Report
on
‘Laplace Expansions in Markov Chain Monte Carlo Algorithm’

ZiXia Huang

2024-03-04

1. Summary of the contributions

Complex hierarchical model:

$$f(X|\theta_S, S)\pi(\theta_S|S)p(S|\lambda)h(\lambda)$$

where S is the parameter of interests, θ_S is the nuisance parameters.

The researcher propose to decrease the number of parameters to simulate at each iteration to improve the convergence of the most commonly used Gibbs sampler by using a Laplace approximation on the nuisance parameters.

1. Summary of the contributions

- *Traditional Gibbs algorithm*

1. $\lambda^t \sim h(\lambda | S^{t-1})$,

2. $S^t \sim p(S | X, \lambda^t, \theta_S^{t-1})$,

3. $\theta_S^t \sim \pi(\theta_S | X, S^t, \lambda^t)$.

- *Proposed Laplace algorithm*

1. $\lambda^t \sim h(\lambda | S^{t-1})$,

2. $S^t \sim \widehat{p}(S | X,\ \lambda^t) \propto \widehat{g}(X | S) p(S |\ \lambda^t)$,

where $\widehat{g}(X | S)$ is the Laplace approximation of $g(X | S) = \int f(X | \theta_S, S) \pi(\theta_S | S)\, d\theta_S$.

It turns out that this proposed Laplace algorithm converges to the stationary distribution more rapidly and has a much more shorter computational time.

The paper also provides a theoretical study: This Laplace approximation gets close to the true posterior distribution as the numbers of observation N goes to infinity.And in the simulation study later, it shows this algorithm has good performance even with just a reasonable number of observations.

2. Simplified model

In the paper, researches have considered the complete model of the general case for instance latent variable models in finite state spaces with discrete or continuous latent variables or curve estimation.

Here, in this report we will simplified the model by assuming we know all the prior distribution of the parameters, and the nuisance parameter θS has only one dimension.

2. Simplified model

Denote $\hat{\theta}_S$ as the properly chosen local maximizer of the joint negative log-likelihood function, with J equals to the second order derivative of the joint-likelihood function evaluated at $\hat{\theta}_S$,then the Laplace approximation is:

$$\hat{g}(X|S) = 2\pi^{1/2} J^{-1/2} \pi(\hat{\theta}_S) \prod_{i=1}^{N} f(X_i|\hat{\theta}_S)$$

Suppose we know that is a normal distribution with $N(\hat{\theta}_S, S)$; $\pi(\theta_S|S)$ is a exponential distribution $exp(S)$; $p(S|\lambda = \{\alpha, \beta\})$ is a gamma distribution and finally $\alpha \sim unif(0, S)$, $\beta \sim unif(0, S)$.

2. Simplified model

We will use Newton's method to find a local maximizer $\widehat{\theta}_S$ for the negative log-likelihood. Given an initial estimate of the parameter $\theta_S^{(0)}$, the algorithm proceeds by iterative updating:

$$\theta_S^{(t+1)} \leftarrow \theta_S^{(t)} - [h^{(2)}(\theta_S^{(t)})]^{-1} h^{(1)}(\theta_S^{(t)})$$

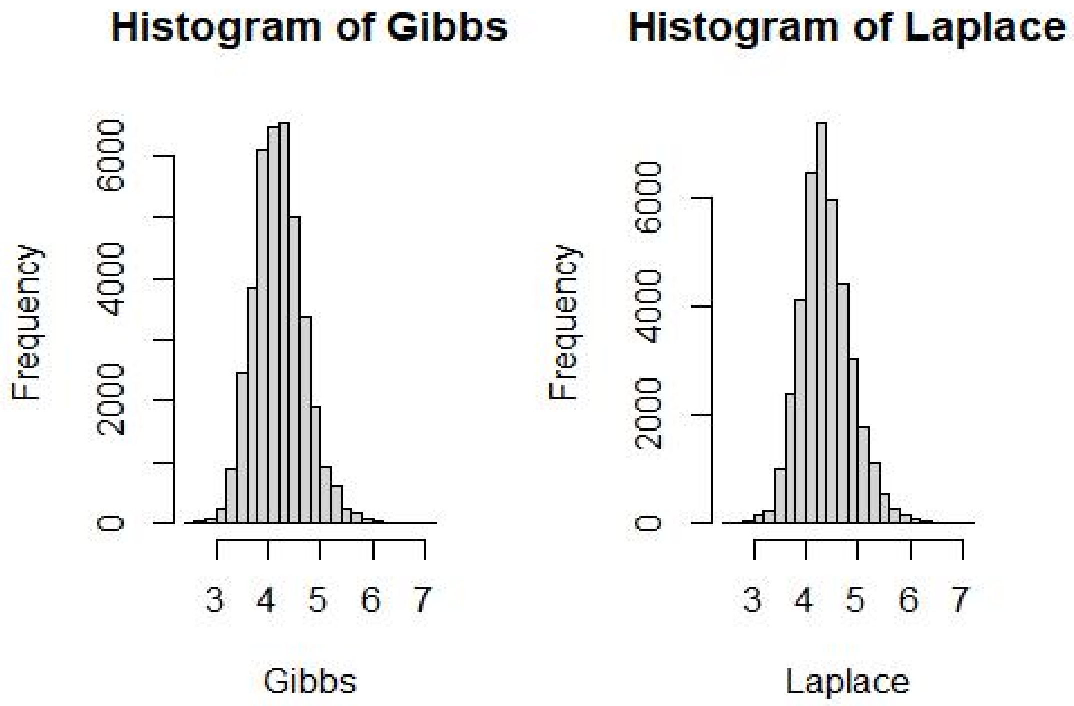
Here we use 20 iterative updates to find a local maximum.

3. Simulation study results

After solving for the posterior distribution in each literative updates, the complexity of the final form is significantly reduced using Laplace approximation. However we still need to use Metropolis hastings algorithm with log-normal and normal proposal to sample from posterior.

3. Simulation study results

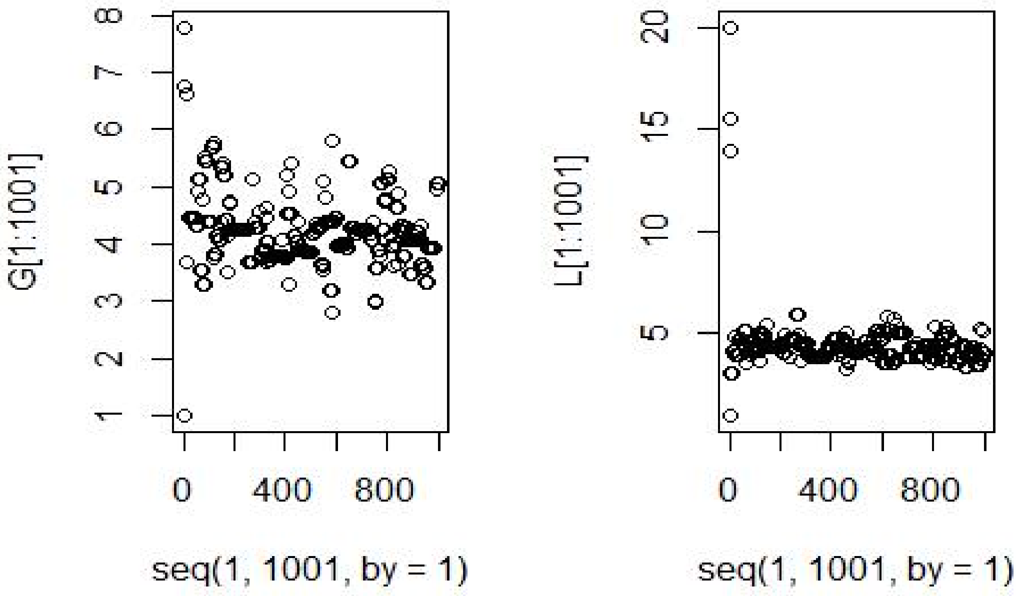
- Histogram: n=100, true value S=4, total iterative update: 40000, burn in period: 1000



Histograms from using Classical Gibbs algorithm and simplified Laplace algorithm suggest that with just a reasonable number of observations, the posterior distribution is very well approximated by the simplified Laplace algorithm. But the 95% credible interval is wider using the Laplace algorithm.

3. Simulation study results

- Convergency rate



Scatter plots of the first 1000 iterations of the Classical Gibbs algorithm and the simplified Laplace algorithm.

Those plots suggest that the chain indeed converges more rapidly to the stationary distribution using Laplace algorithm.

4. Thoughts

- **Good method**

This algorithm could be used in many applied studies where the large computation time is a real problem, as an improvement of the classical Gibbs algorithm in MCMC method.

- **Complex poterior approximation**

Even with smaller iterative cycles, Metropolis-Hastings algorithm is still involved.

- **Simplified model vs Full model**

Nuisance parameters in latent variable models in finite state spaces with discrete or continuous latent variables or curve estimation.

- **Interesting results**

Faster convergency + shorter computational time + reasonable observation numbers