Parallel Adaptive Markov chain Monte Carlo with applications

Mauro Bernardi and Lea Petrella

Abstract Adaptive Markov chain Monte Carlo methods have been applied successfully to many Bayesian statistical problems. These algorithms are specifically designed to automatically adjust the proposal parameters to match the shape of the posterior distribution during the simulation process. In this paper we first introduce a new adaptive Markov chain Monte Carlo algorithm where the posterior distribution is approximated by a mixture of multivariate t-distributions whose parameters are updated at each iteration. Then we extend the proposed sampler by allowing multiple interacting chains to run in parallel. We compare the prosed algorithms in a simulation experiment showing the superiority of the interacting scheme.

Key words: Markov chain Monte Carlo, adaptive Monte Carlo, Bayesian analysis, mixture model.

1 Introduction

In recent years, Markov chain Monte Carlo algorithms (MCMC) have revealed their enormous potential in solving the difficult problem of simulating from the posterior distribution arising in complex stochastic models. Thanks to the ability of the simulation methods to propose an efficient solution to the inferential problem, Bayesian statistics has become very popular. Sometimes MCMC methods show poor performances when high dimensional and multimodal target distributions are involved. In fact, even well designed algorithms can find difficulties in approximating posterior distributions in a reasonable computational time. Adaptive sampling methods

Mauro Bernardi
Dip. MEMOTEF, Sapienza University of Rome, e-mail: mauro.bernardi@uniroma1.it

Lea Petrella
Dip. MEMOTEF, Sapienza University of Rome, e-mail: lea.petrella@uniroma1.it

have been proposed to overcome these problems: they use previous draws of the simulation process to tailor the proposal distribution on the features of the target distribution. Adaptive samplers have been proposed in different contexts: Douc *et al.* [6] extend the population Monte Carlo algorithm of Douc *et al.* [5] and introduce mixture of Gaussians as general proposal distributions where the mixture parameters are fitted using the past history of the particles. In the MCMC context the seminal work of Haario *et al.* [7] on adaptive random walk Metropolis algorithms and the most famous work of Haario *et al.* [8] that introduces adaptation into an independent Metropolis-Hastings algorithm have been followed by a vast literature; for example Haario *et al.* [9], Atchadé and Rosenthal [3], Atchadé and Fort [4], Andrieu and Moulines [1], see also the recent review by Andrieu and Thoms [2] and references therein.

In this paper we propose a mixture of adaptive independent Metropolis kernels. The aim of the method is to overcome the main drawback of the Metropolis-type algorithms that combine fast and good local exploration properties with sub-optimal scaling factors, especially when high dimensional multimodal distributions are involved. The mixture of Independent Metropolis-Hastings kernels can improve the ability of the sampler to explore different regions of the state space.

The second major contribution of the paper consists in extending the adaptive sampling algorithm by considering multiple interacting chains with the aim of reducing the initial learning time. Parallel MCMC algorithms have been proposed by Jasra *et al.* [10] to deal with the problem of simulating from high dimensional multimodal target distributions. In its basic version these methods consist in running different chains in parallel allowing interactions among them that improve the global exploration of the target. The use of multiple interacting chains in conjunction with the flexibility of the proposal distribution guarantees a rapid mixing of the chain and reduces the computational time.

2 Adaptive MCMC

Suppose we are interested on sampling from a target distribution π having support in $\mathscr{X} \subset \mathbb{R}^d$, known up to a normalizing constant. The proposed adaptive-MCMC algorithm is based on a Independent Metropolis Mixture proposal kernel having the following general form, at iteration i:

$$q\left(x, \tilde{\Theta}, \Theta_i\right) = \lambda \mathsf{T}_d\left(x|\tilde{\Theta}\right) + (1-\lambda)\sum_{k=1}^{K} \omega_{k,i} \mathsf{T}_d\left(x|\Theta_{k,i}\right) \tag{1}$$

where $\mathsf{T}_d\left(x|\Theta_i\right)$ is the probability density of a $d-$variate Student-t distribution with parameters $\Theta_i = (\mu_i, \Sigma_i, \nu_i)$, $\omega_{k,i}$ are the mixture weights satisfying the constraints $\omega_{k,i} > 0, \forall\, 1 \leq k \leq K$, and $\sum_{k=1}^{K} \omega_{k,i} = 1$. $\lambda \in (0,1)$ is the weight associated to the non-adapted distribution $\mathsf{T}_d\left(x|\tilde{\Theta}\right)$ which parameters are fixed during iterations. The rationale behind the presence of the fixed component is twofold: first it guarantees

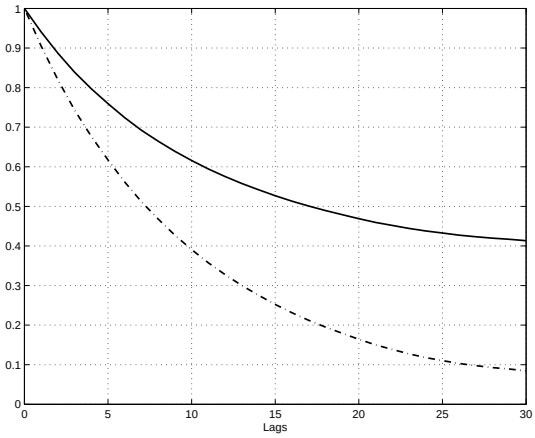

Fig. 1 Autocorrelation function of the posterior draws for the parameter μ_1 of the mixture model defined in equation (2). The adaptive-MCMC output (solid line) is compared with that of the parallel version (dotted line).

the convergence of the algorithm even in the case where the proposal parameter space is unbounded, see Andrieu and Moulines [1] and Haario *et al.* [8]. Second, it is very useful in exploring multimodal posterior distributions when the number of modes of the proposal is smaller than that of the target distribution. The form of the mixture representation is updated at each iteration i when a new observation becomes available from the simulation process. To adapt the proposal mixture parameters we minimize the Kullback-Leibler divergence from the target distribution using the stochastic approximation algorithm of Robbins and Monro [11].
As a second step, the proposed single chain adaptive MCMC is extended by allowing interactions between multiple chains running in parallel. In particular, we consider a population of N chains $X^{(j)} = \left\{x_i^{(j)}\right\}_{i\in\mathbb{N}}$, $j = 1,2,\ldots,N$, with the $j-$th chain having transition kernel equal to $q_i^{(j)}\left(x|\tilde{\Theta},\Theta_i\right)$ as in equation (1). To speed up the learning mechanism of the posterior parameters we consider three types of population moves as described in Jasra *et al.* [10]: mutation, exchange and crossover.

3 Numerical example

In this section we carry out a comparison between the proposed single chain adaptive MCMC algorithm and its parallel interacting extension. We simulate a sample of $1,000$ observations from the following bivariate mixture of two Normal distributions

$$\frac{1}{3}\mathsf{N}_2\left(\mu_1,\Sigma_1\right)+\frac{2}{3}\mathsf{N}_2\left(\mu_2,\Sigma_2\right) \tag{2}$$

with $\mu_1=(0,0)^\mathsf{T}$, $\mu_2=(10,10)^\mathsf{T}$, and $\Sigma_1=\Sigma_2=0.1\mathbb{I}_2$, and we consider $35,000$ iterations and the same proposal distributions for both algorithms.

In Figure 1 the autocorrelation function is plotted with 50 lags for the first component of the bivariate Metroposlis-Hastings chain. It is evident that the autocorrelation function of the interacting scheme MCMC algorithm is substantially lower as compared to that of the single chain.

References

1. Andrieu, C., Moulines, É.: On the ergodicity properties of some adaptive MCMC algorithms. Annals of Applied Probability. **16**, 1462–1505 (2006).
2. Andrieu, C., Thoms, J.: A tutorial on adaptive MCMC, Statistical Computing, **18**, 343–373 (2008).
3. Atchadé, Y.F., Rosenthal, J.S.: On adaptive Markov chain Monte Carlo algorithms. Bernoulli. **11**, 815–828 (2005).
4. Atchadé, Y. F., Fort, G.: Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. Bernoulli. **16**, 116–154 (2010).
5. Douc, R., Guillin, A., Marin, J.-R., and Robert, C.P.: Minimum Variance Importance Sampling via Population Monte Carlo. The Annals of Statistics. **35**, 420.-448 (2004).
6. Douc, R., Guillin, A., Marin, J.-R., Robert, C.P.: Convergence of Adaptive Mixtures of Importance Sampling Schemes. The Annals of Statistics. **35**, 420–448 (2007).
7. Haario, H., Saksman, E. and Tamminen, J.: Adaptive proposal distribution for Random Walk Metropolis algorithm, Computational Statistics. **14**, 375–395 (1999).
8. Haario, H., Saksman, E., Tamminen, J.: An Adaptive Metropolis algorithm. Bernoulli. **14**, pp. 223–242 (2001).
9. Haario, H., Saksman, E., Tamminen, J.: Componentwise adaptation for high dimensional MCMC. Computational Statistics. **20**, 265–274 (2005).
10. Jasra, A., Stephens, D.A., Holmes, C.C.: , On population-based simulation for static inference. Statistical Computing. **17**, 263–279 (2007).
11. Robbins, H., Monro, S.: A stochastic approximation method. Annals of Mathematical Statistics. **22**, 400–407 (1951).