

Generalized Method of Moments for Stochastic Reaction Networks in Equilibrium

Michael Backenköhler¹, Luca Bortolussi^{1,2}, and Verena Wolf¹(✉)

¹ Computer Science Department, Saarland University,
Saarbrücken, Germany
wolf@cs.uni-saarland.de

² Department of Mathematics and Geosciences,
University of Trieste, Trieste, Italy

Abstract. Calibrating parameters is a crucial problem within quantitative modeling approaches to reaction networks. Existing methods for stochastic models rely either on statistical sampling or can only be applied to small systems. Here we present an inference procedure for stochastic models in equilibrium that is based on a moment matching scheme with optimal weighting and that can be used with high-throughput data like the one collected by flow cytometry. Our method does not require an approximation of the underlying equilibrium probability distribution and, if reaction rate constants have to be learned, the optimal values can be computed by solving a linear system of equations. We evaluate the effectiveness of the proposed approach on three case studies.

1 Introduction

Stochastic models have proven to be a powerful tool for the analysis of biochemical reaction networks. Especially when chemical species are present in low copy numbers, a stochastic approach provides important insights on the randomness inherent to the system when compared to deterministic approaches. For the inference of parameters based on experimentally observed samples, more detailed descriptions given by stochastic models can substantially improve the quality of the estimation [18].

The arguably most popular stochastic modeling approach to chemical kinetics is based on a description in terms of continuous time Markov chains (CTMC) [10]. In this case, the exact time evolution of the entire probability distribution is given by the chemical master equation (CME). Although, this description is exact up to the numerical precision of the integration scheme, its solution is only feasible for simple systems with small molecular populations [22]. Therefore, the applicability of inference approaches based on a maximum likelihood estimation (MLE) is limited to this class of networks since they require an approximation of the probability distribution, i.e., a solution of the CME [2, 3]. An alternative to ease the computational burden is to use stochastic simulation to estimate the likelihood function or to learn parameters in a Bayesian setting, e.g. by ABC methods [33]. However, the total number of simulations to be performed is huge, still resulting in a computationally intensive approach.

A computationally more feasible approach is to consider the statistical moments (such as the expected value or the variance) instead of the entire probability distribution. Moment-based analysis approaches rely on a derivation of a system of equations for the time-derivative of the moments [1, 6, 30]. Since the exact time evolution of the moments of order k may depend on moments of higher order, a closure method has to be applied to arrive at a finite system of equations. However, moment-based methods complicate the application of MLE since a reconstruction of the distribution is computationally expensive and may be inaccurate depending on the shape of the distribution [4].

In this paper we propose a parameter estimation approach that does not rely on MLE and distribution approximations, but on the generalized method of moments (GMM), which has been a widely used inference method in econometrics for over 30 years [12, 14]. We consider the case in which experimentally observed samples are drawn when the process is in equilibrium. Population snapshot data of equilibrium processes are considered, for instance, if the (possibly multi-stable) steady-state expression in a gene regulatory network is investigated [7, 17] or if the steady-state behavior of a mutant is compared the behavior of the wild type [19, 27]. Modern high-throughput experimental techniques, like flow cytometry, deliver a large amount of measurements from a population of cells at steady state and thus give detailed information about the distribution of proteins and RNAs [13, 16, 25]. The idea of the GMM is to consider constraints of the form $\mathbb{E}[f(\mathbf{Y}_i, \boldsymbol{\theta}_0)] = 0$ where \mathbf{Y}_i is a sample and $\boldsymbol{\theta}_0$ the parameter vector. We propose to choose f as the time derivatives of the statistical moments of the model, which can directly be derived from the CME. This follows from the fact that the time derivatives will become equal to zero when the process is in equilibrium. A major advantage, given the availability of steady state samples, is that, compared to time depended observations, no moment closure approximations are necessary. Instead exact equations for the steady state moments can be used. If the propensities are linear in the unknown parameters, as is the case for mass action kinetics, a closed linear form is possible. This results in an extremely fast inference procedure since no numerical optimization is needed. In case of propensities that are non-linear in the parameters numerical optimization is necessary. Still, no numerical integration of moment equations or probabilities is needed since the objective function corresponds to the right side of the steady state moment equations.

The moment equations may also contain moments of species whose quantity is hard to measure (e.g. the state of a promoter). Instead of treating these latent variables as unknown (probably non-linear) parameters, here we propose a clustering approach that estimates promoter states in a preprocessing step. Then, a closed linear solution is still possible, which again enables an accurate estimation in very short time.

We analyse the effectiveness of the GMM approach for the p53 oscillator model [9] and two variants of the genetic toggle switch [8, 20]. Our results show that using moments of up to at least second order yields accurate estimates. The inclusion of higher order moments (higher than three) can lead to a further

decrease of the estimators variances, but for the p53 model and only few samples the estimation becomes worse. Nevertheless, even for comparatively small sample sizes (100) the estimates are usually tightly distributed around the true parameter value when moments up to order two or three are considered.

The paper is organized as follows. We first provide some background on the model in Sect. 2 and present our inference approach based on GMM in Sect. 3. We discuss the inference results for the case studies in Sect. 4 and conclude the paper in Sect. 6.

2 Stochastic Chemical Kinetics

A stochastic model of a network of chemical reactions is usually specified by a set of n species, which are represented by a set of symbols S_1, \dots, S_n . We are interested in the system state, i.e., the number of individuals of the species, and thus consider state space $\mathcal{S} \subseteq \mathbb{N}_{\geq 0}^n$. Furthermore, a set of J reactions is given describing the interactions between the different molecular populations. For $j \in \{1, \dots, J\}$ reaction R_j is specified by its stoichiometry

$$R_j : S_1 \nu_{j,1}^- + \dots + S_n \nu_{j,n}^- \xrightarrow{c_j} S_1 \nu_{j,1}^+ + \dots + S_n \nu_{j,n}^+, \quad (1)$$

where the vectors $\boldsymbol{\nu}_j^-$ and $\boldsymbol{\nu}_j^+ \in \mathbb{N}_{\geq 0}^n$ with entries $\nu_{j,i}^-$ and $\nu_{j,i}^+$ for $i \in \{1, \dots, n\}$ specify how many molecules are consumed (produced) of each type, respectively. The vector $\boldsymbol{\nu}_j = \boldsymbol{\nu}_j^+ - \boldsymbol{\nu}_j^-$ is called the *change vector* of R_j . The propensity functions α_j are such that $\alpha_j : \mathcal{S} \times \Theta \rightarrow \mathbb{R}_{\geq 0}$, where Θ is the parameter space. If mass action kinetics are assumed, then α_j is the product of the rate constant c_j and the number of possible combinations of reactant molecules, i.e., $\alpha_j(\mathbf{x}, \boldsymbol{\theta}) = c_j \prod_{i=1}^n \binom{x_i}{\nu_{j,i}^-}$ for $\mathbf{x} \in \mathcal{S}$. Here, we do not restrict to mass action kinetics but only impose certain regularity conditions on the propensity functions, such as continuity and the existence of certain expected values. If a reaction does not follow mass action propensities, we give the propensity function separately from the stoichiometry (1).

Under the assumption of well-stirredness and thermal equilibrium such a system can be accurately described by a continuous-time Markov chain (CTMC) $\mathbf{X}(t) = (X_1(t), \dots, X_n(t))$ over the state space \mathcal{S} [10]. The time evolution of the probability distribution is given by the chemical master equation (CME):

$$\frac{d}{dt} P(\mathbf{X}(t) = \mathbf{x}) = \sum_{j=1}^J P(\mathbf{X}(t) = \mathbf{x} - \boldsymbol{\nu}_j) \alpha_j(\mathbf{x} - \boldsymbol{\nu}_j, \boldsymbol{\theta}) - \sum_{j=1}^J P(\mathbf{X}(t) = \mathbf{x}) \alpha_j(\mathbf{x}, \boldsymbol{\theta}) \quad (2)$$

Due to the largeness of the state space the integration of $\frac{d}{dt} P$ is computationally infeasible, especially if we have to integrate until convergence to determine the equilibrium distribution. Given (2) it is straight-forward to compute the time derivative of the expectation of some polynomial function $g : \mathcal{S} \rightarrow \mathbb{R}$ [6]:

$$\frac{d}{dt} \mathbb{E}[g(\mathbf{X})] = \sum_{j=1}^J \mathbb{E}[(g(\mathbf{X} + \boldsymbol{\nu}_j) - g(\mathbf{X})) \alpha_j(\mathbf{X}, \boldsymbol{\theta})], \quad (3)$$

where we omit the dependence of \mathbf{X} on t . Here we are concerned with the *population moments* of the distribution, which are monomials $\mathbf{X}^{\mathbf{m}}$ over \mathbf{X} where we use the multi-index notation $\mathbf{x}^{\mathbf{m}} = x_1^{m_1} \cdots x_n^{m_n}$ for the vectors $\mathbf{m} = (m_1, \dots, m_n) \in \mathbb{N}_{\geq 0}^n$ and $\mathbf{x} = (x_1, \dots, x_n)$.¹ The order of a moment is given by the sum $m_1 + \cdots + m_n$. The first order moment of the i -th population, for example, is obtained from (3) by setting $g(\mathbf{x}) = x_i$:

$$\frac{d}{dt} \mathbb{E}[X_i] = \sum_{j=1}^J \nu_{j,i} \mathbb{E}[\alpha_j(\mathbf{X}, \boldsymbol{\theta})]. \quad (4)$$

In general, the equation of a moment of a certain order may depend on moments of higher order, except if α is constant or linear, i.e., of the form $\mathbf{c}_j^T \mathbf{x} + b_j$ for some constant $\mathbf{c}_j \in \mathbb{R}^n$ and $b_j \in \mathbb{R}$. Here, we do not aim at finding a finite system of ODEs to approximate the moments but we rather propose to use the exact moment equations when the system is in equilibrium. The *equilibrium probability* of a state \mathbf{x} is defined as the limit of $P(\mathbf{X}(t) = \mathbf{x})$ when $t \rightarrow \infty$, i.e.,

$$\pi(\mathbf{x}) = \lim_{t \rightarrow \infty} P(\mathbf{X}(t) = \mathbf{x}). \quad (5)$$

and is uniquely defined for ergodic processes \mathbf{X} . Since the equilibrium distribution is independent of time, the expected values in (3) are also time-independent when $t \rightarrow \infty$. Thus, we can use the right side of (3) to estimate propensity parameters given samples from the equilibrium distribution.

3 GMM Conditions at Equilibrium

We propose to use the moments of the equilibrium distribution as an input for a GMM inference, which is a very generic framework for parameter estimation [12, 23]. It is most popular in econometrics, where often the exact distribution of a model is not known. In this case MLE cannot be used since it needs a sufficiently accurate description of the distribution for its optimality properties to hold. As opposed to this, the GMM is based on the construction and minimization of certain cost functions, called *moment conditions*, which relate the population and sample moments. A moment condition is given by a function whose expected value is zero for the true parameter value $\boldsymbol{\theta}_0$. Given independent samples $\mathbf{Y}_1, \dots, \mathbf{Y}_N$ of the process \mathbf{X} in equilibrium, a vector of moment conditions is given by

$$\mathbb{E}[\mathbf{f}(\mathbf{Y}, \boldsymbol{\theta}_0)] = 0, \quad (6)$$

where we omit the index of the samples whenever they appear within the expectation operator since $\mathbf{Y}_1, \dots, \mathbf{Y}_N$ are identically distributed according to the equilibrium distribution π . Moreover, let \mathbf{f} be a vector of q different functions,

¹ The existence and convergence of moments is treated Gupta et al. [11]. It can be proved for the models in Sect. 4 with positive rate constants.

i.e., $\mathbf{f} : (\mathcal{S} \times \Theta) \rightarrow \mathbb{R}^q$. The sample equivalent of (6) for the vector \mathbf{f} of moment conditions is given by

$$\mathbf{f}_N(\boldsymbol{\theta}_0) = \frac{1}{N} \sum_{i=1}^N \mathbf{f}(\mathbf{Y}_i, \boldsymbol{\theta}_0) = 0. \quad (7)$$

Depending on the number of such conditions q and the number of parameters to be estimated p , we distinguish for the estimated value the *non-identified case* ($q < p$), the *exactly identified case* ($p = q$) and the *over-identified case* ($q > p$). In the exactly identified case, assuming (7) has a unique solution, we have Pearson's classical *method of moments* [28].

Since we are considering the system at equilibrium, the right-hand side of (3) must equal zero. In principle, it is possible to use any polynomial g meeting certain regularity conditions [12]. However, using population moments, i.e., monomials of \mathbf{Y} is a natural choice that leads to the moment conditions

$$\frac{d}{dt} \mathbb{E}[\mathbf{Y}^{\mathbf{m}}] = \sum_{j=1}^J \mathbb{E}[(\mathbf{Y} + \boldsymbol{\nu}_j)^{\mathbf{m}} - \mathbf{Y}^{\mathbf{m}}] \alpha_j(\mathbf{Y}, \boldsymbol{\theta}_0) = 0 \quad (8)$$

for the estimation of $\boldsymbol{\theta}$. Therefore the moment condition vector \mathbf{f} in (6) is determined by the functional form (8) of a selection of different vectors \mathbf{m} , i.e., the entry in the vector \mathbf{f} that corresponds to \mathbf{m} is

$$f_{\mathbf{m}}(\mathbf{Y}, \boldsymbol{\theta}) = \sum_{j=1}^J ((\mathbf{Y} + \boldsymbol{\nu}_j)^{\mathbf{m}} - \mathbf{Y}^{\mathbf{m}}) \alpha_j(\mathbf{Y}, \boldsymbol{\theta}).$$

Typically, we choose these vectors such that their entries correspond to the moments up to some fixed order. If, for example, we use first order moments only, the i -th entry of \mathbf{f} is equal to $\sum_j \nu_{i,j} \alpha_j(\mathbf{Y}, \boldsymbol{\theta})$ for $i = 1, 2, \dots, n$ (see Eq. 4). For the moments of order two we extend \mathbf{f} with entries according to the right side in (8) of the second order moments and so forth.

We may choose as many moment conditions as there are parameters to exactly identify the estimate. However, the inclusion of further information on the distribution may lead to a more accurate estimation. GMM provides a framework to deal with over-identified estimation problems. The estimator is given by

$$\hat{\boldsymbol{\theta}}_N = \arg \min_{\boldsymbol{\theta} \in \Theta} Q_N(\boldsymbol{\theta}), \quad (9)$$

where $Q_N(\boldsymbol{\theta})$ is the objective function

$$Q_N(\boldsymbol{\theta}) = \mathbf{f}_N(\boldsymbol{\theta})^\top W \mathbf{f}_N(\boldsymbol{\theta}). \quad (10)$$

Here, W is some positive semi-definite matrix containing weights for each pair of moment conditions. Under certain regularity conditions [12], this estimator is asymptotically normal and consistent, i.e., the estimator converges in probability to $\boldsymbol{\theta}_0$. These regularity conditions mostly consist of the existence of expectations

of \mathbf{f} and $\partial\mathbf{f}/\partial\boldsymbol{\theta}$ and their continuity w.r.t. $\boldsymbol{\theta}$ on the parameter space Θ . Assuming convergence to equilibrium moments the validity of these conditions depends solely on the propensity functions. They hold for mass action and Hill's propensities, as they are smooth functions of the parameters. The parameter space itself is assumed to be bounded, which in practice can be done by either fixing a biologically relevant space or assuming a sufficiently large Θ [12]. A further necessary condition for normality is that $\boldsymbol{\theta}_0$ is a unique interior point of Θ such that $\mathbb{E}[\mathbf{f}(\mathbf{Y}, \boldsymbol{\theta}_0)] = 0$. However, if we have only samples from the steady state distribution this property may not hold if one tries to estimate all parameters at once. The reason is that often for a fixed steady-state distribution there is an infinite number of ergodic Markov chains having this steady-state distribution and the system is not fully identifiable.

Although the estimator's normality holds for all positive semi-definite weighting matrices, a good choice of W reduces the asymptotic variance of the estimator. It can be shown, that the asymptotically most efficient matrix W_0 is given by the inverse of $\lim_{N \rightarrow \infty} \text{Var}(\sqrt{N}\mathbf{f}_N(\boldsymbol{\theta}_0))$ [12, 23]. In case of independent and identically distributed samples, W_0 can be estimated as follows [23]:

$$\hat{W}_N = \left(\frac{1}{N} \sum_{i=1}^N \mathbf{f}(\mathbf{Y}_i, \boldsymbol{\theta}_0) \mathbf{f}(\mathbf{Y}_i, \boldsymbol{\theta}_0)^\top \right)^{-1}. \quad (11)$$

Since this estimate depends on $\boldsymbol{\theta}_0$, which is unknown, GMM is usually applied in an iterative manner: A first estimate $\hat{\boldsymbol{\theta}}_1$ is computed using some positive-definite weight matrix, such as the identity matrix. The estimate $\hat{\boldsymbol{\theta}}_1$ is consistent, but likely asymptotically inefficient. This estimate is then used to approximate (11). The procedure of estimating $\boldsymbol{\theta}_0$ and computing \hat{W}_N can be iteratively applied until some convergence criterion is met. Since W is constant at each iterative estimation, the solution to (9) can, under some restrictions on the propensities, be expressed as a linear system (cf. Sect. 3.1).

Beyond this iterative estimation scheme, the *continuously updating GMM* (CUGMM) [15] is a popular variant of the GMM estimator. Instead of recomputing the weight estimate between minimizations, the weight estimation (11) is substituted into the objective function (10). The resulting estimator is thus given by

$$\hat{\boldsymbol{\theta}}_{CU,N} = \arg \min_{\boldsymbol{\theta} \in \Theta} \mathbf{f}_N(\boldsymbol{\theta})^\top \left(\frac{1}{N} \sum_{i=1}^N \mathbf{f}(\mathbf{Y}_i, \boldsymbol{\theta}) \mathbf{f}(\mathbf{Y}_i, \boldsymbol{\theta})^\top \right)^{-1} \mathbf{f}_N(\boldsymbol{\theta}). \quad (12)$$

This estimator is often associated with improved finite sample properties and more reliable test statistics [23]. However, a closed form solution for linear propensities as described in Sect. 3.1 is not possible in a majority of cases. This necessitates numerical optimization to approximate (12).

3.1 Linear Propensities

In general, the minimization problem (9) can be solved using numerical optimization algorithms. However, depending on the rate functions, this may not be

necessary, because a closed form solution, i.e., a linear system, can be obtained for many relevant cases, including mass action kinetics. This system results from the first order condition of the minimization $\partial Q_N(\hat{\boldsymbol{\theta}}_N)/\partial \boldsymbol{\theta} = 0$ which yields [12]

$$\mathbf{0} = \frac{\partial \mathbf{f}_N(\hat{\boldsymbol{\theta}}_N)}{\partial \boldsymbol{\theta}}^\top W \mathbf{f}_N(\hat{\boldsymbol{\theta}}_N). \quad (13)$$

We now compute (13) under the condition that propensities are linear in $\boldsymbol{\theta}$ and W is constant, as is the case in iterative GMM. To this end, let $\mathcal{R} \subseteq \{1, \dots, J\}$ be the index set of functions α_j whose propensity is dependent on $\boldsymbol{\theta}$. Further let $\bar{\mathbf{f}}$ be the part of \mathbf{f} independent of $\boldsymbol{\theta}$ such that the i -th entry equals

$$\bar{f}_{m_i}(\mathbf{Y}) = \sum_{j \in \mathcal{R}} \alpha_j(\mathbf{Y}) ((\mathbf{Y} + \boldsymbol{\nu}_j)^{m_i} - \mathbf{Y}^{m_i}). \quad (14)$$

By computation of the matrix product (13) and splitting the moment condition based on (14), we get

$$\left(\frac{\partial \mathbf{f}_N}{\partial \boldsymbol{\theta}}^\top W \mathbf{f}_N \right)_i = \underbrace{\sum_{h=1}^p \theta_h \sum_{\ell, k=1}^q \frac{\partial f_{N, m_k}}{\partial \theta_i} W_{k, \ell} \frac{\partial f_{N, m_\ell}}{\partial \theta_h}}_{(A\boldsymbol{\theta})_i} + \underbrace{\sum_{\ell, k=1}^q \frac{\partial f_{N, m_k}}{\partial \theta_i} W_{k, \ell} \bar{f}_{N, m_\ell}}_{-b_i}.$$

Note, that the sample derivatives $\partial \mathbf{f}_N / \partial \theta_i$ are independent of $\boldsymbol{\theta}$. In vector notation this gives us the linear system $A\boldsymbol{\theta}_N = \mathbf{b}$ as a solution to (13) where

$$A_{i,j} = \frac{\partial \mathbf{f}_N}{\partial \theta_i}^\top W \frac{\partial \mathbf{f}_N}{\partial \theta_j} \quad b_i = -\frac{\partial \mathbf{f}_N}{\partial \theta_i}^\top W \bar{\mathbf{f}}_N. \quad (15)$$

Analogous to the general iterative scheme, we now solve (15) and use the estimate to in turn estimate W using (11). In the following discussion we will refer to this as the *closed form GMM* (CFGMM). One sees immediately that this method is far more efficient than numerically optimizing Q_N .

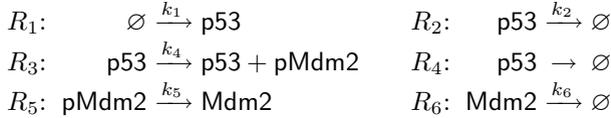
4 Case Studies

We evaluate the GMM estimation on three chemical reaction networks. Samples of the equilibrium distribution were generated by Gillespie's stochastic simulation algorithm (SSA) [10] and drawn by equidistant sampling after the initial warm-up period. For each case study 10^7 samples were generated and sample sets of different sizes were drawn at random from this large set. For each sample size considered, the estimation procedure was carried out on 100 random sample sets, in order to estimate the variance of the estimator.

4.1 P53 System

We first consider Model IV proposed in [9], that describes the interactions of the tumor suppressor p53. This system describes a negative feedback loop between p53 and the oncogene Mdm2, where pMdm2 is a Mdm2 precursor [9]. We chose the same parameter values as in [1], that is, $k_1=90$, $k_2=0.002$, $k_3=1.7$, $k_4=1.1$, $k_5=0.93$, $k_6=0.96$, $k_7=0.01$.

Model 1 (p53 System).



The degradation rate of p53 is in part influenced by Mdm2 and is given by $\alpha_4(\mathbf{x}, \boldsymbol{\theta}) = (k_3 x_{\text{p53}} x_{\text{Mdm2}}) / (x_{\text{p53}} + k_7)$. Terms of species with stoichiometric constant zero are omitted as well as stoichiometric constants equal to one.

We estimated the four parameters k_3 , k_4 , k_5 , and k_6 using the CFGMM as proposed in Sect. 3.1. Note that $\alpha_4(\cdot, \cdot)$ is linear in k_4 . We fixed k_1 and k_2 to ensure identification as well as k_7 to avoid a time-consuming numerical optimization. The iterations were continued until either the parameter vector converged or the maximum number of four iterations was reached. The plot in Fig. 1 (left) shows that the best results were obtained already after the second step for moderate and large sample sizes, while for a small sample size of 100 further iterations were beneficial. It is important to note that for the first iteration, \hat{W}_N is chosen as the identity matrix such that identical weights are assigned and mixed terms are not considered. Hence, the general idea of assigning appropriate weights gives significantly more accurate results compared to an estimation with identical weights.

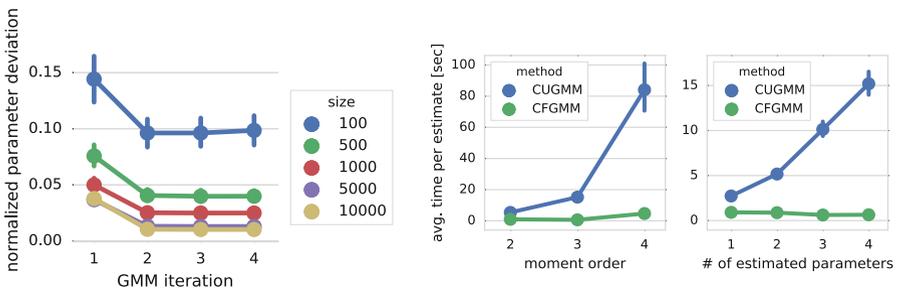


Fig. 1. p53 System: (left) The normalized parameter deviation $\|\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0\| / \|\boldsymbol{\theta}_0\|$ over GMM iterations for different sample sizes. Moment conditions up to order two were used. (right) Comparison of the average running time for a single estimation, as a function of the number of parameters (maximal moment order three) and of the maximum order of moment conditions used (estimation for four parameters), for a sample size of 100.

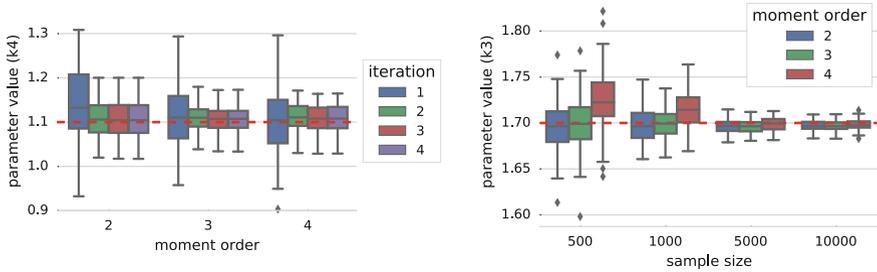


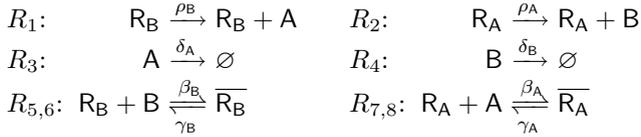
Fig. 2. p53 System: (left) Estimate of k_4 over GMM iterations (sample size 1000) (right) Estimates of parameter k_3 in relation to the sizes of the sample sets and the maximum order of used moment conditions. Results are presented as box plots (whiskers with a maximum of 1.5 IQR).

In Fig. 1 (right) we compare the running times of the CUGMM (using a numerical optimization scheme, the L-BFGS-B algorithm [36]) and of the iteration based method. The reported times are the average of 100 runs for a single estimate for different moment orders and different numbers of estimated parameters. As we can see, the iteration based method for linear propensities not only outperforms CUGMM, but also is essentially insensitive to including higher order moments and to increasing the number of estimated parameters. For CUGMM, an optimization is carried out since (12) is not linear in θ and this optimization becomes more costly when more moment conditions or parameters are considered. The advantage of CFGMM is that the Jacobian and \bar{f}_N is only computed once for all iterations of a sample and no numerical optimization is needed.

In Fig. 2 we show the distribution of the estimate quality for different maximum moment orders against (left) different numbers of iterations for CFGMM and (right) for different sample sizes. The quality of the results is excellent for large sample sizes, while increasing the moment order beyond two does not result in significant improvements or may even (for small sample sizes) significantly decrease the quality (see Fig. 2 (right)). This bias may occur if the degree of overidentification ($q - p$) is increased too much. It can be caused by the estimation of W and the dependence on the previous estimates and decreases proportional to N^{-1} [12, 26]. In our evaluation estimators based on a maximal order of two and three showed the most reliable performance. Moreover, identical weights in the first step of the iteration lead to a very high variance of the corresponding estimator, as shown in Fig. 2 (left). In Fig. 2 (right) we also see that, when the number of samples is increased, the variance of the estimator becomes small.

4.2 Toggle Switch

The toggle switch is a widely known gene regulatory network [8, 20] that models the production of two proteins A and B. Each protein can bind to the promoter of the opposite protein and thereby repress its production.

Model 2 (Explicit Toggle Switch). [20]

Note that, given appropriate starting values, the conservation law $R_X + \overline{R_X} = 1$ holds ($X \in \{A, B\}$). In our study we focus on two cases, that are high binding-/unbinding rates and low binding-/unbinding rates with respect to the production and degradation of proteins.

Slow Binding Toggle Switch. In the case of low binding-/unbinding rates several attractor regions can arise that directly correspond to a given DNA state. Here, we use the parameters $\rho_X = 3$, $\delta_X = 0.5$, $\beta_X = 10^{-6}$, $\gamma_X = 3 \times 10^{-4}$, which are identical for $X = A$ and $X = B$. During the inference procedure, however, we did not make use of the information that the parameters are symmetric. For these parameters we get three distinct attractor regions corresponding to either one of the repressors being bound and both repressors being free².

Currently, our GMM-based approach requires all variables to be observed, which is in general unfeasible for the DNA state. One possible solution, when only proteins are observed, is to cluster the samples of the proteins using the k-Means algorithm (cf. Fig. 3 (left) for an example of a clustering of samples of the toggle switch). Then we can infer the state of the latent DNA state by assigning each cluster to a specific combination of DNA states and by looking at the cluster centroids, as illustrated in Fig. 3 (left). For low binding-/unbinding rates, the attractors are well separated and this approach is feasible, though more sophisticated approaches may be required when clusters overlap. After reconstruction of the state of the unobserved variables, we used the GMM estimation with the closed form solution for linear propensities. Results comparing different sample sizes are shown in Fig. 3 (right). The estimation quality is very good even in the case of only few samples, provided enough iterations are carried out. It is important to note that for these results, we excluded moment conditions corresponding to mixed moments involving the state of the gene as their moment conditions have very similar values. Including them leads to severe numerical instabilities (the matrix of the linear system for linear propensities becomes quasi-singular). However, ill-conditioned matrices are detected automatically when their determinant is calculated during the computation. Then, those entries responsible for the numerical instabilities can be excluded.

Fast Binding Toggle Switch. Often, it can be assumed that the repressor ($R_{A,B}$) binding and unbinding ($R_{5,6}$ and $R_{7,8}$) happens a lot faster than the

² The case of both repressors being bound, would result in samples around the origin, which can be neglected if there are no such samples.

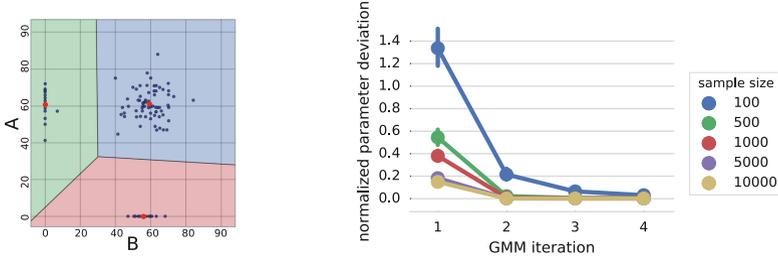


Fig. 3. Slow Switching Toggle Switch: (left) Clustering of a sample (size 100) using k-Means. (right) The normalized parameter deviation $\|\hat{\theta}_N - \theta_0\|/\|\theta_0\|$ over GMM iterations for different sample sizes given the toggle switch with k-Means clustering. Moment conditions up to order 3 were used and 4 parameters were estimated.

protein production. Then, a *Michaelis-Menten* approximation is possible [20]. Therein the time derivative of the repressors is assumed to be zero. Applying this assumption to the mean-field equations of Model 2 yields the implicit toggle switch (Model 3). In this case, we no longer need the repressor state of each sample.

Model 3 (Implicit Toggle Switch).



The rate function of reactions R_1 and R_2 resulting from the Michaelis-Menten approximation are

$$\alpha_1(\mathbf{x}, \boldsymbol{\theta}) = \frac{\rho_A}{1 + k_B x_B} \quad \alpha_2(\mathbf{x}, \boldsymbol{\theta}) = \frac{\rho_B}{1 + k_A x_A}$$

where $\boldsymbol{\theta}$ is the vector of all parameters, $\mathbf{x} = (x_A, x_B)$, and $k_X = \frac{\beta_X}{\gamma_X}$ is the quotient of the binding and unbinding rate, $X \in \{A, B\}$.

The toggle switch exhibits bistability if the binding happens significantly faster than the unbinding, i.e., $k_A, k_B \gg 1$ [20]. However, the estimation of k_A and k_B is inherently difficult because switching between the attractors is a rare event.

In this case study, we simulated the explicit model using the symmetric constants $\beta_X = 100.0$, $\gamma_X = 50.0$, $\rho_X = 0.2$ and $\delta_X = 0.005$, assuming we could observe only the two proteins. Thus, we estimated the parameters k_X and δ_X of the implicit model and fixed ρ_X to ensure identification. Due to non-linear dependency of production rates on k_X , we cannot rely anymore on the method for linear propensities of Sect. 3.1, hence we resort to a numerical minimization routine, namely the L-BFGS-B algorithm [36], for the CUGMM scheme. The initial guess was chosen at random from $[0, 1]^p$. For detection of unsuccessful optimizations we used the J-Test statistic [14], which states that under the null hypothesis of a correctly specified model, $NQ_N(\hat{\boldsymbol{\theta}}_N)$ converges to the χ_{q-p}^2 distribution. A confidence threshold of 90 % was fixed and the optimization was repeated for at

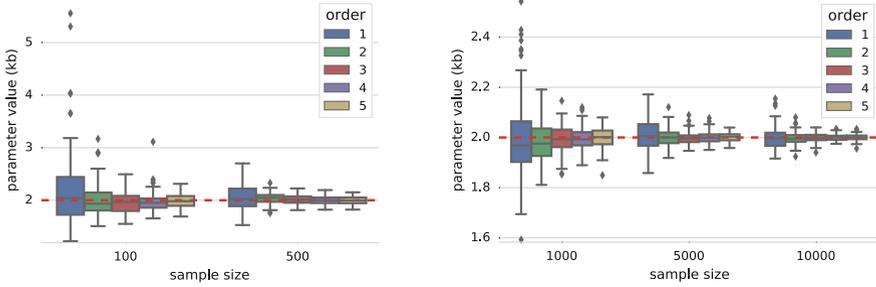


Fig. 4. Fast Binding Toggle Switch: Estimates of parameter k_B in relation to the sizes of the sample sets and the maximum order of used moment conditions. Only the parameter k_B was estimated. Results are presented as box plots (whiskers with a maximum of 1.5 IQR).

most four times until the threshold was met. The use of numerical optimization increased the cost of a single estimate: For a sample size of 10,000 observations and order 2, the computation takes 1–2 of minutes.

In Fig. 4, we give statistics on the quality of estimates based on 100 runs of independently generated datasets. More specifically, we show how the quality of estimates varies with the maximum order of moments considered in the method and with sample size. For a fixed sample size, increasing the order from 1 to 2 improves considerably the quality of results. Use of higher order moments significantly reduces the variance of the estimator, in particular for the case of few samples.

5 Related Work

In the context of stochastic chemical kinetics, parameter inference methods are either based on Bayesian inference [5, 32, 34] or maximum likelihood estimation [2, 3, 29, 31]. The advantage of the latter method is that the corresponding estimators are, in a sense, the most informative estimates of unknown parameters and have desirable mathematical properties such as unbiasedness, efficiency, and normality. On the other hand, the computational complexity of maximum likelihood estimation is high. If an analytic solution of the MLE is not possible, then, as a part of the non-linear optimization problem, the likelihood and its derivatives have to be calculated. Monte-Carlo simulation has been used to estimate the likelihood [31]. During the repeated random sampling it is difficult to explore those parts of the state space that are unlikely under the current rate parameters. Thus, especially if the rates are very different from the true parameters, many simulation runs are necessary to calculate an accurate approximation of the likelihood.

Therefore methods using computationally far more attractive moment expansion approximations have been proposed. Kügler [18] uses results of the moment

closure approximations to apply an ad-hoc weighted least squares estimator. Milner et al. [24] construct a multi-variate normal distribution based on low order moments obtained from a moment closure approximation in order to apply MLE. Another approach based on moment closure and MLE relies on a normal distribution based on sample means and variances [35].

All of the aforementioned moment-based inference methods are, in contrast to the scenario discussed in this paper, based on samples of the transient distribution before equilibrium is reached. Therefore they have to rely on moment closure approximations, which is not necessary in our approach based on the equilibrium distribution. Recently, the performance of GMM estimators has been studied for transient (non-equilibrium) data [21] together with a (hybrid) moment closure approach.

6 Conclusion

Parameter inference methods for stochastic models of reaction networks require 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 in equilibrium 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 no statistical sampling or numerical integration of master or moment equations is necessary. The proposed approach gives accurate results in seconds when the parameters are linear because a closed form of the solution is available. For non-linear parameters, a global optimization problem must be solved and therefore the inference takes longer but is still fast compared to other approaches based on the numerical computation of likelihoods.

Our results show that the GMM estimator yields accurate results, where its variance decreases when moments of higher order are considered. We found that when moments of order higher than three are included, the results become slightly worse in case of the p53 system while for the toggle switch quality improved (variance decreased). A general strategy could be to start with as many cost functions as unknown parameters and increase the maximal order until appropriate statistical tests suggest that higher orders do not lead to an improvement.

Currently, a major drawback of the method is that all species must be observed in order to apply it. For populations of at most one individual, the proposed clustering approach circumvents the problem that such species can usually not be observed. In general, however, the clustering may not always be possible and there may be other species that can not be observed. To deal with such cases, we plan to develop an extension of the method that treats the moments of such species as (additional) unknown parameters. Moreover, we will investigate how measurement errors could be taken into account within the GMM framework.

References

1. Ale, A., Kirk, P., Stumpf, M.: A general moment expansion method for stochastic kinetic models. *J. Chem. Phys.* **138**(17), 174101 (2013)
2. Andreychenko, A., Mikeev, L., Spieler, D., Wolf, V.: Parameter identification for markov models of biochemical reactions. In: Gopalakrishnan, G., Qadeer, S. (eds.) CAV 2011. LNCS, vol. 6806, pp. 83–98. Springer, Heidelberg (2011)
3. Andreychenko, A., Mikeev, L., Spieler, D., Wolf, V.: Approximate maximum likelihood estimation for stochastic chemical kinetics. *EURASIP J. Bioinf. Syst. Biol.* **9**, 1–14 (2012)
4. Andreychenko, A., Mikeev, L., Wolf, V.: Model reconstruction for moment-based stochastic chemical kinetics. *ACM Trans. Model. Comput. Simul. (TOMACS)* **25**(2), 12 (2015)
5. Boys, R., Wilkinson, D., Kirkwood, T.: Bayesian inference for a discretely observed stochastic kinetic model. *Stat. Comput.* **18**, 125–135 (2008)
6. Engblom, S.: Computing the moments of high dimensional solutions of the master equation. *Appl. Math. Comput.* **180**(2), 498–515 (2006)
7. Fournier, T., Gabriel, J.-P., Mazza, C., Pasquier, J., Galbete, J.L., Mermoud, N.: Steady-state expression of self-regulated genes. *Bioinformatics* **23**(23), 3185–3192 (2007)
8. Gardner, T.S., Cantor, C.R., Collins, J.J.: Construction of a genetic toggle switch in *escherichia coli*. *Nature* **403**(6767), 339–342 (2000)
9. Geva-Zatorsky, N., Rosenfeld, N., et al.: Oscillations and variability in the p53 system. *Mol. Syst. Biol.* **2**(1) (2006)
10. Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**(25), 2340–2361 (1977)
11. Gupta, A., Briat, C., Khammash, M.: A scalable computational framework for establishing long-term behavior of stochastic reaction networks. *PLoS Comput. Biol.* **10**(6), e1003669 (2014)
12. Hall, A.R.: *Generalized Method of Moments*. Oxford University Press, New York (2005)
13. Hanley, M.B., Lomas, W., Mittar, D., Maino, V., Park, E.: Detection of low abundance RNA molecules in individual cells by flow cytometry. *PLoS one* **8**(2), e57002 (2013)
14. Hansen, L.P.: Large sample properties of generalized method of moments estimators. *Econometrica* **50**(4), 1029–1054 (1982)
15. Hansen, L.P., Heaton, J., Yaron, A.: Finite-sample properties of some alternative GMM estimators. *J. Bus. Econ. Stat.* **14**(3), 262–280 (1996)
16. Hasenauer, J., Waldherr, S., Doszczak, M., Radde, N., Scheurich, P., Allgöwer, F.: Identification of models of heterogeneous cell populations from population snapshot data. *BMC Bioinf.* **12**(1), 1 (2011)
17. Isaacs, F.J., Hastay, J., Cantor, C.R., Collins, J.J.: Prediction and measurement of an autoregulatory genetic module. *PNAS* **100**(13), 7714–7719 (2003)
18. Kügler, P.: Moment fitting for parameter inference in repeatedly and partially observed stochastic biological models. *PLoS one* **7**(8), e43001 (2012)
19. Lee, Y.J., Holzapfel, K.L., Zhu, J., Jameson, S.C., Hogquist, K.A.: Steady-state production of il-4 modulates immunity in mouse strains and is determined by lineage diversity of INKT cells. *Nat. Immunol.* **14**(11), 1146–1154 (2013)
20. Lipshtat, A., Loinger, A., Balaban, N.Q., Biham, O.: Genetic toggle switch without cooperative binding. *Phys. Rev. Lett.* **96**(18), 188101 (2006)

21. Lück, A., Wolf, V.: Generalized method of moments for estimating parameters of stochastic reaction networks. ArXiv e-prints, May 2016
22. Mateescu, M., Wolf, V., Didier, F., Henzinger, T.A.: Fast adaptive uniformisation of the chemical master equation. *IET Syst. Biol.* **4**(6), 441–452 (2010)
23. Mátyás, L.: Generalized Method of Moments Estimation, vol. 5. Cambridge University Press, New York (1999)
24. Milner, P., Gillespie, C.S., Wilkinson, D.J.: Moment closure based parameter inference of stochastic kinetic models. *Stat. Comput.* **23**(2), 287–295 (2013)
25. Munsky, B., Fox, Z., Neuert, G.: Integrating single-molecule experiments and discrete stochastic models to understand heterogeneous gene transcription dynamics. *Methods* **85**, 12–21 (2015)
26. Newey, W.K., Smith, R.J.: Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica* **72**(1), 219–255 (2004)
27. Nishihara, M., Ogura, H., Ueda, N., Tsuruoka, M., Kitabayashi, C., et al.: IL-6-gp130-STAT3 in T cells directs the development of IL-17+ Th with a minimum effect on that of Treg in the steady state. *Int. Immunol.* **19**(6), 695–702 (2007)
28. Pearson, K.: Contributions to the mathematical theory of evolution. *Philos. Trans. R. Soc. Lond. A* **185**, 71–110 (1894)
29. Reinker, S., Altman, R.M., Timmer, J.: Parameter estimation in stochastic biochemical reactions. *IEEE Proc. Syst. Biol* **153**, 168–178 (2006)
30. Singh, A., Hespanha, J.P.: Lognormal moment closures for biochemical reactions. In: 2006 45th IEEE Conference on Decision and Control, pp. 2063–2068. IEEE (2006)
31. Tian, T., Xu, S., Gao, J., Burrage, K.: Simulated maximum likelihood method for estimating kinetic rates in gene expression. *Bioinformatics* **23**, 84–91 (2007)
32. Toni, T., Welch, D., Strelkova, N., Ipsen, A., Stumpf, M.: Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* **6**(31), 187–202 (2009)
33. Toni, T., Welch, D., Strelkova, N., Ipsen, A., Stumpf, M.P.H.: Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* **6**(31), 187–202 (2009)
34. Wilkinson, D.J.: *Stochastic Modelling for Systems Biology*. C & H, Sesser (2006)
35. Zechner, C., Ruess, J., Krenn, P., Pelet, S., Peter, M., Lygeros, J., Koepl, H.: Moment-based inference predicts bimodality in transient gene expression. *PNAS* **109**(21), 8340–8345 (2012)
36. Zhu, C., Byrd, R.H., Lu, P., Nocedal, J.: Algorithm 778: L-BFGS-B: fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw. (TOMS)* **23**(4), 550–560 (1997)



<http://www.springer.com/978-3-319-45176-3>

Computational Methods in Systems Biology
14th International Conference, CMSB 2016, Cambridge,
UK, September 21-23, 2016, Proceedings
Bartocci, E.; Liò, P.; Paoletti, N. (Eds.)
2016, XIII, 356 p. 108 illus., Softcover
ISBN: 978-3-319-45176-3