METHODOLOGY ARTICLE **Open Access**

CrossMark

Generalized method of moments for estimating parameters of stochastic reaction networks

Alexander Lück and Verena Wolf[*]

Abstract

Background: Discrete-state stochastic models have become a well-established approach to describe biochemical reaction networks that are influenced by the inherent randomness of cellular events. In the last years several methods for accurately approximating the statistical moments of such models have become very popular since they allow an efficient analysis of complex networks.

Results: We propose a generalized method of moments approach for inferring the parameters of reaction networks based on a sophisticated matching of the statistical moments of the corresponding stochastic model and the sample moments of population snapshot data. The proposed parameter estimation method exploits recently developed moment-based approximations and provides estimators with desirable statistical properties when a large number of samples is available. We demonstrate the usefulness and efficiency of the inference method on two case studies.

Conclusions: The generalized method of moments provides accurate and fast estimations of unknown parameters of reaction networks. The accuracy increases when also moments of order higher than two are considered. In addition, the variance of the estimator decreases, when more samples are given or when higher order moments are included.

Keywords: Biochemical reaction network, Stochastic model, Parameter estimation, Generalized method of moments

Background

A widely-used approach in systems biology research is to design quantitative models of biological processes and refine them based on both computer simulations and wet-lab experiments. While a large amount of sophisticated parameter inference methods have been proposed for deterministic models, only few approaches allow the efficient calibration of parameters for large discrete-state stochastic models that describe stochastic interactions between molecules within a single cell. Since research progress in experimental measurement techniques that deliver single-cell and single-molecule data has advanced, the ability to calibrate such models is of key importance. For instance, the widely-used flow cytometric analysis delivers data from thousands of cells which yields sample means and sample variances of molecular populations.

Here, we focus on the most common scenario: a discrete stochastic model of a cellular reaction network with unknown reaction rate constants and population snapshot data such as sample moments of a large number of observed samples. The state of the model corresponds to the vector of current molecular counts, i.e., the number of molecules of each chemical species, and chemical reactions trigger state transitions by changing the molecular populations. A system of ordinary differential equations, the chemical master equation [1], describes the evolution of the state probabilities over time.

A classical maximum likelihood (ML) approach, in which the likelihood is directly approximated, is possible if all populations are small [2] or if the model shows simple dynamics (e.g. multi-dimensional normal distribution with time-dependent mean and covariance matrix) such that the likelihood can be approximated by a normal distribution [3]. In this case, the likelihood (and its derivatives) can usually be approximated efficiently and global optimization techniques are employed to find parameters

*Correspondence: verena.wolf@uni-saarland.de
Department of Computer Science, Saarland University, Campus E 13, 66123
Saarbrücken, Germany

that maximize the likelihood. However, if large populations are present in the system then direct approximations of the likelihood are unfeasible since the underlying system of differential equations contains one equation for each state and the main part of the probability mass of the model distributes on an intractably large number of states. Similarly, if the system shows complex dynamics such as multimodality, approximations of the likelihood based on Gaussian distributions become inaccurate.

In the last years several methods have been developed to accurately simulate the moments of the underlying probability distribution up to a certain order k over time [4–6]. The complexity of these simulation methods is therefore independent of the population sizes but, for large k, the corresponding differential equations may become stiff and lead to poor approximations. However, reconstructions of complex distributions from their moments show that for many systems already for small k (e.g. $k \in \{4, \ldots, 8\}$) the moments contain sufficient information about the distribution such as the strength and location of regions of attraction (i.e. regions of the state space containing a large proportion of the probability mass) [7].

For models with complex distributions such as multiple modes or oscillations, the accuracy and the running time of the moment approximation can be markedly improved, when conditional moments are considered in combination with the probabilities of appropriately chosen system modes such as the activity state of the genes in a gene regulatory network [8–11]. Recently a full derivation of the conditional moment equations was derived and numerical results show that when the maximum order of the considered moments is high, the number of equations that have to be integrated is usually much smaller for the conditional moments approach and the resulting equations are less stiff [12]. In addition, the approximated (unconditional) moments are more accurate when the same maximal order is considered.

An obvious parameter inference approach is the matching of the observed sample moments with those of the moment-based simulation of the model. Defining the differences between sample and (approximated) population moments as cost functions that depend on the parameters, an approach that minimizes the sum of the squared cost functions seems reasonable. However, in a simple least-squares approach low moments such as means and (co-)variances contribute equally to the sum of squared differences as higher moments, whose absolute magnitudes are much higher (even if they are centralized). Moreover, correlations between the different cost functions may exist and thus necessitate an approach where also products of two different cost functions are considered.

The generalized method of moments (GMM) that is widely used in econometrics provides an estimator that is computed after assigning appropriate weights to the different cost function products [13]. The GMM estimator has, similar as the ML estimator, desirable statistical properties such as being consistent and asymptotically normally distributed. Moreover, for optimally chosen weights it is an asymptotically efficient estimator, which implies that (asymptotically) it has minimum variance among all estimators for the unknown parameters.

In this paper we explore the usefulness of the GMM for moment-based simulations of stochastic reaction networks. We focus on two particular estimators that are commonly used in econometrics: the two-step estimator of Hansen [13] and the demean estimator [14]. We study the accuracy and variance of the estimator for different maximal moment orders and different sample sizes by applying the GMM to two case studies. In addition, we show that poor approximations of some higher order moments have a strong influence on the quality of the estimation. Interestingly, we see that the additional information about the covariances of the cost functions can lead to identification of all parameters. In addition, the variance of the estimator becomes smaller when higher order moments are included. Compared to the simple least-squares approach, the GMM approach yields very accurate estimates.

Methods
Stochastic chemical kinetics
Our inference approach relies on a Markov modeling approach that follows Gillespie's theory of stochastic chemical kinetics. We consider a well-stirred mixture of n molecular species in a volume with fixed size and fixed temperature and represent it as a discrete-state Markov process $\{\mathbf{X}(t), t \geq 0\}$ in continuous-time [15]. The random vector $\mathbf{X}(t) = (X_1(t), \ldots, X_n(t))$ describes the chemical populations at time t, i.e., $X_i(t)$ is the number of molecules of type $i \in \{1, \ldots, n\}$ at time t. Thus, the state space of $\mathbf{X}$ is $\mathbb{Z}_+^n = \{0, 1, \ldots\}^n$. The state changes of $\mathbf{X}$ are triggered by the occurrences of chemical reactions. Each of the m different reaction types has an associated non-zero change vector $\mathbf{v}_j \in \mathbb{Z}^n$ $(j \in \{1, \ldots, m\})$, where $\mathbf{v}_j = \mathbf{v}_j^- + \mathbf{v}_j^+$ such that $\mathbf{v}_j^-$ $(\mathbf{v}_j^+)$ contains only non-positive (non-negative) entries and specifies how many molecules of each species are consumed (produced) if an instance of the reaction occurs, respectively. Thus, if $\mathbf{X}(t) = \mathbf{x}$ for some $\mathbf{x} \in \mathbb{Z}_+^n$ with $\mathbf{x} + \mathbf{v}_j^-$ being non-negative, then $\mathbf{X}(t+dt) = \mathbf{x}+\mathbf{v}_j$ is the state of the system after the occurrence of the j-th reaction within the infinitesimal time interval $[\,t, t+dt)$. W.l.o.g. we assume here that all vectors $\mathbf{v}_j$ are distinct.

We use $\alpha_1, \ldots, \alpha_m$ to denote the propensity functions of the reactions, where $\alpha_j(\mathbf{x}) \cdot dt$ is the probability that, given $\mathbf{X}_t = \mathbf{x}$, one instance of the j-th reaction occurs within $[\,t, t + dt)$. Assuming law of mass action kinetics, $\alpha_j(\mathbf{x})$ is chosen proportional to the number of distinct

reactant combinations in state $\mathbf{x}$. An example is given in Table 1, where the first reaction gives as change vectors, for instance, $\mathbf{v}_1^- = (-1, 0, 0), \mathbf{v}_1^+ = (0, 1, 0), \mathbf{v}_1 = (-1, 1, 0)$. Note that, given the initial state $\mathbf{x} = (1, 0, 0)$, at any time either the DNA is active or not, i.e. $x_1 = 0$ and $x_2 = 1$, or $x_1 = 1$ and $x_2 = 0$. Moreover, the state space of the model is infinite in the third dimension. Although our inference approach can be used for any model parameter in the sequel we simply assume that the proportionality constants c_j are unknown and have to be estimated based on experimental data.

For $\mathbf{x} \in \mathbb{Z}_+^n$ and $t \geq 0$, let $p_t(\mathbf{x})$ denote the probability $P(\mathbf{X}(t) = \mathbf{x})$. Assuming fixed initial conditions p_0 the evolution of $p_t(\mathbf{x})$ is given by the chemical master equation (CME) [1]

$$\frac{\partial}{\partial t} p_t(\mathbf{x}) = \sum_{j: \mathbf{x} - \mathbf{v}_j^- \geq 0} \alpha_j(\mathbf{x} - \mathbf{v}_j) p_t(\mathbf{x} - \mathbf{v}_j) - \alpha_j(\mathbf{x}) p_t(\mathbf{x}),$$

which is an ordinary first-order differential equation that has a unique solution under certain mild regularity conditions. Since for realistic systems the number of states is very large or even infinite, applying standard numerical solution techniques to the CME is infeasible. If the populations of all species remain small (at most a few hundreds) then the CME can be efficiently approximated using projection methods [16, 17] or fast uniformization methods [18, 19]. Otherwise, i.e., if the system contains large populations, then analysis methods with running times independent of the population sizes have to be used such as moment closure approaches [4–6] or methods based on van Kampen's system size expansion [20, 21]. For both approaches, accurate reconstructions of the underlying probability distribution, i.e., the solution of the CME, are possible [7, 21].

Moment-based analysis

From the CME it is straightforward to derive the following equation for the derivative of the mean of a polynomial function $T : \mathbb{Z}_+^n \to \mathbb{R}$ on $\mathbf{X}(t)$.

$$\frac{d}{dt} E[T(\mathbf{X}(t))] = \sum_{j=1}^{m} E\left[\alpha_j(\mathbf{X}(t)) \cdot \left(T(\mathbf{X}(t) + v_j) - T(\mathbf{X}(t))\right)\right] \quad (1)$$

Table 1 Simple gene expression model [44]: The evolution of the molecular populations DNA$_{ON}$, DNA$_{OFF}$, and mRNA is described by the random vector $\mathbf{X}(t) = (X_1(t), X_2(t), X_3(t))$, respectively

Reactions	Propensities	Intervals
DNA$_{ON}$ $\to$ DNA$_{OFF}$	$\alpha_1(\mathbf{x}) = b \cdot x_1$	$b \in [0, 0.5]$
DNA$_{OFF}$ $\to$ DNA$_{ON}$	$\alpha_2(\mathbf{x}) = a \cdot x_2$	$a \in [0, 0.5]$
DNA$_{ON}$ $\to$ DNA$_{ON}$+ mRNA	$\alpha_3(\mathbf{x}) = c \cdot x_1$	$c \in [0, 0.5]$

Omitting the argument t of $\mathbf{X}$ and choosing $T(\mathbf{X}) = X_i, X_i^2, \ldots$ yields the following equations for the (exact) time evolution of the k-th moment $E[X_i^k]$ of the distribution for the i-th species.

$$\frac{d}{dt} E[(X_i)^k] = \sum_{j=1}^{m} E\left[\alpha_j(\mathbf{X}) \cdot \left((X_i + v_{ji})^k - (X_i)^k\right)\right], \quad (2)$$

where v_{ji} refers to the i-th component of the change vector $\mathbf{v}_j$. In a similar way, equations for mixed moments are derived.

If all reactions are at most monomolecular ($1 \geq \sum_i |v_{ji}^-|$ for all j), then no moments of order higher than k appear on the right side (also in the mixed case) and we can directly integrate all equations for moments of at most order k. However, most systems *do* contain bimolecular reactions (in particular those with complex behavior such as multistability). In this case we consider a Taylor expansion of the multivariate function

$$f(\mathbf{X}) = \alpha_j(\mathbf{X}) \cdot \left(T(\mathbf{X} + v_j) - T(\mathbf{X})\right)$$

about the mean $\mu := E[\mathbf{X}]$. It is easy to verify that, when applying the expectation to the Taylor sum, the right side only contains derivatives of f at $\mathbf{X} = \mu$, which are multiplied by central moments of increasing order. For instance, for $k = 1$ and a single species system with $n = 1$, Eq. (2) becomes

$$\frac{d}{dt} E[(X_i)] = \sum_{j=1}^{m} v_{ji} E[\alpha_j(\mathbf{X})]$$
$$= \sum_{j=1}^{m} v_{ji} \left(\alpha_j(\mu) + \frac{E[(\mathbf{X} - \mu)]}{1!} \cdot \frac{\partial}{\partial x} \alpha_j(\mu) + \frac{E[(\mathbf{X} - \mu)^2]}{2!} \cdot \frac{\partial^2}{\partial x^2} \alpha_j(\mu) + \ldots\right)$$

In the expansion, central moments of higher order may occur. For instance, in the case of bimolecular reactions, the equations for order k moments involve central moments of order $k + 1$ since second order derivatives are non-zero. By converting the non-central moments to central ones and truncating the expansion at some fixed maximal order k, we can close the system of equations when we assume that higher order central moments are zero. A full derivation of the moment equations using multi-index notation (as required for $n > 1$) can be found in [6].

The accuracy of the inference approach that we propose in the sequel depends not only on the information given by the experimental data but also on the accuracy of the approximated moments. Different closure strategies have been suggested and compared in the last years showing that the accuracy can be improved by making assumptions about the underlying distribution (e.g.

approximate log-normality) [22, 23]. In addition, the accuracy of moment-closure approximations has been theoretically investigated [24].

Hybrid approaches

Compared to deterministic models that describe only average behaviors, stochastic models provide interesting additional information about the behavior of a system. Although this comes with additional computational costs, it is in particular for systems with complex behavior, such as multimodality or oscillations, of great importance. Often the underlying source of multiple modes are discrete changes of gene activation states that are described by chemical species whose maximal count is very small (e.g. 1 for the case that the gene is either active, state 1, or inactive, state 0). Then the moment-based approaches described above can be improved (both in terms of accuracy and computation time) by considering conditional moments instead [8–10, 12, 25]. The idea is to split the set of species into species with small and large populations and consider the moments of the large populations conditioned on the current count of the small populations. For the small populations, a small master equation has to be solved additionally to the moment equations to determine the corresponding discrete distribution. More specifically, if $\hat{\mathbf{x}}$ is the subvector of $\mathbf{x}$ that describes the small populations and $\tilde{\mathbf{x}}$ is the subvector of the large populations (i.e. $\mathbf{x} = (\hat{\mathbf{x}}, \tilde{\mathbf{x}})$), then for the distribution of $\hat{\mathbf{x}}$ we have

$$\frac{d}{dt} p_t(\hat{\mathbf{x}}) = \sum_{j:\hat{\mathbf{x}} - \hat{\mathbf{v}}_j \geq 0} E[\alpha_j(\mathbf{X}) \mid \hat{\mathbf{X}} = \hat{\mathbf{x}} - \hat{\mathbf{v}}_j] \, p_t(\hat{\mathbf{x}} - \hat{\mathbf{v}}_j)$$
$$- \sum_j E[\alpha_j(\mathbf{X}) \mid \hat{\mathbf{X}} = \hat{\mathbf{x}}] \, p_t(\hat{\mathbf{x}})$$

where $\hat{\mathbf{v}}_j$ is the corresponding subvector of $\mathbf{v}_j$. Using Taylor expansion, the conditional expectations of the propensities can, as above, be expressed in terms of conditional moments of the large populations. In addition, equations for the conditional moments of the large populations can be derived in a similar way as above. For instance, the partial mean $E[\tilde{X}_i \mid \hat{x}] \, p_t(\hat{x})$ follows the time evolution

$$\frac{\partial}{\partial t} \left(E[\tilde{X}_i \mid \hat{x}] \, p_t(\hat{x}) \right)$$
$$= \sum_{j:\hat{\mathbf{x}} - \hat{\mathbf{v}}_j \geq 0} E[(\tilde{X}_i + v_{ij}) \alpha_j(\mathbf{X}) \mid \hat{\mathbf{X}} = \hat{\mathbf{x}} - \hat{\mathbf{v}}_j] \, p_t(\hat{\mathbf{x}} - \hat{\mathbf{v}}_j)$$
$$- \sum_j E[\tilde{X}_i \alpha_j(\mathbf{X}) \mid \hat{\mathbf{X}} = \hat{\mathbf{x}}] \, p_t(\hat{\mathbf{x}})$$

where on the right side again Taylor expansion can be used to replace unknown conditional expectations by conditional moments. As above a dependence on higher conditional moments may arise and a closure approach has to be applied to arrive at a finite system of equations. Unconditional moments can then be derived by summing up the weighted conditional moments. It is important to note

that if $p_t(\hat{x}) = 0$ then algebraic equations arise turning the equation system into a system of differential-algebraic equations, which renders its solution more difficult (see [12, 26] for details).

In Fig. 1 we give an example for a comparison of the accuracy of the hybrid approach and the standard moment closure (assuming that all central moments above a fixed maximal order are zero) for one of our case studies. As "exact" moment values we chose the average of 500,000 samples generated by the stochastic simulation algorithm (SSA) [27] and considered the absolute difference to the approximated moments of one chemical population until a maximal order of four. Since for our case studies we assumed 10,000 samples we additionally plot the (approximated) standard deviation of the 50 sample means taken from batches of 10,000 samples. The moments computed based on the hybrid approach show a smaller error than those computed using the standard moment closure and lie within the deviations given by the sample moments. For the example in Fig. 1 we have 126 equations for the standard approach up to an order of four. In the hybrid case there are 14 moment equations and one equation for the mode probability per mode leading to a total number of 45 equations. However, reductions are possible for the standard approach when the model structure is exploited [28]. We do not make use of these reductions here but choose the hybrid approach mainly because it gives more accurate results for the (unconditional) moments. This strongly improves the quality of the estimated parameters as demonstrated in the "Results" section.

Generalized method of moments

We assume that observations of a biochemical network were made using single-cell analysis that gives population
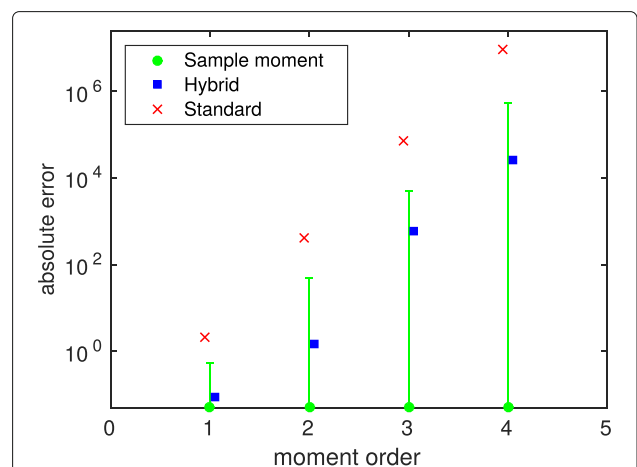
Fig. 1 Absolute error of the first four moments of P_1 for the exclusive switch model, where the moments are either computed based on a standard moment closure approach or a hybrid approach. The maximal order of the considered moments is 5

snapshot data (e.g. from flow cytometry measurements). Typically, large numbers (about 5,000–10,000 [29–31]) of independent samples can be obtained where each sample corresponds to one cell. It is possible to simultaneously observe one or several chemical populations at a time in each single cell. In the sequel, we first describe the inference procedure for a single observation time point and a single chemical species that is observed. Later, we extend this to several time points and species.

For a fixed measurement time t and a fixed index i of the observed population we can define the r-th order sample moment as

$$\hat{m}_r = \frac{1}{N} \sum_{\ell=1}^{N} Y_\ell^r,$$

where Y_ℓ is the ℓ-th sample of the observed molecular count of the i-th species at time t and there are N samples in total. For large N, the sample moments are asymptotically unbiased estimators of the population moments.

Let θ be a vector of, say, $q \leq m$ unknown reaction rate constants[1], for which some biologically relevant range is known. Moreover, let m_r be the r-th theoretical moment, i.e., $m_r(\theta) := E[Y_\ell^r]$. In the sequel we also simply write Y instead of Y_ℓ whenever Y appears inside the expectation operator or when the specific index of the sample is not relevant. An obvious inference approach would be to consider the ordinary least squares estimator

$$\hat{\theta} = \arg \min_\theta \sum_{r=1}^{k} \left(\hat{m}_r - m_r(\theta)\right)^2, \tag{3}$$

where k is the number of moment constraints. Under certain conditions related to the identification of the parameters as discussed below, this estimator is consistent (converges in probability to the true value of θ) and asymptotically normal. However, its variance may be very high. This is due to the fact that for increasing order the variance of the sample moments increases and so does the variance of the estimator. This problem can be mitigated by choosing appropriate weights for the summands in (3). Moreover, since correlations between the cost functions

$$g_r(\theta) = \hat{m}_r - m_r(\theta)$$

exist, a more general approach that considers mixed terms is needed. This leads to a class of estimators, called generalized method of moments (GMM) estimators that have been introduced by Hansen [13]. The idea is to define the estimator as

$$\hat{\theta} = \arg \min_\theta \mathbf{g}(\theta)' W \mathbf{g}(\theta) \tag{4}$$

where $\mathbf{g}(\theta)$ is the column vector with entries $g_r(\theta), r = 1, \ldots, k$, and W is a positive semi-definite weighting

matrix. Note that by defining $f_r(Y, \theta) = Y^r - m_r(\theta)$ we see that

$$g_r(\theta) = \frac{1}{N} \sum_\ell f_r(Y_\ell, \theta) = \frac{1}{N} \sum_\ell Y_\ell^r - m_r(\theta)$$

is the sample counterpart of the expectation $E[f_r(Y, \theta)]$. The latter satisfies

$$\theta_0 = \arg \min_\theta E[\mathbf{f}(Y, \theta)]' \, W E[\mathbf{f}(Y, \theta)],$$

where $\mathbf{f}(Y, \theta)$ is the column vector with entries $f_r(Y, \theta)$ and θ_0 is the true value of θ. Note that the choice $W = I$ gives the least-squares estimator with k terms while for general W there are $\frac{k \cdot (k+1)}{2}$ terms in the objective function (with k being the dimension of $\mathbf{g}(\theta)$). In addition, we remark that in general W may depend on θ and/or the samples Y_ℓ.

Here we assume that identification of θ is possible, i.e., we require that $q \leq k$, i.e., the number of the moment constraints used is at least as large as the number of unknown parameters and

$$E[\mathbf{f}(Y, \theta)] = \mathbf{0} \text{ if and only if } \theta = \theta_0.$$

In addition, the theoretical moments $m_r(\theta)$ should not be functionally dependent (see Chapter 3.3 in [32]) to ensure that the information contained in the moment conditions is sufficient for successfully identifying the parameters.

By applying the central limit theorem to the sample moments, it is possible to show that the GMM estimator is consistent and asymptotically normally distributed and that its variance becomes asymptotically minimal if the matrix W is chosen such that it is proportional to the inverse of the covariances between the Y_ℓ^r [13]. This result is intuitive since usually higher moments might be more volatile than others and, thus, it makes sense to normalize the errors in the moments by the corresponding covariance. Formally, we define $\mathbf{Y}_\ell$ as the random vector with entries $(Y_\ell)^r$ for $r = 1, \ldots, k$ and, as before, omit the subindex ℓ if it is not relevant. Then,

$$F(\theta_0) = COV[\mathbf{Y}, \mathbf{Y}] = E[\mathbf{f}(Y, \theta_0)\mathbf{f}(Y, \theta_0)^T]$$

and choosing $W \propto F^{-1}$ will give an estimator with smallest possible variance, i.e., it is asymptotically efficient in this class of estimators [13, 32].

Since F depends on the true value θ_0, a two-step updating procedure has been suggested [13] during which W is chosen as the identity matrix I in the first step such that an initial estimate $\tilde{\theta}$ is computed. In a second step, F is estimated by the sample counterpart of $E[\mathbf{f}(Y, \tilde{\theta})\mathbf{f}(Y, \tilde{\theta})^T]$, i.e.,

$$\hat{F}_1(\tilde{\theta}) = \frac{1}{N} \sum_{\ell=1}^{N} \mathbf{f}(Y_\ell, \tilde{\theta})\mathbf{f}(Y_\ell, \tilde{\theta})^T. \tag{5}$$

If, however, the model is "misspecified", i.e., there is no θ_0 for which

$$E[\mathbf{f}(Y, \theta_0)] = \mathbf{0},$$

then the above estimator is no longer consistent. In particular, if the theoretical moments are poorly approximated, it is likely that also the accuracy of the resulting estimates is poor. An estimator for F that is consistent is then given by [32]

$$\hat{F}_2 = \frac{1}{N} \sum_{\ell=1}^{N} (\mathbf{Y}_\ell - \overline{\mathbf{Y}})(\mathbf{Y}_\ell - \overline{\mathbf{Y}})^T, \qquad (6)$$

where $\overline{\mathbf{Y}}$ is the vector with entries $\frac{1}{N} \sum_{\ell=1}^{N} \mathbf{Y}_\ell^r$. In the sequel we refer to the estimator based on (6) as the *demean estimator*. This estimator removes the inconsistencies in the covariance matrices estimated from the sample moments by "demeaning". Since moment-based analysis methods usually give approximations of the moments and not the exact values, we consider both, the demean estimator defined by (6) and the estimator of the 2-step procedure in (5) for our numerical results.

The estimation procedure described above can be generalized to several dimensions by also using mixed sample moments instead of only $\hat{m}_r$ and mixed theoretical moments instead of only $m_r(\theta)$. For instance, for moments up to order two and two simultaneously observed species X and Y, we use the cost functions

$$\begin{aligned}
g_1(\theta) &= \frac{1}{N} \sum_{\ell=1}^{N} X_\ell - E[X_\ell \mid \theta] \\
g_2(\theta) &= \frac{1}{N} \sum_{\ell=1}^{N} Y_\ell - E[Y_\ell \mid \theta] \\
g_3(\theta) &= \frac{1}{N} \sum_{\ell=1}^{N} X_\ell Y_\ell - E[X_\ell Y_\ell \mid \theta] \\
g_4(\theta) &= \frac{1}{N} \sum_{\ell=1}^{N} X_\ell^2 - E[X_\ell^2 \mid \theta] \\
g_5(\theta) &= \frac{1}{N} \sum_{\ell=1}^{N} Y_\ell^2 - E[Y_\ell^2 \mid \theta].
\end{aligned}$$

In the same way, we can extend the estimators $\hat{F}_1$ and $\hat{F}_2$ to several dimensions. For instance, the covariance between $X_\ell Y_\ell$ and X_ℓ^2 can be estimated as

$$\frac{1}{N} \sum_{\ell=1}^{N} (X_\ell Y_\ell - \overline{XY})(X_\ell^2 - \overline{X^2}),$$

where again we use $\overline{*}$ to denote the sample mean operator.

If, instead of snapshot data for a single observation time, independent samples for different times are available then the GMM estimator can also be easily generalized to

$$\hat{\theta} = \arg\min_\theta \sum_{t=t_0}^{t_f} \mathbf{g}^{(t)}(\theta)' W^{(t)} \mathbf{g}^{(t)}(\theta). \qquad (7)$$

Here, for each time point $t \in \{t_0, \dots, t_f\}$ the vector of cost functions $\mathbf{g}^{(t)}$ is calculated as before and the minimum is taken over the sum of these uncorrelated cost functions. Note that for each observation time point a weight matrix $W^{(t)}$ has to be computed. In the two-step approach, the initial weight matrices are all equal to the identity matrix and then in the second step different weight matrices may arise since the estimator of F depends on Y, which in turn depends on the distribution of the model at the specific time t.

Results

To analyze the performance of the GMM we consider two case studies (see Additional file 1), the simple gene expression model in Table 1 and a network of two genes with mutual repression, called exclusive switch [33]. The reactions of the exclusive switch are listed in Table 2. All propensities follow the law of mass action. For the parameters that we chose, the corresponding probability distribution is bi-modal.

For fixed reaction rate constants and initial conditions, we used the SSA to generate trajectories of the systems and record samples of the size of the corresponding protein/mRNA populations. In addition, we used the software tool SHAVE [34] to generate moment equations both for the standard moment closure and for the hybrid approach. In SHAVE the partial moments are integrated instead of the conditional moments such that the differential-algebraic equations are transformed into a system of (ordinary) differential equations after truncating modes with insignificant probabilities. Then an accurate approximation of the solution using standard numerical integration methods can be obtained. The system of moment equations is always closed by setting all central moments of order $> k$ to zero. We used for the inference approach only the moments up to order $k - 1$ since the precision of the moments of highest order k is often poor. SHAVE allows to export the (hybrid) moment equations as a MATLAB-compatible m-file. We then used MATLAB's ode45 solver, which is based on a fifth order Runge-Kutta method, to integrate the (hybrid) moment equations. Note that for the gene expression example, the moment equations are exact since all propensities are linear. Thus, even an analytic solution is possible for this system.

Table 2 Exclusive switch model [33]: Two different proteins P_1 and P_2 can bind to a promotor region on the DNA. If P_1 is bound to the promotor the production of P_2 is inhibited and vice versa. In the free state both proteins can be produced

	Reactions, $i = 1, 2$		Rate constant	Interval
DNA	$\rightarrow$	DNA + P_i	p_i	[0.5,1.5]
DNA.P_i	$\rightarrow$	DNA.P_i + P_i	p_i	[0.5,1.5]
P_i	$\rightarrow$	$\emptyset$	d_i	[0,0.05]
DNA + P_i	$\rightarrow$	DNA.P_i	b_i	[0,0.1]
DNA.P_i	$\rightarrow$	DNA + P_i	u_i	[0,0.1]

We then used MATLAB's Global Search routine to minimize the objective function in Eq. (4). Global Search is a method for finding the global minimum by starting a local solver from multiple starting points that are chosen according to a heuristic [35]. Therefore the total running time of our method depends on the tightness of the intervals that we use as constraints for the unknown parameters as well as on the starting points of the Global Search procedure. The running times for one local solver call (using the hybrid approach for computing moments) were about 2 s (demean estimator) and 40 s (2-Step estimator) for the gene expression model. For the exclusive switch the average running time for a local solver call was about 2 min (demean) and 10 min (2-Step). Note that the total running time depends on the amount of local solver calls carried out by Global Search, which varied between 2 and 50. For all experiments we chose a single initial point that is located far away from the true values and allowed Global Search to choose 500 (potential) further starting points. Different initial points yielded similar results except if the initial points is chosen close to the true values (then the results are significantly better in particular in the case of only few moment constraints).

The intervals that we used as constraints for the parameters are all listed in Tables 1 and 2.

Standard vs. hybrid moment-based analysis

In Fig. 2 we plot the results of a comparison between the standard and the hybrid moment closure when it is performed during the optimization procedure of the GMM inference approach. We chose the exclusive switch model for this since for this model the accuracy of the standard approach is poor. As an estimator for F we used (6), which is based on demeaning (demean). Results for the 2-step procedure show similar differences when standard and hybrid moment closure are compared. We fixed the degradation rates to ensure that identification of p_1 and p_2 is possible when the two protein populations are measured at only a single observation time point. To simultaneously identify all parameters (including p_1 and p_2) several observation time points are necessary (see Additional file 2).

The true values of the six unknown parameters are plotted against the means and standard deviations of the estimated values for a maximal moment order of 2 and 3, where for each of the six unknown parameters 50 estimations based on 10,000 samples each were used.
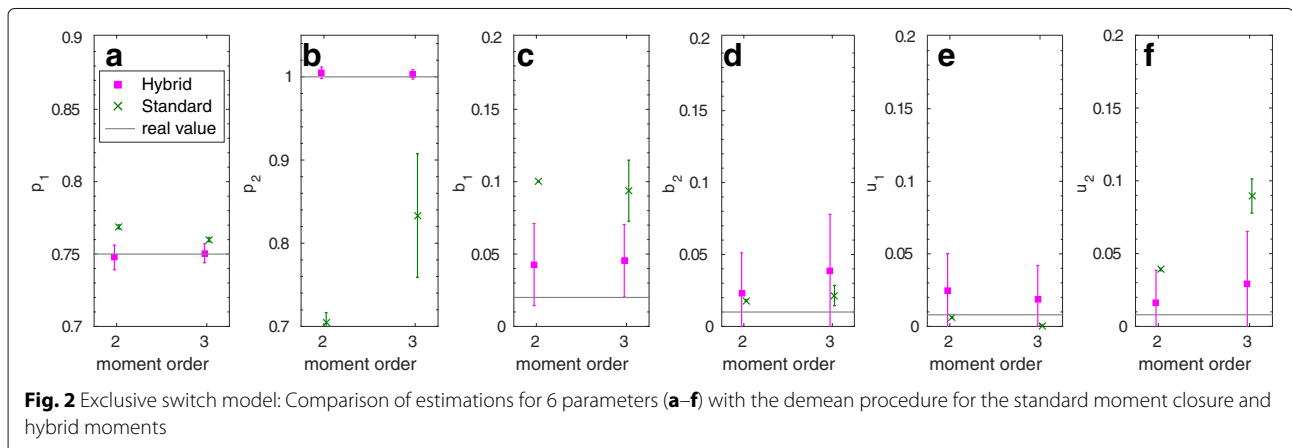
We see that the inaccurately approximated moments of the standard approach lead to severe problems in the inference approach. Nearly all parameters are estimated more accurately when the hybrid moment closure is used. For parameter b_1 most of the optimization runs converged to the upper limit of the given interval (0.1) when the standard approach was used. For the results in the sequel, we only used the hybrid moment closure.
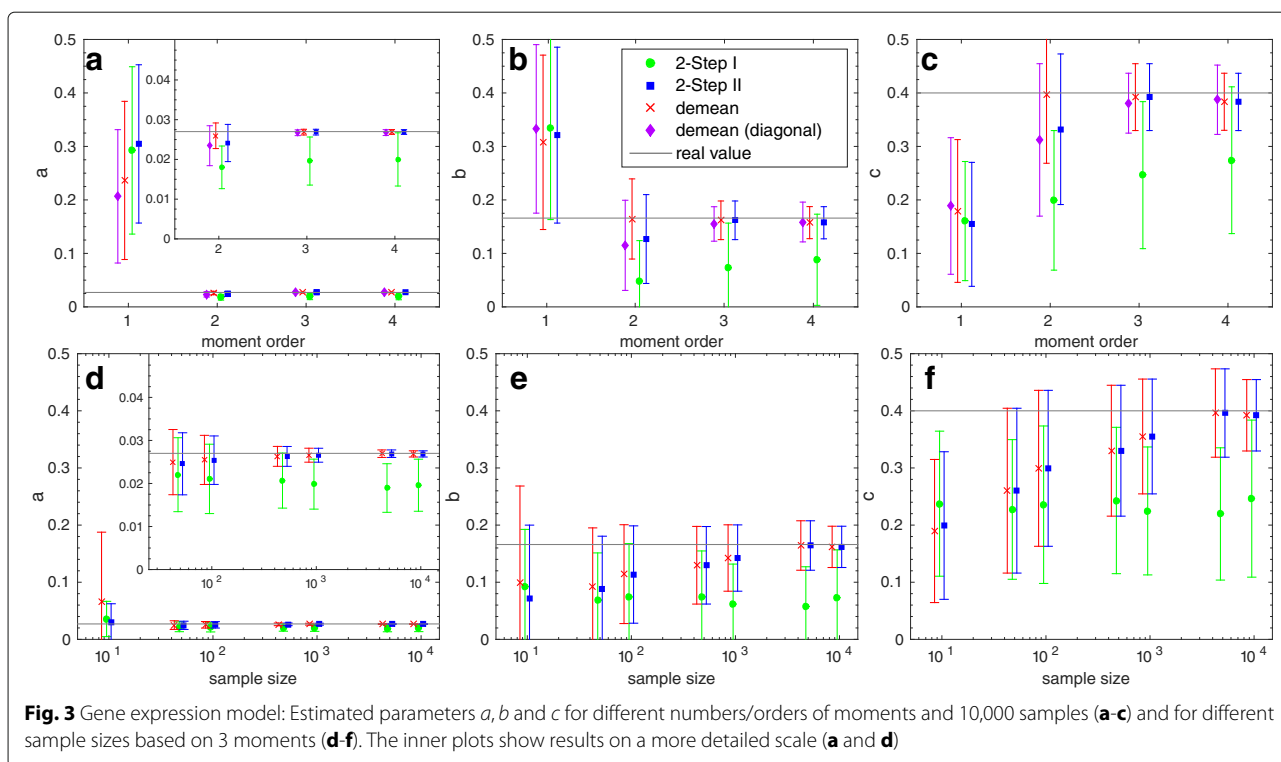
Two-step vs. demean approach

In Figs. 3 and 4 we plot results of the GMM approach applied to the two example networks, where we compare the performance of the two-step estimator in Eq. (5) with the demean estimator in Eq. (6). We plot the true values of the parameters against the estimated values, where 2-Step I is the result of the first step of the two-step procedure (with $W = I$) and 2-Step II that of the second step (with $W = \hat{F}_1$ and $\hat{F}_1$ as defined in Eq. (5)).

For the results in Fig. 3 only one population (mRNA) was observed at $t = 100$ where the initial conditions were such that $DNA_{OFF} = 1$, $DNA_{ON} = 0$ and 10 mRNA molecules were present in the system. For three parameters the means and standard deviations of the estimated values are plotted, again based on 50 repetitions of the inference procedure.

In the first row of Fig. 3 the accuracy of the estimation is compared with respect to the number/order of moments considered, where again for each of the 50 estimated values 10,000 samples were used. We see that if only one moment is considered or if equal weights are used for the first two moments, only a rough estimate is possible since identification is not possible. The accuracy is

Fig. 2 Exclusive switch model: Comparison of estimations for 6 parameters (**a**–**f**) with the demean procedure for the standard moment closure and hybrid moments

Fig. 3 Gene expression model: Estimated parameters *a*, *b* and *c* for different numbers/orders of moments and 10,000 samples (**a-c**) and for different sample sizes based on 3 moments (**d-f**). The inner plots show results on a more detailed scale (**a** and **d**)

markedly improved when the weights are chosen according to the demean approach. Here, it is important to note that for a maximal order of $k = 2$, in W we also consider, besides the squared cost functions $g_1(\theta)^2$ and $g_2(\theta)^2$, the mixed term $g_1(\theta)g_2(\theta)$. This additional term significantly improves the quality of the estimation such that it is possible to achieve a good estimation of the parameters with only the sample mean and the sample second moment. To further investigate the positive influence of the mixed term, we additionally plot results for the case that only variances are estimated, referred to as 'demean (diagonal)', i.e., the weight matrix is the inverse of a diagonal matrix that contains the variances estimated based on the demean approach.

However, the variance of the estimator for a maximum order of two is relatively high but decreases significantly when also the third (and fourth) moment is considered. Here, demean and the second step of the two-step procedure perform equally well and also demean (diagonal) gives very good results. Opposed to this $W = I$ (first step of two-step procedure) gives poor results and a high variance also if higher moments are considered.
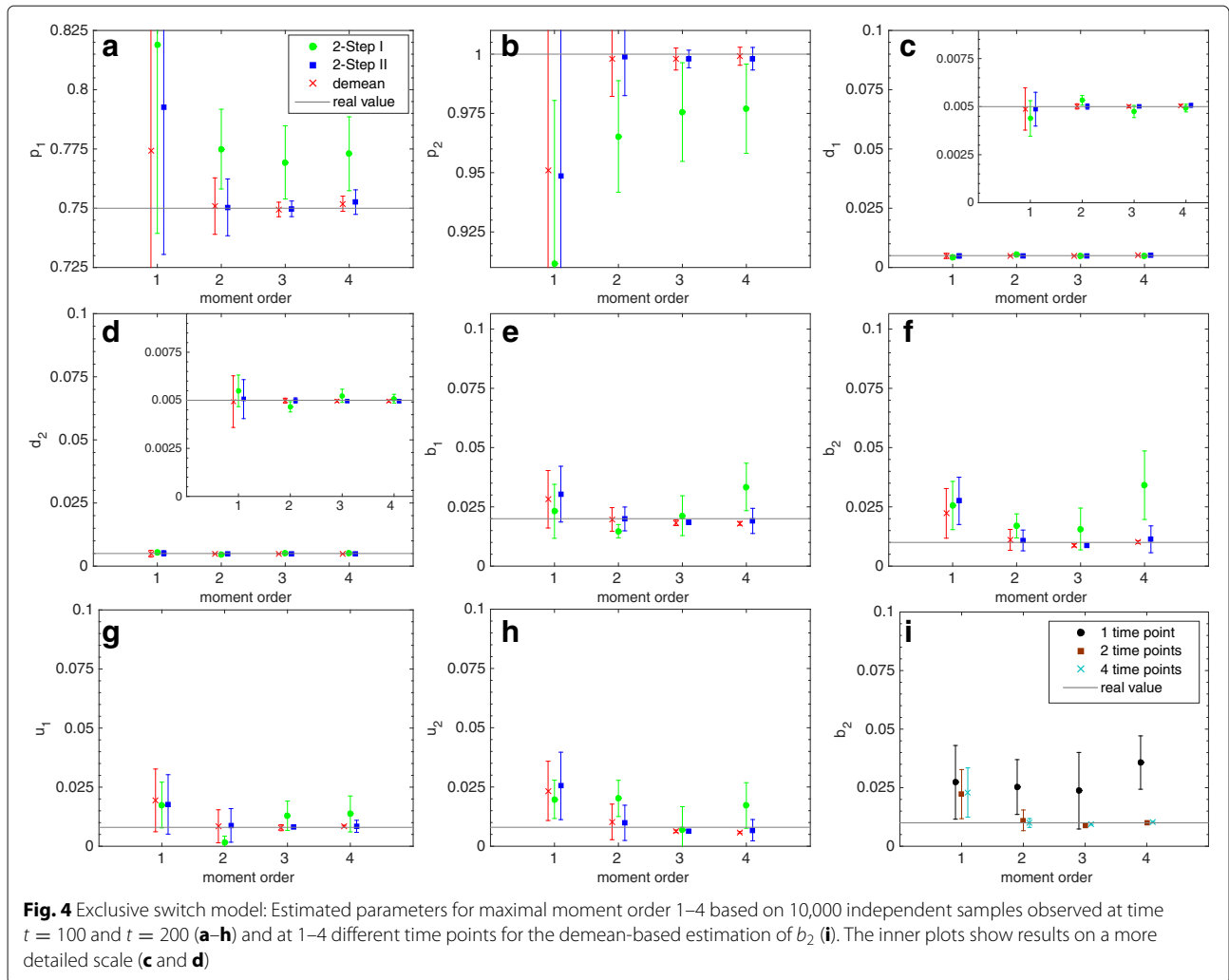
In Table 3 we give an example for the (normalized) matrix W as used for demean and 2-Step II. The two methods choose nearly identical weights and the mean has the highest weight. Then, the mixed cost function for mean and second moment has a (negative) weight of about $2 \cdot (-4.95)$ % since these moments are negatively correlated (and so are the second and third moment). All terms

that involve the third moment have a very small weight as their covariances are high.

It is important to note that also if the number of moment constraints, k, is equal to the number of parameters, q, 2-Step I performs poor (see results for maximal order $k = 3$ in the first row of Fig. 3). The reason is that in this example identification is not possible if only three terms are used due to functional dependencies between the parameters of the first two reactions and due to the fact that only at a single time point measurements were made. If identification was possible and the computed population moments were exact, the results should be independent of the choice of W for the case that q equals k.

Thus, the weights given by the estimators for F in (5) and (6) substantially increase the accuracy of the results and allow identification, because additional information about the covariances between the Y^r are used. Moreover, due to the off-diagonal entries of W additional mixed terms are part of the objective function.

In the second row in Fig. 3, we compare the accuracy for different samples sizes where the first three moments were considered. While 2-Step I does not show a systematic improvement when the number of samples increases, we see for 2-Step II and demean not only significantly improved estimates but also smaller variances. However, in the case of few samples, demean gives in particular for parameter a a high variance. This comes from the fact that the corresponding estimator uses the sample mean

Fig. 4 Exclusive switch model: Estimated parameters for maximal moment order 1–4 based on 10,000 independent samples observed at time $t = 100$ and $t = 200$ (**a**–**h**) and at 1–4 different time points for the demean-based estimation of b_2 (**i**). The inner plots show results on a more detailed scale (**c** and **d**)

instead of the theoretical mean and therefore the weight matrix is far from optimal if N is small.

In Fig. 4, **a**–**h**, we plot results for the exclusive switch model where all eight parameters were estimated based on observations of the two protein populations of P_1 and P_2 at two time points. On the x-axis the maximal order

of moments used is plotted. For the orders 1, 2, 3 and 4 there are in total 2, 5, 9 or 14 moments, respectively. Again, 2-Step II and demean both give accurate results from a maximal order of two on, whereas 2-Step I gives poor results. In addition, the variance of the estimator decreases with increasing maximal order. However, the values for 2-Step II become slightly worse and have higher variance for a maximal order of four since these moments are not approximated very accurately. Also the accuracy of the demean estimator does not improve when the maximum order is increased from three to four. Thus, the cost functions of order four moments do not lead to any significant improvement in this example and should be excluded.

Further estimators

For our results we focused on the most popular GMM estimators, that is, demean and two-step. However, we also implemented two additional variants of estimators that are frequently described in the GMM literature

Table 3 Weight matrices for the two-step and demean procedure with moment order 3 for the gene expression model

	W		
	1	-0.0495	0.0007
Two-step	−0.0495	0.0025	$-3.86e^{-5}$
	0.0007	$-3.86e^{-5}$	$6.11e^{-7}$
	1	-0.0494	0.0007
Demean	−0.0494	0.0025	$-3.85e^{-5}$
	0.0007	$-3.85e^{-5}$	$6.09e^{-7}$

The entries are normalized with respect to the weight for the mean and rounded (the original weight matrices are both positive semi-definite)

[14, 36]. One is the estimator that results from further iterations of the two-step procedure (iterated GMM estimator [36]). However, in our examples we did not see an increase in accuracy after the second iteration. The second approach is the continuously updating GMM estimator [36], where we use in Eq. (4) the weight matrix $W(\theta) = (\hat{F}_1(\theta))^{-1}$ of Eq. (5) and the argument θ is not fixed for the optimization but optimized simultaneously with the argument of $\mathbf{g}(\theta)$. The results for this approach did not show increased accuracy, also when we used results of the other estimators (e.g. demean) as starting points for the optimization. Moreover, for large weight matrices, the recomputation in each step of the optimization resulted in longer running times.

Overall, our experiments show that for sufficiently large N the demean estimator usually yields the best results, while two-step performs better for small N. Moreover, choosing three as the maximum order gave the best results (accurate average value and small standard deviations) for the examples that we considered.

Discussions

In the context of stochastic chemical kinetics, parameter inference methods are either based on Markov chain Monte Carlo schemes [37–40], on approximate Bayesian computation techniques [41–43] or on maximum likelihood estimation using a direct approximation of the likelihood [2, 44] or a simulation-based estimate [45, 46]. Maximum likelihood estimators are, in a sense, the most informative estimates of unknown parameters [47] and have desirable mathematical properties such as unbiasedness, efficiency, and normality. On the other hand, the computational complexity of maximum likelihood estimation is high as it requires a simulation-based or numerical solution of the CME for many different parameter instances. Since the applicability of these methods is limited, approaches based on moment closure [3, 23, 48–51] or linear noise approximations [52–54] have been developed. An approximation of the likelihood of order-two sample moments is maximized in [23, 48, 49, 51]. The approach exploits that for large numbers of samples these sample moments are asymptotically normally distributed. The negative log-likelihood leads to an optimization problem where the differences between the sample and theoretical moments up to order two are weighted and minimized as well. As opposed to the GMM, the weight matrix in [48, 49] is estimated based on the theoretical moments of the model up to order four and independent of the samples while in the GMM approach this matrix depends on the samples (and theoretical moments up to order two). Moreover, the objective function contains an additional summand, which is the logarithm of the determinant of the estimated covariance matrix. In [23], Bogomolov et al. insert sample instead of theoretical

moments in the derived formulas for the covariances of moment conditions up to order two. A comparison for the two examples that we consider in the previous section yields that when the theoretical moments are used to estimate covariances, similar to the continuously updating GMM, optimization was slow and sometimes failed to return the global optimum due to a much more complex landscape of the objective function. When sample moments are considered as suggested in [23], the results are similar to those of the GMM demean estimator for a maximum order of two. In [51], only variances are considered (weight matrix is diagonal) and estimated based on the samples. Therefore, it does not exploit the information contained in the mixed terms, which lead to improved estimates in our examples (see results for 'demean (diagonal)' in Fig. 3).

A similar approach is used in [3] where the moment equations are closed by a Gaussian approximation. The parameter estimation is based on using a ML estimator and a Markov chain Monte-Carlo approach. In [50] the importance of higher moment orders when using least square estimators is shown. Weights for terms that correspond to different moments are chosen ad hoc and not based on any statistical framework.

Here, we present results for the general method of moments that assigns optimal weights to the different moment conditions for an arbitrary maximal moment order and number of species. We showed that trivial weights (e.g. identity matrix) give results whose accuracy can be strongly increased when optimal weights are chosen. In the very common case that functional dependencies between parameters exist (e.g. degradation and production of the same species) and identification is difficult, the GMM estimator allows to accurately identify the parameters. Moreover, our results indicate that the accuracy of the estimation increases when moments of order higher than two are included. A general strategy could be to start with $k = q$ cost functions (equal to the number of unknown parameters) and increase the maximal order until tests for over-identifying restrictions (e.g. the Hansen test [13]) suggest that higher orders do not lead to an improvement. In this way, cost functions that do not improve the quality of the estimation, such as the fourth order cost functions for the results in Fig. 4, can be identified.

We also found that an accurate approximation of the moments is crucial for the performance of the GMM estimator. Thus, hybrid approaches such as the method of conditional moments [12] or sophisticated closure schemes (e.g. [23]) should be preferred. If all propensities in the network are linear, the moment equations are exact and model misspecification is not an issue. However, for most networks the moments can only be approximated, since the propensities are nonlinear, and hence the model

is potentially misspecified. Again, statistical tests can be used to detect model misspecification [32] and equations for higher order moments may be added to the (conditional) moment equations to improve the approximation of the lower order moments.

Finally, we note that the GMM framework can also be applied when the observed molecular counts are subject to measurement errors. It is straight forward to extend the GMM framework to the case of samples $Y_\ell + \varepsilon$ where the error term ε is independent and normally distributed with mean zero.

Conclusion

Parameter inference for stochastic models of cellular processes demands huge computational resources. The proposed approach based on the generalized method of moments is based on an adjustment of the statistical moments of the model and therefore does not require the computation of likelihoods. This makes the approach appealing for complex networks where stochastic effects play an important role, since the integration of the moment equations is typically fast compared to other computations such as the computation of likelihoods. The method does not make any assumptions about the distribution of the process (e.g. Gaussian) and complements the existing moment-based analysis approaches in a natural way.

Here, we used a multistart gradient-based minimization scheme, but the approach can be combined with any global optimization method. We found that the weights of the cost functions computed by the GMM estimator yield clearly more accurate results than trivial (identical) weights. In particular, the variance of the estimator decreases when moments of higher order are considered. We focused on the estimation of reaction rate constants and, as future work, we plan to investigate how well Hill coefficients and initial conditions are estimated.

An important advantage of the proposed method is that in the economics literature the properties of GMM estimators have been investigated in detail over decades and several variants and related statistical tests are available. We will also check how accurate approximations for the variance of the GMM estimator are [32]. Since we found that when moments of order higher than three are included, the results become slightly worse, we will in addition explore the usefulness of statistical tests for over-identifying moment conditions. In this way, we can ensure that only moments conditions are included that improve the estimation.

Endnote

[1] It is straightforward to adapt the approach that we present in the sequel to the case that other unknown continuous parameters have to be estimated.

Additional files

Abbreviations
CME: Chemical master equation; GMM: Generalized method of moments; ML: Maximum likelihood; SSA: Stochastic simulation algorithm

References

1. McQuarrie DA. Stochastic approach to chemical kinetics. J Appl Probab. 1967;4:413–78.
2. Andreychenko A, Mikeev L, Spieler D, Wolf V. Approximate maximum likelihood estimation for stochastic chemical kinetics. EURASIP J Bioinforma Syst Biol. 2012;9:1–14.
3. Milner P, Gillespie CS, Wilkinson DJ. Moment closure based parameter inference of stochastic kinetic models. Stat Comput. 2013;23(2):287–95.
4. Singh A, Hespanha JP. Lognormal moment closures for biochemical reactions. In: Decision and Control, 2006 45th IEEE Conference On. Piscataway: IEEE; 2006. p. 2063–8.
5. Ale A, Kirk P, Stumpf MPH. A general moment expansion method for stochastic kinetic models. J Chem Phys. 2013;138(17):174101.
6. Engblom S. Computing the moments of high dimensional solutions of the master equation. Appl Math Comput. 2006;180(2):498–515.
7. Andreychenko A, Mikeev L, Wolf V. Model Reconstruction for Moment-Based Stochastic Chemical Kinetics. ACM TOMACS. 2015;25(2):1–19.
8. Henzinger T, Mateescu M, Mikeev L, Wolf V. Hybrid numerical solution of the chemical master equation. In: Proc. of CMSB'10. New York: ACM DL; 2010.
9. Menz S, Latorre JC, Schutte C, Huisinga W. Hybrid stochastic–deterministic solution of the chemical master equation. SIAM MMS. 2012;10(4):1232–62.
10. Hellander A, Lötstedt P. Hybrid method for the chemical master equation. J Comput Phys. 2007;227(1):100–22.

11. Jahnke T. On reduced models for the chemical master equation. SIAM MMS. 2011;9(4):1646–76.
12. Hasenauer J, Wolf V, Kazeroonian A, Theis FJ. Method of conditional moments (MCM) for the chemical master equation. J Math Biol. 2013;69(3):687–735.
13. Hansen LP. Large sample properties of generalized method of moments estimators. Econometrica. 19821029–54.
14. Hall AR, et al. Generalized Method of Moments. Oxford: Oxford University Press; 2005.
15. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J Phys Chem. 1977;81(25):2340–61.
16. Henzinger T, Mateescu M, Wolf V. Sliding window abstraction for infinite Markov chains. In: Proc. CAV. LNCS. Heidelberg: Springer; 2009.
17. Munsky B, Khammash M. The finite state projection algorithm for the solution of the chemical master equation. J Chem Phys. 2006;124:044144.
18. Mateescu M, Wolf V, Didier F, Henzinger TA. Fast adaptive uniformisation of the chemical master equation. IET Syst Biol. 2010;4(6):441–52.
19. Sidje R, Burrage K, MacNamara S. Inexact uniformization method for computing transient distributions of Markov chains. SIAM J Sci Comput. 2007;29(6):2562–80.
20. Van Kampen NG, Vol. 1. Stochastic Processes in Physics and Chemistry. Amsterdam: Elsevier; 1992.
21. Thomas P, Grima R. Approximate probability distributions of the master equation. Phys Rev E. 2015;92(1):012120.
22. Schnoerr D, Sanguinetti G, Grima R. Comparison of different moment-closure approximations for stochastic chemical kinetics. J Chem Phys. 2015;143(18):185101.
23. Bogomolov S, Henzinger TA, Podelski A, Ruess J, Schilling C. Adaptive moment closure for parameter inference of biochemical reaction networks. In: Proc. of CMSB'15. Heidelberg: Springer; 2015. p. 77–89.
24. Grima R. A study of the accuracy of moment-closure approximations for stochastic chemical kinetics. J Chem Phys. 2012;136(15):154105.
25. Soltani M, Vargas-Garcia CA, Singh A. Conditional moment closure schemes for studying stochastic dynamics of genetic circuits. IEEE Trans Biomed Circ Syst. 2015;9(4):518–26. doi:10.1109/TBCAS.2015.2453158.
26. Kazeroonian A, Fröhlich F, Raue A, Theis FJ, Hasenauer J. CERENA: Chemical reaction network analyzer: A toolbox for the simulation and analysis of stochastic chemical kinetics. PloS ONE. 2016;11(1):0146732.
27. Gillespie DT. A general method for numerically simulating the time evolution of coupled chemical reactions. J Comput Phys. 1976;22:403–34.
28. Ruess J. Minimal moment equations for stochastic models of biochemical reaction networks with partially finite state space. J Chem Phys. 2015;143(24):244103.
29. Hanley MB, Lomas W, Mittar D, Maino V, Park E. Detection of low abundance RNA molecules in individual cells by flow cytometry. PLoS ONE. 2013;8(2):1–8. doi:10.1371/journal.pone.0057002.
30. Hasenauer J, Waldherr S, Doszczak M, Radde N, Scheurich P, Allgöwer F. Identification of models of heterogeneous cell populations from population snapshot data. BMC Bioinforma. 2011;12(1):1–15. doi:10.1186/1471-2105-12-125.
31. Munsky B, Fox Z, Neuert G. Integrating single-molecule experiments and discrete stochastic models to understand heterogeneous gene transcription dynamics. Methods. 2015;85:12–21. Inferring Gene Regulatory Interactions from Quantitative High-Throughput Measurements. doi:10.1016/j.ymeth.2015.06.009.
32. Hall AR. Generalized Method of Moments. Advanced Texts in Econometrics. Oxford: Oxford University Press; 2004.
33. Loinger A, Lipshtat A, Balaban NQ, Biham O. Stochastic simulations of genetic switch systems. Phys Rev E. 2007;75:021904.
34. Lapin M, Mikeev L, Wolf V. SHAVE – Stochastic hybrid analysis of Markov population models. In: Proc. of HSCC'11. ACM International Conference Proceeding Series. New York: ACM; 2011.
35. Ugray Z, Lasdon L, Plummer J, Glover F, Kelly J, Martí R. Scatter search and local nlp solvers: A multistart framework for global optimization. INFORMS JOC. 2007;19(3):328–40.
36. Hansen LP, Heaton J, Yaron A. Finite-sample properties of some alternative GMM estimators. J Bus Econ Stat. 1996;14(3):262–80.
37. Boys R, Wilkinson D, Kirkwood T. Bayesian inference for a discretely observed stochastic kinetic model. Stat Comp. 2008;18:125–35.
38. Wilkinson DJ. Stochastic Modelling for Systems Biology. Boca Raton: CRC press; 2011.
39. Golightly A, Wilkinson DJ. Bayesian parameter inference for stochastic biochemical network models using particle markov chain monte carlo. Interface focus. 2011;1(6):807–20.
40. Daigle BJ, Roh MK, Petzold LR, Niemi J. Accelerated maximum likelihood parameter estimation for stochastic biochemical systems. BMC Bioinforma. 2012;13(1):68.
41. Drovandi CC, Pettitt AN. Estimation of parameters for macroparasite population evolution using approximate bayesian computation. Biometrics. 2011;67(1):225–33.
42. Fearnhead P, Prangle D. Constructing summary statistics for approximate bayesian computation: semi-automatic approximate bayesian computation. J R Stat Soc Series B Stat Methodol. 2012;74(3):419–74.
43. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf M. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J R Soc Interface. 2009;6(31):187–202.
44. Reinker S, Altman RM, Timmer J. Parameter estimation in stochastic biochemical reactions. IEEE Proc Syst Biol. 2006;153:168–78.
45. Tian T, Xu S, Gao J, Burrage K. Simulated maximum likelihood method for estimating kinetic rates in gene expression. Bioinformatics. 2007;23:84–91.
46. Uz B, Arslan E, Laurenzi I. Maximum likelihood estimation of the kinetics of receptor-mediated adhesion. J Theor Biol. 2010;262(3):478–87.
47. Higgins JJ. Bayesian inference and the optimality of maximum likelihood estimation. Int Stat Rev. 1977;45(1):9–11.
48. Zechner C, Ruess J, Krenn P, Pelet S, Peter M, Lygeros J, Koeppl H. Moment-based inference predicts bimodality in transient gene expression. PNAS. 2012;109(21):8340–5.
49. Ruess J, Lygeros J. Moment-based methods for parameter inference and experiment design for stochastic biochemical reaction networks. ACM TOMACS. 2015;25(2):8.
50. Kügler P. Moment fitting for parameter inference in repeatedly and partially observed stochastic biological models. PloS ONE. 2012;7(8): 43001.
51. Fröhlich F, Thomas P, Kazeroonian A, Theis FJ, Grima R, Hasenauer J. Inference for stochastic chemical kinetics using moment equations and system size expansion. PLoS Comput Biol. 2016;12(7):1–28. doi:10.1371/journal.pcbi.1005030.
52. Komorowski M, Finkenstädt B, Harper CV, Rand DA. Bayesian inference of biochemical kinetic parameters using the linear noise approximation. BMC Bioinforma. 2009;10(1):1–10. doi:10.1186/1471-2105-10-343.
53. Zimmer C, Sahle S. Deterministic inference for stochastic systems using multiple shooting and a linear noise approximation for the transition probabilities. IET Syst Biol. 2015;9(5):181–92.
54. Bergmann FT, Sahle S, Zimmer C. Piecewise parameter estimation for stochastic models in COPASI. Bioinformatics. 2016;32(10):1586–8.