The Multiple-Try Metropolis and its Variations

Xu Chen
Department of Statistical Science
xu.chen2@duke.edu

Menglan Jiang
Department of Statistical Science
menglan.jiang@duke.edu

Abstract

Markov Chain Monte Carlo (MCMC) has been extensively applied in many complicated computational problems to sample from an arbitrary distribution. The fundamental idea is to generate a Markov chain whose invariant distribution is the target distribution. The traditional Metropolis-Hastings algorithm (MH) based on local search may suffer from slow converging problem since the sampler may get stuck in a local mode especially for multimodal parameter spaces. Multiple-try Metropolis (MTM) was proposed to overcome this difficulty by proposing multiple trial points and then sampling based on their importance. The numerical experiments illustrate that the sampler can efficiently explore the parameter space. This project will prove the validity of MTM and implement the algorithm and its variations including Griddy-Gibbs Multiple-Try Metropolis (MTM-Gibbs) and Langevin-within-MTM on artificial data and real dataset. Comparisons are made to show the superiority of the algorithm over traditional MH algorithms.

1 Background

Markov Chain Monte Carlo (MCMC) has been extensively applied in many complicated computational problems in the bayesain paradigm. The conceptual understanding about this algorithm is to sample randomly from a defined distribution and to obtain an invariant distribution which is also the target distribution.

The traditional Metropolis-Hastings algorithm (MH) is one of the representatices of MCMC. The algorithm is defined as follows,

$$r = \min\left\{1, \frac{\pi(x_{t+1})T(x_{t+1}, x_t)}{\pi(x_t)T(x_t, x_{t+1})}\right\}$$

where $\pi(x_i)$ is the probability distribution of x_i; $T(.)$ is a markov transition function to generate a proposal x_{t+1} given the current status of x_t; r is the probability of accepting x_{t+1}, also known as MH ratio.

Although compared to other bayesian methods, the standard MH has already been more computational approachable, it is still of probablistic. Its random proposed local move may arouse low convergence concern. In addition, trapping in local mode is also a major imperfection of this algorithm. A possible solution to enlarge the area for the proposal while it often comes along a small MH ratio. Hence, it is an inefficient algorithm. Liu, Liang and Wong (2000) proposed a series of variations of the standard MH algorithm and proved their effectiveness.

The improved algorithm based on the traditional MH is called Multiple-try-Metropolis (MTM). This method differs from the classical MH by proposing a mutiple candidate points rather than a simple point for each iteration. As a result, the algorithm is more robust for effectively searching the optima.

This paper intends to research the MTM class to a further extent. The MTM and a variation combined with gibbs sampler will be tested first. These two algorithms are theretically discussed in

previous literature. We will testify these two methods based on both artificial data or real dataset. An original variation which incorporates hamiltonian dynamics is proposed in the end. We hope that by doing empirical study on the algorithms proposed, we can enhance the power of MTM algorithm

2 The algorithm and its variations

In this section, we will introduce the multiple-try metropolis as well as its variations. In the traditional MH algorithm, we have the following settings. Firstly, a proposal function $T(x_i, x_{i+1})$ is defined, which clarifies the relationship between the tth trial and the $(t+1)$th candidate parameters. Next, a function used for evaluating the likelihood of a trial is defined as follows,

$$w(x_i, x_{i+1}) = \pi(x_i) T(x_i, x_{i+1}) \lambda(x_i, x_{i+1})$$

where $\lambda(x_i, x_{i+1})$ is an adjustment function to enhance the power of the algorithm, which is non-negative and symmetric.

Two basic requirements should be satisfied.
1. $T(x_i, x_{i+1}) > 0$ if and only if $T(x_{i+1}, x_i) > 0$
2. If $T(x_i, x_{i+1}) > 0$, then $\lambda(x_i, x_{i+1}) > 0$.
The choices of $\lambda(x_i, x_{i+1})$ vary in different situations.

2.1 Multiple-try Metropolis

2.1.1 Algorithm

The MTM algorithm can be achieved through 4 steps. The symbols are in line with the standard MH algorithm.
1. Sample k iid trials $x_{t+1}^1, ..., x_{t+1}^k$ from $T(x_i, .)$. Compute $w(x_{t+1}^j, x_t)$ for $j = 1, ..., k$.
2. Select $\mathbf{X} = x_{t+1}$ among the proposal set $\{x_{t+1}^1, ..., x_{t+1}^k\}$ with probability propotional to $\{w(x_{t+1}^1, x_i), ..., w(x_{t+1}^k, x_i)\}$.
3. Sample $x_\star^1, ..., x_\star^{k-1}$ from the distribution $T(x_{t+1}, .)$, and let $x_\star^k = x_t$.
4. Accept x_{t+1} with probability

$$r = \min\left\{1, \frac{w(x_{t+1}^1, x_t) + \cdots + w(x_{t+1}^k, x_t)}{w(x_\star^1, x_{t+1}) + \cdots + w(x_\star^k, x_{t+1})}\right\}$$

and reject it with probability $1 - r$. The quantity r is called generlaized MH ratio.

2.1.2 Model Specification: Mixture Gaussian distribution

The follwoing settings are used in our testification of the algorithm.

Initial state: $x_0^i = \text{Uni}(-5, 5) \quad i = 1, 2$

The proposal function $T(.)$: $\begin{pmatrix} x_{t+1}^1 \\ x_{t+1}^2 \end{pmatrix} = \begin{pmatrix} x_t^1 \\ x_t^2 \end{pmatrix} + \mathbf{N}_2(\mathbf{0}, \sigma^2 \mathbf{I}_2)$

The probability density function:

$$\pi(x) = \frac{1}{3}\mathbf{N}_2(\mathbf{0}, \mathbf{I}_2) + \frac{1}{3}\mathbf{N}_2\left\{\begin{pmatrix} -6 \\ -6 \end{pmatrix}, \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}\right\} + \frac{1}{3}\mathbf{N}_2\left\{\begin{pmatrix} 4 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix}\right\}$$

The adjustment function $\lambda(x_i, x_{i+1}) = 1$

2.2 Griddy Gibbs Multiple-Try Metropolis

2.2.1 Algorithm

One of the major characteristic of the Gibbs sampler is that it is constructed based on the conditional distribution. However, in many cases, it is intracable to derive the conditional distribution for sampling. The griddy Gibbs sampler (Ritter and tanner (1992)) solves the computational challenge.

Combined with MTM method, the MTM-Gibbs algorithm is designed to enhance the performance of the MTM.

Assume $\mathbf{X} = \mathbf{x}$ where $\mathbf{x} = (x(1), ..., x(d))$, the algorithm of MTM-Gibbs is illustrated as follows,

1. Select any element of $\mathbf{x}$, say xi. Sample $y_1, ..., y_k$ iid from a transition function $T(x(i), .)$ in line with the direction of $x(i)$, and calculate

$$w(y_j, x(i)) = \pi(x(1), ..., x(i-1), y_j, x(i+1), ..., x(d))T(y_j, x(i))\lambda(y_j, x(i)),$$

for j = 1,2,...,k.

2. Select $y = y_j$ with probability proportional to $w(y_j, x(i))$. Draw $k-1$ iid samples from T(y,.), denote by $s_1, ..., s_{k-1}$. Make $s_k = x(i)$.

3. Compute the generalized Metropolis ratio r defined above, accept y with probability r and reject with $1 - r$.

2.2.2   Model Specification: Non-linear model

The model used in testing the MTM-Gibbs is defined as follows,

$$y_i = \theta_1(1 - \exp(-\theta_2 x_i)) + \varepsilon_i$$

where y is the BOD at time x with independent normal errors with constant variance σ^2.

The following settings are used in our testification of the algorithm.

Initial state: $\theta_i = \text{Uni}(-1, 1) \quad i = 1, 2$

The proposal function $T(.)$: $\theta_{t+1}^i = \theta_{t+1}^i + \mathbf{N}(0, \sigma^2) \quad i = 1, 2$

The adjustment function $\lambda(x_i, x_{i+1}) = 1$

2.3   Langevin-within-MTM

2.3.1   Algorithm

In this method, we combined the hamiltonian dynamics with MTM, in which langevin is one of the representatives of halmiltonian dynamics.

2.3.2   Model Specification:

3   Implementation

4   Optimization and high performance computing

5   Experimental results and comparisons

6   Conclusions

References

[1] Faming Liang, Chuanhai Liu & Raymond Carroll (2011) *Advanced Markov chain Monte Carlo methods: learning from past samples* vol:714 John Wiley & Sons

[2] Jun S. Liu (2001) *Monte Carlo Strategies in Scientific Computing* Statistics, Springer-Verlag, New York

[3] Jun S. Liu, Faming Liang & Wing Hung Wong (2001) The Multiple-try method and local optimization in Metropolis Sampling. *Journal of the American Statistical Association*, 95:449, pp. 121-134

[4] Radford M. Neal (2011) MCMC using Hamiltonian dynamics. In Steve Brooks et al. (eds.) *Handbook of Markov chain Monte Carlo* Chapter 5, Chapman & Hall/CRC Press