Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC-PapersOnLine 49-26 (2016) 089–094

Preconditioned Metropolis sampling as a strategy to improve efficiency in Posterior exploration ⋆

Stefan Engblom * **Vikram Sunkara** **

* *Division of Scientific Computing, Department of Information Technology, Uppsala University, SE-751 05 Uppsala, Sweden.(email:stefane@it.uu.se)*

** *Computational Systems Biology, Konrad-Zuse-Zentrum for Informationstechnik, D-10557 Berlin, Germany. (email: sunkara@zib.de)*

Abstract: In the low copy number regime, the dynamics of chemically reacting systems is accurately modeled as a continuous-time Markov chain and the associated probability density obeys the chemical master equation. Parameter inference in such models is very challenging for various reasons: large levels of noise implies that large amount of data is required for identification, the presence of transient phases may shadow subsets of the parameters, and accurate likelihood estimation requires the solutions to master equations. The latter is itself a computational very challenging problem and although many approximate computational methods have been proposed previously, the final implied accuracy in estimated rate parameters is difficult to assess.

In this paper we look at the problem from the perspective of the Markov chain Monte Carlo method. Assuming the existence of a practically exact, but expensive, master equations solver, together with a cheaper, approximate alternative, we pick up the idea of *preconditioned Metropolis sampling*. Here the solutions of full master equations almost always imply an accepted step in the Markov chain, and consequently, step rejections are much cheaper. We investigate the properties of this technique theoretically and via illustrative examples. Whenever a suitable preconditioner is available, large savings in computational times are possible while the accuracy in deduced parameters is identical to using the exact likelihood.

Keywords: Chemical Master Equation, Preconditioning Markov Chain Monte Carlo, Metropolis Hastings.

1. INTRODUCTION

Systems biology is working towards describing complex interactions and processes in biology by Bio-Chemical Reaction Networks (Biological Networks for short). Through advances in imaging and sequencing technologies, biologists are able to scope deeper and describe critical biological processes as paths on a complex biological network. Current systems biology has been able to represent metabolic processes of cells Kholodenko (2000), stem cell fate paths MacArthur et al. (2009), gene transcriptions Srivastava et al. (2002) and translation regulation processes as biological networks. Blake et al. (2003) In the last three decades it was shown by experimentalists that the observed variation in the data can be attributed to the inherent stochasticity in parts of the network where low copy numbers are present.

The modeling of biological networks with intrinsic stochasticity gives realistic forecasts of the system behavior. However, to go further and use experimental data to infer the reaction rates of the underlying model is a computationally and mathematically challenging task. The critical aspect is the computation of the *likelihood function*. The likelihood function describes the conditional probability of observing the data for a given parameter value. In engineering settings where the noise is understood to be Gaussian and additive, the likelihood function can be easily approximated by a mean dynamics of the system and some fixed variance. However, if the system has intrinsic stochasticity, then computing the likelihood requires the solutions to equations such as the *Chemical Master Equation* (CME) or the *Fokker Planck equation*. These equations are difficult to solve numerically as they are prone to the *curse of dimensionality* Higham (2008); Engblom (2009a).

While many numerical methods have been investigated to solve the CME for given set of parameters, the accuracy of these methods for the purpose of inferring parameters from some given data is unclear. Let us see why: let $\mathcal{L}_{CME}(O|\sigma)$ be the likelihood function, given by the solutions of the CME, that data O is observed for the parameter σ. Similarly, let $\mathcal{L}_\star(O|\sigma)$ denote an approximate likelihood constructed using some approximation of the CME. We

⋆ The work was supported by the Swedish Research Council within the UPMARC Linnaeus center of Excellence (S. Engblom) and by BMBF (Germany) project PrevOp-OVERLOAD, grant number 01EC1408H (V. Sunkara).

10.1016/j.ifacol.2016.12.108

say that the likelihood $\mathcal{L}_\star$ is biased if

$$\|\mathcal{L}_{CME}(O|\sigma) - \mathcal{L}_\star(O|\sigma)\| = f(\sigma), \tag{1}$$

where $f(\cdot)$ is not a constant function. For many numerical approximations of the CME the function f is unfeasible to compute. With this in mind, when we use a biased likelihood for finding a posterior distribution, there is no simple rule on how the error in the likelihood translates across into the error of the posterior. For scenarios when we have a biased likelihood and computing f in (1) is unfeasible, one must get samples from the true posterior to verify the accuracy of the posterior constructed by the biased likelihood. Computing approximations of the CME which have a uniform error over the parameter space is computationally demanding. This motivates the proposal for a *preconditioning* MCMC algorithm (pcMCMC) Efendiev et al. (2006) [1] . In summary, the pcMCMC has two proposal steps in series for the same proposed state. The first proposal step uses the biased likelihood to accept or reject the new state. If the state is accepted in the first proposal step, then that state needs to be accepted in the second proposal step; this then involves the computation of a likelihood made up of unbiased CME approximations. The state which is accepted by both proposals in series is a true sample of the posterior distribution of the unbiased likelihood function.

The principle idea being that if the biased likelihood is close to the unbiased likelihood (locally), then the acceptance rate of the second proposal should be higher than the first. Since the CMEs are only being computed in the second proposal, having a higher acceptance rate would imply that we are minimizing the amount of CMEs we need to compute to find true samples of the posterior distribution. The acceptance rate of the second proposal is also an indicator of how close the biased likelihood is to the unbiased likelihood in some local region of the parameter space. In the sections below, we introduce the CME and the parameter inference problem. Then, we give an overview of the preconditioning MCMC method followed by some illustrative examples.

2. CHEMICAL MASTER EQUATION

The population of $N_s \in \mathbb{N}$ species undergoing $N_r \in \mathbb{N}$ reactions is described by the following sum of inhomogeneous Poisson processes,

$$X_\sigma(t) := X(0) + \sum_{r=1}^{N_r} \mathcal{P}_r\left(\int_0^t \alpha_r(X_\sigma(s),\sigma)\right)\rho_r. \tag{2}$$

The state space of $X_\sigma(t)$ is denoted by Ω and is a subset of $\mathbb{N}_0^{N_s}$. The variable σ is an element of our parameter space $\Sigma \subset \mathbb{R}_+^{N_r}$. The function $\alpha_r : \Omega \times \Sigma \to [0,\infty)$ describes the propensity/intensity at which the r^{th} reaction occurs. The stoichiometric vector, $\rho_r \in \mathbb{M}_{N_s\times 1}$, gives the change in population induced by the r^{th} reaction. Many biological processes' populations are described by (2). Historically, Thomas Kurtz investigated the convergence and analysis of (2) applied to the field of stochastic epidemiology, and for this reason we refer to (2) as the *Kurtz process* Ethier and Kurtz (2009).

[1] also refereed to as delayed acceptance by the statistics community Golightly et al. (2015)

To find the probability of observing $X_\sigma(t)$ in a state $x \in \Omega$ at a time point t, we need to substitute the Kurtz process into the Chapman–Kolmogorov equation. This will lead to the evolution of the probability over the state space being governed by the *Chemical Master Equation* (CME),

$$\frac{\partial P_\sigma(x;t)}{\partial t} = \sum_{r=1}^{N_r} \alpha_r(x-\rho_r,\sigma)P_\sigma(x-\rho_r;t) - \sum_{r=1}^{N_r}\alpha_r(x,\sigma)P_\sigma(x;t). \tag{3}$$

Verbosely, the change in probability of observing a state x at time t is equal to the transition probability of coming from an adjacent state into x, minus the transition probability of leaving the state x . The CME can be solved by formulating a linear initial value problem (IVP):

$$\frac{dp_\sigma(t)}{dt} = \mathcal{A}_\sigma p_\sigma(t) \quad \text{i.c.} \quad p(0), \tag{4}$$

where $p_\sigma(t)$ is a vector indexed by states in the state space and $\mathcal{A}_\sigma$ is an infinitesimal generator with columns summing to zero.

Broadly speaking, there are three major strategies for numerically approximating the solution to (4): Domain reduction, Galerkin methods and Tensor decomposition. An example of a domain reduction method is aggregation, where the idea is to aggregate states where the distribution has a shallow spatial gradient reducing the number of equations to solve Munsky and Khammash (2006); Sunkara and Hegland (2010). In Galerkin methods, the distribution is projected on a finite dimensional Hilbert space spanned by a chosen set of basis functions. This changes the IVP of the probability distribution to an IVP of the weights of the basis representation of the probability distribution Engblom (2009b); Jahnke and Udrescu (2010). Like in the continuous Galerkin methods, if the distribution has some inherent regularity, choosing the right basis functions can give a significantly smaller IVP to solve. Lastly, a Tensor decomposition method exploits possible tensor structure of the infinitesimal generator $\mathcal{A}_\sigma$ Kazeev et al. (2014). It has been shown that representing the generators of particular systems in a tensor format can partially overcome the curse of dimensionality. All methods exploit some inherent structure to achieve a significant speed-up over constructing an empirical distribution using trajectory based methods. Since our focus is not on any particular solver, we use the notational convention of writing a $\star$ when we are referring to an arbitrary approximation method of the CME. Similarly, we denote $P_\sigma^\star(x;t)$ as the corresponding solution to the approximate CME.

2.1 Random Variables and Likelihood

In this section we familiarize ourselves with the notational convention that will be used through this paper describing random variables and their likelihoods. We begin with the notational convention for the parameters we wish to infer. As in earlier sections, we denote the parameter to be inferred as $\sigma \in \Sigma$. The set Σ is a closed subset of $\mathbb{R}_0^{N_p}$, where N_p is the number of parameters we are to infer.

Let O_n^t (the data) be the n^{th} *random variable* governed by the stochastic process $X_\sigma(t)$ with probability $p_\sigma(t)$. For simplicity we assume data from only two time points, that is, snapshot data. The term O_0^0 denotes the initial