Particle Markov Chain Monte Carlo

Pierre Jacob

CREST-CEREMADE
funded by AXA Research
Supervisor: Christian P. Robert

30 Mars 2010

Outline

Outline

Introduction

Particle Markov Chain Monte Carlo
is an article by Andrieu, Doucet, Holenstein,
to appear in JRSS B.

Motivation

Bayesian inference in state space models.

State Space Models

In these models:

- we observe some data $Y_{1:T} = (Y_1, \dots Y_T)$,
- we suppose they are linked with some hidden states $X_{1:T}$.

A particular case of SSM: Hidden Markov models.

Quick Reminder on Hidden Markov Models

Hidden states $(X_{1:T})$, observations $(Y_{1:T})$.
Transition probability density:

$$\forall n \in [1, T] \quad (X_{n+1}|X_{1:n} = x_{1:n}) \sim (X_{n+1}|X_n = x_n) \sim f_\theta(\cdot|x_n)$$
$$\text{and } X_1 \sim \mu_\theta(\cdot)$$

Measure probability density:

$$Y_n|(X_1, \ldots X_n = x_n, \ldots X_T) \sim g_\theta(\cdot|x_n)$$

Parameter: $\theta \in \Theta$.

Quick Reminder on Hidden Markov Models

Some interesting distributions

Bayesian inference focuses on:

$$\pi(\theta | y_{1:T})$$

Filtering focuses on:

$$\forall t \in [1, T] \quad \pi(x_{1:t} | y_{1:t}, \theta)$$

also denoted by $p_\theta(x_{1:t} | y_{1:t})$

PMCMC methods will provide a sample from:

$$\pi(\theta, x_{1:T} | y_{1:T})$$

Motivating example. . . ?

An application in finance is the stochastic volatility model:

$$y_t | x_t \sim \mathcal{N}(0, e^{x_t})$$
$$x_t = \mu + \rho(x_{t-1} - \mu) + \sigma\varepsilon_t$$

Here the parameter is $\theta = (\mu, \rho, \sigma)$.

Outline

Markov Chain Monte Carlo

A popular method to sample from a distribution π.

Algorithm 1 Metropolis-Hastings algorithm

1: Set some x_1
2: **for** $i = 2$ to P **do**
3: Propose $x^* \sim q(\cdot|x_{i-1})$
4: Compute the ratio:

$$\alpha = \min\left(1, \frac{\pi(x^*)}{\pi(x_{i-1})}\frac{q(x_{i-1}|x^*)}{q(x^*|x_{i-1})}\right)$$

5: Set $x_i = x^*$ with probability α, otherwise set $x_i = x_{i-1}$
6: **end for**

Markov Chain Monte Carlo

The output is that, after some "burn-in":

$$x_{1:P} \sim \pi$$

Note: if $q(\cdot|x_{i-1})$ does not depend on x_{i-1} it is called Independent Metropolis-Hastings.

Back to HMM...

We would like to use this algorithm to sample from $\pi(\theta|y_{1:T})$ but it is very expensive to compute this density pointwise... ARGH!

Or maybe to use this algorithm to sample from $\pi(x_{1:T}|y_{1:T}, \theta)$ but it is very hard to choose a good proposal q... ARGH!

$\Rightarrow$ Gibbs sampling is an option but it is very model-specific.

Outline

Back to HMM

Suppose we are interested in $p_\theta(x_{1:T}|y_{1:T})$, when θ is known.

Bayes formula

$$p_\theta(x_{1:T}|y_{1:T}) \propto p_\theta(y_{1:T}|x_{1:T})p_\theta(x_{1:T})$$

which leads to:

$$p_\theta(x_{1:T}|y_{1:T}) \propto \left(\prod_{n=1}^{T} g_\theta(y_n|x_n)\right)\left(\mu_\theta(x_1)\prod_{n=2}^{T} f_\theta(x_n|x_{n-1})\right)$$

and also

$$\propto p_\theta(x_{1:T-1}|y_{1:T-1})g_\theta(y_T|x_T)f_\theta(x_T|x_{T-1})$$

Particle filter

Definition

A particle filter is just a collection of weighted points, called particles.

Writing

$$(w^{(i)}, X^{(i)})_{i=1}^{P} \sim \pi$$

means that the empirical distribution of $(w^{(i)}, X^{(i)})_{i=1}^{P}$:

$$\sum_{i=1}^{P} w^{(i)} \delta_{X^{(i)}}(dx)$$

converges towards π when $P \to +\infty$.

Particle filter

Another way of seeing it...
Writing

$$(w^{(i)}, X^{(i)})_{i=1}^{P} \sim \pi$$

means that

$$\sum_{i=1}^{P} w_i \varphi(X_i) \xrightarrow[P\to\infty]{p.s.} \mathbb{E}_\pi(\varphi)$$

for every function φ such that $\mathbb{E}_\pi(\varphi)$ exists.

Next...

...let us build a particle filter $(w_n^{(i)}, X_{1:n}^{(i)})_{i=1}^{P}$ that follows $p_\theta(x_{1:n}|y_{1:n})$, using importance sampling.

Importance Sampling for particle filters

If:

$$(w_1^{(i)}, X^{(i)})_{i=1}^{P} \sim \pi_1$$

and if we write:

$$w_2^{(i)} = w_1^{(i)} * \frac{\pi_2(X^{(i)})}{\pi_1(X^{(i)})}$$

then

$$(w_2^{(i)}, X^{(i)})_{i=1}^{P} \sim \pi_2$$

under some assumptions on π_1 and π_2.

Back to HMM

Suppose

$$(w_n^{(i)}, X_{1:n}^{(i)})_{i=1}^P \sim p_\theta(x_{1:n}|y_{1:n})$$

We want

$$(w_{n+1}^{(i)}, X_{1:n+1}^{(i)})_{i=1}^P \sim p_\theta(x_{1:n+1}|y_{1:n+1})$$

Propose $X_{n+1}^{(i)} \sim q(\cdot|X_{1:n}^{(i)}, y_{n+1})$.
Set $X_{1:n+1}^{(i)} := (X_{1:n}^{(i)}, X_{n+1}^{(i)})$.
Then:

$$(w_n^{(i)}, X_{1:n+1}^{(i)})_{i=1}^P \sim p_\theta(x_{1:n}|y_{1:n})q(x_{n+1}|x_{1:n}^{(i)}, y_{n+1})$$

Using Importance Sampling

Importance sampling to correct the weights:

$$w_{n+1}^{(i)} = w_n^{(i)} \times \frac{f_\theta(X_{n+1}^{(i)}|X_n^{(i)})g_\theta(y_{n+1}|X_{n+1}^{(i)})}{q(x_{n+1}|X_{1:n}^{(i)}, y_{n+1})}$$

Then we get what we wanted:

$$(w_{n+1}^{(i)}, X_{1:n+1}^{(i)})_{i=1}^P \sim p_\theta(x_{1:n+1}|y_{1:n+1})$$

But iterating leads to weight degeneracy.

Sequential Monte Carlo

Algorithm 2 Sequential Monte Carlo algorithm

1: Propose $X_1^i \sim q_1$
2: Compute weights w_1^i
3: **for** $n = 2$ to T **do**
4: Resampling: $(w_n^i, X_n^i) \leftarrow (1/P, X_n^j)$ for a certain $j \in [1, P]$
5: Propose $X_n^i \sim q_n$, let $X_{1:n}^i = (X_{1:n-1}^i, X_n^i)$
6: Update weights w_n^i
7: **end for**

Sequential Monte Carlo

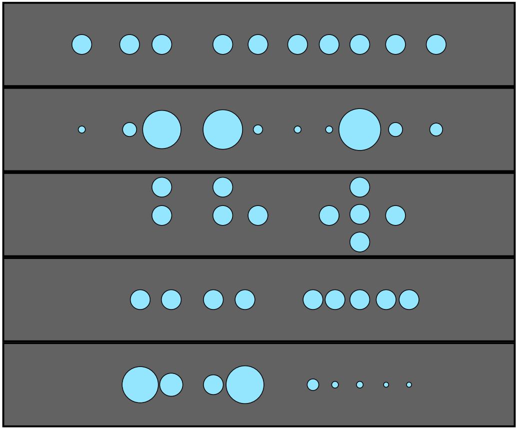

Figure: Weighting and resampling

SMC

In the end we get particles:

$$(w_T^i, X_{1:T}^i)_{i=1}^P$$

such that:

$$\hat{p}_\theta(dx_{1:T}|y_{1:T}) := \sum_{i=1}^{P} w_T^i \delta_{X_{1:T}^i}(dx_{1:T})$$

is an approximation of $p_\theta(dx_{1:T}|y_{1:T})$.

SMC: a side effect

A side effect of the SMC algorithm is that we can get the marginal likelihood $p_\theta(y_{1:T})$ "for free" with the following estimate:

$$\prod_{n=1}^{T}\left(\frac{1}{P}\sum_{i=1}^{P} w_n^i\right) \xrightarrow[P\to\infty]{\mathbb{P}} p_\theta(y_{1:T})$$

We will denote this estimate by $\hat{Z}$.

Outline

Introducing PMCMC

In short...

The main idea is to use Sequential Monte Carlo methods to propose some $X_{1:T}$ in a Metropolis-Hastings algorithm targeting the posterior distribution of $X_{1:T}$.

Note: there is also a Gibbs sampling version of PMCMC.

PIMH

First we look at the Particle Independent Metropolis-Hastings algorithm. The parameter θ is known.

Goal

Filtering: sample from the posterior distribution of hidden states.

$$X_{1:T} \sim p_\theta(dx_{1:T} | y_{1:T})$$

PIMH algorithm

Algorithm 3 PIMH algorithm

1: **for** $i \geq 1$ **do**
2: Run a SMC algorithm, sample $X^*_{1:T} \sim \hat{p}_\theta(dx_{1:T}|y_{1:T})$
3: Compute $\hat{Z}^*$
4: With probability
$$\alpha = \min\left(1, \frac{\hat{Z}^*}{\hat{Z}^i}\right)$$
5: Set $X^i_{1:T} = X^*_{1:T}$, $\hat{Z}^i = \hat{Z}^*$
6: Otherwise set $X^i_{1:T} = X^{i-1}_{1:T}$, $\hat{Z}^i = \hat{Z}^{i-1}$
7: **end for**

PIMH explanations

The tricky part here is to prove that the acceptance ratio:

$$\alpha = \min\left(1, \frac{\hat{Z}^*}{\hat{Z}^i}\right)$$

indeed leads to the target distribution $p_\theta(dx_{1:T}|y_{1:T})$.

Sequential Monte Carlo. . . again!

Algorithm 4 Sequential Monte Carlo algorithm

1: Propose $X_1^i \sim q_1$
2: Compute weights w_1^i
3: **for** $n = 2$ to T **do**
4: Resampling: $(w_n^i, X_n^i) \leftarrow (1/P, X_n^j)$ for a certain $j \in [1, P]$
5: Propose $X_n^i \sim q_n$, let $X_{1:n}^i = (X_{1:n-1}^i, X_n^i)$
6: Update weights w_n^i
7: **end for**

PIMH explanations

During the SMC algorithm various random variables are used:

- during the resampling we sample $T-1$ variables:

$$A_1, \dots A_{T-1} \in (1, \dots P)^{P(T-1)}$$

(these can be multinomial variables for instance)

- during the "propagation" step we sample from q_n:

$$\bar{X}_1, \dots \bar{X}_T \in \mathcal{X}^{PT}$$

What if we consider the joint distribution of all those random variables?

PIMH explanations

This distribution is denoted by ψ and has an explicit form which depends on $\left(q_n, \bar{X}_n, w_n, A_n\right)_{n=1}^{T}$. Thus if we sample from

$$\sum_{i=1}^{P} w_i \delta_{X^i_{1:T}}(dx_{1:T})$$

at the end of the SMC by choosing an index k in $[1, P]$ with probability w_T^k, the distribution of the joint variables is

$$q^P(k, \bar{x}_1, \ldots \bar{x}_T, a_1, \ldots a_{T-1}) = w_T^k \psi(\bar{x}_1, \ldots \bar{x}_T, a_1, \ldots a_{T-1})$$

PIMH explanations

Now that we have the proposal density of the PIMH algorithm, we can prove that the target density is

$$\tilde{\pi}^P(k, \bar{x}_1, \ldots \bar{x}_T, a_1, \ldots a_{T-1})$$

from which the distribution of interest $p(dx_{1:T}|y_{1:T})$ is a marginal. To prove this we establish that:

$$\frac{\hat{Z}}{Z} = \frac{\tilde{\pi}^P(k, \bar{x}_1, \ldots \bar{x}_T, a_1, \ldots a_{T-1})}{q^P(k, \bar{x}_1, \ldots \bar{x}_T, a_1, \ldots a_{T-1})}$$

PMMH

Next we look at the Particle Marginal Metropolis-Hastings algorithm. The parameter θ is unknown.

Goal

Sample:

$$(\theta, X_{1:T}) \sim p(d\theta, dx_{1:T} | y_{1:T})$$

Let us put a prior on θ, denoted by π_0.

PMMH algorithm

Algorithm 5 PMMH algorithm

for $i \geq 1$ **do**
  Propose $\theta^* \sim q(\cdot|\theta(i-1))$
  Run a SMC algorithm given θ^*, sample $X^*_{1:T} \sim \hat{p}_{\theta^*}(\cdot|y_{1:T})$,
  Compute $\hat{Z}^*$
  With probability

$$\alpha = \min\left(1, \frac{\hat{Z}^* \pi_0(\theta^*) q(\theta(i-1)|\theta^*)}{\hat{Z}^i \pi_0(\theta(i)) q(\theta^*|\theta(i-1))}\right)$$

  Set $\theta(i) = \theta^*$, $X^i_{1:T} = X^*_{1:T}$
  Otherwise set $\theta(i) = \theta(i-1)$, $X^i_{1:T} = X^{i-1}_{1:T}$
end for

Comment on PMMH

This is nearly the same algorithm as the PIMH, but this one is much more interesting in practice, for we get a sample:

$$(\theta^i, X^i_{1:T})_{i=1}^N \sim p(d\theta, dx_{1:T}|y_{1:T})$$

The proof is nearly the same in the PIMH case.
In fact it is a standard IMH algorithm, therefore under some conditions we have a convergence theorem:

$$||\mathcal{L}^N((\theta^i, X^i_{1:T}) \in \cdot) - p(\cdot, \cdot|y_{1:T})|| \xrightarrow[i\to\infty]{} 0$$

where $\mathcal{L}^N$ is the marginal distribution of $(\theta^i, X^i_{1:T})$.

Comment on PMMH

- This holds for any number of particles P,
- moreover if $P \longrightarrow \infty$, the acceptance ratio converges towards 1 because $\hat{Z}$ is a consistent estimate of Z,
- in practice the computational time is linear in P and N, but the SMC part is easy to parallelize.

Applications

An application in finance is the stochastic volatility model:

$$y_t | x_t \sim \mathcal{N}(0, e^{x_t})$$
$$x_t = \mu + \rho(x_{t-1} - \mu) + \sigma\varepsilon_t$$

Here the parameter is $\theta = (\mu, \rho, \sigma)$. We put some prior distribution on the parameters.
Let us try PMMH on 500 observations with

- 100 particles,
- 10.000 iterations.

Results on simulated data

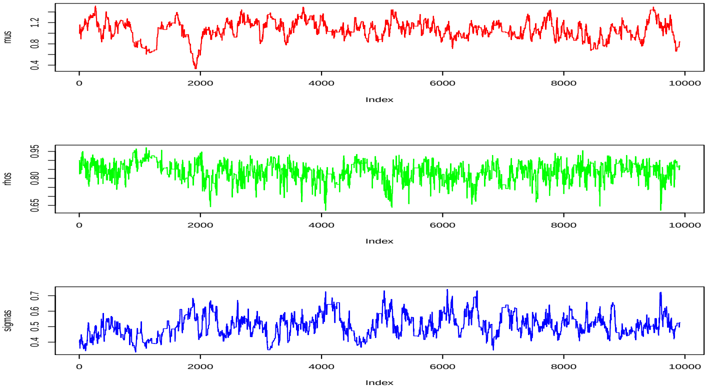

Figure: Posterior parameters, real values: $(1, 0.9, 0.5)$

Results on simulated data

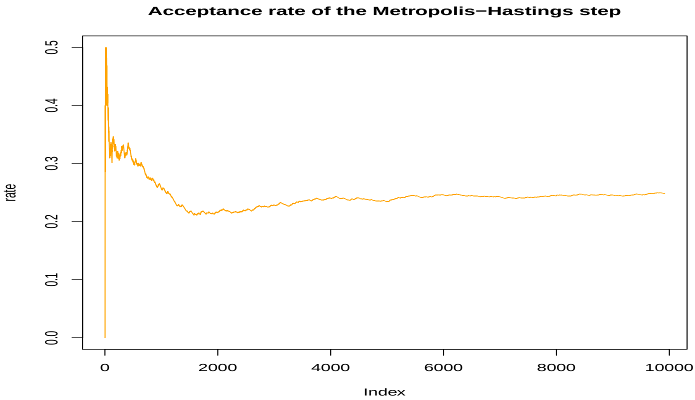

Figure: Acceptance rate

Results on simulated data

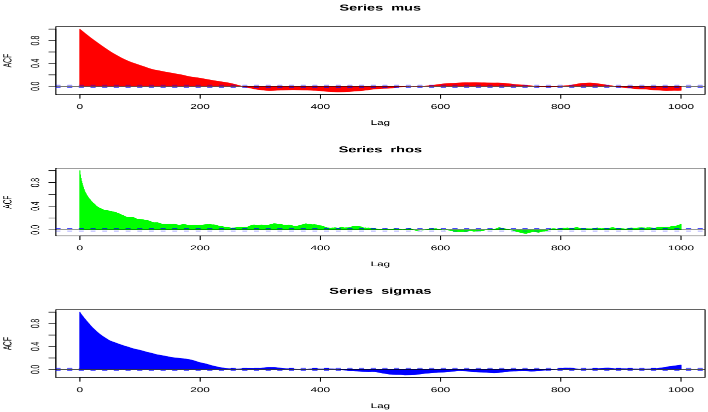

Figure: Autocorrelation function

Conclusion

- It is a algorithm to target posterior distributions in any SSM.
- It clearly combines MCMC and SMC in a novel way.
- It can be more efficient than Gibbs sampling for some models.

Questions please!