Bayesian numerical inference for Markovian models – Application to tropical forest dynamics

Fabien Campillo, R. Rakotozafy, Vivien Rossi

International Conference on Approximation
Methods and Numerical Modelling in
Environment and Natural Resources
Granada (Spain), July 11-13, 2007

Bayesian numerical inference for markovian models. Application to tropical forest dynamics.

Fabien Campillo[1], **Rivo Rakotozafy**[2] **and Vivien Rossi**[3]

[1]INRIA/IRISA
Campus de Beaulieu, 35042 Rennes Cedex, France
e-mail: Fabien.Campillo@inria.fr
[2] University of Fianarantsoa
BP 1264, Andrainjato, 301 Fianarantsoa, Madagascar
e-mail: rrakotozafy@uni-fianar.mg
[3] CIRAD
Campus International de Baillarguet, 34398 Montpellier cedex 5, France
e-mail: vivien.rossi@cirad.fr

Keywords: Markov chain Monte Carlo method, interacting chains, hidden Markov model, forest dynamic model

Abstract. *Bayesian modelling is fluently employed to assess natural ressources. It is associated with Monte Carlo Markov Chains (MCMC) to get an approximation of the distribution law of interest. Hence in such situations it is important to be able to propose N independent realizations of this distribution law. We propose a strategy for making N parallel Monte Carlo Markov Chains interact in order to get an approximation of an independent N-sample of a given target law. In this method each individual chain proposes candidates for all other chains. We prove that the set of interacting chains is itself a MCMC method for the product of N target measures. Compared to independent parallel chains this method is more time consuming, but we show through example that it possesses many advantages. This approach will be applied to a forest dynamic model.*

1 Introduction

In many models arising in environment (ecology, population dynamics, renewable resource management etc.), measurements $\mathsf{y}_1, \ldots, \mathsf{y}_T$ are collected yearly or monthly so that the real-time constraint is not relevant even if the underlying law features a temporal structure. State-space modeling of these data consists in proposing a Markov process $(\mathsf{x}_t, \mathsf{y}_t)_{t=1\cdots T}$, where the state process x_t is not observed and y_t are the associated observation process. This process usually depends on some unknown parameter θ with given a priori law. The goal of the Bayesian inference is to determine the a posteriori law of $(\mathsf{x}_{1:T}, \theta)$ given the measurements $\mathsf{y}_{1:T}$. In the case of general state-space modeling, we need to utilize approximation procedures. The success of the

Bayesian inference is mainly due to the development of efficient Monte Carlo approximation techniques [6]. Among them, MCMC methods allow us to sample from almost any prescribed distribution law [3]. Still high dimensional or "tricky" distribution laws are barely tackled by these techniques and should be approached with realistic expectations. Together with numerical Bayesian inference, hidden Markov models for general state-space have been recently used in environment sciences and ecology, see e.g. [5]

MCMC algorithms [6] allow us to draw samples from a probability distribution $\pi(x)\,dx$ known up to a multiplicative constant. This consists in sequentially simulating a single Markov chain whose limit distribution is $\pi(x)\,dx$. There exist many techniques to speed up the convergence toward the target distribution by improving the mixing properties of the chain.

In practice one however can make use of several chains in parallel. It is then tempting to exchange information between these chains to improve mixing properties of the MCMC samplers. A general framework of "Population Monte Carlo" has been proposed in this context [2]. In this paper we propose an interacting method between parallel chains which provides an independent sample from the target distribution. Contrary to papers previously cited, the proposal law in our work is given and does not adapt itself to the previous simulations. Hence, the problem of the choice of the proposal law still remains.

2 Parallel/interacting Metropolis within Gibbs (MwG) algorithm

Let $\pi(x)$ be the probability density function of a target distribution defined on $(I\!R^n, \mathcal{B}(I\!R^n))$. For $\ell = 1, \ldots, n$, we define the conditional laws:

$$\pi_\ell(x_\ell | x_{\neg\ell}) \stackrel{\text{def}}{=} \pi(x_{1:n}) / \int \pi(x_{1:n})\, dx_{\neg\ell}\,. \tag{1}$$

where $\neg\ell \stackrel{\text{def}}{=} \{m = 1 : n; m \neq \ell\}$. When we know to sample from (1), we are able to use the Gibbs sampler. It is possible to adapt our interacting method to parallel Gibbs sampler. But very often we do not know how to sample from (1) and therefore we consider proposal conditional densities $\pi_\ell^{\text{prop}}(x_\ell)$ defined for all ℓ. In this case, we use MwG algorithm.

One iteration $X \to Z$ of the parallel/interacting MwG method consists in updating the components X_ℓ successively for $\ell = 1, \ldots, n$, i.e. $[X_{1:n}] \to [Z_1 X_{2:n}] \to [Z_{1:2} X_{3:n}] \cdots [Z_{1:n-1} X_n] \to [Z_{1:n}]$. For each ℓ fixed, the subcomponents X_ℓ^i are updated sequentially for $i = 1 : N$ in two steps:

1. *Proposal step:* We sample independently N candidates $Y_\ell^j \in I\!R$ according to:

$$Y_\ell^j \sim \pi_{i,j}^{\ell,\text{prop}}(\xi | [\![Z, X_\ell^i, X]\!]_\ell^i)\, d\xi\,, \qquad 1 \leq j \leq n$$

where $[\![Z, \xi, X]\!]_\ell^i \stackrel{\text{def}}{=} \left[\begin{array}{c|c|c} & Z_\ell^1 & \\ & \vdots & \\ & Z_\ell^{i-1} & \\ Z_{1:\ell-1} & \xi & X_{\ell+1:n} \\ & X_\ell^{i+1} & \\ & \vdots & \\ & X_\ell^N & \end{array}\right]$.

We also use the following lighter notation: $\pi_{i,j}^{\ell,\text{prop}}(\xi|\xi') = \pi_{i,j}^{\ell,\text{prop}}(\xi | [\![Z, \xi', X]\!]_\ell^i)$.

2. *Selection step:* The subcomponent X_ℓ^i could be replaced by one of the N candidates $Y_\ell^{1:N}$ or stay unchanged according to a multinomial sampling, the resulting value is called Z_ℓ^i,

i.e.:

$$Z_\ell^i \leftarrow \begin{cases} Y_\ell^1 & \text{with probability } \frac{1}{N}\,\alpha_\ell^{i,1}(X_\ell^i, Y_\ell^1)\,, \\ \vdots & \\ Y_\ell^N & \text{with probability } \frac{1}{N}\,\alpha_\ell^{i,N}(X_\ell^i, Y_\ell^N)\,, \\ X_\ell^i & \text{with probability } \tilde{\rho}_\ell^i(X_\ell^i, Y_\ell^{1:N}) \end{cases}$$

where:

$$\alpha_\ell^{i,j}(\xi,\xi') \stackrel{\text{def}}{=} \frac{\pi_\ell(\xi'|X_{-\ell}^i)}{\pi_\ell(\xi|X_{-\ell}^i)}\,\frac{\pi_{i,j}^{\ell,\text{prop}}(\xi|\xi')}{\pi_{i,j}^{\ell,\text{prop}}(\xi'|\xi)} \wedge 1\,, \qquad \tilde{\rho}_\ell^i(X_\ell^i, Y_\ell^{1:N}) \stackrel{\text{def}}{=} 1 - \tfrac{1}{N}\textstyle\sum_{j=1}^N \alpha_\ell^{i,j}(X_\ell^i, Y_\ell^j)\,.$$

Proposition Let $P(X, dZ)$ the Markov kernel on $I\!R^{n\times N}$ associated with the MwG algorithm, the measure $\Pi(dX) = \pi(X^1)\,dX^1 \cdots \pi(X^N)\,dX^N$ is invariant for the kernel P, *i.e.* $\Pi P = \Pi$.

The proof is detailed in [1].

3 Numerical tests

Our aim is to show that making parallel samplers interact could speed up the convergence toward the stationary distribution.

An hidden Markov model. We apply the parallel/interacting MwG sampler to a toy problem where a good estimate $\hat{\pi}$ of the target distribution π is available. Consider

$$\mathsf{s}_{\ell+1} = \mathbf{a}\,\mathsf{s}_\ell + \mathsf{w}_\ell\,, \qquad\qquad \mathsf{y}_\ell = \mathbf{b}\,\mathsf{s}_\ell + \mathsf{v}_\ell$$

for $\ell = 1 \cdots n$, where $\mathsf{s}_1 \sim \mathcal{N}(\bar{\mathsf{s}}_1, Q_1)$, $\mathsf{w}_{1:n}$ and $\mathsf{v}_{1:n}$ are centered white Gaussian noises with variances σ_w^2 and σ_v^2. Suppose that $\mathbf{b}$ is known and $\mathbf{a} = \theta$ is unknown with a priori law $\mathcal{N}(\mu_\theta, \sigma_\theta^2)$. We also suppose that $\mathsf{w}_{1:n}$, $\mathsf{v}_{1:n}$, s_1 and θ are mutually independent.

The state variable is $\mathsf{x}_{1:n+1} \stackrel{\text{def}}{=} (\mathsf{s}_{1:n}, \theta)$ and the target law is $\pi(s_{1:n}, \vartheta)\,ds_{1:n}\,d\vartheta \stackrel{\text{def}}{=} \text{law}(\mathsf{s}_{1:n}, \theta|\mathsf{y}_{1:n} = y_{1:n})$. We will perform two algorithms: (*i*) N parallel/interacting MwG samplers and (*ii*) N parallel/independent MwG samplers. For methods (*i*) and (*ii*) we perform $N = 50$ parallel samplers so we compute $\epsilon^{(i)}$ and $\epsilon^{(ii)}$. $\epsilon^{(\cdot)}$ is an indicator which must decrease and remain close to a small value when convergence toward the stationary distribution occurs.

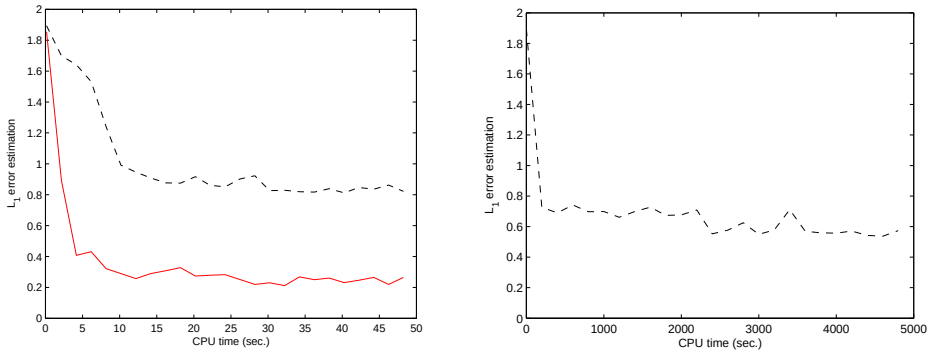

Figure 1: **Left**: Evolution of the indicators $\epsilon^{(ii)}$ for the parallel/independent MwG sampler (- -), and $\epsilon^{(i)}$ for the parallel/interacting MH sampler (–). **Right**: Evolution of the indicator $\epsilon^{(ii)}$ for the parallel/independent MwG sampler (- -). After 5000 sec. CPU time, the convergence of this method is still unsatisfactory.

$\epsilon^{(\cdot)}$ is based on the L^1 errors between $\hat{\pi}$ and the kernel density estimates of the target density based on the final values of method (*i*) or (*ii*) (see [1] for details).

To compare fairly the parallel/independent MwG algorithm and the parallel/interacted MwG algorithm, we represent on Figure 1 the indicators $\epsilon^{(i)}$ and $\epsilon^{(ii)}$ not as a function of the iteration number of the algorithm but as a function of the CPU time. In Figure 1 (left) we see that even if one iteration of algorithm (*i*) needs more CPU than one of (*ii*), still the first algorithm converges more rapidly than the second one. This shows the inefficiency of parallel/independent MwG on this simple model.

Real case study: a tropical forest. Forest dynamics models have emerged to assess the sustainability of the management of tropical forest. Among various types of models, matrix models rely on a description of the forest stand by a discrete diameter distribution. The state of a tree corresponds to its belonging to a diameter class and the shifts from one class to another occur by discrete time step. Matrix model of forest dynamics rely on four hypotheses (see [4] for details) of which Markov's hypothesis. So the parallel/interacted MwG algorithm can be used to calibrate a matrix model. An illustration on a data set obtained from experimental forest plots in a tropical rain forest of French Guiana will be presented.

4 Conclusion

This work showed that making parallel MCMC chains interact could improve their convergence properties. We presented the basic properties of the MCMC method, we did not prove that the proposed strategy speeds up the convergence. This difficult point is related to the problem of the rate of the convergence of the MCMC algorithms. Through simple examples we saw that the MwG strategy could be a poor strategy. In this situation our strategy improved the convergence properties.

REFERENCES

[1] F. Campillo and V. Rossi. *Parallel and interacting Markov chains Monte Carlo method*. Research report n.6008, INRIA, (2006), 30p. `http://hal.inria.fr/inria-00103871`.

[2] O. Cappé, A. Guillin, J.-M. Marin, and C.P. Robert.*Population Monte Carlo*. Journal of Computational and Graphical Statistics, Vol. 13(4), (2004), pp. 907–929.

[3] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors.*Markov Chain Monte Carlo in practice*. Chapman & Hall, London, 1995.

[4] N. Picard, A. Bar-Hen, Y. Guédon. *Modelling diameter class distribution with a second-order matrix model*. For. Ecol. Manage., Vol. 180, (2002), pp.389-400.

[5] E. Rivot, E. Prevost, E. Parent, and J.-L. Blaginière. *A Bayesian state-space modelling framework for fitting a salmon stage-structured population dynamic model to multiple time series of field data*. Ecological Modelling, Vol. 179, (2004), pp.463-485.

[6] C.P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer–Verlag, Berlin, 2nd edition, 2004.