Inférences bayésiennes en absence de vraisemblance calculable (Bayesian Inference for intractable likelihoods)

Anthony Pettitt

Bayesian Inference for models with intractable likelihoods

Tony Pettitt

School of Mathematical Sciences,
Queensland University of Technology, GPO Box 2434,
Brisbane, Australia
a.pettitt@qut.edu.au

Keywords: intractable likelihood; normalizing constant; random graph model; Ising model; imaging

Abstract

Cet exposé fait la revue des développements récents des techniques déterministes et de Monte Carlo permettant l'obtention d'inférences bayésiennes de paramètres inconnus lorsque la vraisemblance a une fonction normalisatrice inconnue ou lorsqu' elle est incalculable par limitation des ordinateurs. C'est le cas des données en treillis spatial dont la vraisemblance implique des modèles d'Ising, Potts ou autologistiques. En effet, la constante de normalisation peut ne pas être traitable si le treillis est trop large. Dans de tels cas, Møller Pettitt, Reeves and Berthelsen (2006) ont introduit un algorithme utilisant la méthode de Monte Carlo qui permet d'éviter le calcul de la constante de normalisation à la condition que le modèle simule parfaitement une variable auxiliaire. De ce point de départ, des approximations et améliorations de cet algorithme et algorithmes apparentés vont être décrites, ainsi que des applications, dont (1) une approche varationnelle de Bayes d'analyse d'image et (2) des distributions de graphes aléatoires (ex. réseaux sociaux). Certaines de ces applications vont être utilisées pour comparer ces différentes approches.

We will review recent developments of Monte Carlo and deterministic techniques for making Bayesian inferences for unknown parameters when the likelihood function has an unknown normalizing function. For spatial lattice data with likelihoods involving the Ising, Potts or autologistic models, the normalizing constant can be intractable if the lattice is too large. For this situation Møller Pettitt, Reeves and Berthelsen (2006) introduced a Markov chain Monte Carlo algorithm which avoids computation of the normalizing constant but at the expense of requiring perfect simulation from the model for an auxiliary variable. Starting with this development, approximations and improvements to this and related algorithms will be described and applications given including (1) a variational Bayes approach to image analysis and (2) distributions on random graphs (eg social networks). Some applications will be given to compare and contrast the various approaches.

Introduction.

We review recent developments involving Monte Carlo and deterministic techniques for making Bayesian inferences for unknown parameters when the likelihood function has an unknown normalizing function or which is completely intractable from a computational point of view. We start by considering developments using Markov chain Monte Carlo (MCMC) algorithms involving auxiliary variables and applied to the Ising model. We then investigate how these algorithms can be applied to general exponential family models applied to random graphs. Finally, in order to provide solutions which are computationally efficient we consdier variational Bayes approximations which utilise deterministic approaches.

Auxiliary variable Monte Carlo

For binary data spatial lattice data with likelihoods involving the Ising model, the normalizing constant can be intractable if the lattice is too large. For this situation Møller, Pettitt, Reeves and Berthelsen (2006) introduce an MCMC algorithm which avoids computation of the normalizing constant but at the expense of requiring perfect simulation from the model for an auxiliary variable, so the method is sometimes known as single auxilairy variable MCMC or SVMCMC. The unknown normalizing constant cancels in the Metropolis Hastings ratio by using a proposal distribution for the auxiliary variable which is from the Ising model but the distribution of the auxiliary variable remains arbitrary. A solution to the problem is to take the auxiliary distribution as an Ising model with its parameter estimated by pseudolikelihood but this can lead to a very sticky MCMC. In the context of a probabilistic nearest neighbour algorithm, Cucala, Marin, Robert and Titterington (2009) develop an adaptive version of the SVMCMC where the unknown parameter is learnt by using historical values in the MCMC to estimate the posterior mean of the unknown parameter. They additionally replace perfect sampling from the Ising model by the final sample from an MCMC output for the Ising model. The resulting algorithm has improved computational and convergence properties over the original algorithm of Møller and others (2006).

A variant on the SVMCMC approach is an algorithm which has its origins in Murray (2007) but the proof as an MCMC algorithm is rather obscure. It can be viewed in terms again where the target distribution (the posterior) is extended to include both a copy of the parameter and data from the Ising model and then a ratio involving the unknown normalizing constant is naively estimated by importance sampling. The Metropolis-Hastings acceptance probabilty is then derived correctly from seeing the proposal as an exchange of states as in population MCMC. This algorithm still requires in theory perfect sampling from the Ising model but has improved convergence properties compared with the SVMCMC algorithm. The need to define a distribution for the auxiliary copy parameter as close to the posterior is not required, but instead a distribution for the auxiliary parameter can be taken as a random walk centred on the target parameter.

Application to Exponential Random Graph Models

The methods such as the SVMCMC and exchange algorithm are becoming popular in the analysis of random graph models which describe networks. Network models are useful for dealing with the kinds of statistical dependence induced by a variety of relationships between entities, varying from social relationships between people to proteins. Much recent effort has been focused on inference for social network models and applications are in fields such as epidemiology, with the spread of diseases, business and political science. The key idea is to represent complex relationships and interactions between objects of interest (e.g. people, regions, proteins) by a network graph comprising nodes connected via edges representative of node relationships. This can result in extremely complex graphs, the advantage being that typically complex real-world settings can be better represented (see, for example, Read, Eames and Edmunds (2008)). The standard statistical models for random network structure are exponential random graph models (ERGMs), which are in the exponential family and have a long history in the networks literature (Hunter, Goodreau and Handcock, 2008). Current methods for calculating approximate maximum likelihood estimates (MLEs) of the ERGM parameters given an observed network are conceptually simple, but their practical implementation for relatively large social networks has proven largely elusive because the models have intractable normalizing constants in exactly the same manner as the Ising model. The primary MLE approach used for social networks is based on that of Geyer and Thompson (1992) and involves an off-line MCMC approximation of the normalizing constant. Obviously the SVMCMC and exchange algorithm can be applied to exponential random graph models and examples are given.

Application to image analysis

To make sense of the image data that results from a medical scan, it is first turned into a statistical map. After this, a statistical classification or segmentation of the pixels of the image can be carried out to take account of the noise present in the images. The aim of a statistical analysis might be to classify regions of the image as being from an area where there is or is not a tumour present, for example. Naturally, the statistical approach used for classification will influence the conclusions that are drawn from the image. Therefore, it is important to find analytical methods that are reliable and yet are also fast and therefore feasible to implement.

The approach of Friel, Pettitt, Reeves and Wit (2008) develops a model based on the Ising model to define classes in a hidden layer and the pixel data are observed with noise. The SVMCMC approach is used to obtain a Bayesian classification of an image however the algorithm is very slow. Friel and others (2008) develop deterministic approximations for the normalizing constant based on the efficient recursion of Reeves and Pettitt (2004) for small arrays and apply the method to large arrays by incorporating the approximation into MCMC.

For practical implemenation with larger arrays, MCMC is generally too slow and variational Bayes (VB) is an emerging fast, deterministic approach to Bayesian inference. It involves finding a close approximation to the intractable posterior distribution in a Bayesian analysis. The approximation is based on the idea of minimising the Kullback-Leibler (KL) divergence between the variational approximating distribution and the target joint posterior distribution. Approximate methods like VB can lead to highly time efficient solutions to inference problems with only a small reduction in accuracy when compared to alternative MCMC-based approaches. McGrory and Titterington (2009) highlight the great potential for the use of VB in image analysis where schemes must be time efficient in order to be feasibly applied to the large datasets involved and the length of time required for an MCMC analysis is prohibitively long. In McGrory, Titterington, Reeves and Pettitt (2009) the variational Bayes approach is developed for a hidden Markov random field given by an Ising model plus Gaussian noise for the observed data using a determinstic approximation for the normalizing constant. Applications using this approach are provided.

Ackowledgements: This paper involves joint work with Dr Nial Friel and Dr Clare McGrory and supported by the Australian Reserach Council.

References

Cucala, L., Marin, J.M, Robert, C. and Titterington, D.M. (2009). A Bayesian reassessment of nearest-neighbour classification. *J. Amer. Statist. Assoc.*

Friel, N., Pettitt, A.N. , Reeves, R. and Wit, E. (2008). Bayesian inference in hidden Markov random fields for binary data defined on large lattices. Submitted for publication.

Geyer, C.J., and Thompson, E.A., (1992). Constrained Monte Carlo maximum likelihood calculations (with discussion). *Journal of the Royal Statistical Society, Series B, **54***, 657–699.

Hunter, D.R., Goodreau, S.M. and Handcock, M.S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association.* **481,** 248–258.

McGrory, C.A. and Titterington, D.M. (2009). Bayesian analysis of hidden Markov models using variational approximations. *Australian and New Zealand Journal of Statistics*. To appear.

McGrory, C.A., Titterington, D.M., Reeves, R. and Pettitt, A.N. (2009). Variational Bayes for Hidden Markov random field analysis. *Statistics and Computing*.

Møller, J., Pettitt, A.N., Reeves, R. and Berthelsen, K.K. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika,* **93**, 451-458.

Murray, I. (2007). *Advances in Markov Chain Monte Carlo Methods*. PhD Thesis, Gatsby Computational Neuroscience Unit, University College London

Read, J.M., Eames, K.T.D. and Edmunds, J.W. (2008). Dynamic social networks and the implications for the spread of infectious disease. *Journal of the Royal Society Interface,* **5**, 1001-1007.

Reeves, R, and Pettitt, A.N. (2004). Efficient Recursions for General Factorisable Models. *Biometrika*, **91**, 751-757.