Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Model comparison for Gibbs random fields using noisy reversible jump Markov chain Monte Carlo

Lampros Bouranis *, Nial Friel, Florian Maire

School of Mathematics and Statistics & Insight Centre for Data Analytics, University College Dublin, Ireland

ARTICLE INFO

Article history:
Received 9 January 2018
Received in revised form 12 July 2018
Accepted 12 July 2018
Available online xxxx

Keywords:
Bayes factors
Intractable likelihoods
Markov random fields
Noisy MCMC

ABSTRACT

The reversible jump Markov chain Monte Carlo (RJMCMC) method offers an across-model simulation approach for Bayesian estimation and model comparison, by exploring the sampling space that consists of several models of possibly varying dimensions. A naive implementation of RJMCMC to models like Gibbs random fields suffers from computational difficulties: the posterior distribution for each model is termed doubly-intractable since computation of the likelihood function is rarely available. Consequently, it is simply impossible to simulate a transition of the Markov chain in the presence of likelihood intractability. A variant of RJMCMC is presented, called noisy RJMCMC, where the underlying transition kernel is replaced with an approximation based on unbiased estimators. Based on previous theoretical developments, convergence guarantees for the noisy RJMCMC algorithm are provided. The experiments show that the noisy RJMCMC algorithm can be much more efficient than other exact methods, provided that an estimator with controlled Monte Carlo variance is used, a fact which is in agreement with the theoretical analysis.

1. Introduction

Model selection is a problem of great importance in statistical science. The aim is to choose which model among a set of possible ones best describes the data $y \in \mathcal{Y}$. From a Bayesian perspective, the prior beliefs for each model are reflected by a prior distribution and this information is then updated subjectively when data are observed. This step is typically carried out by calculating the marginal likelihood or *evidence* for each model, which is defined as the integrated likelihood with respect to the prior measure. In many cases, this quantity cannot be derived analytically and thus needs to be estimated.

This paper considers the problem of Bayesian model comparison of *doubly-intractable* distributions. The motivating application is Gibbs random fields (GRFs), which are discrete-valued Markov random fields where an intractable normalising constant that depends on the model parameters, $z(\theta)$, is present for the tractable un-normalised likelihood $q(y \mid \theta)$. The likelihood density, given a vector of parameters $\theta \in \Theta \subseteq \mathbb{R}^d$ and a vector of statistics $s(y) \in \mathcal{S} \subseteq \mathbb{R}^d_+$ that are sufficient for the likelihood, is

$$f(y \mid \theta) = \frac{q(y \mid \theta)}{z(\theta)} = \frac{\exp\left\{\theta^\top s(y)\right\}}{\sum_{y \in \mathcal{Y}} \exp\left\{\theta^\top s(y)\right\}}. \quad (1)$$

Posterior parameter estimation for GRFs, which have found applicability in areas like image analysis and disease mapping (Friel and Rue, 2007), genetic analysis (François et al., 2006) and social network analysis (Wasserman and Pattison, 1996), is termed a doubly-intractable problem because the normalisation terms of both the likelihood function and the posterior

* Corresponding author.
E-mail addresses: lampros.bouranis@ucdconnect.ie (L. Bouranis), nial.friel@ucd.ie (N. Friel), florian.maire@ucd.ie (F. Maire).

https://doi.org/10.1016/j.csda.2018.07.005

Please cite this article in press as: Bouranis L., et al., Model comparison for Gibbs random fields using noisy reversible jump Markov chain Monte Carlo. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.07.005.

distribution $\pi(\theta \mid y) \propto f(y \mid \theta)p(\theta)$ are intractable. Bayesian model comparison of such models has attracted the attention of researchers and several methods have been considered, relying on likelihood simulations (Friel, 2013; Caimo and Friel, 2013; Everitt et al., 2017a), approximations to the intractable likelihood (Bouranis et al., 2018) and likelihood-free simulation techniques like the Approximate Bayesian Computation (ABC) algorithm (Grelaud et al., 2009) .

In this paper we explore trans-dimensional Markov chain Monte Carlo (MCMC) for GRFs, focusing on the direct approach of a single across-model Markov chain using the celebrated reversible jump MCMC (RJMCMC) technique (Green, 1995). The advantage of across-model approaches is that they avoid the need for computing the evidence for each competing model by treating the model indicator as a parameter, where the chain explores simultaneously the model set and the parameter space. In the context of GRFs, however, RJMCMC techniques simply cannot be implemented because the likelihood normalising constant $z(\theta)$ cannot be computed point-wise. Below we present a summary of the main contributions of this paper.

A variant of the RJMCMC algorithm is developed, where the intractable ratio of normalising constants that the acceptance probability depends on is approximated by an unbiased estimator. The resulting algorithm falls in the *noisy MCMC* framework (Alquier et al., 2016) as it simulates a chain that is not invariant for the target distribution.

As pointed out in Alquier et al. (2016), noisy MCMC is connected to pseudo-marginal algorithms (Andrieu and Roberts, 2009), where an unbiased and positive estimate of the target density is required. In the presence of an intractable normalising constant, the pseudo-marginal approach requires an unbiased estimate of $1/z(\theta)$. However, the reciprocal $1/\hat{z}(\theta)$ yields a biased approximation of $1/z(\theta)$ and so is not directly applicable to inference using pseudo-marginal techniques. Recently, Lyne et al. (2015) addressed this bias with Russian Roulette sampling, which yields an asymptotically exact algorithm. The implementation of the algorithm is computationally expensive, however, creating difficulties when inferring the model parameters for GRFs. Non-negative unbiased estimators have also been studied in Jacob and Thiery (2015), which showed that finding such an estimate is very challenging.

Consequently, we do not pursue such a computational approach in this paper. Instead, motivated by the inefficiency of a standard RJMCMC algorithm, we develop a noisy RJMCMC sampler that targets an approximated posterior distribution, rather than the desired one. We extend the theoretical analysis of noisy MCMC algorithms proposed in Alquier et al. (2016) to trans-dimensional kernels, providing bounds on the total variation between the Markov chain of a noisy RJMCMC algorithm and a Markov chain with the desired target distribution under certain conditions.

We show that noisy RJMCMC algorithms are only useful when the estimator of the ratio of normalising constants has a small variance. Motivated by Gelman and Meng (1998), we propose a smoother transition path between different models and resort to an alternative estimator with lower variance. We demonstrate empirically that this idea simultaneously: (i) improves the mixing of the RJ Markov chain and (ii) decreases the asymptotic bias between the exact RJ Markov chain and its noisy approximation. Finally, we construct efficient jump proposal distributions for random walk (RW) noisy RJMCMC, which could be useful in the context of a large number of nested competing models.

The outline of the article is as follows. Section 2 introduces the reader to basic concepts regarding Bayesian model comparison and to the reversible jump MCMC formulation. In Section 3 we discuss the extension of the reversible jump methodology to doubly-intractable posterior distributions and present theoretical properties and practical aspects of the RJMCMC samplers under such computational difficulties. Section 4 presents some proposal tuning strategies for noisy RJMCMC. In Section 5 we study the theoretical behaviour of the noisy RJMCMC algorithm and derive convergence bounds. We investigate the performance of noisy RJMCMC with a detailed numerical study that focuses on social network analysis in Section 6. We conclude the paper in Section 7 with final remarks.

2. Preliminaries

Suppose a finite set of competing models $\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \ldots\}$ are under consideration to describe the data y. In the Bayesian setting each model $\mathcal{M}_m$, where $m \in \{1, 2, 3, \ldots\}$, is characterised by a likelihood function $f(y \mid \theta_m, \mathcal{M}_m) = f_m(y \mid \theta_m) \propto q_m(y \mid \theta_m)$, parameterised by an unknown parameter vector $\theta_m \in \Theta_m$. Each model is also associated with a prior distribution $p(\theta_m \mid \mathcal{M}_m) = p_m(\theta_m)$, used to express the beliefs about the parameter vector prior to observing the data y. The focus of interest in Bayesian inference for each competing model is the posterior distribution

$$\pi(\theta_m \mid y, \mathcal{M}_m) = \frac{f_m(y \mid \theta_m)p_m(\theta_m)}{\pi(y \mid \mathcal{M}_m)}. \tag{2}$$

The prior beliefs for each model are expressed through a prior distribution $p(\mathcal{M}_m)$, such that $\sum_{m\in\mathcal{M}} p(\mathcal{M}_m) = 1$.

2.1. Bayesian model comparison

The marginal likelihood or model *evidence* for model $\mathcal{M}_m$ is

$$\pi(y \mid \mathcal{M}_m) = \int_{\Theta_m} f_m(y \mid \theta_m)p_m(\theta_m)\,\mathrm{d}\theta_m$$

Please cite this article in press as: Bouranis L., et al., Model comparison for Gibbs random fields using noisy reversible jump Markov chain Monte Carlo. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.07.005.