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 Introduction

2 The algorithm and its variations

In this section, we will introduce the multiple-try metropolis as well as its variations. In the traditional M-H 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 $\pi(x_i)$ is the probability distribution of x_i, $\lambda(x_i, x_{i+1})$ is an adjusting function to enhance the power of the algorithm, which is nonnegative 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

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

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 M-H ratio.

2.2 Griddy Gibbs Multiple-Try Metropolis

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 $1 - r$.

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