005].

One can introduce such *extrinsic variability* in the form of reaction rates that differ between the cells. We assume here that the differences between the cells do not change within the time course of an experiment such that the reaction rates can be modeled by a constant random vector Z governed by a multidimensional probability distribution P_Z. In principle, it is also possible to consider time-varying extrinsic variability by allowing the reaction rates to be governed by stochastic processes [Shahrezaei et al. 2008; Ruess et al. 2013]. However, the time evolution of the moments of the chemical species then depends on the specific form of these stochastic processes, and no results that are generally valid for any type of time-varying extrinsic variability can be stated.

Under the assumption of constant extrinsic variability, the propensities are (for fixed values of x) also constant random variables that are linear in Z. The time evolution of the probability distribution for a single cell can then be described by a CME that is conditioned on the realization z of Z in the cell:

$$\dot{p}(x,t|z) = -p(x,t|z)\sum_{k=1}^{K} a_k(x,z) + \sum_{k=1}^{K} p(x-\nu_k,t|z)a_k(x-\nu_k,z), \tag{4}$$

where, for cases where only some of the reaction rates vary from cell to cell, the dependence of the propensities on the rates that are not varying is again omitted in the notation.

To compute population statistics from this conditional CME, one could take samples z from P_Z and then solve Equation (4) for all the samples [Toni and Tidor 2013]. It is, however, also possible to directly derive a system of moment equations that allows one to approximately compute the population moments up to some order L without having to sample the extrinsic variables. Based on the results in Zechner et al. [2012], we show how equations for moments up to order 2 can be obtained in the presence of one variable reaction rate (i.e., we assume that Z is one-dimensional). Equations for higher-order moments and multiple variable rates can be derived in the same fashion. Denote by $\mu(t)$ the moments of $X(t)$ up to order 2, by $\mu_Z = [\mathbb{E}[Z]\ \mathbb{E}[Z^2]]^T$ the first two moments of Z, by $\mu_{X,Z}(t)$ the cross-moments of $X(t)$ and Z of order 2, that is,

$$\mu_{X,Z}(t) = [\mathbb{E}[X^1(t)Z]\ \cdots\ \mathbb{E}[X^m(t)Z]]^T,$$

and by $\bar{\mu}(t)$ the vector containing all moments of the joint distribution of $X(t)$ and Z of order 3 or higher. Then we can state the following theorem.

THEOREM 3.1. *Consider a biochemical reaction network modeled by a CTMC $X(t)$ where the dynamics depend on the value of a random reaction rate $Z \sim P_Z$ that is distributed according to P_Z and a vector of reaction rates $\tilde{\theta}$ that are the same for all the cells of the population. Then, the moment equations of $X(t)$ up to order 2 can be written as*

$$\dot{\mu}(t) = A(\tilde{\theta})\mu(t) + B(\tilde{\theta})\mu_{X,Z}(t) + C(\tilde{\theta})\mu_Z + D(\tilde{\theta})\bar{\mu}(t), \tag{5}$$

$$\dot{\mu}_{X,Z}(t) = E(\tilde{\theta})\mu_{X,Z}(t) + F(\tilde{\theta})\mu_Z + G(\tilde{\theta})\bar{\mu}(t), \tag{6}$$

where $\mu(t) = [\mu_1(t)\ \mu_2(t)]^T$ and $\bar{\mu}(t)$ contains moments of the joint distribution of $X(t)$ and Z of order at most 4, which contain at most two powers of Z. The matrices $A(\tilde{\theta})$, $B(\tilde{\theta})$, $C(\tilde{\theta})$, $D(\tilde{\theta})$, $E(\tilde{\theta})$, $F(\tilde{\theta})$, and $G(\tilde{\theta})$ are determined by the model structure.

A proof of this theorem is given in the appendix. We can now invoke some moment closure technique and replace the third- and fourth-order moments by a function of the lower-order moments, that is, $\bar{\mu}(t) = f(\mu(t), \mu_{X,Z}(t), \mu_Z)$. The resulting approximate system of population moments will then depend on the reaction rates that are not varying in the population $\tilde{\theta}$ and moments up to order 2 of Z. Consequently, the model is now parameterized by $\gamma := [\tilde{\theta}, \mu_Z]$ and we can (approximately) compute the moments $\mu(t)$ of $X(t)$ as a function of γ.

4. PARAMETER INFERENCE

4.1. Maximum-Likelihood Estimation

In real applications, it is usually impossible or very hard to measure parameter values such as reaction rates. The goal of parameter inference is then to estimate the unknown parameters from the available data, which typically consists of measurements relating

to the quantities of some of the chemical species. In the following, depending on the context, we will regard the data either as a random variable Y (to stress its random nature) or as a realization y of Y (if we refer to specific measurements that have been collected in an experiment). As an estimator for the model parameters, we consider the maximum-likelihood estimator (MLE) $\widehat{\gamma}$. The MLE can be obtained by solving

$$\widehat{\gamma}(y) = \arg\max_{\gamma}\{p_Y(y,\gamma)\}, \tag{7}$$

where $p_Y(y,\gamma)$ is the likelihood, that is, the probability of measuring y given that γ are the model parameters.

For simplicity, we assume that only one species X^i of the reaction network is measured during some transient dynamics of the system. The time course data we consider is of the form $\{X_1^i(t_s),\ldots,X_n^i(t_s),\ s=1,\ldots,S\}$, where n is the number of measured cells, $t_s,\ s=1,\ldots,S$, are the measurement time points, and all the different measured molecule counts are statistically independent. We will refer to experiments that give such data as single-cell population experiments and have flow cytometry as the currently most common source of such data in mind. In flow cytometry experiments, the independence of all the measured molecule counts stems from the fact that the cells are discarded after being measured and consequently two different measurements can never come from the same cell. Experimental techniques such as time-lapse microscopy that track individual cells and lead to measurements that are correlated in time are not considered in this article. Identifying parameters from such data requires one to evaluate path likelihoods of $X(t)$ [Andreychenko et al. 2011; Zechner et al. 2014], which cannot be obtained from the moment equations unless further assumptions are imposed [Komorowski et al. 2011]. In our case, owing to the independence of all the measured samples, the likelihood factorizes, that is,

$$\begin{aligned} p_Y(y,\gamma) &= \prod_{s=1}^{S}\prod_{j=1}^{n} P\big(X_j^i(t_s) = x_j^i(t_s)|\gamma\big) \\ &= \prod_{s=1}^{S}\prod_{j=1}^{n} p^i\big(x_j^i(t_s), t_s|\gamma\big), \end{aligned}$$

where $p^i(\cdot,t_s|\gamma)$ is the distribution of species i at time point t_s, given that γ are the model parameters. For models where no extrinsic variability is present, $p^i(\cdot,t_s|\gamma)$ can be computed by solving the CME. This makes computing the MLE theoretically possible. Practically, however, approximating the solution of the CME to sufficient accuracy may be computationally too expensive for solving Equation (7) in reasonable time. Further, if extrinsic variability is present in the model, the only way to evaluate the likelihood is through extensive simulation using SSA, and for most systems this will not lead to a tractable optimization problem in Equation (7).

4.2. Moment-Based Inference

A more tractable optimization problem can be obtained if moments of the measured samples are taken as measurements instead of the entire samples. Throughout this article, we take only two moments, namely, sample mean and sample variance, as our measurements. Higher-order moments could be included in a similar way, but evaluating the cost function of the resulting optimization problem would be computationally more expensive.

Let $S_M := \{S_M(t_1),\ldots,S_M(t_S)\}$, $S_M(t_s) := \frac{1}{n}\sum_{j=1}^{n} X_j^i(t_s)$, $s=1,\ldots,S$ be the sample means and $S_V := \{S_V(t_1),\ldots,S_V(t_S)\}$, $S_V(t_s) := \frac{1}{n-1}\sum_{j=1}^{n}(X_j^i(t_s)^2 - S_M(t_s)^2)$, $s=1,\ldots,S$

be the sample variances at the different measurement time points. Then, we can consider all the sample means and variances $S_J := \{S_J(t_1), \ldots, S_J(t_S)\}$, where $S_J(t_s) = [S_M(t_s)\ \ S_V(t_s)]^T$, $s = 1, \ldots, S$, as our new data Y.

THEOREM 4.1. *Let $s \in \{1, \ldots, S\}$. Given model parameters γ, denote by $\mu_1^i(t_s)$ the mean and by $\mu_l^i(t_s),\ l = 2, \ldots, 4$ the centered moments of order l of $p^i(\cdot, t_s|\gamma)$, the distribution of the ith species at time t_s. If the size of the measured sample n is sufficiently large, the distribution of $S_J(t_s)$ is approximately a normal distribution $p_J(\cdot, t_s|\gamma) = \mathcal{N}(\mu(t_s), \Sigma(t_s))$, where*

$$\mu(t_s) = [\mu_1^i(t_s)\ \mu_2^i(t_s)]^T \quad \textit{and} \tag{8}$$

$$\Sigma(t_s) = \frac{1}{n}\begin{bmatrix} \mu_2^i(t_s) & \mu_3^i(t_s) \\ \mu_3^i(t_s) & \mu_4^i(t_s) - \frac{n-3}{n-1}\left(\mu_2^i(t_s)\right)^2 \end{bmatrix}. \tag{9}$$

A proof of this theorem is given in the appendix.

Remark 4.2. Note that we have switched to using mean and centered moments of order $2, \ldots, 4$. Previously, we used uncentered moments since these are more convenient for deriving the moment equations. The centered moments up to any order L are uniquely determined by the uncentered moments up to order L, and we can therefore solve the moment equations for the uncentered moments and subsequently convert them into centered moments.

Remark 4.3. A sufficiently large sample size n is required, because then approximate normality of $S_J(t_s)$ follows from the central limit theorem. How large the sample size needs to be in a specific situation depends on how irregular the distribution $p^i(\cdot, t_s|\gamma)$ is. In practice, a surprisingly small sample size is often already sufficient (in the order of 10 to at most 30 samples for the distribution of $S_M(t_s)$, possibly some more for $S_J(t_s)$). It can be safely assumed that flow cytometry, which can provide measurements of thousands of cells or even more, will always lead to large enough sample sizes. In general, if this assumption is in question, it can be verified by first determining the distribution of the sample moments using bootstrapping and then testing it for normality (see, for instance, Zechner et al. [2012]). The hidden assumption that the same number of cells n is measured at each time point only serves to simplify the notation and can directly be replaced by a more realistic scenario where the number of measured cells differs between the time points.

COROLLARY 4.4. *Taking S_J as the data, the likelihood is a product of two-variate normal distributions that are parameterized by the first four moments of the underlying process:*

$$\begin{aligned} p_Y(y, \gamma) &= \prod_{s=1}^{S} P(S_J(t_s) = s_J(t_s) \mid \gamma) \\ &= \prod_{s=1}^{S} p_J(s_J(t_s), t_s \mid \gamma), \end{aligned}$$

where $p_J(\cdot, t_s|\gamma) = \mathcal{N}(\mu(t_s), \Sigma(t_s))$ is given by Theorem 4.1.

PROOF. This is a direct consequence of the independence of the samples at different time points and Theorem 4.1. □

This likelihood can be computed from the first four moments, which can be approximated through moment closure from the moment equations. Consequently, we can

Table I. Advantages and Disadvantages of Different Approaches

	Moment Based	FSP Based	SSA Based
Computational cost	low	varies	high
Scalability in species number	high	low	high
Scalability in molecule numbers	high	low	low
Complexity	high	medium	low
Tolerates extrinsic variability	yes	no	yes
Tolerates non-mass-action kinetics	varies	yes	yes

maximize the likelihood and compute the MLE without having to compute the entire solution of the CME. The downside of this approach is that the higher-order moments of the data are not used for the inference, leading to information loss (unless sample means and variances form jointly sufficient statistics). Further, moment closure methods are approximate methods that do not give error guarantees. Consequently, the computed MLE will contain errors of unknown size and should be validated a posteriori. This can, for instance, be done by stochastically simulating the system with the inferred maximum-likelihood parameters. The simulation results can then be compared, on the one hand, to moment closure approximations to evaluate the accuracy of the used moment closure method and, on the other hand, to the measured data to check whether the inferred parameters are really in agreement with the measured data.

4.3. Comparison to Other Methods

As already discussed, a number of methods for parameter inference have been proposed lately. None of these methods is universally applicable, and which method is the most suitable for a given situation strongly depends on the considered reaction network. The aim of this section is to give a brief qualitative comparison of the strengths and weaknesses of approaches that are based on stochastic simulation (SSA) [Poovathingal and Gunawan 2010; Lillacci and Khammash 2013], finite state projection (FSP) [Munsky et al. 2009; Andreychenko et al. 2011], and the moment equations [Zechner et al. 2012; Kügler 2012], respectively.

The general advantages and disadvantages of the different approaches are summarized in Table I. The main advantage of approaches based on SSA is that they are always applicable and straightforward to use. The drawback is that this generality comes at the price of a possibly prohibitively large computational cost. This tends to be especially severe in cases where some species are present in large amounts such that the reaction propensities are large and simulating a trajectory of the system is expensive because many reactions take place. If in addition to this some other species are present in low amounts such that rare events are important and SSA estimates converge slowly, SSA-based approaches are often infeasible with the currently available computational power. Approaches based on FSP are generally very useful for small systems but scale badly in both the number of species and the amounts of molecules of each individual species that are present in the system. Besides this, FSP can only be used if all random variables that play a role for the system have a discrete state space. This means that FSP-based approaches are not applicable if extrinsic variability is included in the form of continuous distributions on the reaction rates. Consequently, such approaches can only be used in a very limited number of cases. The main advantage of moment-based approaches is that they scale very well in the number of molecules (the computational cost does not depend at all on how large the molecule counts are) and reasonably well in the number of different chemical species. Besides this, there are cases where the moments can be computed exactly but only approximations or estimates of the entire distribution are available (for instance, the example considered

in Section 6.3). Clearly, in such cases moment-based approaches should be used. The disadvantages of moment-based methods are that often a lot of expertise is required to derive and solve the moment equations and that in some cases moment closure will fail to provide good approximations. Furthermore, contrary to approaches based on FSP or SSA, it is often difficult to determine a priori whether a moment-based approach will be successful.

5. EXPERIMENT DESIGN

5.1. The Fisher Information

The first thing we need to design experiments is a way of judging the quality of an experiment. If we are interested in estimating the model parameters γ, a good experiment should be one that allows us to estimate γ to high precision. Regarding the data as a random variable Y, estimators of the parameters are also random variables and the quality of the estimators can be characterized in terms of their variances. The Cramér-Rao inequality tells us that the variance of no unbiased estimator can be smaller than the inverse of the Fisher information. Consequently, the Fisher information determines the precision to which the model parameters can at best be estimated in a given experimental setting and therefore provides a way to judge the quality of an experiment. When there is more than one unknown parameter, the Fisher information is a positive semidefinite matrix whose size is equal to the number of unknown parameters and its elements are given by

$$(I_Y(\gamma))_{k,l} = \mathbb{E}\left[\left(\frac{\partial}{\partial \gamma_k}\log p_Y(y;\gamma)\right)\left(\frac{\partial}{\partial \gamma_l}\log p_Y(y;\gamma)\right)\right]. \tag{10}$$

The expectation in Equation (10) is taken over all possible realizations of the data and consequently $I_Y(\gamma)$ does not depend on any measurements and can be used to evaluate the quality of different experiments before they are performed. However, $I_Y(\gamma)$ is a local measure and depends on the true parameters γ, which are unknown. In practice, the Fisher information is often evaluated at some initial estimates of the parameters in the hope that this will lead to an information matrix that is close to the true one [Hagen et al. 2013]. Alternatively, it was proposed to evaluate $I_Y(\gamma)$ at several values, sampled from a prior distribution on the parameters, and to take averages of the results [Pronzato and Walter 1985; Ruess et al. 2013]. While such an approach can be expected to be more robust with respect to uncertainty in the parameters, it leads to a large increase in the computational cost. Here, we evaluate the Fisher information at some fixed initial estimates without taking the uncertainty of these estimates into account. In the following, we show how $I_Y(\gamma)$ can be computed if either the entire sample or the sample moments are taken as measurements. As in the previous section, we assume for simplicity that only species X^i is measured in n cells.

THEOREM 5.1. *Consider a biochemical reaction network modeled by a CTMC that is parameterized by* $\gamma = [\gamma_1 \cdots \gamma_R]^T$. *Assume that data* $Y = \{Y(t_s),\ s = 1, \ldots, S\}$, *where* $Y(t_s) = \{X_1^i(t_s), \ldots, X_n^i(t_s)\},\ s = 1, \ldots, S$ *can be obtained from an experiment. Then, the Fisher information is given by*

$$I_Y(\gamma) = \sum_{s=1}^{S} I_{Y(t_s)}(\gamma), \quad \textit{where} \tag{11}$$

$$\left(I_{Y(t_s)}(\gamma)\right)_{k,l} = n \sum_{\{x^i \in \mathbb{N} \mid p^i(x^i,t_s|\gamma)\neq 0\}} \frac{\frac{\partial p^i(x^i,t_s|\gamma)}{\partial \gamma_k}\frac{\partial p^i(x^i,t_s|\gamma)}{\partial \gamma_l}}{p^i(x^i,t_s|\gamma)}. \tag{12}$$

PROOF. The first equality follows from the fact that the samples at different time points are independent, because the joint Fisher information matrix of independent random variables is equal to the sum of the Fisher information matrices of the individual random variables. The second equation follows directly from the definition of the Fisher information and the discrete nature of the likelihood:

$$\begin{aligned}\left(I_{X_1^i(t_s)}(\gamma)\right)_{k,l} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\gamma_k}\log p^i(x^i,t_s|\gamma)\right)\left(\frac{\partial}{\partial\gamma_l}\log p^i(x^i,t_s|\gamma)\right)\right]\\ &= \sum_{\{x^i\in\mathbb{N}|p^i(x^i,t_s|\gamma)\neq 0\}} p^i(x^i,t_s|\gamma)\cdot\frac{1}{p^i(x^i,t_s|\gamma)^2}\frac{\partial p^i(x^i,t_s|\gamma)}{\partial\gamma_k}\cdot\frac{\partial p^i(x^i,t_s|\gamma)}{\partial\gamma_l}\\ &= \sum_{\{x^i\in\mathbb{N}|p^i(x^i,t_s|\gamma)\neq 0\}} \frac{\frac{\partial p^i(x^i,t_s|\gamma)}{\partial\gamma_k}\frac{\partial p^i(x^i,t_s|\gamma)}{\partial\gamma_l}}{p^i(x^i,t_s|\gamma)},\end{aligned}$$

where $I_{X_1^i(t_s)}(\gamma)$ is the Fisher information of $X_1^i(t_s)$. The theorem follows because $X_1^i(t_s),\ldots,X_n^i(t_s)$ are i.i.d. and therefore $I_{Y(t_s)}(\gamma)=n\cdot I_{X_1^i(t_s)}(\gamma)$. □

THEOREM 5.2. *Consider a biochemical reaction network modeled by a CTMC that is parameterized by $\gamma=[\gamma_1\cdots\gamma_R]^T$. Assume that data $Y=S_J=\{S_J(t_1),\ldots,S_J(t_S)\}$ consisting of the first two moments of the samples $Y(t_s),\ s=1,\ldots,S$ can be obtained from an experiment with a sufficiently large number n of measured cells. Then, the Fisher information is up to a term that is constant in n given by*

$$I_Y(\gamma)=\sum_{s=1}^{S} I_{S_J(t_s)}(\gamma),\quad \textit{where} \tag{13}$$

$$\begin{aligned}&\left(I_{S_J(t_s)}(\gamma)\right)_{k,l}\\ &\quad\approx n\frac{\frac{\partial\mu_1^i(t_s)}{\partial\gamma_k}\frac{\partial\mu_1^i(t_s)}{\partial\gamma_l}}{\mu_2^i(t_s)}+n\frac{\left(\mu_2^i(t_s)\frac{\partial\mu_2^i(t_s)}{\partial\gamma_k}-\frac{\partial\mu_1^i(t_s)}{\partial\gamma_k}\mu_3^i(t_s)\right)\left(\mu_2^i(t_s)\frac{\partial\mu_2^i(t_s)}{\partial\gamma_l}-\frac{\partial\mu_1^i(t_s)}{\partial\gamma_l}\mu_3^i(t_s)\right)}{\left(\mu_2^i(t_s)\right)^2\left(\mu_4^i(t_s)-\left(\mu_2^i(t_s)\right)^2\right)-\mu_2^i(t_s)\left(\mu_3^i(t_s)\right)^2}.\end{aligned} \tag{14}$$

PROOF. The first equality in Theorem 5.2 again follows from the independence of the measurements at different time points. To show the second approximate equality, first note that $S_J(t_s)$ has approximately a Gaussian distribution $S_J(t_s)\sim\mathcal{N}(\mu(t_s),\Sigma(t_s))$ according to Theorem 4.1. We can then make use of the fact that for Gaussian random variables, the expression for the Fisher information simplifies [Komorowski et al. 2011] and we obtain

$$\left(I_{S_J(t_s)}(\gamma)\right)_{k,l}=\frac{\partial\mu(t_s)^T}{\partial\gamma_k}\Sigma(t_s)^{-1}\frac{\partial\mu(t_s)}{\partial\gamma_l}+\frac{1}{2}\mathrm{tr}\left(\Sigma(t_s)^{-1}\frac{\partial\Sigma(t_s)}{\partial\gamma_k}\Sigma(t_s)^{-1}\frac{\partial\Sigma(t_s)}{\partial\gamma_l}\right). \tag{15}$$

Substituting Equations (8) and (9) into Equation (15), we obtain after some basic calculations that the first term in Equation (15) is given by

$$\begin{aligned}&\frac{\partial\mu(t_s)^T}{\partial\gamma_k}\Sigma(t_s)^{-1}\frac{\partial\mu(t_s)}{\partial\gamma_l}\\ &\quad\approx n\frac{\frac{\partial\mu_1^i(t_s)}{\partial\gamma_k}\frac{\partial\mu_1^i(t_s)}{\partial\gamma_l}}{\mu_2^i(t_s)}+n\frac{\left(\mu_2^i(t_s)\frac{\partial\mu_2^i(t_s)}{\partial\gamma_k}-\frac{\partial\mu_1^i(t_s)}{\partial\gamma_k}\mu_3^i(t_s)\right)\left(\mu_2^i(t_s)\frac{\partial\mu_2^i(t_s)}{\partial\gamma_l}-\frac{\partial\mu_1^i(t_s)}{\partial\gamma_l}\mu_3^i(t_s)\right)}{\left(\mu_2^i(t_s)\right)^2\left(\mu_4^i(t_s)-\left(\mu_2^i(t_s)\right)^2\right)-\mu_2^i(t_s)\left(\mu_3^i(t_s)\right)^2}.\end{aligned}$$

The theorem follows because the second term is approximately constant in the sample size n and thus the information is dominated by the first term for sufficiently large n. □

5.2. Optimal Design

With the formulae of the previous subsection, we can compute the Fisher information matrix for given experiments before they are performed. If we consider a set of possible experiments $\mathcal{E}$, the goal of experiment design is to find the most informative experiment $e^* \in \mathcal{E}$. An experiment in $\mathcal{E}$ could, for instance, involve selecting a set of measurement times as well as one of many possible perturbations that can be applied to a cell culture (e.g., nutrient shifts [Menolascina et al. 2011], stress signals [Uhlendorf et al. 2012], or light [Milias-Argeitis et al. 2011]). To optimize the informativeness of the experiment, we first need to summarize the information encoded in the Fisher information matrix in a scalar. There are multiple ways in which this can be done [Franceschini and Macchietto 2008]. Arguably, the most common is to take the determinant, giving rise to so-called D-optimality. The determinant can be thought of as a measure of the inverse of the volume of a confidence region for the parameter estimates. Other possibilities are the trace of the inverse of the Fisher information (A-optimality corresponding to the sum of the variances of the marginal distributions of the parameter estimators) or the minimal eigenvalue (E-optimality corresponding to the information about the least identifiable parameter combination). Also of interest for biological application may be D_s-optimality where the parameter vector can be divided into parameters of interest and nuisance parameters [Hunter et al. 1969; Walter and Pronzato 1990]. Here, we use the determinant and aim at solving the optimization problem

$$e^* = \arg\max_{e \in \mathcal{E}} \{\det I(\gamma, e)\}. \tag{16}$$

Computing the information of the entire sample using Theorem 5.1 requires a marginal of the solution of the CME and its derivatives with respect to the model parameters. This makes evaluating the cost function in the experiment design optimization problem (16) even harder than computing the likelihood in Equation (7). If we use Theorem 5.2 instead to compute the information, we need to compute the moments up to order 4 and the parameter sensitivities of the moments up to order 2 to evaluate the cost function. These can again be approximated using moment closure and some solver for ordinary differential equations that also returns sensitivities such as CVODES of the SUNDIALS toolbox [Hindmarsh et al. 2005]. How well sensitivities of the moments can be approximated with moment closure methods has not yet been studied much in the literature. First attempts have been made in Ale et al. [2013]. We will study some further cases in the next section.

6. CASE STUDIES

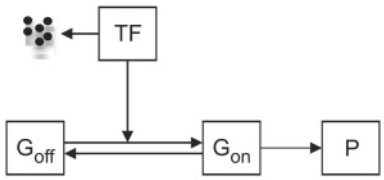
In this section, we demonstrate the use of the theory outlined previously in three case studies. The goal of the first case study is to test how well sensitivities can be approximated using moment closure and to investigate how approximation errors in the moments and the sensitivities propagate to the computation of the Fisher information. The second case study is intended to demonstrate how moment-based inference can be practically implemented in a situation that is not directly compliant with moment closure techniques. Finally, as a third case study, we consider an example where the initial distribution is unknown and that also includes extrinsic variability.

6.1. A Model of Transient Gene Expression

As a first case study, we consider a model of transient gene expression, which was studied in Zechner et al. [2012] and Ruess and Lygeros [2013]. The moment dynamics

of this system are nonclosed and moment closure is required to approximate the information. We choose the parameters in this example such that it is possible to accurately approximate the entire probability distribution using finite state projection [Munsky and Khammash 2006]. Consequently, the true information can be computed up to a small error and we can evaluate the quality of our moment-closure-based approximate results. In Zechner et al. [2012], moment closure was used for parameter inference for this model, whereas in Ruess and Lygeros [2013], it was shown that the first two moments contain sufficient information for identifying all the parameters. Here, we investigate how much information can be gained by optimally placing the measurement times. The model consists of four species, a transcription factor TF, the gene G_{off}, the gene with bound transcription factor G_{on}, and the produced protein P and the following reactions:

$$
\begin{aligned}
TF &\xrightarrow{a} \emptyset \\
TF + G_{off} &\xrightarrow{b} G_{on} \\
G_{on} &\xrightarrow{c} TF + G_{off} \\
G_{on} &\xrightarrow{d} G_{on} + P
\end{aligned}
$$



As parameters, we take the same values as in Ruess and Lygeros [2013], that is, $a = 0.15, b = 0.008, c = 0.01$, and $d = 0.4$. We assume that initially the gene is in the off state, that is, $G_{off} = 1, G_{on} = 0$, and that no protein molecules and 20 TF-molecules are present. Further, we assume that the amount of protein can be measured by flow cytometry at several time points, which can be either equally spaced or placed optimally as part of the experiment design; specific examples are provided later. In Figure 1(A) we see that, even though the moment equations are not closed, second-order low-dispersion moment closure (setting third-centered moments to zero) approximates the time evolution of the moments up to order 2 rather well. Figure 1(B), however, shows that two of the four sensitivities of the variance are not approximated well. On the contrary, if we use a fourth-order closure and set the fifth-centered moments to zero, the moments up to order 4 and the sensitivities of the moments up to order 2 are all approximated very well. This shows that if sensitivities are to be approximated in addition to moments, a higher-order closure might be required. Note that to compute the information using Theorem 5.2, we need the moments up to order 4. Consequently, we would anyway have to use a fourth-order closure, unless we impose the additional assumption that the marginal distribution of the measured species has some form for which the third and fourth moments are uniquely determined by mean and variance.

In Figure 1(C) and 1(D) 95% confidence regions for pairs of parameters computed from the approximated information matrix are compared to the true ones. Thereby, we assumed that measurements are taken at equally spaced time points $t = 25, t = 50, t = 75$, and $t = 100$. The confidence regions computed with the fourth-order closure (red) agree very well with the true ones (blue), while the second-order closure coupled with the additional assumption that the measured protein has a Gaussian distribution leads to somewhat inaccurate results (green). To test whether this inaccuracy is a result of the additional Gaussian assumption or of the poorly approximated sensitivities, we also consider a fictitious scenario where we approximate the sensitivities with the second-order closure but assume that the correct third and fourth moments of the protein are

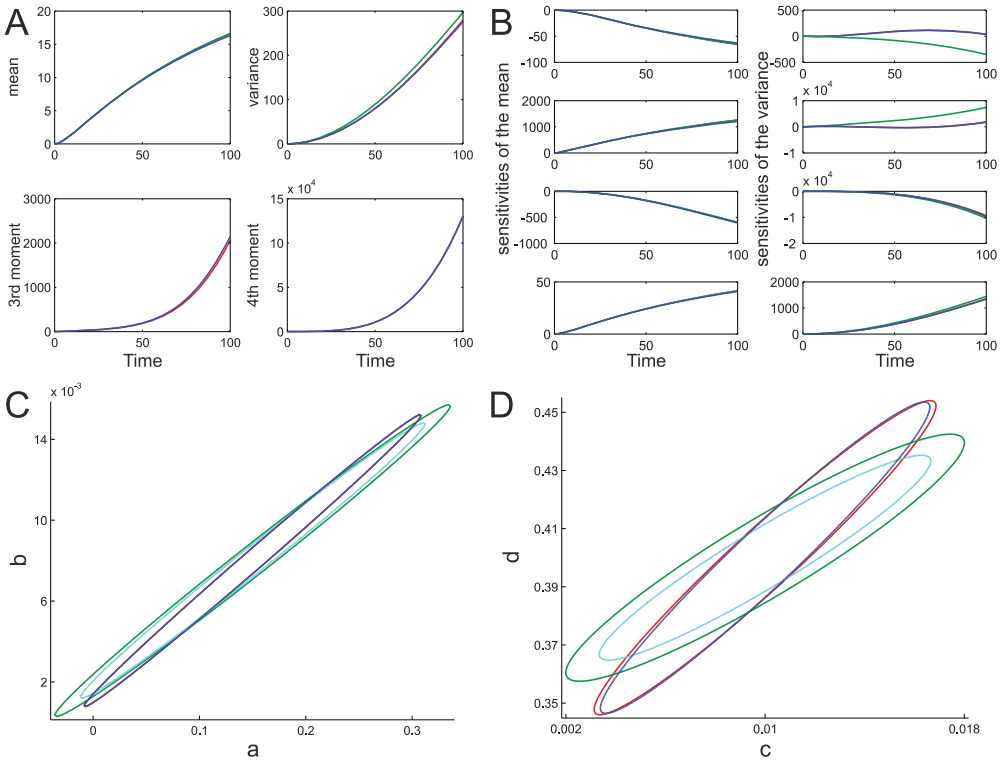


Fig. 1. (A) Mean and centered moments of the protein up to order 2 approximated with second-order low-dispersion moment closure (green) and up to order 4 approximated with fourth-order low-dispersion moment closure (red) are compared to moments computed from the solution of the CME, which was approximated using finite state projection with a very large truncation (blue). In cases where certain colors are not visible, the approximations agree perfectly with the solution obtained by finite state projection. (B) Approximated sensitivities of the protein mean and variance with the same color coding as above. (C,D) Approximated 95% confidence regions for pairs of parameters with the same color coding as above. The confidence regions corresponding to the second-order closure were computed, on the one hand, with the assumption that the measured protein has a Gaussian distribution (green) and, on the other hand, with the assumption that the true third and fourth moments are known (cyan).

available. The results (cyan) suggest that the wrong orientation of the confidence regions stems from poorly approximated sensitivities, whereas the Gaussian assumption influences the size of the confidence regions.

To investigate how much information can be gained by placing the measurements at optimal time points, we proceed as in Ruess et al. [2013]. Starting with one measurement fixed at the final time of the experiment $t = 100$, we grid the time axis and sequentially place the next measurement at the time point where it yields maximal information under the constraint that it is at least five time units apart from all the previous measurements. Figure 2 shows the resulting measurement times and the decrease of the confidence regions if the optimal measurement times are used instead of equally spaced measurement times. It can be seen that the previously rather large confidence regions shrink significantly. The determinant of the Fisher information increases by a factor of approximately 5.7 when the measurement times are placed using this greedy heuristic.

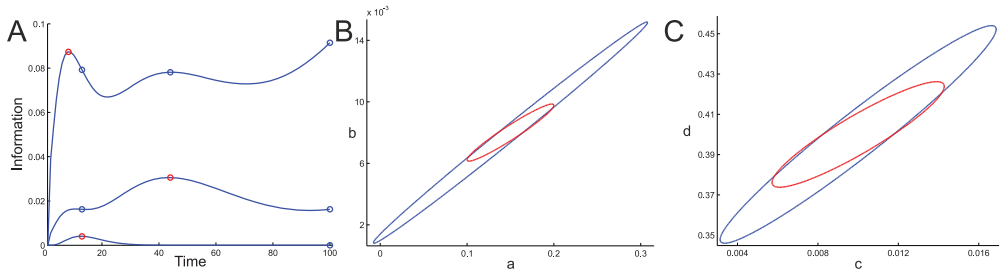
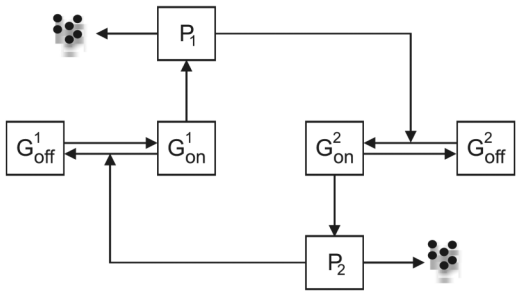


Fig. 2. (A) Optimal placement of the measurement times at N time points. The different lines correspond to $N = 2, 3, 4$ measurements. On each line, the blue circles correspond to the $N - 1$ measurement time points that were already placed, and the red circle corresponds to the best choice for the Nth measurement time point. (B,C) 95% confidence regions for pairs of parameters resulting from equally spaced (blue) and greedily placed (red) measurement times.

6.2. A Two-Gene Network

As a second case study, we consider a system of two interacting genes that constitute a negative feedback loop. The reactions of the system are given as follows:

$$
\begin{aligned}
G^1_{off} &\xrightarrow{c_1} G^1_{on} \\
G^1_{on} + P_2 &\xrightarrow{c_2} G^1_{off} + P_2 \\
G^1_{on} &\xrightarrow{c_3} G^1_{on} + P_1 \\
P_1 &\xrightarrow{d} \emptyset \\
G^2_{off} + P_1 &\xrightarrow{c_4} G^2_{on} + P_1 \\
G^2_{on} &\xrightarrow{c_5} G^2_{off} \\
G^2_{on} &\xrightarrow{c_6} G^2_{on} + P_2 \\
P_2 &\xrightarrow{d} \emptyset
\end{aligned}
$$



The two genes G_1 and G_2 can be either in a state where protein is produced, G^1_{on} and G^2_{on}, or in a state where no protein production is possible, G^1_{off} and G^2_{off}. The produced proteins P_1 and P_2 mediate the switching of the genes between the on and off states such that G_2 can only be switched on if P_1 is present and G_1 can only be switched off if P_2 is present. We assume that initially the genes are switched off and that no protein molecules are present. This leads to dynamics where at first the gene G_1 is switched on. The following production of P_1 enables gene G_2 to switch on, resulting in production of P_2, which leads to the gene G_1 being switched off again. We assume that the parameters are given by $c_1 = 0.1, c_2 = 0.01, c_3 = 10, c_4 = 0.005, c_5 = 0.1, c_6 = 10$ and that the degradation rates are known and equal to $d = 0.1$. This would be the case, for example, if both proteins are very stable and degradation is determined purely by the known dilution rate. The remaining parameters are considered to be unknown and need to be estimated from measurements. Figure 3(A) shows an example of a possible trajectory of the system for the parameter values given earlier.

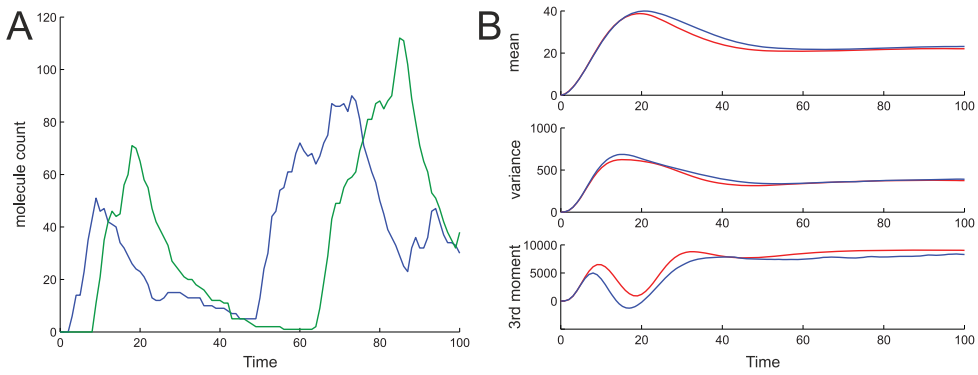


Fig. 3. (A) Trajectories of the two proteins P_1 (blue) and P_2 (green) for a single realization of the system. (B) First three moments of protein P_1 computed using 50,000 SSA trajectories (blue) and moment closure (red) where a third-order derivative matching closure was used.

The first thing to note about this system is that there are conservation laws. Since each of the genes can only be either on or off $G^i_{off} + G^i_{on} = 1,\ i = 1, 2$ at all times. Consequently, the system can be simplified such that only one of the gene states is left as a species for each of the genes. Accordingly, we eliminate the species $G^i_{off},\ i = 1, 2$ from the model. The next thing to note is that the remaining gene states $G^i_{on},\ i = 1, 2$ are Bernoulli distributed for all time points (with time varying parameters) and hence $(G^i_{on})^2 = G^i_{on},\ i = 1, 2$. This means that the moment equations can be simplified by replacing all the moments, which include squares or higher powers of one of the gene states by lower-order moments. Note, however, that this does not eliminate all the higher-order cross-moments and the resulting system of moment equations will still be nonclosed. Consequently, to perform moment-based inference or experiment design, we need to choose a closure method. It is a priori unclear which of the existing closure methods is most suitable. At first glance, unfortunately, most common closure methods appear to be rather ill-suited for this problem. A Gaussian closure might fail because the molecule counts of all the species can become very low and thus the distributions can become highly skewed. The use of derivative matching is known to be problematic if molecule counts are likely to become zero [Hespanha 2008].

To test whether some closure method can nevertheless be applied, we randomly sampled five parameter sets from an assumed prior distribution and evaluated zero-cumulant, low-dispersion, and derivative matching moment closure (all up to fourth order and implemented with the StochDyn toolbox [Hespanha 2006]) by comparing the approximated moments to results from 10,000 SSA runs. For derivative matching, since it involves divisions by moments that are initially zero, we set all the initial moments to very small values. For all closure orders, zero-cumulant and low-dispersion moment closure produced good approximations for only one (and sometimes a second) of the parameter sets and either lead to inaccurate results or even unstable systems for the remaining parameter sets. Derivative matching provided the best approximations and the third-order closure resulted in reasonably good approximations for all parameter sets. Fourth order derivative matching was only applicable if the initial moments were set to values far from zero. This, however, resulted in a large approximation error in the early transient dynamics. Consequently, we decided to rely on third-order derivative matching. Figure 3 shows that this leads to good approximations for the true parameters that were not part of the tested parameter sets.

Since the third-order closure does not return fourth-order moments, we need a further assumption if we want to compute the information according to Theorem 5.2 (for

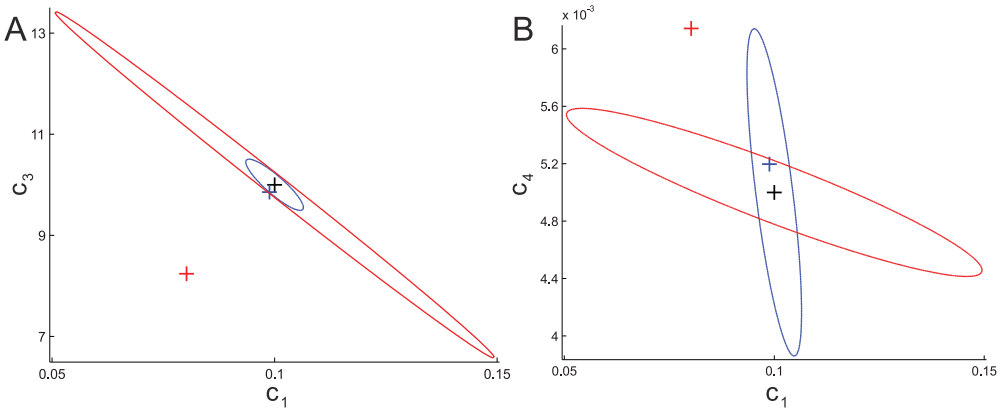


Fig. 4. 95% confidence regions, computed using the true parameters, for the pairs $c1 - c3$ (A) and $c1 - c4$ (B) for the scenarios where time course data of P_1 and the stationary mean of P_2 (blue) and time course data of P_2 and the stationary mean of P_1 (red) are measured. The crosses show the true parameters (black) and the parameters that were estimated from simulated data using moment-based inference (blue and red).

performing moment-based inference, no further assumption is required if the third and fourth moments of the measured species are computed from the data as in Zechner et al. [2012]). Specifically, we require the fourth moments of the marginal distributions of the species P_1 and P_2 because these will in the following be the measured species. We verified on the test parameter sets that the assumption that the marginal distributions of the proteins are such that $\mu_4^i = 3(\mu_2^i)^2$ leads to reasonable agreement with the simulation results. Note that the moments of the marginal distributions of P_1 and P_2 are not required to close the moment equations and hence this is an additional assumption that is independent of the closure assumption.

Having chosen the closure method, we can start with moment-based inference. We assume that the amount of either P_1 or P_2 can be measured in 10,000 cells by flow cytometry at the fixed times $t_1 = 10, \ldots, t_{10} = 100$. According to the information computed using Theorem 5.2, neither means and variances of P_1 nor means and variances of P_2 alone provide sufficient information to accurately estimate all the parameters. On the contrary, a scenario where time course data of P_1 is measured and subsequently only the average amount of P_2 is quantified at the end of the experiment leads to satisfactory results (see Figure 4).

To verify these results, we simulated data using stochastic simulation and performed moment-based inference for the different measurement scenarios. We randomly initialized the Markov chain Monte Carlo (MCMC) algorithm, which was used in Zechner et al. [2012], at 100 different starting points and recorded the parameters that attained the highest likelihood in each run. For the scenarios where only one of the proteins is measured, the resulting parameters differed significantly between the runs. This appears to be because for these scenarios, the likelihood seems to have a multitude of separated local maxima, which all lead to similar agreement of model predictions and data. Furthermore, the parameters that lead to the highest likelihood among all the 100 runs differed significantly from the true parameters. On the contrary, if in addition to measurements of the time course of one of the proteins the mean of the other protein at the end of the experiment is measured, the situation improves significantly, especially if time course data of P_1 and the steady-state mean of P_2 are measured. In this case, the MCMC algorithm consistently converges to parameters close to the true ones, which indicates that only a small region of high likelihood around the true parameters is left. This is in agreement with the predicted information

Table II. Estimated Parameters for the Different Measurement Scenarios

Measured time course of: and stationary mean of:	P_1 -	P_2 -	P_1 P_2	P_2 P_1	True Values
c_1	0.0947	0.1427	0.0988	0.0802	0.1
c_2	0.0028	0.0061	0.0083	0.0052	0.01
c_3	10.0632	41.4386	9.8558	8.2390	10
c_4	0.0070	0.0007	0.0052	0.0061	0.005
c_5	0.0633	0.1320	0.0541	0.1623	0.1
c_6	19.3029	11.2129	7.4606	11.7167	10

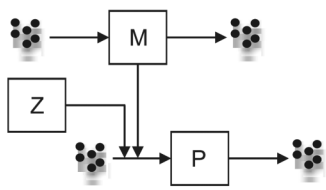
values and the tight confidence regions shown in Figure 4. It should be noted, however, that only some of the parameters that attained the highest likelihood among the 100 MCMC runs (shown in Table II and as crosses in Figure 4) fall into the 95% confidence regions in Figure 4. This stems from the approximation error of moment closure, which leads, on the one hand, to only approximate confidence regions and, on the other hand, to the result that, even in a noise-free scenario, the true parameters do not coincide with the maximum of the approximated likelihood. Strictly speaking, we should also note that the confidence regions correspond to a Gaussian distribution with the inverse of the Fisher information as covariance matrix, which is, for the finite sample sizes considered here, not necessarily the true distribution of the estimators.

To summarize this example, we can say that successful identification of the model parameters using the moments is possible but requires careful a priori tests of the different closure methods. This may seem to be a daunting task at first glance, but all our a priori investigations could be automated and completed within minutes. In general, there is no guarantee that testing the closure methods on some randomly sampled parameters is sufficient to predict the performance of the methods for the true parameters. However, our results suggest that, at least to some extent, the performance of closure methods is determined by the model structure. Consequently, an automated a priori closure checking procedure seems to be reasonable. The additional computational cost of our a priori investigations is more than compensated for in the following parameter inference where we evaluated the likelihood using the moment equations close to 1 million times, a task that would have been infeasible using, for instance, stochastic simulation.

6.3. Gene Expression with Extrinsic Variability

As a third case study, we consider a simple two-stage model of gene expression that does not include any promoter states. The objective of this case study is twofold. On the one hand, it serves as an example that also includes extrinsic variability; on the other hand, we also include unknown initial distributions of the species and investigate how well parameters of these distributions can be identified. The network we consider is given by the following reactions:

$$\emptyset \xrightarrow{c_1} M$$
$$M \xrightarrow{c_2} \emptyset$$
$$M \xrightarrow{Z} M + P$$
$$P \xrightarrow{c_4} \emptyset$$



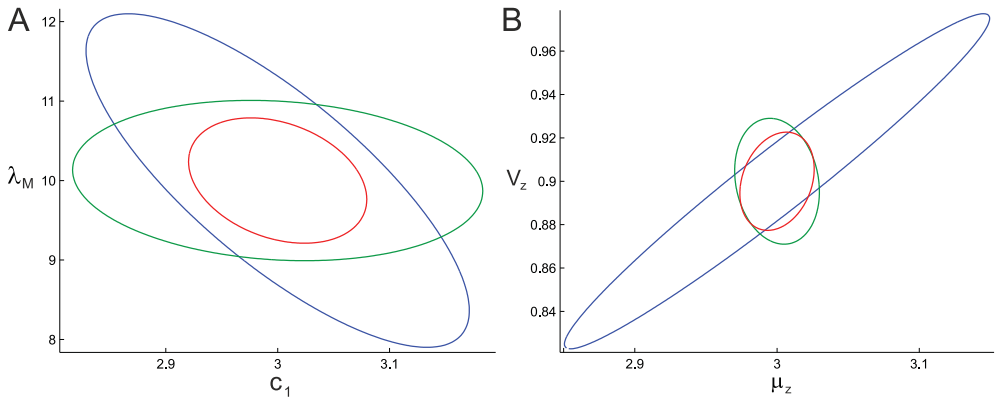


Fig. 5. 95% confidence regions, computed using the true parameters, for the pairs $c1 - \lambda_M$ (A) and $\mu_z - V_z$ (B) for equally spaced (blue) and optimally placed (red) measurements. The green ellipses are obtained when all measurements are placed at the earliest possible time points under the constraint that they are all two time units apart from each other.

Z is an extrinsic variable that could, for instance, represent variability in the number of ribosomes. We assume that Z has a log-normal distribution with unknown parameters, that $M(0)$ and $P(0)$ have unknown Poisson distributions, and that $M(0)$ and $P(0)$ are independent from each other and from Z. We further assume that $c_1 = 3$, $c_2 = 0.1$, $c_4 = 0.1$, $\lambda_M := \mathbb{E}[M(0)] = 10$, $\lambda_P := \mathbb{E}[P(0)] = 100$, $\mu_z := \mathbb{E}[Z] = 3$, and $V_z := Var(Z) = 0.9$ and that P can be measured in 10,000 cells at $t = 0$ and at 10 point in the time interval $(0, 50]$. λ_P can then be obtained from the measurement at $t = 0$ and does not have to be included as an unknown parameter. Consequently, there are six unknown parameters, $\gamma = [c_1\ c_2\ c_4\ \mu_z\ V_z\ \lambda_M]^T$.

It is easy to see that this example is equivalent to a reaction network where the third reaction is replaced by

$$M + Z \xrightarrow{1} M + Z + P$$

and the extrinsic variable Z is included as a third species. Since this is a bimolecular reaction, the moment equations are nonclosed. It turns out, however, that it is nevertheless possible to derive an exact and closed system of equations for moments up to any desired order L. This requires including some selected moments up to order $2L-1$ in the equations and using the fact that $M(t)$ and Z are independent for all $t \in [0, 50]$. We are interested in the moments up to order 4, and this requires including six moments of order 5, three moments of order 6, and one moment of order 7 in the equations (a MATLAB script for solving these moment equations is available upon request).

It is intuitively clear that at least some of the measurements should be placed at very early time points, because otherwise the influence of the initial condition on the measurements will be too small for a successful identification of λ_M. However, since λ_M is only one of the unknown parameters, it is not clear how many of the measurements should be placed at early time points. We investigated this by optimally placing the measurement times under the assumption that all measurements have to be at least two time units apart from each other. We found that it is best to place seven of the measurements at early time points, three at late time points, and none at intermediate time points. The information obtained with the optimal measurement times is visualized in Figure 5. Besides this, we also computed the information that is obtained if the measurements are placed either equally spaced in $(0, 50]$ or all at very early time points. Figure 5 shows that in all measurement scenarios, the resulting confidence regions are

very tight, which indicates that optimal placement of the measurements is not really necessary here. This is because we chose a rather large number of measurements. The situation would be different if fewer measurements were available.

To test moment-based inference, we simulated data using stochastic simulation. Since the optimal measurement times were obtained with the true parameters and would not be available in a realistic situation, we assumed that the measurements are taken at equally spaced time points. As in the previous example, we performed the parameter identification using an MCMC algorithm, which we initialized at 10 random starting points. The parameters that attained the highest likelihood in all the runs are given by $c_1 = 2.9230$, $c_2 = 0.1026$, $c_4 = 0.0971$, $\mu_z = 3.0700$, $V_z = 0.9912$, and $\lambda_M = 9.3630$. They agree almost perfectly with the true values, which confirms that equally spaced measurement times are sufficient here.

Concluding this example, we can say that in principle it is possible to jointly identify unknown reaction rates, extrinsic parameters, and initial conditions. However, the example we used to demonstrate this contains only one unknown parameter λ_M for the initial condition and only one additional parameter V_z, which stems from extrinsic variability. It should not come as a surprise that for larger models that contain unknown initial conditions for multiple hidden species and possibly also variability in multiple reaction rates, the data we considered in this article will not be informative enough to identify all parameters. In such cases, a successful identification of the parameters will require multiple experiments in different conditions [Zechner et al. 2012] and possibly also adding measurements obtained with different experimental technologies such as fluorescence microscopy.

As a final note, we would like to point out that we chose Poisson distributions for the initial conditions of the species and a log-normal distribution for Z because this is possibly the most reasonable choice from a biological perspective. Changing these distributions would not change anything about the moment equations and the fact that the moments of this reaction network can be computed exactly. Only the number of unknown parameters and how these parameters enter the moment equations would change. Consequently, we could equally well choose other, for instance, bimodal, distributions according to some parameterization and identify the parameters of these distributions from moments of P.

7. CONCLUSIONS

The interplay of mathematical modeling and experimentation is at the heart of the field of systems biology, yet for stochastic biochemical reaction networks, mathematical modeling in practice has remained limited to rather small systems. The major limiting factor is that the chemical master equation is intractable for all but the simplest systems. While it is foreseeable that more efficient simulation methods and improved numerical approximation techniques will provide some progress on that front, it also has to be expected that the advancement of biological measurement techniques will lead to the need for ever more complex models; for instance, mass cytometry can already now produce data that is so plentiful that it is way out of the scope of the existing approaches for stochastic modeling.

Instead of trying to develop better numerical approximation techniques, we developed methods that avoid the chemical master equation altogether and resort to the moment equations instead. This bypasses the previously limiting step and enables stochastic modeling for larger systems. Our approach raises the need for an accurate moment closure method that may or may not be available for a given system. This makes the study and development of moment closure methods an alternative future direction for advancing stochastic modeling of biochemical reaction networks. For instance, guarantees on the performance of a closure method for the unknown true model

parameters, given its performance for a set of sampled parameters, would be useful. Our numerical results suggest that the performance of a closure method for sampled parameters can indeed be an indicator of the performance for the true parameters and thus that closure methods can be tested a priori to ensure that an appropriate method is used in the parameter inference. Other possible future research directions are to extend the framework to delayed CTMCs [Guet et al. 2012] and to consider the problem of model selection to also be able to address cases where the model structure (i.e., which species and reactions are present) is not known.

Finally, more effort should be dedicated to investigating parameter identifiability. Whether or not the parameters can be identified from available data obviously depends on the data and the number of unknown parameters but may also depend on the true values of the parameters. Intuitively, we cannot hope to recover very fast switching rates between promoter states from measurements of comparably slow protein dynamics, whereas slow promoter switching might pose no problem. This shows that the local nature of the Fisher information may be a weakness: if we compute the information only locally at imprecise initial parameter estimates when we design experiments, we may overlook identifiability problems. A second drawback of the presented theory is that the moment equations do not allow one to compute path likelihoods of the CTMC and therefore cannot be used for parameter inference when data from experiments such as fluorescence microscopy, in which individual cells are tracked in time, is available. Such data usually contains smaller numbers of measured cells compared to flow cytometry but has the advantage that time correlations are also included. Whether or not this additional information outweighs the smaller sample sizes is unclear, but it has to be expected that there exist situations in which time correlation information will be necessary to identify all parameters. So far this question has only been addressed for one specific example under a Gaussian approximation of the CTMC [Komorowski et al. 2011] and it would be of interest to investigate it further.

APPENDIX

Proof of Theorem 3.1

PROOF. Multiplying both sides of Equation (4) by the ith component x^i of the state vector x, summing over all states $x \in \mathbb{N}^m$, and integrating over all possible values of Z with respect to the probability measure P_Z yields

$$\int \sum_{x\in\mathbb{N}^m} \frac{\mathrm{d}}{\mathrm{d}t} p(x,t|z)x^i dP_Z = \int \sum_{x\in\mathbb{N}^m} \sum_{k=1}^{K} \left[p(x-\nu_k,t|z)a_k(x-\nu_k,z)x^i - p(x,t|z)a_k(x,z)x^i\right] dP_Z.$$

The left-hand side is then the time derivative of the marginal mean of X^i, whereas the right-hand side can be simplified by a change of variables as follows:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X^i(t)] &= \int \sum_{k=1}^{K} \sum_{x\in\mathbb{N}^m} \left[p(x,t|z)a_k(x,z)(x^i+\nu_{k_i}) - p(x,t|z)a_k(x,z)x^i\right] dP_Z \\ &= \sum_{k=1}^{K} \nu_{k_i} \int \sum_{x\in\mathbb{N}^m} p(x,t|z)a_k(x,z)dP_Z = \sum_{k=1}^{K} \nu_{k_i}\mathbb{E}[a_k(X(t),Z)],\end{aligned}$$

where ν_{k_i} is the ith component of ν_k. Since the propensities are at most quadratic in X and linear in Z, this can give moments of order 3 or less where the cross-moments contain at most one power of Z.

Multiplying both sides of Equation (4) by $x^i x^j$ instead of x^i and performing the same calculations yields the time derivatives of the second-order moments:

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X^i(t)X^j(t)] \\
&\quad= \int \sum_{k=1}^{K} \sum_{x\in\mathbb{N}^m} \left[p(x,t|z)a_k(x,z)(x^i+\nu_{k_i})(x^j+\nu_{k_j}) - p(x,t|z)a_k(x,z)x^i x^j\right] dP_Z \\
&\quad= \sum_{k=1}^{K} \left(\nu_{k_i}\mathbb{E}[X^j(t)a_k(X(t),Z)] + \nu_{k_j}\mathbb{E}[X^i(t)a_k(X(t),Z)] + \nu_{k_i}\nu_{k_j}\mathbb{E}[a_k(X(t),Z)]\right).
\end{aligned}$$

Here, with the same reasoning as earlier, the first two terms in the sum can give moments of order 4 or less while the third term gives moments of order 3 or less where for all terms the cross-moments contain at most one power of Z.

The time derivatives of the cross-moments $\mathbb{E}[X^i(t)Z]$ can be obtained with the same calculations as earlier by multiplying both sides of Equation (4) by $x^i z$:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X^i(t)Z] &= \int \sum_{k=1}^{K} \sum_{x\in\mathcal{X}} \left[p(x,t|z)a_k(x,z)(x^i+\nu_{k_i})z - p(x,t|z)a_k(x,z)x^i z\right] dP_Z \\
&= \sum_{k=1}^{K} \nu_{k_i}\mathbb{E}[Za_k(X(t),Z)].
\end{aligned}$$

This can give moments of order up to 4 where for all terms the cross-moments contain at most two powers of Z, which concludes the proof. □

Proof of Theorem 4.1

Proof. If the number of measured cells n is large enough, approximate normality of $S_J(t_s)$ is a direct consequence of the central limit theorem. $\mathbb{E}\left[S_J(t_s)\right] = [\mu_1^i(t_s)\ \mu_2^i(t_s)]^T$ is trivial. It remains to be shown that the covariance matrix of $S_J(t_s)$ is given by Equation (9). To this end, assume without loss of generality that $\mu_1^i(t_s) = 0$ (else a shifted version of $X^i(t_s)$ can be considered). For ease of notation throughout the proof, we will write X_j instead of $X_j^i(t_s)$.

The top left entry of $\Sigma(t_s)$ is the variance of $S_M(t_s)$. Consequently, we have to show that $Var(S_M(t_s)) = \frac{1}{n}\mu_2^i(t_s)$.

$$Var(S_M(t_s)) = \mathbb{E}\left[S_M(t_s)^2\right] = \mathbb{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right)^2\right] = \frac{1}{n^2}\mathbb{E}\left[\sum_{j=1}^{n} X_j^2 + \sum_{h\neq j} X_h X_j\right] = \frac{1}{n}\mu_2^i(t_s).$$

Thereby, the last equality holds because

$$\mathbb{E}\left[\sum_{h\neq j} X_h X_j\right] = \sum_{h\neq j}\mathbb{E}[X_h X_j] = \sum_{h\neq j}\mathbb{E}[X_h]\mathbb{E}[X_j] = \sum_{h\neq j}\mu_1^i(t_s)^2 = 0,$$

where we used that $X_j,\ j = 1,\ldots,n$ are i.i.d. with mean $\mu_1^i(t_s) = 0$.

The off-diagonal entry of $\Sigma(t_s)$ is the covariance of $S_M(t_s)$ and $S_V(t_s)$. Consequently, we have to show that $Cov(S_M(t_s), S_V(t_s)) = \frac{1}{n}\mu_3^i(t_s)$.

$$\begin{aligned} Cov(S_M(t_s), S_V(t_s)) &= \mathbb{E}[S_M(t_s)S_V(t_s)] \\ &= \frac{1}{n(n-1)}\mathbb{E}\left[\left(\sum_{j=1}^n X_j\right)\left(\sum_{j=1}^n X_j^2\right) - \frac{1}{n}\left(\sum_{j=1}^n X_j\right)\left(\sum_{j=1}^n X_j^2 + \sum_{h\neq j}^n X_h X_j\right)\right] \\ &= \frac{1}{n-1}\mu_3^i(t_s) - \frac{1}{n(n-1)}\mu_3^i(t_s) = \frac{1}{n}\mu_3^i(t_s), \end{aligned}$$

where we used the same arguments as earlier.

The bottom right entry of $\Sigma(t_s)$ is the variance of $S_V(t_s)$. Consequently, we have to show that $Var(S_V(t_s)) = \frac{1}{n}(\mu_4^i(t_s) - \frac{n-3}{n-1}\left(\mu_2^i(t_s)\right)^2)$.

$$\mathbb{E}\left[S_V(t_s)^2\right] = \frac{1}{(n-1)^2}\left(\underbrace{\mathbb{E}\left[\left(\sum_{j=1}^n X_j^2\right)^2\right]}_{\textbf{Term 1}} - \underbrace{\frac{2}{n}\mathbb{E}\left[\left(\sum_{j=1}^n X_j^2\right)\left(\sum_{j=1}^n X_j\right)^2\right]}_{\textbf{Term 2}} + \underbrace{\frac{1}{n^2}\mathbb{E}\left[\left(\sum_{j=1}^n X_j\right)^4\right]}_{\textbf{Term 3}}\right)$$

$$\textbf{Term 1} = \mathbb{E}\left[\sum_{j=1}^n X_j^4 + \sum_{h\neq j} X_h^2 X_j^2\right] = n\mu_4^i(t_s) + n(n-1)\left(\mu_2^i(t_s)\right)^2,$$

$$\begin{aligned} \textbf{Term 2} &= \frac{2}{n}\mathbb{E}\left[\left(\sum_{j=1}^n X_j^2\right)\left(\sum_{j=1}^n X_j^2 + \sum_{h\neq j} X_h X_j\right)\right] = \frac{2}{n}\mathbb{E}\left[\sum_{j=1}^n X_j^4 + \sum_{h\neq j} X_h^2 X_j^2\right] \\ &= 2\mu_4^i(t_s) + 2(n-1)\left(\mu_2^i(t_s)\right)^2, \end{aligned}$$

$$\begin{aligned} \textbf{Term 3} &= \frac{1}{n^2}\mathbb{E}\left[\left(\sum_{j=1}^n X_j^2 + \sum_{h\neq j} X_h X_j\right)^2\right] = \frac{1}{n^2}\mathbb{E}\left[\sum_{j=1}^n X_j^4 + \sum_{h\neq j} X_h^2 X_j^2 + 2\sum_{h\neq j} X_h^2 X_j^2\right] \\ &= \frac{1}{n}\mu_4^i(t_s) + \frac{3(n-1)}{n}\left(\mu_2^i(t_s)\right)^2. \end{aligned}$$

It follows that

$$\begin{aligned} Var(S_V(t_s)) &= \mathbb{E}[S_V(t_s)^2] - \mathbb{E}[S_V(t_s)]^2 \\ &= \frac{n^2-2n+1}{(n-1)^2 n}\mu_4^i(t_s) + \frac{n^2-2n+3}{(n-1)n}\left(\mu_2^i(t_s)\right)^2 - \left(\mu_2^i(t_s)\right)^2 \\ &= \frac{1}{n}\left(\mu_4^i(t_s) - \frac{n-3}{n-1}\left(\mu_2^i(t_s)\right)^2\right). \quad \square \end{aligned}$$

ACKNOWLEDGMENTS

This article summarizes the theory behind the methods that were originally proposed in Zechner et al. [2012] and Ruess et al. [2013]. We are grateful to the coauthors of these papers for their contributions to the work.

REFERENCES

M. Acar, J. Mettetal, and A. van Oudenaarden. 2008. Stochastic switching as a survival strategy in fluctuating environments. *Nature Genetics* 40, 2008, 471–475. DOI:http://dx.doi.org/10.1038/ng.110

A. Ale, P. Kirk, and M. Stumpf. 2013. A general moment expansion method for stochastic kinetic models. *Journal of Chemical Physics* 138, 2013, 174101. DOI:http://dx.doi.org/10.1063/1.4802475

A. Andreychenko, L. Mikeev, D. Spieler, and V. Wolf. 2011. Parameter identification for Markov models of biochemical reactions. In *Proceedings of the 23rd International Conference on Computer Aided Verification (CAV'11).* (2011), 83–98. DOI:http://dx.doi.org/10.1007/978-3-642-22110-1_8

G. Balazsi, A. van Oudenaarden, and J. Collins. 2011. Cellular decision making and biological noise: From microbes to mammals. *Cell* 144, 2011, 910–925. DOI:http://dx.doi.org/10.1016/j.cell.2011.01.030

A. Colman-Lerner, A. Gordon, E. Serra, T. Chin, O. Resnekov, D. Endy, G. Pesce, and R. Brent. 2005. Regulated cell-to-cell variation in a cell-fate decision system. *Nature* 437, 7059 (2005), 699–706. DOI:http://dx.doi.org/10.1038/nature03998

M. Elowitz, A. Levine, E. Siggia, and P. Swain. 2002. Stochastic gene expression in a single cell. *Science* 297, 5584 (2002), 1183–1186. DOI:http://dx.doi.org/10.1126/science.1070919

S. Engblom. 2006. Computing the moments of high dimensional solutions of the master equation. *Applied Mathematics and Computation* 180, 2 (2006), 498–515. DOI:http://dx.doi.org/10.1016/j.amc.2005.12.032

G. Franceschini and S. Macchietto. 2008. Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science* 63, 19 (2008), 4846–4872. DOI:http://dx.doi.org/10.1016/j.ces.2007.11.034

D. Gillespie. 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics* 22, 4 (1976), 403–434. DOI:http://dx.doi.org/10.1016/0021-9991(76)90041-3

D. Gillespie. 1992. A rigorous derivation of the chemical master equation. *Physica A* 188, 1–3 (1992), 404–425. DOI:http://dx.doi.org/10.1016/0378-4371(92)90283-v

A. Gonzalez, J. Uhlendorf, J. Schaul, E. Cinquemani, G. Batt, and G. Ferrari-Trecate. 2013. Identification of biological models from single-cell data: A comparison between mixed-effects and moment-based inference. In *Proceedings of the 12th European Control Conference.*

J. Goutsias and G. Jenkinson. 2013. Markovian dynamics on complex reaction networks. *Physics Reports* 529 (2013), 199–264. DOI:http://dx.doi.org/doi:10.1016/j.physrep.2013.03.004

C. Guet, A. Gupta, T. A. Henzinger, M. Mateescu, and A. Sezgin. 2012. Delayed continuous-time Markov chains for genetic regulatory circuits. In *Proceedings of the 24th International Conference on Computer Aided Verification (CAV'12).* 294–309. DOI:http://dx.doi.org/10.1007/978-3-642-31424-7_24

D. Hagen, J. White, and B. Tidor. 2013. Convergence in parameters and predictions using computational experimental design. *Interface Focus* 3 (2013), 20130008. DOI:http://dx.doi.org/10.1098/rsfs.2013.0008

J. Hasenauer, V. Wolf, A. Kazeroonian, and F. Theis. 2014. Method of conditional moments (MCM) for the chemical master equation. *Journal of Mathematical Biology* 69, 3 (2013), 687–735. DOI:http://dx.doi.org/10.1007/s00285-013-0711-5

J. Hasty, J. Pradines, M. Dolnik, and J. Collins. 2000. Noise-based switches and amplifiers for gene expression. *Proceedings of the National Academy of Sciences of the U.S.A.* 97, 5 (2000), 2075–2080. DOI:http://dx.doi.org/10.1073/pnas.040411297

T. A. Henzinger, L. Mikeev, M. Mateescu, and V. Wolf. 2010. Hybrid numerical solution of the chemical master equation. *Proceedings of the 8th International Conference on Computational Methods in Systems Biology (CMSB'10).* ACM, New York, 55–65. DOI:http://dx.doi.org/10.1145/1839764.1839772

J. Hespanha. 2006. StochDynTools - A MATLAB toolbox to compute moment dynamics for stochastic networks of bio-chemical reactions. Available at http://www.ece.ucsb.edu/~hespanha/software.

J. Hespanha. 2008. Moment closure for biochemical networks. In *Proceedings of the 3rd International Symposium on Communications, Control and Signal Processing.* 142–147. DOI:http://dx.doi.org/10.1109/ISCCSP.2008.4537208

A. Hindmarsh, P. Brown, K. Grant, S. Lee, R. Serban, D. Shumaker, and C. Woodward. 2005. SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software (TOMS)* 31, 3 (2005), 363–396. DOI:http://dx.doi.org/10.1145/1089014.1089020

A. Hjartarson, J. Ruess, and J. Lygeros. 2013. Approximating the solution of the chemical master equation by combining finite state projection and stochastic simulation. In *Proceedings of the IEEE 52nd Annual Conference on Decision and Control (CDC'13).*

W. Hunter, W. Hill, and T. Henson. 1969. Designing experiments for precise estimation of all or some of the constants in a mechanistic model. *Canadian Journal of Chemical Engineering* 47, 1 (1969), 76–80. DOI:http://dx.doi.org/10.1002/cjce.5450470114

C. Ko, Y. Yamada, D. Welsh, E. Buhr, A. Liu, E. Zhang, M. Ralph, S. Kay, D. Forger, and J. Takahashi. 2010. Emergence of noise-induced oscillations in the central circadian pacemaker. *PLoS Biology* 8 (2010), e1000513. DOI:http://dx.doi.org/10.1371/journal.pbio.1000513

M. Komorowski, M. Costa, D. Rand, and M. Stumpf. 2011. Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. *Proceedings of the National Academy of Sciences of the U.S.A.* 108, 21 (2011), 8645–8650. DOI:http://dx.doi.org/10.1073/pnas.1015814108

P. Kügler. 2012. Moment fitting for parameter inference in repeatedly and partially observed stochastic biological models. *PLoS ONE* 7, 8 (2012), e43001. DOI:http://dx.doi.org/10.1371/journal.pone.0043001

G. Lillacci and M. Khammash. 2013. The signal within the noise: Efficient inference of stochastic gene regulation models using fluorescence histograms and stochastic simulations. *Bioinformatics* 29, 18 (2013), 2311–2319. DOI:http://dx.doi.org/10.1093/bioinformatics/btt380

M. Mateescu, V. Wolf, F. Didier, and T. A. Henzinger. 2010. Fast adaptive uniformisation of the chemical master equation. *IET Systems Biology* 4, 6 (2010), 441–452. DOI:http://dx.doi.org/10.1049/iet-syb.2010.0005

H. McAdams and A. Arkin. 1997. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the U.S.A.* 94, 3 (1997), 814–819.

F. Menolascina, M. di Bernardo, and D. di Bernardo. 2011. Analysis, design and implementation of a novel scheme for in-vivo control of synthetic gene regulatory networks. *Automatica* 47, 6 (2011), 1265–1270.

L. Mikeev and V. Wolf. 2012. Parameter estimation for stochastic hybrid models of biochemical reaction networks. In *Proceedings of the 15th ACM International Conference on Hybrid Systems: Computation and Control*. ACM, New York. (2012), 155–166. DOI:http://dx.doi.org/10.1145/2185632.2185657

A. Milias-Argeitis, S. Summers, J. Stewart-Ornstein, I. Zuleta, D. Pincus, H. El-Samad, M. Khammash, and J. Lygeros. 2011. In silico feedback for in vivo regulation of a gene expression circuit. *Nature Biotechnology* 29 (2011), 1114–1116. DOI:http://dx.doi.org/10.1038/nbt.2018

B. Munsky and M. Khammash. 2006. The finite state projection algorithm for the solution of the chemical master equation. *Journal of Chemical Physics* 124 (2006), 044104. DOI:http://dx.doi.org/10.1063/1.2145882

B. Munsky, B. Trinh, and M. Khammash. 2009. Listening to the noise: Random fluctuations reveal gene network parameters. *Molecular Systems Biology* 5, 1 (2009), 318. DOI:http://dx.doi.org/10.1038/msb.2009.75

S. Poovathingal and R. Gunawan. 2010. Global parameter estimation methods for stochastic biochemical systems. *BMC Bioinformatics* 11, 1 (2010), 414–425. DOI:http://dx.doi.org/10.1186/1471-2105-11-414

L. Pronzato and E. Walter. 1985. Robust experimental design via stochastic approximation. *Mathematical Biosciences* 75 (1985), 103–120. DOI:http://dx.doi.org/10.1016/0025-5564(85)90068-9

J. Raser and E. O'Shea. 2005. Noise in gene expression: Origins, consequences, and control. *Science* 309, 5743 (2005), 2010–2013. DOI:http://dx.doi.org/10.1126/science.1105891

J. Ruess and J. Lygeros. 2013. Identifying stochastic biochemical networks from single-cell population experiments: A comparison of approaches based on the Fisher information. In *Proceedings of the IEEE 52nd Annual Conference on Decision and Control (CDC'13)*.

J. Ruess, A. Milias-Argeitis, and J. Lygeros. 2013. Designing experiments to understand the variability in biochemical reaction networks. *Journal of the Royal Society Interface* 10, 88 (2013), 20130588. DOI:http://dx.doi.org/10.1098/rsif.2013.0588

J. Ruess, A. Milias-Argeitis, S. Summers, and J. Lygeros. 2011. Moment estimation for chemically reacting systems by extended Kalman filtering. *Journal of Chemical Physics* 135 (2011), 165102. DOI:http://dx.doi.org/10.1063/1.3654135

M. Samoilov and A. Arkin. 2006. Deviant effects in molecular reaction pathways. *Nature Biotechnology* 24, 10 (2006), 1235–1240. DOI:http://dx.doi.org/10.1038/nbt1253

V. Shahrezaei, J. Ollivier, and P. Swain. 2008. Colored extrinsic fluctuations and stochastic gene expression. *Molecular Systems Biology* 4, 196 (2008). DOI:http://dx.doi.org/10.1038/msb.2008.31

A. Singh and J. Hespanha. 2006. Lognormal moment closures for biochemical reactions. In *Proceedings of the IEEE 45th Annual Conference on Decision and Control (CDC'06)*. 2063–2068. DOI:http://dx.doi.org/10.1109/CDC.2006.376994

A. Singh and J. Hespanha. 2011. Approximate moment dynamics for chemically reacting systems. *IEEE Transactions on Automatic Control* 56, 2 (2011), 414–418. DOI:http://dx.doi.org/10.1109/TAC.2010.2088631

T. Toni and B. Tidor. 2013. Combined model of intrinsic and extrinsic variability for computational network design with application to synthetic biology. *PLoS Computational Biology* 9 (2013), 3. DOI:http://dx.doi.org/10.1371/journal.pcbi.1002960

J. Uhlendorf, A. Miermont, T. Delaveau, G. Charvin, F. Fages, S. Bottani, G. Batt, and P. Hersen. 2012. Long-term model predictive control of gene expression at the population and single-cell

levels. *Proceedings of the National Academy of Sciences of the U.S.A.* 109, 35 (2012), 14271–14276. DOI:http://dx.doi.org/10.1073/pnas.1206810109

D. Volfson, J. Marciniak, W. Blake, N. Ostroff, L. Tsimring, and J. Hasty. 2005. Origins of extrinsic variability in eukaryotic gene expression. *Nature* 439, 7078 (2005), 861–864. DOI:http://dx.doi.org/10.1038/nature04281

E. Walter and L. Pronzato. 1990. Qualitative and quantitative experiment design for phenomenological models—a survey. *Automatica* 26, 2 (1990), 195–213. DOI:http://dx.doi.org/10.1016/0005-1098(90)90116-Y

P. Whittle. 1957. On the use of the normal approximation in the treatment of stochastic processes. *Journal of the Royal Statistical Society Series B Statistical Methodology* 19 (1957), 268–281.

V. Wolf, R. Goel, M. Mateescu, and T. A. Henzinger. 2010. Solving the chemical master equation using sliding windows. *BMC Systems Biology* 4 (2010), 42. DOI:http://dx.doi.org/10.1186/1752-0509-4-42

C. Zechner, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koeppl. 2012. Moment-based inference predicts bimodality in transient gene expression. *Proceedings of the National Academy of Sciences of the U.S.A.* 109, 21 (2012), 8340–8345. DOI:http://dx.doi.org/10.1073/pnas.1200161109

C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koeppl. 2014. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nature Methods* 11, 2 (2014), 197–202. DOI:http://dx.doi.org/10.1038/nmeth.2794

Received January 2014; revised June 2014; accepted July 2014