Computational Statistics and Data Analysis 71 (2014) 402–416

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Use of SAMC for Bayesian analysis of statistical models with intractable normalizing constants

Ick Hoon Jin [a], Faming Liang [b,*]

[a] Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030-4009, USA

[b] Department of Statistics, Texas A & M University, College Station, TX 77843-3143, USA

ARTICLE INFO

Article history:
Received 20 December 2011
Received in revised form 9 June 2012
Accepted 5 July 2012
Available online 13 July 2012

Keywords:
Bayesian analysis
Intractable normalizing constant
Markov chain Monte Carlo
Stochastic approximation

ABSTRACT

Statistical inference for the models with intractable normalizing constants has attracted much attention. During the past two decades, various approximation- or simulation-based methods have been proposed for the problem, such as the Monte Carlo maximum likelihood method and the auxiliary variable Markov chain Monte Carlo methods. The Bayesian stochastic approximation Monte Carlo algorithm specifically addresses this problem: It works by sampling from a sequence of approximate distributions with their average converging to the target posterior distribution, where the approximate distributions can be achieved using the stochastic approximation Monte Carlo algorithm. A strong law of large numbers is established for the Bayesian stochastic approximation Monte Carlo estimator under mild conditions. Compared to the Monte Carlo maximum likelihood method, the Bayesian stochastic approximation Monte Carlo algorithm is more robust to the initial guess of model parameters. Compared to the auxiliary variable MCMC methods, the Bayesian stochastic approximation Monte Carlo algorithm avoids the requirement for perfect samples, and thus can be applied to many models for which perfect sampling is not available or very expensive. The Bayesian stochastic approximation Monte Carlo algorithm also provides a general framework for approximate Bayesian analysis.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

In statistics, many models possess an intractable normalizing constant. Examples of such models include the Ising model used in statistical physics; the Potts model used in image analysis (Hurn et al., 2003; Johnson and Piert, 2009), the autologistic, autonormal, and pairwise interaction models used in spatial statistics (Besag, 1974; Bognar, 2005); the exponential random graph model used in social networks (Robins et al., 2007a,b; Snijders et al., 2006), among others. The problem of Bayesian inference for these models can be posed as follows.

Suppose that we have a dataset generated from a model with the density/mass function given by

$$f(x|\theta) = \frac{p(x,\theta)}{\kappa(\theta)}, \quad x \in \mathcal{X},\ \theta \in \Theta, \tag{1}$$

where θ denotes the vector of parameters, and $\kappa(\theta)$ is the normalizing constant which depends on θ and is not available in closed form. Let $\pi(\theta)$ denote the prior density of θ. The posterior density of θ given the data $Z = \boldsymbol{z}$ can then be expressed as

$$\pi(\theta|\boldsymbol{z}) \propto \frac{1}{\kappa(\theta)} p(\boldsymbol{z},\theta)\pi(\theta). \tag{2}$$

Since the closed form of $\kappa(\theta)$ is not available, inference for θ has put a great challenge on current statistical methods.

* Corresponding author. Tel.: +1 979 845 8885.
E-mail addresses: ijin@mdanderson.org (I.H. Jin), fliang@stat.tamu.edu (F. Liang).

0167-9473/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2012.07.005

The Metropolis–Hasting (MH) algorithm cannot be directly applied to simulate from the posterior $\pi(\theta|\boldsymbol{z})$, as the acceptance probability would include an unknown normalizing constant ratio $\kappa(\theta)/\kappa(\theta')$, where θ' denotes the proposed value. To avoid this obstacle, various approximation-based methods have been proposed in the literature. Besag (1974) proposed to approximate the likelihood function by a product of a series of conditional likelihood functions by ignoring the dependence between components of X. This method is simple, but its performance is often not satisfactory, particularly for the problems for which the dependence between components of X is strong. Geyer and Thompson (1992) proposed to approximate $\kappa(\theta)$ using Monte Carlo samples generated from a distribution $f(x|\theta_0)$, where θ_0 is an initial estimate of θ. To be precise, let $x^{(1)}, \ldots, x^{(m)}$ denote random samples drawn from $f(x|\theta_0)$, which can be obtained via Markov chain Monte Carlo (MCMC) simulations, then the log-likelihood function can be approximated by

$$\log f_m(\boldsymbol{z}|\theta) = \log p(\boldsymbol{z}, \theta) - \log \kappa(\theta_0) - \log\left(\frac{1}{m}\sum_{i=1}^{m} p(x^{(i)}, \theta)/p(x^{(i)}, \theta_0)\right), \tag{3}$$

which approaches $\log f(\boldsymbol{z}|\theta)$ as $m \to \infty$. The estimator $\widehat{\theta} = \arg\max_\theta \log f_m(\boldsymbol{z}|\theta)$ is called the Monte Carlo maximum likelihood estimator (MCMLE) of θ. The performance of the MCMLE method depends on the choice of θ_0. If θ_0 is near the true maximum likelihood estimate, the method usually produces a good estimate of θ. Otherwise, the method may converge to a suboptimal solution or even fail to converge. To alleviate this difficulty, Geyer and Thompson (1992) recommended an iterative approach, which works as follows:

(a) Simulate m auxiliary samples $x_t^{(1)}, \ldots, x_t^{(m)}$ from $f(x|\theta_t)$ using MCMC.
(b) Find $\theta_{t+1} = \arg\max_\theta \log f_m(\boldsymbol{z}|\theta)$.

Even with this iterative approach, non-convergence is still quite common if θ_0 is far from the true MLE. Liang (2007) proposed an alternative Monte Carlo approach to approximate $\kappa(\theta)$, where $\kappa(\theta)$ is viewed as a marginal density function of the unnormalized distribution $p(x, \theta)$ and estimated using an adaptive kernel smoothing approach with Monte Carlo draws. On the other hand, analytical approximations to the normalizing constant function have also been developed for some specific models, e.g., the Markov random field models defined on the cylinder or lattice. See Pettitt et al. (2003), Friel and Rue (2007), Friel et al. (2009), among others.

As an alternative to the approximation-based methods, a class of auxiliary variable MCMC algorithms have been proposed for sampling from the posterior distribution (2). Møller et al. (2006) proposed to augment the posterior distribution $\pi(\theta|\boldsymbol{z})$ to $\pi(\theta, y|\boldsymbol{z})$ by including an auxiliary variable Y. With an appropriate choice of the auxiliary distribution $f(y|\theta, \boldsymbol{z})$ and the proposal distribution, the normalizing constant ratio $\kappa(\theta)/\kappa(\theta')$ can be canceled in simulations. The exchange algorithm (Murray et al., 2006) works in a similar way to the Møller algorithm; it cancels the normalizing constants ratio in simulations by including an auxiliary variable in its proposal. Although the Møller and exchange algorithms work well for some discrete models, such as the Ising and autologistic models, they cannot be applied to many other models for which perfect sampling is not available. Even for the Ising and autologistic models, perfect sampling can be very expensive when the temperature is near or below the critical point. Recently, Liang (2010) proposed the double Metropolis–Hastings (MH) algorithm, which is to approximate expensive perfect samples using cheap MCMC samples. Although this algorithm works well for many problems, it lacks a theoretical justification for the consistency of the resulting estimator. It is worth mentioning that Walker (2011) introduced a latent variable method for sampling from the posterior distribution (2). In Walker's method, the likelihood function $f(x|\theta)$ is augmented with a varying number of latent variables such that the joint likelihood function of x and the latent variables is tractable, and then the joint likelihood function is sampled using the reversible jump MCMC algorithm (Green, 1995). This method is mathematically sound, but, due to the difficulty in dimension jumping moves, it can be very difficult to be applied to the problems for which the dimension of the model (i.e., the dimension of x) is high.

In this paper, we propose a new algorithm, the Bayesian Stochastic Approximation Monte Carlo (BSAMC) algorithm, for tackling the intractable normalizing constant problem. The BSAMC algorithm works by simulating from a sequence of approximate distributions, which are obtained using the stochastic approximation Monte Carlo (SAMC) algorithm (Liang et al., 2007). Let $\pi_t(\theta|\boldsymbol{z})$, $t = 1, 2, \ldots$, denote the sequence of approximate distributions. Let θ_t denote a sample simulated from $\pi_t(\theta|\boldsymbol{z})$. Under mild conditions, we show that for any bounded measurable function $\varphi(\theta)$, $\sum_{t=1}^{n} \varphi(\theta_t)/n$ converges almost surely to the posterior mean of $\varphi(\theta)$ as n goes to infinity. One significant advantage of BSAMC over the auxiliary variable MCMC methods is that it avoids the requirement for perfect samples, and thus can be applied to many models for which perfect sampling is not available or very expensive. Compared to the MCMLE method, BSAMC is very robust to the choice of θ_0 due to the powerful ability of SAMC in sample space exploration. Finally, we note that although BSAMC works based on SAMC, SAMC itself cannot be directly applied to sample from the posterior $\pi(\theta|\boldsymbol{z})$. Hence, BSAMC represents an extension of SAMC for Bayesian analysis. BSAMC also provides a general framework for approximate Bayesian analysis through simulating from a sequence of approximate distributions with their average converging to the target posterior distribution.

The remainder of this paper is organized as follows. In Section 2, we describe the BSAMC algorithm and explore its theoretical properties. In Section 3, we apply BSAMC to Ising models along with a comparison with the MCMLE method. The numerical results show that BSAMC performs robustly to the initial guess of θ. In Section 4, we apply BSAMC to autologistic and autonormal models. In Section 5, we conclude the paper with a brief discussion.

Download English Version:

https://daneshyari.com/en/article/6870413

Download Persian Version:

https://daneshyari.com/article/6870413

Daneshyari.com