Computational Statistics and Data Analysis 71 (2014) 392–401

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Computation of marginal likelihoods with data-dependent support for latent variables☆

Sarah E. Heaps*, Richard J. Boys, Malcolm Farrow

School of Mathematics & Statistics, Herschel Building, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom

ARTICLE INFO

Article history:
Received 9 June 2012
Received in revised form 22 July 2013
Accepted 23 July 2013
Available online 31 July 2013

Keywords:
Annealed importance sampling
Latent variables
Linked importance sampling
Marginal likelihood
Power posterior method
Spatial count data

ABSTRACT

Several Monte Carlo methods have been proposed for computing marginal likelihoods in Bayesian analyses. Some of these involve sampling from a sequence of intermediate distributions between the prior and posterior. A difficulty arises if the support in the posterior distribution is a proper subset of that in the prior distribution. This can happen in problems involving latent variables whose support depends upon the data and can make some methods inefficient and others invalid. The correction required for models of this type is derived and its use is illustrated by finding the marginal likelihoods in two examples. One concerns a model for competing risks. The other involves a zero-inflated over-dispersed Poisson model for counts of centipedes, using latent Gaussian variables to capture spatial dependence.

1. Introduction

The marginal likelihood, also known as the integrated likelihood or the evidence, plays an important role in Bayesian inference, particularly in model selection and model averaging, where it is used in the computation of Bayes factors and posterior model probabilities.

Consider data $\mathbf{y}$ and a statistical model $p(\mathbf{y}|\boldsymbol{\theta})$ which depends on unknowns $\boldsymbol{\theta}$. The marginal likelihood is defined as $p(\mathbf{y}) = \int p(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}$, where $\pi(\boldsymbol{\theta})$ is the prior density. Typically this integral cannot be evaluated in closed form and so we turn to numerical approximation; see, for example, Friel and Wyse (2012), for a recent review. It is convenient to use methods which involve Markov chain Monte Carlo (MCMC) sampling. In particular, this allows auxiliary or latent variables to be included in the unknowns $\boldsymbol{\theta}$ and sampled along with the model parameters. We focus primarily on latent variable problems in this paper.

Amongst the Monte Carlo methods particularly suitable for latent variable problems are Chib's method (Chib, 1995; Chib and Jeliazkov, 2001) and techniques which we term *intermediate-density-methods*. Chib's method is based on a rearrangement of Bayes Theorem to express the marginal likelihood in terms of the prior density, likelihood and posterior density. Evaluating or approximating each term in the resulting identity at a single point in the parameter space then yields the marginal likelihood approximation. Intermediate-density-methods connect the unnormalised prior and posterior densities through a sequence of intermediate densities labelled by an index $t \in [0, 1]$. They are derived from more general approaches for computing ratios of normalising constants and include the power posterior method (Friel and Pettitt, 2008; Friel et al., 2012), annealed importance sampling (AIS) (Neal, 2001) and linked importance sampling (LIS) (Neal, 2005). The

* Corresponding author. Tel.: +44 0 191 222 7245.
E-mail addresses: sarah.heaps@newcastle.ac.uk (S.E. Heaps), richard.boys@newcastle.ac.uk (R.J. Boys), malcolm.farrow@newcastle.ac.uk (M. Farrow).

0167-9473/$ – see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.07.033

unnormalised intermediate density with index t may then be, for example, the product of the unnormalised prior and the likelihood raised to the power t. An advantage of Chib's and, in particular, intermediate-density-methods, is the ease with which they can be programmed, often simply by rearranging code for sampling from the posterior distribution. Both types of methods can also be very effective. For instance, Germain (2010) found them to provide an easily implemented and accurate approximation to the marginal likelihoods of hidden Markov models with different numbers of states.

Latent variable problems can have the property that the support of the prior and posterior distributions do not coincide because the support for the latent variables changes when data are observed. For models with this property, some of the intermediate-density-methods, such as the power posterior approach, cannot be directly applied whilst others, like AIS, are likely to be very inefficient in cases where the prior probability of the posterior support is small, such as in multivariate probit models. Data-dependent support can also present problems for Chib's method if the likelihood ordinate (typically the observed data likelihood) is difficult to evaluate. This paper addresses the former of these issues and describes a general two-stage procedure to correct, or improve the efficiency of, intermediate-density-methods in problems involving data-dependent support, whilst also highlighting the situations in which implementation of the proposed approach is likely to be simpler than Chib's method.

We review intermediate-density-methods for computing marginal likelihoods in Section 2. In Section 3 we discuss the change of support problem and derive our two-stage approximation procedure. Next, in Section 4 we consider two examples. Section 4.1 concerns a simple model for competing risks and Section 4.2 applies our two-stage procedure to a zero inflated–over-dispersed Poisson model for a set of centipede count data. In this model, latent Gaussian variables capture the spatial dependences between the presence and the abundance of centipedes and we compare three variants of the model which use different parametric forms for the covariance matrix. Finally Section 4.3 provides a numerical comparison between our proposed method and other, related methods for marginal likelihood approximation.

2. Computing marginal likelihoods using sequences of densities

Consider a pair of density functions $p_t(\boldsymbol{\theta})$, $t = 0, 1$, with $p_t(\boldsymbol{\theta}) = q_t(\boldsymbol{\theta})/z_t$ for $\boldsymbol{\theta} \in \Theta_t$, where $q_t(\boldsymbol{\theta})$ is the unnormalised density, z_t is a normalising constant and Θ_t is the support of p_t. Several techniques for computing marginal likelihoods are special cases of more general methods for computing ratios of normalising constants, $r = z_1/z_0$. Let $p_0(\boldsymbol{\theta})$ be the prior density, $\pi(\boldsymbol{\theta})$, and let $p_1(\boldsymbol{\theta})$ be the posterior density, $\pi(\boldsymbol{\theta}|\mathbf{y})$. Then, if $q_1(\boldsymbol{\theta}) = p(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$, where $p(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood, z_1 is the marginal likelihood, $p(\mathbf{y})$. Typically, the normalising constant of the prior distribution will be known and we can assume that $z_0 = 1$. Then the ratio $r = z_1/z_0 = z_1 = p(\mathbf{y})$.

2.1. Computing ratios of normalising constants

Provided that $\Theta_1 \subseteq \Theta_0$, it appears that we might approximate the ratio z_1/z_0 using *simple importance sampling*:

$$\frac{z_1}{z_0} = \mathrm{E}_{p_0}\left\{\frac{q_1(\boldsymbol{\theta})}{q_0(\boldsymbol{\theta})}\right\} \simeq \frac{1}{M}\sum_{i=1}^{M}\frac{q_1(\boldsymbol{\theta}^{[i]})}{q_0(\boldsymbol{\theta}^{[i]})}, \tag{1}$$

where E_{p_0} denotes expectation with respect to p_0 and $\boldsymbol{\theta}^{[1]}, \ldots, \boldsymbol{\theta}^{[M]}$ are a sample drawn from p_0. However this method works poorly when the overlap of p_0 and p_1 is small, as will typically be the case if p_0 and p_1 represent the prior and posterior and the posterior is very concentrated relative to the prior.

In response to this problem, *bridge sampling* (Meng and Wong, 1996) uses an unnormalised density $q_{0.5}$, with support $\Theta_0 \cap \Theta_1$, to provide a "bridge" between p_0 and p_1. This leads to the identity

$$\frac{z_1}{z_0} = \frac{\mathrm{E}_{p_0}\left\{\frac{q_{0.5}(\boldsymbol{\theta})}{q_0(\boldsymbol{\theta})}\right\}}{\mathrm{E}_{p_1}\left\{\frac{q_{0.5}(\boldsymbol{\theta})}{q_1(\boldsymbol{\theta})}\right\}}, \tag{2}$$

in which the ratios in the numerator and denominator are each approximated using simple importance sampling, as in (1). Whereas simple importance sampling requires $\Theta_1 \subseteq \Theta_0$, bridge sampling only requires $\int_{\Theta_0\cap\Theta_1} p_0(\boldsymbol{\theta})p_1(\boldsymbol{\theta})\,d\boldsymbol{\theta} > 0$.

When there is little overlap between p_0 and p_1, bridge sampling with a single intermediate density will perform poorly. However we can improve performance by introducing a sequence of intermediate densities, $p_{t_i}(\boldsymbol{\theta}) = q_{t_i}(\boldsymbol{\theta})/z_{t_i}$, $\boldsymbol{\theta} \in \Theta_{t_i}$, $i = 0, \ldots, n$, between p_0 and p_1, with $0 = t_0 < t_1 < \cdots < t_n = 1$. Then the ratio z_1/z_0 can be expressed as

$$\frac{z_1}{z_0} = \prod_{i=1}^{n}\frac{z_{t_i}}{z_{t_{i-1}}}. \tag{3}$$

Each of the ratios $z_{t_i}/z_{t_{i-1}}$ can then be approximated by simple importance sampling or by bridge sampling using an unnormalised bridging density $q_{t_{i-0.5}}$. Provided that each pair $p_{t_{i-1}}, p_{t_i}$ displays sufficient overlap, this can provide substantial improvement over standard importance or bridge sampling. In the remainder of this paper, methods based on these ideas will be called *extended importance sampling* and *extended bridge sampling* techniques.