ARTICLE IN PRESS

Computational Statistics and Data Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

On the use of marginal posteriors in marginal likelihood estimation via importance sampling

Konstantinos Perrakis [a], Ioannis Ntzoufras [a,*], Efthymios G. Tsionas [b]

[a] Department of Statistics, Athens University of Economics and Business, Greece

[b] Department of Economics, Athens University of Economics and Business, Greece

ARTICLE INFO

Article history:
Received 4 July 2013
Received in revised form 8 March 2014
Accepted 8 March 2014
Available online xxxx

Keywords:
Finite normal mixtures
Importance sampling
Marginal posterior
Marginal likelihood estimation
Random effect models
Rao–Blackwellization

ABSTRACT

The efficiency of a marginal likelihood estimator where the product of the marginal posterior distributions is used as an importance sampling function is investigated. The approach is generally applicable to multi-block parameter vector settings, does not require additional Markov Chain Monte Carlo (MCMC) sampling and is not dependent on the type of MCMC scheme used to sample from the posterior. The proposed approach is applied to normal regression models, finite normal mixtures and longitudinal Poisson models, and leads to accurate marginal likelihood estimates.

1. Introduction

The problem of estimating the marginal likelihood has received considerable attention during the last two decades. The topic is of importance in Bayesian statistics as it is associated with the evaluation of competing hypotheses or models via Bayes factors and posterior model odds. Consider, briefly, two competing models M_1 and M_2 with corresponding prior probabilities $\pi(M_1)$ and $\pi(M_2) = 1 - \pi(M_1)$. After observing a data vector $\mathbf{y}$, the evidence in favor of M_1 (or against M_2) is evaluated through the odds of the posterior model probabilities $p(M_1|\mathbf{y})$ and $p(M_2|\mathbf{y})$, that is,

$$\frac{p(M_1|\mathbf{y})}{p(M_2|\mathbf{y})} = \frac{m(\mathbf{y}|M_1)}{m(\mathbf{y}|M_2)} \times \frac{\pi(M_1)}{\pi(M_2)}.$$

The quantity $B_{12} = m(\mathbf{y}|M_1)/m(\mathbf{y}|M_2)$ is the ratio of the marginal likelihoods or prior predictive distributions of M_1 and M_2 and is called the Bayes factor of M_1 versus M_2. The Bayes factor can also be interpreted as the ratio of the posterior odds to the prior odds. When M_1 and M_2 are assumed to be equally probable a-priori, the Bayes factor is equal to the posterior odds.

The marginal likelihood of a given model M_k associated with a parameter vector $\boldsymbol{\theta}_k$ is essentially the normalizing constant of the posterior $p(\boldsymbol{\theta}_k|\mathbf{y}, M_k)$, obtained by integrating the likelihood function $l(\mathbf{y}|\boldsymbol{\theta}_k, M_k)$ with respect to the prior density $\pi(\boldsymbol{\theta}_k|M_k)$, i.e.

$$m(\mathbf{y}|M_k) = \int l(\mathbf{y}|\boldsymbol{\theta}_k, M_k)\pi(\boldsymbol{\theta}_k|M_k)\mathrm{d}\boldsymbol{\theta}_k. \tag{1}$$

* Corresponding author. Tel.: +30 210 8203968.
E-mail addresses: dinosperrakis@gmail.com (K. Perrakis), ntzoufras@aueb.com (I. Ntzoufras), tsionas@aueb.gr (E.G. Tsionas).

http://dx.doi.org/10.1016/j.csda.2014.03.004

The integration in (1) may be evaluated analytically for some elementary cases. Most often, it is intractable thus giving rise to the marginal likelihood estimation problem. Numerical integration methods can be used as an approach to the problem, but such techniques are of limited use when sample sizes are moderate to large or when vector $\boldsymbol{\theta}_k$ is of large dimensionality. In addition, the simplest Monte Carlo (MC) estimate, which is given by

$$\widehat{m}(\mathbf{y}|M_k) = N^{-1} \sum_{n=1}^{N} l(\mathbf{y}|\boldsymbol{\theta}_k^{(n)}, M_k), \quad (2)$$

using draws $\{\boldsymbol{\theta}_k^{(n)} : n = 1, 2, \ldots, N\}$ from the prior distribution, is extremely unstable when the posterior is concentrated in relation to the prior. This scenario is frequently met in practice when flat, low-information prior distributions are used to express prior ignorance. A detailed discussion regarding Bayes factors and marginal likelihood estimation is provided by Kass and Raftery (1995).

It is worth noting that the problem of estimating (1) can be bypassed by considering model indicators as unknown parameters. This option has been investigated by several authors (e.g. Green, 1995, Carlin and Chib, 1995 and Dellaportas et al., 2002) who introduce MCMC algorithms which sample simultaneously over parameter and model space and deliver directly posterior model probabilities. However, implementation of these methods can get quite complex since they require enumeration of all competing models and specification of tuning constants or "pseudopriors" (depending upon approach) in order to ensure successful mixing in model space. Moreover, since these methods focus on the estimation of posterior model probabilities, accurate estimation of the marginal likelihoods and/or Bayes factors will not be feasible in the cases where a dominating model exists in the set of models under consideration; tuning, following the lines of Ntzoufras et al. (2005), might be possible but is typically inefficient and time consuming.

In contrast, "direct" methods provide marginal likelihood estimates by utilizing the posterior samples of separate models. These methods are usually simpler to implement and are preferable in practice when the number of models under consideration is not large, namely when it is practically feasible to obtain a posterior sample for each of the competing models. Work along these lines includes the Laplace–Metropolis method (Lewis and Raftery, 1997), the harmonic-mean and the prior/posterior mixture importance sampling estimators (Newton and Raftery, 1994), bridge-sampling methods (Meng and Wong, 1996), candidate's estimators for Gibbs sampling (Chib, 1995) and Metropolis–Hastings sampling (Chib and Jeliazkov, 2001), annealed importance sampling (Neal, 2001), importance-weighted marginal density estimators (Chen, 2005) and nested sampling approaches (Skilling, 2006). More recently, Raftery et al. (2007) presented a stabilized version of the harmonic-mean estimator, while Friel and Pettitt (2008) and Weinberg (2012) proposed new approaches based on power posteriors and Lebesgue integration theory, respectively. It is worth mentioning that Bayesian evidence evaluation is also of particular interest in the astronomy literature where nested sampling is commonly used for marginal likelihood estimation (e.g. Feroz et al., 2009 and Feroz et al., 2011). Recent reviews comparing popular methods based on MCMC sampling can be found in Friel and Wyse (2012) as well as in Ardia et al. (2012). Alternative approaches for marginal likelihood estimation include sequential Monte Carlo (Del Moral et al., 2006) and variational Bayes (Parise and Welling, 2007) methods.

In this paper we propose using the marginal posterior distributions on importance sampling estimators of the marginal likelihood. The proposed approach is particularly suited for the Gibbs sampler, but it is also feasible to use for other types of MCMC algorithms. The estimator can be implemented in a straightforward manner and it can be extended to multi-block parameter settings without requiring additional MCMC sampling apart from the one used to obtain the posterior sample.

The remainder of the paper is organized as follows. The proposed estimator and its variants are discussed in Section 2. In Section 3 the method is applied to normal regression models, to finite normal mixtures and also to hierarchical longitudinal Poisson models. Concluding remarks are provided in Section 4.

2. The proposed estimator

In the following we first introduce the proposed estimator in a two block setting. The more general multi-block case is considered next, explaining why the estimator will be useful in such cases. We further present details concerning the implementation of the proposed approach when the model formulation includes latent variables or nuisance parameters that are not of prime interest for model inference. The section continues with a description of the different estimation approaches of the posterior marginal distributions used as importance functions. We conclude with a note on a convenient implementation of the estimator for models where the posterior distribution becomes invariant under competing diffuse priors and brief remarks about the calculation of numerical standard errors. In the remaining of the paper, the dependence to the model indicator M_k (introduced in the previous section) is eliminated for notational simplicity.

2.1. Introducing the estimator in a two-block setting

Let us consider initially the 2-block setting where $l(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\phi})$ is the likelihood of the data conditional on parameter vectors $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T$ and $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_q)^T$, which can be either independent, i.e. $\pi(\boldsymbol{\theta}, \boldsymbol{\phi}) = \pi(\boldsymbol{\theta})\pi(\boldsymbol{\phi})$, or dependent, e.g. $\pi(\boldsymbol{\theta}, \boldsymbol{\phi}) = \pi(\boldsymbol{\theta}|\boldsymbol{\phi})\pi(\boldsymbol{\phi})$, a-priori. In general, one can improve the estimator in (2) by introducing a proper importance