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

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Fast computation of the deviance information criterion for latent variable models

Joshua C.C. Chan [a,*], Angelia L. Grant [b]

[a] *Research School of Economics, Australian National University, Australia*

[b] *Centre for Applied Macroeconomic Analysis, Australian National University, Australia*

A R T I C L E I N F O

Article history:
Received 18 December 2013
Received in revised form 28 July 2014
Accepted 29 July 2014
Available online xxxx

Keywords:
Bayesian model comparison
State space
Factor model
Vector autoregression
Semiparametric model

A B S T R A C T

The deviance information criterion (DIC) has been widely used for Bayesian model comparison. However, recent studies have cautioned against the use of certain variants of the DIC for comparing latent variable models. For example, it has been argued that the conditional DIC – based on the conditional likelihood obtained by conditioning on the latent variables – is sensitive to transformations of latent variables and distributions. Further, in a Monte Carlo study that compares various Poisson models, the conditional DIC almost always prefers an incorrect model. In contrast, the observed-data DIC – calculated using the observed-data likelihood obtained by integrating out the latent variables – seems to perform well. It is also the case that the conditional DIC based on the maximum a posteriori (MAP) estimate might not even exist, whereas the observed-data DIC does not suffer from this problem. In view of these considerations, fast algorithms for computing the observed-data DIC for a variety of high-dimensional latent variable models are developed. Through three empirical applications it is demonstrated that the observed-data DICs have much smaller numerical standard errors compared to the conditional DICs. The corresponding MATLAB code is available upon request.

1. Introduction

Hypothesis testing, and more generally model comparison, has long been an important problem in statistics and econometrics. Bayesian model comparison has traditionally been performed using the Bayes factor, which is defined to be the ratio of the marginal likelihoods of the two competing models. This model comparison criterion has a natural interpretation and is often easy to compute for a wide range of simple models (see, e.g., Kroese and Chan, 2014, pp. 251–254). However, the development of Markov chain Monte Carlo (MCMC) methods has made it possible to fit increasingly flexible and complex models, and estimating the marginal likelihoods of these typically high-dimensional models is often difficult. In fact, there is a vast and growing literature on marginal likelihood estimation using MCMC methods (see, e.g., Gelfand and Dey, 1994; Chib and Jeliazkov, 2001; Friel and Pettitt, 2008; Bauwens and Rombouts, 2012; Chan and Eisenstat, forthcoming, among many others). Despite these recent advances, computing the marginal likelihood remains a difficult problem in practice, which often involves nontrivial programming efforts and heavy computation. In addition, the values of the Bayes factor are often found to be sensitive to the choice of prior distributions.

* Correspondence to: Research School of Economics, ANU College of Business and Economics, LF Crisp Building 26, The Australian National University, Canberra ACT 0200, Australia. Tel.: +61 2 612 57358; fax: +61 2 612 50182.

E-mail addresses: joshuacc.chan@gmail.com, joshua.chan@anu.edu.au (J.C.C. Chan).

http://dx.doi.org/10.1016/j.csda.2014.07.018
0167-9473/

These considerations have motivated the search for alternative model selection criteria. In particular, since Spiegelhalter et al. (2002) introduced the concept in their seminal paper, the deviance information criterion (DIC) has been widely used for Bayesian model comparison. Its popularity is further enhanced by the introduction of a number of alternative definitions of the DIC – many of them easy to compute – for latent variable models in Celeux et al. (2006). In addition, DIC computation is implemented in standard software packages, including WinBUGS. The DIC has been successfully applied to a wide variety of applications, such as comparing various stochastic volatility models in finance (see, e.g., Berg et al., 2004; Abanto-Valle et al., 2010; Wang et al., 2013), testing functional forms in energy modeling (see, e.g., Xiao et al., 2007), and discriminating between competing models for inflation as well as other macroeconomic time series (see, e.g., Lopes and Salazar, 2006; Chen et al., 2012; Mumtaz and Surico, 2012). A Monte Carlo study comparing the DIC with other Bayesian model selection criteria can be found in Ward (2008).

Nevertheless, some recent studies have cautioned against the use of the DIC for comparing latent variable models. For instance, Li et al. (2012) argue that the DIC should not be used with data augmentation, as the complete-data likelihood of the augmented data is nonregular and hence invalidates the standard asymptotic arguments that are needed to justify the DIC. Moreover, the DIC based on the complete-data likelihood is sensitive to transformations of latent variables and distributional representations. In the context of comparing Poisson models, Millar (2009) provides a Monte Carlo study which shows that the DIC based on the conditional likelihood – obtained by conditioning on the latent variables – almost always prefers the Poisson-gamma model instead of the Poisson-lognormal model, even when data are simulated from the latter. The author concludes that "the DIC is a potentially dangerous tool in the present context". In contrast, he shows that the DIC calculated using the integrated likelihood – obtained by integrating out the latent variables – seems to perform well. This result is not surprising since standard asymptotic arguments for justifying the DIC apply to the DIC based on the integrated likelihood. However, evaluation of the integrated likelihood is typically time-consuming, which is the main reason why it is rarely used in applied work. We take a first step to address these issues by proposing fast methods for computing the DIC based on the integrated likelihood for a variety of high-dimensional latent variable models.

More specifically, the contribution of this paper is twofold. Firstly, we provide analytical expressions for the integrated likelihoods under three popular families of latent variable models: factor models, linear Gaussian state space models and semiparametric models. To evaluate these integrated likelihoods, we draw on recent advances in sparse matrix algorithms, and the computational details are carefully discussed. Secondly, we document the differences in variability of the DICs computed using the complete-data likelihood, the conditional likelihood and the integrated likelihood in three empirical examples. We show that the DICs based on the complete-data and conditional likelihoods generally have large numerical standard errors. On the other hand, the DICs based on the integrated likelihoods are more accurately estimated. This result is intuitive since integrating out the high-dimensional latent variables is expected to reduce the variance in Monte Carlo simulation. Our results provide another practical reason for why DICs based on conditional and complete-data likelihoods should not be used.

The rest of this paper is organized as follows. In Section 2 we introduce the concept of deviance and several definitions of the DIC. Section 3 discusses fast algorithms for computing the DIC based on the integrated likelihood for three classes of latent variable models. In Section 4, the proposed methods are illustrated via three empirical applications, involving returns on stock portfolios, US macroeconomic time series and female body mass index and wages.

2. Deviance information criterion

In complex hierarchical models, basic concepts like parameters and their dimension are not always clear and they may take several equally acceptable definitions. In their seminal paper, Spiegelhalter et al. (2002) introduce the concept of *effective number of parameters* and develop the theory of *deviance information criterion* (DIC) for model comparison. The model selection criterion is based on the *deviance*, which is defined as

$$D(\boldsymbol{\theta}) = -2\log f(\mathbf{y} \mid \boldsymbol{\theta}) + 2\log h(\mathbf{y}),$$

where $f(\mathbf{y} \mid \boldsymbol{\theta})$ is the likelihood function of the parametric model and $h(\mathbf{y})$ is some fully specified standardizing term that is a function of the data alone. Then the effective number of parameters p_D is defined as

$$p_D = \overline{D(\boldsymbol{\theta})} - D(\widetilde{\boldsymbol{\theta}}),$$

where

$$\overline{D(\boldsymbol{\theta})} = -2\mathbb{E}_{\boldsymbol{\theta}}[\log f(\mathbf{y} \mid \boldsymbol{\theta}) \mid \mathbf{y}] + 2\log h(\mathbf{y})$$

is the posterior mean deviance and $\widetilde{\boldsymbol{\theta}}$ is an estimate of $\boldsymbol{\theta}$, which is typically taken as the posterior mean or mode. Then, the deviance information criterion is defined as

$$\mathrm{DIC} = \overline{D(\boldsymbol{\theta})} + p_D.$$

The posterior mean deviance can be used as a Bayesian measure of model fit or adequacy. Hence, the deviance information criterion, which is the sum of the posterior mean deviance and the effective number of parameters, can be viewed as a trade-off between model adequacy and complexity. For model comparison, we set $h(\mathbf{y}) = 1$ for all models. Therefore, the

Please cite this article in press as: Chan, J.C.C., Grant, A.L., Fast computation of the deviance information criterion for latent variable models. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.07.018