Computational Statistics and Data Analysis 118 (2018) 1–17

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Accelerating pseudo-marginal MCMC using Gaussian processes

Christopher C. Drovandi [a,b,*], Matthew T. Moores [c], Richard J. Boys [d]

[a] *Queensland University of Technology, Brisbane, 4000, Australia*
[b] *Australian Centre of Excellence for Mathematical and Statistical Frontiers, Australia*
[c] *University of Warwick, Coventry CV4 7AL, United Kingdom*
[d] *Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom*

ARTICLE INFO

Article history:
Received 29 March 2017
Received in revised form 28 August 2017
Accepted 3 September 2017
Available online 14 September 2017

Keywords:
Gaussian processes
Likelihood-free methods
Markov processes
Particle Markov chain Monte Carlo
Pseudo-marginal methods
State space models

ABSTRACT

The grouped independence Metropolis–Hastings (GIMH) and Markov chain within Metropolis (MCWM) algorithms are pseudo-marginal methods used to perform Bayesian inference in latent variable models. These methods replace intractable likelihood calculations with unbiased estimates within Markov chain Monte Carlo algorithms. The GIMH method has the posterior of interest as its limiting distribution, but suffers from poor mixing if it is too computationally intensive to obtain high-precision likelihood estimates. The MCWM algorithm has better mixing properties, but tends to give conservative approximations of the posterior and is still expensive. A new method is developed to accelerate the GIMH method by using a Gaussian process (GP) approximation to the log-likelihood and train this GP using a short pilot run of the MCWM algorithm. This new method called GP-GIMH is illustrated on simulated data from a stochastic volatility and a gene network model. The new approach produces reasonable posterior approximations in these examples with at least an order of magnitude improvement in computing time. Code to implement the method for the gene network example can be found at http://www.runmycode.org/companion/view/2663.

1. Introduction

Bayesian inference for high-dimensional latent variable models is currently challenging. In particular Markov chain Monte Carlo (MCMC) samplers can suffer from poor mixing due to correlation between the parameter of interest and the latent variables. Beaumont (2003) and Andrieu and Roberts (2009) have introduced pseudo-marginal methods to improve the statistical efficiency of MCMC. These methods work by replacing the actual likelihood with an unbiased likelihood estimate in the Metropolis–Hastings ratio. This allows proposals for the Markov chain to be made directly on the space of the parameter of interest, rather than conditional on the value of a set of the latent variables.

One of these methods, the grouped independence Metropolis–Hastings (GIMH) method by Beaumont (2003), recycles the likelihood estimate for the current value of the chain to the next iteration. Andrieu and Roberts (2009) have shown that the GIMH method has the desired posterior as its limiting distribution, which is why it has received considerable attention in the literature (Andrieu et al., 2010; Doucet et al., 2015). However, a drawback of the GIMH method is that it can suffer from poor mixing if it is too computationally expensive to estimate the likelihood with high precision.

* Correspondence to: School of Mathematical Sciences, Queensland University of Technology, P.O. Box 2434, Brisbane, Queensland, 4001, Australia.
E-mail addresses: c.drovandi@qut.edu.au (C.C. Drovandi), M.T.Moores@warwick.ac.uk (M.T. Moores), richard.boys@newcastle.ac.uk (R.J. Boys).

http://dx.doi.org/10.1016/j.csda.2017.09.002
0167-9473/

The other method, the Markov chain within Metropolis (MCWM, Beaumont, 2003) algorithm, estimates the likelihood at both the current and proposed values of the Markov chain at every iteration. This method generally possesses better mixing properties as it is able to escape an overestimated likelihood value by re-estimating it at the next MCMC iteration. However, the MCWM method does not have the posterior distribution of the parameter of interest as its limiting distribution. Because of this, MCWM has received comparatively less attention.

In this paper we make use of a Gaussian process (GP) to accelerate the GIMH method while at the same time accepting some approximation to the posterior distribution.

Wilkinson (2014) proposes that GPs be used to accelerate approximate Bayesian computation (ABC) methods where the likelihood is approximated by generating many model simulations from each proposed parameter value, and measuring the distance between observed and simulated data through a careful choice of summary statistics. Here GPs are used to emulate the actual (ABC) log-likelihood surface based on noisy estimates obtained through simulation. The method iteratively uses the GP to discard implausible parts of the parameter space, re-trains the GP in the updated not-implausible part of the parameter space and continues this process until the GP fit has been deemed as satisfactory. The final GP is then used within an MCMC method to predict the log-likelihood surface at all proposed values of the parameter of interest. GPs have also been used for ABC by Meeds and Welling (2014), Gutmann and Corander (2016) and Järvenpää et al. (2016).

We follow a similar approach to Wilkinson (2014) to accelerate pseudo-marginal methods. However, one key difference is that we take advantage of the pseudo-marginal literature. In particular, we use a short run of the MCWM method as a natural approach to obtain training samples for the GP in non-negligible regions of the posterior support. The MCWM method is ideal for training the GP as it has better mixing properties and is less prone to sticky periods than the GIMH method. Medina-Aguayo et al. (2016) develop sufficient conditions for the geometric ergodicity and hence the existence of an invariant distribution of MCWM. Our experience with MCWM is that it is generally conservative (inflated posterior variance), allowing the tails of the posterior to be explored. The fitted GP is used instead of expensive likelihood estimates within the GIMH method. We introduce further novelties into our method to make it practically useful.

The paper has the following outline. In Sections 2.1 and 2.2 we provide a brief overview of pseudo-marginal methods and GPs, respectively. In Section 2.3 we present our new method, GP-GIMH, which uses the MCWM algorithm to train the GP and subsequently uses the GP to accelerate the GIMH method. Finally, in Section 4, we conclude with a discussion.

2. Accelerated pseudo-marginal MCMC

In this section we give some background on pseudo-marginal MCMC methods and Gaussian processes before describing how, by emulating the log-likelihood using a GP, we can accelerate pseudo-marginal MCMC.

2.1. Pseudo-marginal MCMC

Suppose we have observed data $\boldsymbol{y}$ in Y which is described by a statistical model with likelihood function $p(\boldsymbol{y}|\boldsymbol{\theta})$ and depends on an unknown parameter $\boldsymbol{\theta}$ in $\mathbb{R}^d$. Prior beliefs about the parameter are summarised by the prior density $p(\boldsymbol{\theta})$. We assume that the model requires, or is facilitated by, an auxiliary variable $\boldsymbol{x}$ in X, whose value is not of direct interest. In this scenario the complete data likelihood is $p(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta})$ and leads to the observed data likelihood $p(\boldsymbol{y}|\boldsymbol{\theta}) = \int_{\mathsf{X}} p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta})d\boldsymbol{x}$. Ideally this observed data likelihood is combined with the prior to make inferences about the parameters via the posterior density $p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$. However, in non-toy problems the observed data likelihood is an analytically intractable integral. Therefore the parameter posterior is accessed as the marginal of the joint posterior for all unknowns, that is, via $p(\boldsymbol{\theta}|\boldsymbol{y}) = \int_{\mathsf{X}} p(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})d\boldsymbol{x}$.

A standard Bayesian approach for fitting such a latent variable model is to develop an MCMC algorithm that samples the joint posterior $p(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})$ and marginalises by ignoring the $\boldsymbol{x}$ samples. A common approach is to develop an MCMC algorithm using two blocks, $\boldsymbol{\theta}$ and $\boldsymbol{x}$, that iteratively samples from the full conditionals $p(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{y})$ and $p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$. A key problem with such algorithms is that they can mix poorly because of high posterior correlation between the blocks $\boldsymbol{\theta}$ and $\boldsymbol{x}$. Further, for non-trivial state space models, $p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ cannot be sampled directly and is difficult to sample efficiently (see Andrieu et al., 2010 for a discussion). In an attempt to overcome the mixing issue, Beaumont (2003) develops algorithms that replace the computationally intractable likelihood $p(\boldsymbol{y}|\boldsymbol{\theta})$ with an unbiased estimate $\hat{p}(\boldsymbol{y}|\boldsymbol{\theta})$. The underpinning mathematics of these pseudo-marginal MCMC algorithms is studied in Andrieu and Roberts (2009) and they develop conditions under which they indeed have the correct posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ as their limiting distribution. A simple example of an unbiased likelihood estimate is one obtained through importance sampling, namely

$$\hat{p}(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} \frac{p(\boldsymbol{y}|\boldsymbol{x}_i, \boldsymbol{\theta})p(\boldsymbol{x}_i|\boldsymbol{\theta})}{g(\boldsymbol{x}_i)},$$

where $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \overset{\text{i.i.d}}{\sim} g(\boldsymbol{x})$ and g is an importance density defined on X. Alternative approaches to obtaining an unbiased likelihood estimate are available. For example, Andrieu et al. (2010) show that when the model of interest is a state-space model, the likelihood $p(\boldsymbol{y}|\boldsymbol{\theta})$ can be estimated unbiasedly using a particle filter with N particles. Such pseudo-marginal methods are referred to as particle Markov chain Monte Carlo (PMCMC). We consider models in the state-space form in Section 3 and use the bootstrap particle filter of Gordon et al. (1993) to obtain an unbiased likelihood estimator.