A POPULATION MONTE CARLO METHOD FOR BAYESIAN INFERENCE AND ITS APPLICATION TO STOCHASTIC KINETIC MODELS

Eugenia Koblents, Joaquín Míguez

Department of Signal Theory and Communications, Universidad Carlos III de Madrid.
Avda. de la Universidad 30, 28911 Leganés, Madrid, Spain.
email: ekoblents@tsc.uc3m.es, joaquin.miguez@uc3m.es

ABSTRACT

We introduce an extension of the population Monte Carlo (PMC) methodology to address the problem of Bayesian inference in high dimensional models. Specifically, we introduce a technique for the selection and update of importance functions based on the construction of Gaussian Bayesian networks. The structure of the latter graphical model enables a sequential sampling procedure that requires drawing only from unidimensional conditional distributions and leads to very efficient PMC algorithms. In order to illustrate the potential of the new technique we have considered the estimation of rate parameters in stochastic kinetic models (SKMs). SKMs are multivariate systems that model molecular interactions in biological and chemical problems. We present some numerical results based on a simple SKM known as predator-prey model.

1. INTRODUCTION

The problem of performing inference in high dimensional spaces appears in many practical applications. For example, it is of increasing interest in the biological sciences to develop new techniques that allow for the efficient estimation of the parameters governing the behavior of complex autoregulatory networks [1]. Another typical example in engineering is the problem of multi-target tracking, which consists of the dynamical estimation of the time-varying parameters of a set of multiple manoeuvering targets. The main difficulty often encountered when tackling this kind of problems is the design of numerical inference algorithms which are stable and have guaranteed convergence in high-dimensional spaces.

A very common strategy, which has been successfully applied in a broad variety of complex problems, is the Monte Carlo methodology. In particular, we have considered a recently proposed technique known as population Monte Carlo (PMC) [5], which is based on an iterative importance sampling approach. The aim of this method is the approximation of probability distributions by way of random measures consisting of samples and associated weights.

Although the PMC algorithm is elegant and simple to understand, its performance depends directly on the choice of the proposal distributions (or importance functions) that are used to generate the samples and compute the weights. These importance functions should be updated (i.e., improved) along the iterations of the algorithm. They should also remain simple, so that both drawing samples and computing weights is numerically tractable, and, finally, they should generate candidate samples in regions where the posterior probability is large. A recently proposed approach for the proposal update is the mixture PMC scheme [4], which models the importance functions as mixtures of transition kernels. However, this method does not perform any partitioning of the space of the variables of interest (they are drawn jointly) and, therefore, it may be inefficient for approximating distributions in high dimensional spaces.

As an alternative to this approach, in this paper we propose to represent the proposal functions at each iteration as the joint density of a Gaussian Bayesian network. The main advantage of this graphical model for our purposes is the fact that it allows for a straightforward sampling procedure in spaces of arbitrarily high dimension. Indeed, in the proposed scheme a topological order of the variables of interest is designed that enables us to draw samples from them sequentially (one variable at a time) using the conditional distribution of each variable given its "ancestors". (This approach is often termed ancestral sampling [9]).

The new algorithm is termed GBN-PMC. It provides (a) a simple proposal update procedure based on the selection of a GBN structure that fits the data adequately and (b) the ancestral sampling technique. We have particularized the proposed method to the problem of estimating the unobserved rate parameters of stochastic kinetic models (SKMs) [11]. Such models describe the time evolution of the population of a set of species or chemical molecules, which evolve according to the mentioned set of rate parameters, and present an autoregulatory behavior. This problem is currently of great interest in a variety of biological and molecular problems, such as complex autoregulatory gene networks. As a simple and intuitive example, yet physically meaningful, we have obtained numerical results for the Lotka-Volterra model, also known as a predator-prey model, consisting of two interacting species related by three reactions with associated unknown rates.

The rest of the paper is organized as follows. In Section 2, we give a formal statement of the class of problems we address. In Section 3, the population Monte Carlo algorithm is described. In Section 4, the formalism of Gaussian Bayesian networks is briefly reviewed. In Section 5, the proposed algorithm is introduced. In Section 6, we describe the practical application of the proposed algorithm to the problem of estimating the rates of a SKM, and present computer simulation results. Section 7 is devoted to the conclusions.

2. PROBLEM STATEMENT

Let $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_K]^\top$ be a vector of K unobserved real random variables with prior density $p(\boldsymbol{\theta})$ and let $\mathbf{y}$ be a vector of real random observations related to $\boldsymbol{\theta}$ by way of a conditional probability density function (pdf) $p(\mathbf{y}|\boldsymbol{\theta})$.

In this paper we address the problem of approximating the posterior pdf of $\boldsymbol{\theta}$, denoted by $p(\boldsymbol{\theta}|\mathbf{y})$, using a random grid of M points, $\{\boldsymbol{\theta}^{(i)}\}_{i=1}^{M}$, in the space of the random vector $\boldsymbol{\theta}$. With this grid, it is simple to approximate any moments of $p(\boldsymbol{\theta}|\mathbf{y})$ (e.g., the posterior mean). However, the generation of useful samples that represent $p(\boldsymbol{\theta}|\mathbf{y})$ adequately when K is large is normally a very difficult task.

The main goal of this work is to devise and assess a sampling scheme that is sequential in the space of $\boldsymbol{\theta}$, i.e., that draws scalar random variates $\theta_1, \ldots, \theta_K$ one at a time and can be used with arbitrary values of K.

3. POPULATION MONTE CARLO

3.1 Importance sampling

The main application of statistical Monte Carlo methods is the approximation of integrals $\mathcal{I}(f)$ by means of empirical sums $\mathcal{I}^M(f)$, which are of the form

$$\mathcal{I}(f) = \int f(\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}, \quad \mathcal{I}^M(f) = \frac{1}{M}\sum_{i=1}^{M} f\left(\boldsymbol{\theta}^{(i)}\right),$$

where f is a real, integrable function of $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta})$ is some pdf of interest (often termed the *target* density), and $\{\boldsymbol{\theta}^{(i)}\}_{i=1}^M$ is a collection of M Monte Carlo samples drawn from $\pi(\boldsymbol{\theta})$. It is straightforward to analyze the convergence of $\mathcal{I}^M(f)$ towards $\mathcal{I}(f)$ [3].

However, in many practical cases it is not possible to sample from $\pi(\boldsymbol{\theta})$ directly. A common approach to overcome this difficulty is to apply an importance sampling (IS) procedure [3]. The key idea is to draw the samples $\{\boldsymbol{\theta}^{(i)}\}_{i=1}^M$ from a (simpler) proposal pdf, or importance function, $q(\boldsymbol{\theta})$, and then compute normalized importance weights of the form

$$w^{(i)} \propto \pi\left(\boldsymbol{\theta}^{(i)}\right) / \, q\left(\boldsymbol{\theta}^{(i)}\right), \quad i = 1, \ldots, M.$$

The integral $\mathcal{I}(f)$ is then approximated by the weighted sum

$$\mathcal{I}^M(f) = \sum_{i=1}^{M} w^{(i)} f\left(\boldsymbol{\theta}^{(i)}\right).$$

The efficiency of an IS algorithm depends heavily on the choice of the proposal, $q(\boldsymbol{\theta})$. However, in order to ensure the asymptotic convergence of the approximation $\mathcal{I}^M(f)$, when M is large enough, it is sufficient to select $q(\boldsymbol{\theta})$ such that $q(\boldsymbol{\theta}) > 0$ whenever $\pi(\boldsymbol{\theta}) > 0$ [3].

3.2 Population Monte Carlo algorithm

The population Monte Carlo (PMC) method [5] is an iterative IS scheme that generates a sequence of proposal pdfs $q_\ell(\boldsymbol{\theta})$, $\ell = 0, \ldots, L$, such that every new proposal is "closer" to the target density $\pi(\boldsymbol{\theta})$ than the previous importance function. Such scheme demands the ability to learn about the target $\pi(\boldsymbol{\theta})$, given the set of particles and weights at the $(\ell-1)$-th iteration, in order to produce the new proposal $q_\ell(\boldsymbol{\theta})$ for the ℓ-th iteration. Taking this ability for granted, the algorithm is simple and can be outlined as follows:

Iteration ($\ell = 0, \ldots, L$):

1. Select a proposal pdf $q_\ell(\boldsymbol{\theta})$.
2. Draw a collection of M i.i.d. (independent and identically distributed) samples $\Theta_\ell^M = \{\boldsymbol{\theta}_\ell^{(i)}\}_{i=1}^M$ from $q_\ell(\boldsymbol{\theta})$.
3. For each sample $\boldsymbol{\theta}_\ell^{(i)}$ compute the normalized weight $w_\ell^{(i)} \propto \pi\left(\boldsymbol{\theta}_\ell^{(i)}\right) / \, q_\ell\left(\boldsymbol{\theta}_\ell^{(i)}\right)$, $i = 1, \ldots, M$.

In our problem, the target density is the posterior pdf of $\boldsymbol{\theta}$, i.e., $\pi(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathbf{y})$, and a straightforward way of initializing the algorithm is to use the prior as the starting proposal, $q_0(\boldsymbol{\theta}) = p(\boldsymbol{\theta})$. However, the key element of the PMC iteration is the proposal update mechanism. We propose to select the importance function as the joint pdf of a Gaussian Bayesian network that represents the conditional dependencies among the θ_k's by means of Gaussian distributions and allows for a simple sampling scheme.

4. GAUSSIAN BAYESIAN NETWORKS

The aim of this section is to show how the proposal functions $q_\ell(\boldsymbol{\theta})$, $\ell = 1, \ldots, L$, can be suitably constructed from the unweighted sample set $\tilde{\Theta}_{\ell-1}^M = \{\tilde{\boldsymbol{\theta}}_{\ell-1}^{(i)}\}_{i=1}^M$, obtained from $\Theta_{\ell-1}^M = \{\boldsymbol{\theta}_{\ell-1}^{(i)}, w_{\ell-1}^{(i)}\}_{i=1}^M$ via a resampling step, using the formalism of Gaussian Bayesian networks (GBNs).

4.1 Bayesian networks

A Bayesian network (BN) is a probabilistic graphical model that encodes the relationships among a set of variables of interest $\Theta = \{\theta_1, \ldots, \theta_K\}$ using a directed acyclic graph (DAG). Here, we consider the case when all the variables θ_k associated with the nodes are continuous and the probabilistic model of the graph is given by a joint pdf $q(\boldsymbol{\theta})$ [9].

The parent set $\Pi_k \subset \Theta$ of a variable θ_k is the subset of nodes in the graph such that there exists an edge from any node $\theta_j \in \Pi_k$ to θ_k. It is always possible to find a topological sorting of the variables such that the parents of a node θ_k have strictly lower indices, i.e., if $\theta_j \in \Pi_k$ then $j < k$.

Moreover, BNs satisfy a conditional independence property which states that each variable is independent of its non-descendants given its parent nodes, i.e., $q(\theta_k|\theta_1, \ldots, \theta_{k-1}) = q(\theta_k|\Pi_k)$. This property allows the factorization of the joint pdf of the BN by way of the chain-rule as the product of the unidimensional conditional pdfs of each variable θ_k, namely $q(\boldsymbol{\theta}) = \prod_{k=1}^K q(\theta_k|\Pi_k)$, allowing for an efficient numerical evaluation of $q(\boldsymbol{\theta})$.

An important reason for the use of BNs as proposal functions is the fact that the topological sorting of the variables allows to sample efficiently from joint densities $q(\boldsymbol{\theta})$ with arbitrarily high dimension K with a simple procedure which requires to draw from unidimensional densities only. This algorithm is known as ancestral sampling [9] and it consists in sampling from the unidimensional conditional distributions $q(\theta_k|\Pi_k^{(i)})$ of θ_k conditioned on the samples of the parent variables $\Pi_k^{(i)}$ in the topological order $k = 1, \ldots, K$. Finally, the set of scalar samples $\{\theta_k^{(i)}\}_{k=1}^K$ corresponds to a vector sample $\boldsymbol{\theta}^{(i)}$ from the joint pdf $q(\boldsymbol{\theta})$.

4.2 Gaussian Bayesian networks

In this work we have considered a particular kind of BNs known as Gaussian BNs (GBN) [7], where the joint probability distribution of $\boldsymbol{\theta}$ is a multivariate normal distribution $q(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta};\ \boldsymbol{\mu}, \boldsymbol{\Sigma})$ with mean vector $\boldsymbol{\mu}$ and positive definite covariance matrix $\boldsymbol{\Sigma}$. Given a topological sorting of the variables this joint density can be factorized into a product of univariate, independent normal conditional pdfs

$$q(\theta_k|\Pi_k) = \mathcal{N}\left(\theta_k;\ \mu_k + \sum_{j=1}^{k-1} c_{k,j}\,(\theta_j - \mu_j),\ \sigma_k^2\right),$$

where μ_k is the unconditional mean of θ_k, σ_k^2 is the variance of θ_k conditioned on values of its parents and $c_{k,j}$ is a linear coefficient reflecting the strength of the relationship between θ_j and θ_k. Thus, a multivariate normal distribution is equivalent to a GBN, and the precision matrix $\mathbf{T} = \boldsymbol{\Sigma}^{-1}$ may be easily obtained from the GBN parameters σ_k^2 and $\mathbf{c}_k = [c_{k,1}, \ldots, c_{k,k-1}]^\top$, $k = 1, \ldots, K$, and viceversa.

4.3 Learning GBNs from data

A number of algorithms have been proposed in the literature to learn the "posterior" network structure of a BN and its parameters, as a combination of prior knowledge and a set of observed data [7]. Such algorithms are based on a scoring metric, that provides a quantitative assessment of the "goodness" of the BN, and a search procedure that explores the (usually large) space of candidate networks and determines the sequence of BNs to be scored.

4.3.1 Scoring metric

A Bayesian measure of the goodness of a network structure B is its posterior probability $q(B|\tilde{\Theta}^M)$ given an unweighted sample set $\tilde{\Theta}^M$ from the joint density $q(\boldsymbol{\theta})$. Its computation is practically intractable even for small networks and the joint density $q(\tilde{\Theta}^M, B) = q(\tilde{\Theta}^M|B)q(B)$ is often used as a score. If all network structures are equally likely a priori the joint density reduces to the likelihood $q(\tilde{\Theta}^M|B)$. The scoring metric of an arbitrary GBN B may be computed based on the metric for a complete network [7].

Different DAG structures may represent the same dependencies among variables, and are then said to be equivalent. In such case, they provide equivalent factorizations of the joint density $q(\boldsymbol{\theta})$, thus yielding the same score [7].

4.3.2 Search procedure

Unless the number of variables is small, it is computationally infeasible to search the maximizing GBN structure by exhaustively considering all DAG patterns. For this reason, heuristic approximate search algorithms over DAGs have been proposed to seek highly likely structures [9].

In this work a greedy search algorithm has been considered [9]. This algorithm starts from an initial DAG structure (e.g. the empty graph) and considers the operations of addition, removal or reversal of an edge in the graph. At each iteration, it computes the scoring metric of every network structure that may be obtained from the current one performing a single operation, and "greedily" selects the one that maximizes the metric. The algorithm stops when no operation increases this score.

In combination with the scoring metric, this algorithm enables us to select the DAG structure B^* that best fits the set of data $\tilde{\Theta}^M_{\ell-1}$, and the associated joint pdf constitutes the importance function $q_\ell(\boldsymbol{\theta})$.

5. GBN-PMC ALGORITHM

In this work, we propose to represent the importance functions in the PMC scheme as the joint densities of a sequence of GBNs. The structure and parameters of the network are adaptively adjusted along the iterations, based on a sequence of unweighted sample sets $\tilde{\Theta}^M_\ell = \{\tilde{\boldsymbol{\theta}}^{(i)}_\ell\}_{i=1}^M$, $\ell = 0, \ldots, L$.

In high dimensional problems (where K is large), it is hard to devise efficient Monte Carlo sampling methods. Indeed, unless the likelihood function $p(\mathbf{y}|\boldsymbol{\theta})$ be very broad (i.e., the data are very noisy), the probability of generating samples with non-negligible likelihood using a proposal distribution is extremely low. In order to circumvent this difficulty, we propose to generate a sequence of models $p_\ell(\boldsymbol{\theta}|\mathbf{y}) \propto p_\ell(\mathbf{y}|\boldsymbol{\theta})p_\ell(\boldsymbol{\theta})$, $\ell = 0, \ldots, L$, that converges to the real posterior $p(\boldsymbol{\theta}|\mathbf{y})$ when L is large enough but are "simpler" (e.g., they have broader likelihoods, broader priors or both) for Monte Carlo approximation. If we keep approximating accurately the models $p_\ell(\boldsymbol{\theta}|\mathbf{y})$ along the iterations of a PMC algorithm, as $p_\ell(\boldsymbol{\theta}|\mathbf{y}) \to p(\boldsymbol{\theta}|\mathbf{y})$ we obtain an adequate Monte Carlo representation of $p(\boldsymbol{\theta}|\mathbf{y})$.

The proposed generic GBN-PMC algorithm for the estimation of the hidden parameters $\boldsymbol{\theta}$ is summarized below.

Initialization ($\ell = 0$):

1. Draw a collection of M samples, $\Theta^M_0 = \{\boldsymbol{\theta}^{(i)}_0\}_{i=1}^M$, from the prior importance function $q_0(\boldsymbol{\theta}) = p(\boldsymbol{\theta})$.
2. Go to step 4.

PMC iteration ($\ell = 1, \ldots, L$):

1. Resample with replacement from the weighted set $\Theta^M_{\ell-1} = \{\boldsymbol{\theta}^{(i)}_{\ell-1}, w^{(i)}_{\ell-1}\}_{i=1}^M$ to obtain the unweighted set $\tilde{\Theta}^M_{\ell-1} = \{\tilde{\boldsymbol{\theta}}^{(i)}_{\ell-1}\}_{i=1}^M$.
2. Apply the greedy search algorithm (Section 4.3.2) to select a GBN structure B^* with high likelihood $q(\tilde{\Theta}^M_{\ell-1}|B^*)$ and estimate its parameters $\{\mu_k, \sigma^2_k, \mathbf{c}_k\}$, $k = 1, \ldots, K$ to construct the new proposal density $q_\ell(\boldsymbol{\theta})$.
3. Apply the ancestral sampling algorithm to draw a collection of M samples $\Theta^M_\ell = \{\boldsymbol{\theta}^{(i)}_\ell\}_{i=1}^M$ from $q_\ell(\boldsymbol{\theta})$.
4. Construct the ℓ-th model $p_\ell(\boldsymbol{\theta}|\mathbf{y})$ based on the set Θ^M_ℓ.
5. Compute the normalized weights for each particle $\boldsymbol{\theta}^{(i)}_\ell$, $i = 1, \ldots, M$,

$$w^{(i)}_\ell \propto \frac{p_\ell\left(\boldsymbol{\theta}^{(i)}_\ell \mid \mathbf{y}\right)}{q_\ell\left(\boldsymbol{\theta}^{(i)}_\ell\right)} \propto \frac{p_\ell\left(\mathbf{y} \mid \boldsymbol{\theta}^{(i)}_\ell\right) p_\ell\left(\boldsymbol{\theta}^{(i)}_\ell\right)}{q_\ell\left(\boldsymbol{\theta}^{(i)}_\ell\right)}. \quad (1)$$

6. EXAMPLE: A STOCHASTIC KINETIC MODEL

In this section, the proposed GBN-PMC algorithm is applied to the problem of estimating the hidden parameters of a simple stochastic kinetic model (SKM), known as the predator-prey model. A SKM is a multivariate continuous-time jump process modeling the interactions among molecules, or species, that take place in chemical reaction networks of biochemical and cellular systems [11, 2].

6.1 Predator-prey model

The Lotka-Volterra, or predator-prey, model is a simple SKM that describes the time evolution of two species x_1 (prey) and x_2 (predator), by means of $K = 3$ reaction equations [2, 10]

$$\begin{aligned} x_1 &\xrightarrow{\theta_1} 2x_1 && \text{prey reproduction} \\ x_1 + x_2 &\xrightarrow{\theta_2} 2x_2 && \text{predator reproduction} \\ x_2 &\xrightarrow{\theta_3} \emptyset && \text{predator death} \end{aligned}$$

The k-th reaction takes place stochastically according to its instantaneous rate $a_k(t) = \theta_k g_k(\mathbf{x}(t))$, where $\theta_k > 0$ is the random rate parameter and $g_k(\cdot)$ is a continuous function of the current state of the system $\mathbf{x}(t) = [x_1(t), x_2(t)]^\top$. We denote by $x_1(t), x_2(t)$ the nonnegative, integer population of each species at time t. In this simple example, the instantaneous rates are of the form

$$a_1(t) = \theta_1 x_1(t),\ a_2(t) = \theta_2 x_1(t) x_2(t),\ a_3(t) = \theta_3 x_2(t).$$

The waiting time to the next reaction is exponentially distributed with parameter $a_0(t) = \sum_{k=1}^K a_k(t)$, and the probability of each reaction type is given by $a_k(t)/a_0(t)$.

We denote by $\mathbf{x}$ the vector containing the population of each species at the occurrence time of each reaction in a time interval $t \in [0, T]$, i.e., $\mathbf{x} = [\mathbf{x}^\top(t_1), \mathbf{x}^\top(t_2), \ldots, \mathbf{x}^\top(t_R)]^\top$, where R is the total number of reactions occurred in the time period of length T.

Assuming that the entire vector $\mathbf{x}$ is observed, the likelihood function for the rate parameter vector $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_K]^\top$ may be computed analytically, and it allows the factorization [11]

$$p(\mathbf{x}|\boldsymbol{\theta}) = \prod_{k=1}^K p(\mathbf{x}|\theta_k) = \prod_{k=1}^K \theta_k^{r_k} \exp\left\{-\theta_k \int_0^T g_k(\mathbf{x}(t))\, dt\right\},$$

where r_k is the total number of reactions of type k occurred in the time interval $[0, T]$.

The structure of this likelihood function allows the selection of a conjugate prior distribution for the rate parameters, comprising independent Gamma components, i.e.,

$$p(\boldsymbol{\theta}) = \prod_{k=1}^K p(\theta_k) = \prod_{k=1}^K \mathcal{G}(\theta_k;\, a_k, b_k),$$

where $a_k, b_k > 0$ are the scale and shape parameters of each component, respectively. Thus, the posterior distribution $p(\boldsymbol{\theta}|\mathbf{x}) = \prod_{k=1}^{K} p(\theta_k|\mathbf{x})$ may be also factorized into a set of independent Gamma components

$$p(\theta_k|\mathbf{x}) = \mathcal{G}\left(\theta_k;\ a_k + r_k, b_k + \int_0^T g_k\left(\mathbf{x}(t)\right) dt\right),$$

which indicates that, in the complete-data scenario, exact inference may be done for each rate constant θ_k separately. However, making inference for complex, high-dimensional and discretely observed SKMs (where $\mathbf{x}$ is not fully observed) is a challenging problem [2].

Exact stochastic simulation of generic SKMs, and predator-prey models in particular, can be carried out by the Gillespie algorithm [8]. This procedure allows to draw samples from the prior pdf of the populations, $p(\mathbf{x}|\boldsymbol{\theta})$, for arbitrarily high-dimensional SKMs.

In this work we restrict ourselves to this simple but representative example of SKM. A generalization of the algorithm to more complex models is straightforward given the efficient sampling procedure and evaluation of the weights in high dimensional problems. An additional complexity of this model relies is the highly dimensional auxiliar random variable $\mathbf{x}$, which makes this problem numerically difficult to tackle.

6.2 GBN-PMC algorithm for SKMs

We assume that a set of J noisy observations of the populations of both species are collected at regular time intervals of length Δ, that is, $\mathbf{y}_j = \mathbf{x}_j + \mathbf{u}_j,\ j = 1, \ldots, J$, where $\mathbf{x}_j = [x_1(j\Delta), x_2(j\Delta)]^\top$ and $\mathbf{u}_j$ is a Gaussian noise component with zero mean vector and covariance matrix $\sigma^2\mathbf{I}$. We denote the complete vector of observations as $\mathbf{y} = [\mathbf{y}_1^\top, \ldots, \mathbf{y}_J^\top]^\top$ with dimension $2J \times 1$. Thus, the likelihood of the populations $\mathbf{x}$ is given by $p(\mathbf{y}|\mathbf{x}) = \prod_{j=1}^{J} \mathcal{N}\left(\mathbf{y}_j;\ \mathbf{x}_j, \sigma^2\mathbf{I}\right)$.

The goal is to estimate the posterior density $p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$, given the prior distribution $p(\boldsymbol{\theta})$ and the likelihood $p(\mathbf{y}|\boldsymbol{\theta})$, using the GBN-PMC method. The model construction and the computation of the weights is now particularized for this concrete application.

In this particular problem, the observations $\mathbf{y}$ are related to the variables $\boldsymbol{\theta}$ through the random vector $\mathbf{x}$. Indeed, the likelihood of $\boldsymbol{\theta}$ has the form

$$p(\mathbf{y}|\boldsymbol{\theta}) = \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\boldsymbol{\theta})d\mathbf{x} = E_{p(\mathbf{x}|\boldsymbol{\theta})}\left[p(\mathbf{y}|\mathbf{x})\right], \qquad (2)$$

where $E_{p(\mathbf{x}|\boldsymbol{\theta})}[\cdot]$ denotes expectation with respect to the pdf in the subscript, and $p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = p(\mathbf{y}|\mathbf{x})$, since the observations are independent of the parameters $\boldsymbol{\theta}$ given the population vector $\mathbf{x}$.

In principle, it is possible to approximate the integral in (2) as an average of the likelihoods $p(\mathbf{y}|\mathbf{x}^{(i)})$ of a set $\{\mathbf{x}^{(i)}\}_{i=1}^{I}$ of exact Monte Carlo samples from the density $p(\mathbf{x}|\boldsymbol{\theta})$, drawn using the Gillespie algorithm, that is,

$$p(\mathbf{y}|\boldsymbol{\theta}) \approx \frac{1}{I}\sum_{i=1}^{I} p\left(\mathbf{y} \mid \mathbf{x}^{(i)}\right).$$

This approach, however, is computationally intractable, because it demands drawing a huge number of samples I to obtain a useful approximation of the posterior $p(\mathbf{y}|\boldsymbol{\theta})$, since the probability of generating a trajectory of populations $\mathbf{x}^{(i)}$ similar to the observations is extremely low.

To overcome this difficulty, we propose a simple approach based on using a standard particle filter [6] to approximate the posterior mean of $\mathbf{x}$ conditioned on $\mathbf{y}$ and a realization $\boldsymbol{\theta}^{(i)}$ of the rate parameters,

$$\hat{\mathbf{x}}^{(i)} = E_{p\left(\mathbf{x}|\mathbf{y},\boldsymbol{\theta}^{(i)}\right)}[\mathbf{x}] = \int \mathbf{x}\, p\left(\mathbf{x} \mid \mathbf{y}, \boldsymbol{\theta}^{(i)}\right) d\mathbf{x}.$$

Complete details on the implementation of the particle filter and the approximation of $\hat{\mathbf{x}}^{(i)}$ can be found in [6].

Thus, for each sample $\boldsymbol{\theta}_\ell^{(i)}$ obtained via ancestral sampling in step 3 of the GBN-PMC scheme, an approximation of the population vector $\hat{\mathbf{x}}_\ell^{(i)}$ is computed via particle filtering and the likelihood of Eq. (2) is approximated as

$$p\left(\mathbf{y} \mid \boldsymbol{\theta}_\ell^{(i)}\right) \approx p\left(\mathbf{y} \mid \hat{\mathbf{x}}_\ell^{(i)}\right).$$

We have considered a model update scheme of the form $p_\ell\left(\boldsymbol{\theta}|\mathbf{y}\right) \propto p_\ell(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ where the likelihood $p_\ell(\mathbf{y}|\boldsymbol{\theta})$ is constructed by a clipping procedure, thus obtaining a flat likelihood in the region of interest. As a consequence, a large enough set of samples obtains non negligible weights, which allows to compute consistent Monte Carlo approximations in high dimensional spaces.

Specifically, the model at iteration ℓ is computed from the likelihood of the sample set $\Theta_\ell^M = \{\boldsymbol{\theta}_\ell^{(i)}\}_{i=1}^{M}$ as

$$p_\ell\left(\mathbf{y} \mid \boldsymbol{\theta}_\ell^{(i)}\right) \propto \min\left\{\mathcal{T}_\ell, p\left(\mathbf{y} \mid \boldsymbol{\theta}_\ell^{(i)}\right)\right\},$$

where the threshold $\mathcal{T}_\ell$ is computed such that the number of samples $\boldsymbol{\theta}_\ell^{(i)}$ that satisfy $p(\mathbf{y}|\hat{\mathbf{x}}_\ell^{(i)}) \geq \mathcal{T}_\ell$ is equal to $M_T < M$.

The threshold $\mathcal{T}_\ell$ converges to the maximum of the likelihood $p(\mathbf{y}|\boldsymbol{\theta})$ as ℓ grows, and thus, also the model $p_\ell\left(\mathbf{y}|\boldsymbol{\theta}\right)$ converges to $p\left(\mathbf{y}|\boldsymbol{\theta}\right)$. We may define a stopping criteria based on the value of the threshold $\mathcal{T}_\ell$ and stop iterating when it reaches a stable value. The parameter M_T must be selected to represent adequately the multidimensional posterior distribution $p_\ell\left(\boldsymbol{\theta}|\mathbf{y}\right)$.

The model update step 4 in the general GBN-PMC algorithm requires the computation of the likelihood $p(\mathbf{y}|\hat{\mathbf{x}}_\ell^{(i)})$ for each particle $\boldsymbol{\theta}_\ell^{(i)}$, $i = 1, \ldots, M$. Then, the threshold $\mathcal{T}_\ell$ must be computed and the ℓ-th model $p_\ell\left(\mathbf{y}|\mathbf{x}\right)$ constructed via the clipping procedure.

Finally, the normalized weights in (1) are of the form

$$w_\ell^{(i)} \propto p_\ell\left(\mathbf{y} \mid \hat{\mathbf{x}}_\ell^{(i)}\right) p\left(\boldsymbol{\theta}_\ell^{(i)}\right) / \ q_\ell\left(\boldsymbol{\theta}_\ell^{(i)}\right).$$

6.3 Computer simulation results

We have applied the proposed GBN-PMC algorithm to the problem of estimating the posterior pdf of the constant rate parameters vector $\boldsymbol{\theta}$ in a simple predator-prey model. The true vector of parameter rates which we aim to estimate has been set to $\boldsymbol{\theta} = [0.5, 0.0025, 0.3]^\top$.

A realization of the populations $\mathbf{x}$ has been generated from the prior distribution $p(\mathbf{x}|\boldsymbol{\theta})$ with initial populations set to $\mathbf{x}(0) = [71, 79]^\top$, and a total length of $T = 40$. The observation vector $\mathbf{y}$ has been obtained with an observation period $\Delta = 1$ and a Gaussian noise variance $\sigma^2 = 100$.

Figure 1 (left) depicts the time evolution of the true populations of both species $\mathbf{x}$, and the corresponding discrete-time noisy observations $\mathbf{y}$. The autoregulatory behavior of the model can be clearly observed on the graph.

The parameters of the prior Gamma distribution $p(\boldsymbol{\theta})$ have been set such that the corresponding mean and standard deviation vectors are $[0.4, 0.0035, 0.4]^\top$ and $[0.1, 0.001, 0.1]^\top$, respectively. That is, they present a bias with respect to the true vector $\boldsymbol{\theta}$. Both the marginal prior

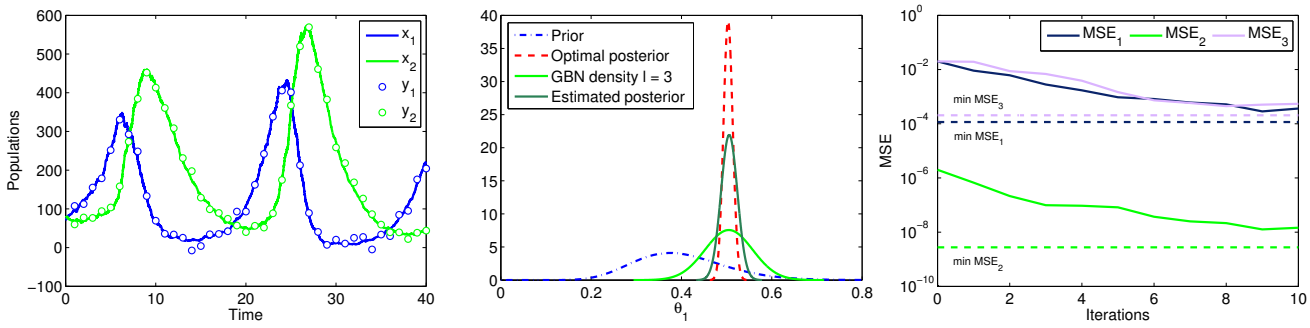

Figure 1: *Left:* Real populations ($\mathbf{x}$) and discrete-time noisy observations ($\mathbf{y}$) in a predator-prey model. *Middle:* Approximate posterior densities generated by the GBN-PMC algorithm. The prior and the optimal posterior (with full knowledge of $\mathbf{x}$) are also displayed for comparison. *Right:* Evolution of the MSE of the rate estimates.

and the optimal posterior computed in this example for θ_1 are shown on the central plot in Figure 1.

We have applied the proposed GBN-PMC algorithm to the approximation of $p(\boldsymbol{\theta}|\mathbf{y})$ in this scenario. We have set the parameters to $M = 100$, $M_T = 50$ and $L = 10$.

We have observed that, given discrete-time and noisy observations of the process $\mathbf{x}$, the rate parameters present posterior dependencies, which have been modeled in terms of a GBN. In this example it is computationally feasible to evaluate the likelihood $q(\tilde{\Theta}_\ell^M|B)$ of every possible GBN structure B given a set of unweighted samples $\tilde{\Theta}_{\ell-1}^M$ since, for $K = 3$, there exist only 11 equivalence classes of DAGs. The network structure that maximizes the metric at each iteration $\ell = 0, \ldots, L$ is a complete network, and we have arbitrarily selected the DAG structure with the factorization $q_\ell(\boldsymbol{\theta}) = q_\ell(\theta_1)q_\ell(\theta_2|\theta_1)q_\ell(\theta_3|\theta_1, \theta_2)$.

Figure 1 (middle) displays the evolution of the marginal proposal densities of the rate parameter θ_1. It can be observed that it smoothly converges to the optimum posterior in a low number of iterations ($L = 10$). The difference of the final estimate with respect to the optimal posterior is due to the discrete-time and noisy nature of the observations $\mathbf{y}$ (the optimal posterior corresponds to the case of complete data, i.e., $\mathbf{x}$ is fully observed). The results obtained for the rest of the parameters are very similar and have been omitted.

We asses the merit of an unweighted sample set $\{\tilde{\theta}_k^{(i)}\}_{i=1}^M$ as an estimator of the parameter θ_k, by means of the mean square error $MSE_k = \frac{1}{M}\sum_{i=1}^M (\tilde{\theta}_k^{(i)} - \theta_k)^2$, $k = 1, 2, 3$.

Figure 1 (right) shows the evolution of the MSE of each parameter, as well as the corresponding lower bound given by the optimal marginal posterior. It can be seen that it smoothly decreases until a final value close to the lower bound, even with $L = 10$ iterations. The convergence of the algorithm can be adjusted by tuning M and M_T.

7. CONCLUSIONS

We have addressed the problem of approximating posterior probability distributions in spaces of potentially high dimension. We have proposed a novel PMC scheme, termed GBN-PMC, which is based on a representation of the proposal distributions as the joint density of a Gaussian Bayesian network. This graphical model provides a straightforward sampling procedure and a factorization of the joint density that enables an efficient evaluation in high-dimensional spaces. In order to illustrate the application of the proposed technique, we have presented numerical results obtained in the estimation of the rate parameters of a simple stochastic kinetic model, known as predator-prey model.

8. ACKNOWLEDGMENTS

E. K. acknowledges the support of *Ministerio de Educación* of Spain (*Programa de Formación de Profesorado Universitario*, ref. AP2008-00469). This work has been partially supported by *Ministerio de Ciencia e Innovación* of Spain (program Consolider-Ingenio 2010 CSD2008-00010 COMONSENS and project DEIPRO TEC2009-14504-C02-01) and a joint program of *Comunidad de Madrid* and *Universidad Carlos III de Madrid* (CCG10-UC3M/TIC-5225 ETORS).

REFERENCES

[1] J. M. Bower and H. Bolouri, editors. *Computational Modeling of Genetic and Biochemical Networks*. MIT Press, Cambridge (MA), 2001.

[2] R. J. Boys, D. J. Wilkinson, and T. B. L. Kirkwood. Bayesian inference for a discretely observed stochastic kinetic model. *Statistics and Computing*, 18(2):125–135, 2008.

[3] G. Casella C. P. Robert. *Monte Carlo Statistical Methods*. Springer, 2004.

[4] O. Cappé, R. Douc, A. Guillin, J. M. Marin, and C. P. Robert. Adaptive importance sampling in general mixture classes. *Statistics and Computing*, 18(4):447–459, 2008.

[5] O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert. Population Monte Carlo. *Statistics and Computing*, 13(4):907–929, 2004.

[6] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Burgallo, and J. Míguez. Particle filtering. *IEEE Signal Processing Magazine*, 20(5):19–38, 2003.

[7] D. Geiger and D. Heckerman. Learning Gaussian networks. In *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence*, 1994.

[8] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[9] R. Neapolitan. *Learning Bayesian Networks*. Prentice-Hall, 2004.

[10] V. Volterra. Fluctuations in the abundance of a species considered mathematically. *Nature*, 118:558–560, 1926.

[11] D. J. Wilkinson. *Stochastic Modelling for Systems Biology*. Chapman & Hall/CRC, Mathematical and Computational Biology Series, 2006.