Stats 230 Final Project Report

David Mwakima and Jizhi Zhang

03/10/2023

1 Introduction

What is MCMC for Bayesian inference?

[Add details here]

The Metropolis-Hastings algorithm developed by Metropolis et al. (1953) is used for MCMC Gelfand and Smith (1990).

The use of MCMC for Large datasets presents a new research frontier.

[Add details here]

In our report we consider a paper by Maire, Friel, and Alquier (2019) that addresses the problem of using MCMC for large datasets. This paper proposes a new methodology, which the authors call *Informed Sub-Sampling MCMC* (ISS-MCMC), for doing Bayesian MCMC approximation of the posterior distribution. This is a scalable version of the Metropolis-Hastings algorithm designed for situations when N is so big that to approximate the posterior distribution takes a very long time.

2 Main ideas of how it works

ISS-MCMC is "informed" because it makes use of a measure of similarity with respect to the full dataset through summary statistics. It is "sub-sampling" because it uses this measure to select a subset of the dataset that will be used by the Markov transition kernel at the k-th iteration of the algorithm. In this way, the Markov chain transition kernel uses only a fraction n/N of the entire dataset. They show using examples that choosing $n \ll N$ can lead to significant reductions in computational run-times while still retaining the simplicity of the standard Metropolis-Hastings algorithm. In the following subsections we consider in more detail the main ideas of how this algorithm works. See section 4.

2.1 Similarity through summary statistics

See Fearnhead and Prangle (2012)

2.2 Transition kernel

2.3 The Algorithm

3 Comparison with other approaches

Other similar approaches to solve the same statistical problems are Quiroz et al. (2018) and the *Confidence Sampler* in Bardenet, Doucet, and Holmes (2017). Both of these approaches use sophisticated "control variates" to get positive unbiased estimators (based on a subset of data) for the likelihoods in the Metropolis-Hastings acceptance ratio. The authors note that these control variates are computationally intensive.

The noisy approaches due to Korattikara, Chen, and Welling (2014) and Alquier et al. (2016)

Maclaurin and Adams (2015) also addresses the computational issue are independent.

Another approach which the authors compare their approach with is an approach based on continuous time Markov processes (Langevin diffusion, Zig-Zag process) in Fearnhead et al. (2018) and Bierkens, Fearnhead, and Roberts (2019). Here, the authors note that the computational hurdle involves calculation of the gradient of the log-likelihood, which may not always be unbiased. Moreover, these approaches depart significantly from the simplicity of the original discrete M-H algorithm.

4 Example with logistic regression

Here we reproduce one example in their paper for the case of logistic regression. See section 6.3

4.1 Compare convergence of ISS-MCMC with M-H.

References

Alquier, Pierre, Nial Friel, Richard Everitt, and Aidan Boland. 2016. "Noisy Monte Carlo: Convergence of Markov Chains with Approximate Transition Kernels." *Statistics and Computing* 26 (1-2): 29–47.

Bardenet, Rémi, Arnaud Doucet, and Christopher C Holmes. 2017. "On Markov Chain Monte Carlo Methods for Tall Data." *Journal of Machine Learning Research* 18 (47).

Bierkens, Joris, Paul Fearnhead, and Gareth Roberts. 2019. "The Zig-Zag Process and Super-Efficient Sampling for Bayesian Analysis of Big Data." *The Annals of Statistics* 47 (3): 1288–1320.

Fearnhead, Paul, Joris Bierkens, Murray Pollock, and Gareth O Roberts. 2018. "Piecewise Deterministic Markov Processes for Continuous-Time Monte Carlo." *Statistical Science* 33 (3): 386–412.

Fearnhead, Paul, and Dennis Prangle. 2012. "Constructing Summary Statistics for Approximate Bayesian Computation: Semi-Automatic Approximate Bayesian Computation." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74 (3): 419–74.

Gelfand, Alan E, and Adrian FM Smith. 1990. "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association* 85 (410): 398–409.

Korattikara, Anoop, Yutian Chen, and Max Welling. 2014. "Austerity in MCMC Land: Cutting the Metropolis-Hastings Budget." In *International Conference on Machine Learning*, 181–89.

Maclaurin, Dougal, and Ryan P. Adams. 2015. "Firefly Monte Carlo: Exact MCMC with Subsets of Data." In *Proceedings of the 24th International Conference on Artificial Intelligence*, 4289–95. AAAI Press.

Maire, Florian, Nial Friel, and Pierre Alquier. 2019. "Informed Sub-Sampling MCMC: Approximate Bayesian Inference for Large Datasets." *Statistics and Computing* 29: 449–82.

Metropolis, Nicholas, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. 1953. "Equation of State Calculations by Fast Computing Machines." *The Journal of Chemical Physics* 21 (6): 1087–92.

Quiroz, Matias, Robert Kohn, Mattias Villani, and Minh-Ngoc Tran. 2018. "Speeding up MCMC by Efficient Data Subsampling." *Journal of the American Statistical Association.*