A new Approximate Bayesian Inference algorithm for Bayesian Lasso: A local approximation correction approach

Yuhao Li

Supervisor: A/Prof. John Ormerod

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science(Honour)(Data Science)

Mathematics and Statistics

THE UNIVERSITY OF SYDNEY

June 2023

Statement of originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Yuhao Li

Abstract

Bayesian Paradigm

Variational Approximation: as a deterministic approximation algorithm for Approximate Bayesian Inference(ABI) of intractable posterior distribution, has been applied prevalently for fast Approximate Bayesian Inference(ABI) among the Bayesian Statistical community, while it is also a faster alternative to Monte Carlo methods such as Markov Chain and Monte Carlo(MCMC). The main idea behind Variational Approximation is: given an assumed distribution set, it will search for an optimal posterior distribution by continuing minimizing the gap between true posterior and estimated posterior such as using Kullback–Leibler divergence(KL-divergence) as a distance metric.

Nevertheless, elegant property in MCMC such as obatining exact posterior if infinite burn-in time period is assigned, doesn't occur in Variational Inference, which means the approximation accuracy will be a pivotal concern as unsatisfied distribution such as underestimating the variance when the correlation of variables becomes large.

In this thesis, we will firstly introduce lasso distribution, which is an invented distribution that could be matched for facilitating local parameter estimate, followed by the introduction of two fast and more accurate Variational Approximation algorithms and their application in the Bayesian Lasso regression problem. By assuming the global parameter assuming Gaussian Approximation, the information of local parameter distribution would be accommodated by the univariate or multivariate lasso distribution so that global distribution would be obtained by product of local distribution and a conditional Gaussian distribution.

The first method involves matching with marginal univariate lasso distribution by updating global parameter for each variable per iteration. Additionally, we propose another algorithm for matching a local bivariate lasso distribution for updating global parameter for each pair of variables per iteration, successfully addressing the issue when initial diagonal covariance matrix is assigned.

To verify the efficiency and accuracy of our algorithm, numerous experiments would be

conducted under real-world datasets such as Hitter Dataset using several evaluation metric such as l_1 accuracy and matrix norm. Our result suggest their high Variational Approximation accuracy with a descent time efficiency, compared with the traditional Monte Carlo methods and Mean-Field Variational Bayes(MFVB).

Acknowledgements

Thanks to Supervisor and family

Finally, I would like to thank my friends my family: my mother, father, brother, grandparents, and cousins. The last couple years has been full of adversity and I could not have overcome it without your support behind the scenes.

Contents

Chapter 1

Introduction

1.1 Motivation

Bayesian and Frequentist Frequentist Bayesian

Bayesian Inferenece

Variational Inference

Stochastic Sampling based Approach Determinstic Approach

Variational Approximation: an optimimization based technique for approximate bayesian inference, obtain interval estimate and error variance. A class of techniques which try to approximate the itnractable posterior distribution with a tractable distribuition. Generally the paraemters of tracatble approximation are chosen to minimize some measure of its distance from the true posterior(KL-divergence)

Bayesian Lasso Problem

1.2 Contribution

Our main contribution could be concluded as the following part:

- Design of a new posterior parameter correction approach based on the posterior estimate of Mean-Field Variation Bayes parameter

1.3 Thesis Organization

This paper will be divided up into 6 chapters. Chapter 1 will briefly illustrate the motivation and application of variational approximation. Section 2 will briefly introduce basic definition and methodlogy in previous work such as MCMC(Monte Carlo Method) and Mean-Field Variational Bayes(MFVB). We will present our main methodlogy of variational correction algorithm in Chapter 3, followed by a comprehensive experiment for testing the effectiveness of algorithm in Chapter 4. In Chapter 5 and Chapter 6, we will briefly discuss and explain our result and potential improvement in the future.

Chapter 2

Definition and Literature Review

2.1 Least Absolute Shrinkage and Selection Operator(LASSO) penalized regression

2.1.1 Lasso penalty formulation

2.1.2 Bayesian Lasso regression

Univariate Lasso Distribution

Multivariate Lasso Distribution

2.2 Bayesian Paradigm

2.3 Variational Inference

2.3.1 Mean Field Variational Bayes

2.4 Expectation Maximization

2.4.1 Bayesian Expectation Maximization

2.5 Markov Chain Monte Carlo(MCMC)

2.5.1 Metropolis–Hastings (MH) Algorithm

2.5.2 Gibbs Sampler

Chapter 3

Methodlogy

3.1 Lasso distribution

3.1.1 Univariate Lasso Distribution

Basic Property

Derivation

3.1.2 Multivariate Lasso Distribution

Basic Property

Derivation

3.2 Local-Global Algorithm

Chapter 4

Experiment Result and Analysis

4.1 Experimental Setting

4.1.1 Parameter selection

4.1.2 Evaluation metric

L1 accuracy

(MORE) Matrix norm Posterior cov and estimated Cov

Objective: Use math to find posterior mode of lasso distribution: Given a,b,c Task: Check if posterior estimate reach Metric: Use Posterior TP/FP Rate(Soft thresholding operator) Check if a local parameter mode is close to 0, compared with true parameteer and use this for Variable Selection

Expectation: posterior mode sparse, posterior mean not sparse

4.1.3 Experimental datasets

toy dataset 3-4 datasets

4.2 Experimental Result

Chapter 5

Discussion and Conclusion

Bibliography