A note on Reversible Jump Markov Chain Monte Carlo

Hedibert Freitas Lopes
Graduate School of Business
The University of Chicago
5807 South Woodlawn Avenue
Chicago, Illinois 60637

February, 1st 2006

1 Introduction

Bayesian model comparison is commonly performed by computing posterior model probabilities. More precisely, suppose that the competing models can be enumerable and are represented by the set $\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, \ldots\}$. Under model $\mathcal{M}_k$, the posterior distribution is

$$p(\theta_k|y,k) = \frac{p(y|\theta_k,k)p(\theta_k|k)}{p(y|k)} \quad (1)$$

where $p(y|\theta_k,k)$ and $p(\theta_k|k)$ represent the probability model and the prior distribution of the parameters of model $\mathcal{M}_k$, respectively. The computation of predictive densities

$$p(y|k) = \int_{\Theta_k} p(y|\theta_k,k)p(\theta_k|k)d\theta_k$$

lies at the heart of the model selection and comparison problem since they are used to compute the Bayes factors and, consequently, posterior odds ratios,

$$\underbrace{\frac{Pr(k|y)}{Pr(k'|y)}}_{\text{Posterior Odds}} = \underbrace{\frac{p(y|k)}{p(y|k')}}_{\text{Bayes Factor}} \times \underbrace{\frac{Pr(k)}{Pr(k')}}_{\text{Prior odds}}$$

A variety of methods are available for computing these marginal data density values – often referred to as the *normalising constant* problem. Some are specific to analysis based on MCMC methods for each individual sub-model, and some are generic and based on analytic and asymptotic arguments. A review of some standard methods appears in Kass and Raftery (1995), where the connections between various methods of approximating Bayes' factors using combinations of analytic and asymptotic arguments are explored. We list here just a few of them: (i) the so-called *candidate formula* (Chib, 1995), (ii) the *harmonic mean estimator* (Newton and Raftery, 1994), (iii) *Gelfand and Dey's estimator* (Gelfand and Dey, 1994), (iv) the *Laplace-Metropolis* estimator (Lewis and Raftery, 1997), and various novel approaches based on the recent innovative developments in (v) *bridge sampling* (Meng and Wong, 1996). Additional useful references in this general area include, for example, Gilks *et al.* (1996), DiCiccio *et al.* (1997) and Godsill (1998), which study comparisons and connections between some of the various methods just referenced. Gamerman and Lopes (2006) provide detailed description and worked examples that utilize the estimators listed here as well as several others that appeared more recently.

We also introduce the *Reversible Jump Markov Chain Monte Carlo* (hereafter RJMCMC, see Green (1995)) algorithm for moving between models with different numbers of factors. RJMCMC approaches avoid the need for computing marginal data densities by treating the number of factors as a parameter, but require ingenuity in designing appropriate jumping rules to produce computationally efficient and theoretically effective methods.

2 Reversible jump MCMC

We present the Reversible Jump algorithm as introduced in Green (1995). Among many others, Richardson and Green (1997), Dellaportas *et al.* (1998), Denison *et al.* (1997), Troughton and Godsill (1997), Insua and Müller (1998), Barbieri and O'Hagan (1996) and Huerta and West (1999) applied the reversible jump sampler to mixture models, variable selection, curve fitting, autoregressive models, neural networks, ARMA models and component structure in AR models, respectively. We also explore its relationship to Carlin and Chib's (1995) pseudo-prior method. Particular attention is given to the *Metropolized Carlin-Chib* algorithm simultaneously introduced by Dellaportas *et al.* (1998) and Godsill (1998). The results presented here are mainly based on the developments from Dellaportas *et al.* (1998) and Godsill (1998). Additional overview and/or further extensions can be found in Chen *et al.* (2000), Section 9.5, and Gamerman and Lopes (2006), Section 7.3.

The RJMCMC algorithm

Suppose that the competing models can be enumerable and are represented by the set $\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, \ldots\}$. Under model $\mathcal{M}_k$, the posterior distribution is

$$p(\theta_k|y, k) \propto p(y|\theta_k, k)p(\theta_k|k) \tag{2}$$

where $p(y|\theta_k, k)$ and $p(\theta_k|k)$ represent the probability model and the prior distribution of the parameters of model $\mathcal{M}_k$, respectively. Then,

$$p(\theta_k, k|y) \propto p(k)p(\theta_k|k, y) \tag{3}$$

The RJMCMC methods involve Metropolis-Hastings type algorithms that move a simulation analysis between models defined by (k, θ_k) to $(k', \theta_{k'})$ with different defining dimensions k and k'. The resulting Markov chain simulations jump between such distinct models and form samples from the joint distribution $p(\theta_k, k)$. The algorithm are designed to be reversible so as to maintain detailed balance of a irreducible and aperiodic chain that converges to the correct target measure. Further details of the general methodology and ideas can be found in Green (1995).

Here we present the algorithm in a schematic form. If the current state of the Markov chain is (k, θ_k), then one possible version of the RJMCMC algorithm is as follows:

Step 1. Propose a visit to model $\mathcal{M}_{k'}$ with probability $J(k \to k')$.

Step 2. Sample u from a proposal density $q(u|\theta_k, k, k')$.

Step 3. Set $(\theta_{k'}, u') = g_{k,k'}(\theta_k, u)$, where $g_{k,k'}(\cdot)$ is a bijection between (θ_k, u) and $(\theta_{k'}, u')$, where u and u' play the role of matching the dimensions of both vectors.

Step 4. The acceptance probability of the new model, $(\theta_{k'}, k')$ can be calculated as the minimum between one and

$$\underbrace{\frac{p(y|\theta_{k'}, k')p(\theta_{k'})p(k')}{p(y|\theta_k, k)p(\theta_k)p(k)}}_{\text{model ratio}} \underbrace{\frac{J(k' \to k)q(u'|\theta_{k'}, k', k)}{J(k \to k')q(u|\theta_k, k, k')} \left| \frac{\partial g_{k,k'}(\theta_k, u)}{\partial(\theta_k, u)} \right|}_{\text{proposal ratio}} \tag{4}$$

Looping through steps 1-4 generates a sample $\{k_l, l = 1, \dots, L\}$ for the model indicators and $Pr(k|y)$ can be estimated by

$$\widehat{P}r(k|y) = \frac{1}{L}\sum_{l=1}^{L} \mathbf{1}_k(k_l) \tag{5}$$

where $\mathbf{1}_k(k_l) = 1$ if $k = k_l$ and zero otherwise. The choice of the model proposal probabilities, $J(k \to k')$, and the proposal densities, $q(u|k, \theta_k, k')$, must be cautiously made, especially in highly parameterized problems.

Independent sampler: If all parameters of the proposed model are generated from the proposal distribution, then $(\theta_{k'}, u') = (u, \theta_k)$ and the Jacobian in (4) is one.

Standard Metropolis-Hastings: When the proposed model k' equals the current model k, the loop through steps 1-4 corresponds to the traditional Metropolis-Hastings algorithm (Metropolis *et al.*, 1995; Hastings, 1970; Peskun, 1973; Chib and Greenberg, 1995).

Posterior densities as proposal densities: If $p(\theta_k|y,k)$ is available in close form for each model $\mathcal{M}_k$, then $q(u'|\theta_{k'},k',k) = p(\theta_k|y,k)$ and the acceptance probability (equation 4) reduces to the minimum between one and

$$\frac{p(k')p(y|k')}{p(k)p(y|k)} \frac{J(k' \to k)}{J(k \to k')} \tag{6}$$

using the fact that $p(y|\theta_k,k)p(\theta_k)p(k) = p(\theta_k,k|y)p(y|k)$. Again, the Jacobian equals one. The predictive density or normalizing constant, $p(y|k)$, is also available in close form. Moreover, if $J(k' \to k) = J(k \to k')$, the acceptance probability is the minimum between one and the posterior odds ratio from model $\mathcal{M}_{k'}$ to model $\mathcal{M}_k$, that is the move is automatically accepted when model $\mathcal{M}_{k'}$ has higher posterior probability than model $\mathcal{M}_k$; otherwise the posterior odds ratio determines how likely is to move to a lower posterior probability model.

Metropolized Carlin and Chib's algorithm

Let $\Theta = (\theta_k, \theta_{-k})$ be the vector containing the parameters of all competing models. Then the joint posterior of (Θ, k) is

$$p(\Theta,k|y) \propto p(k)p(y|\theta_k,k)p(\theta_k|k)p(\theta_{-k}|\theta_k,k) \tag{7}$$

where $p(\theta_{-k}|\theta_k,k)$ are *pseudo-prior* densities (Carlin and Chib, 1995). Carlin and Chib propose a Gibbs sampler where the full posterior conditional distributions are

$$p(\theta_k|y,k,\theta_{-k}) \propto \begin{cases} p(y|\theta_k,k)p(\theta_k|k) & \text{if } k = k' \\ p(\theta_k|k') & \text{if } k = k' \end{cases} \tag{8}$$

and

$$p(k|\Theta,y) \propto p(y|\theta_k,k)p(k) \prod_{m \in \mathcal{M}} p(\theta_m|k) \tag{9}$$

Notice that the pseudo-prior densities and the RJMCMC's proposal densities have similar functions. As a matter of fact, Carlin and Chib suggest using pseudo-prior distributions that are close to the posterior distributions within each competing model.

The main problem with Carlin and Chib's Gibbs sampler is the need of evaluating and drawing from the pseudo-prior distributions at each iteration of the MCMC scheme. This problem can be overwhelmingly exacerbated in large situations where the number of competing models is relatively large (See Clyde, 1999, for applications and discussions in variable selection in regression models).

To overcome this last problem Dellaportas *et al.* and Godsill (1998) proposes "Metropolizing" Carlin and Chib's Gibbs sampler. If the current state of the Markov chain is at (θ_k, k), then they suggest proposing and accepting/rejecting a move to a new model in the following way:

Step 1. Propose a new model $\mathcal{M}_{k'}$ with probability $J(k \to k')$.

Step 2. Generate $\theta_{k'}$ from the pseudo-prior $p(\theta_{k'}|k)$.

Step 3. The acceptance probability of the new model, k' can be calculated as the minimum between one and

$$\frac{p(y|\theta_{k'}, k')p(k')J(k' \to k)\prod_{m\in\mathcal{M}} p(\theta_m|k')}{p(y|\theta_k, k)p(k)J(k \to k')\prod_{m\in\mathcal{M}} p(\theta_m|k)}$$

which can be simplified to

$$\frac{p(y|\theta_{k'}, k')p(k')J(k' \to k)p(\theta_{k'}|k')p(\theta_k|k')}{p(y|\theta_k, k)p(k)J(k \to k')p(\theta_k|k)p(\theta_{k'}|k)} \tag{10}$$

since the other pseudo-prior densities cancel out.

Once again, if $p(\theta_k|y, k)$ is available in close form for each model $\mathcal{M}_k$, and $p(\theta_k|k') = p(\theta_k|y, k)$, then the acceptance probability in (10) reduces to (6). As we have mentioned earlier the pseudo-prior densities and the RJMCMC's proposal densities have similar functions and the closer their are to the competing models' posterior probabilities the better the sampler mixing.

Bibliography

Barbieri, M.M. and O'Hagan, A. (1996) A reversible jump MCMC sampler for Bayesian analysis of ARMA time series. Technical report. Dipartimento di Statistica,Probabilità e Statistiche Applicate, Università La Sapienza,Roma.

Carlin, B.P. and Chib, S. (1995) Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society,* B, **57,** 473–484.

Chen, M.-H., Shao, Q.-M. and Ibrahim, J.G. (2000) *Monte Carlo methods in Bayesian computation.* New York: Springer-Verlag.

Chib, S. (1995) Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, **90,** 1313–1321.

Chib, S. and Greenberg, E. (1995) Understanding the Metropolis-Hastings algorithm. *The American Statistician*, **49,** 327–335.

Clyde, M. (1999) Bayesian model averaging and model search strategies (with discussion). In *Bayesian Statistics 6* (eds J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith), pp. 157–185. John Wiley.

Dellaportas, P., Forster, J.J. and Ntzoufras, I. (1998) On Bayesian model and variable selection using MCMC. Working Paper. Athens University of Economics and Business.

Denison, D.G.T., Mallick, B.K. and Smith, A.F.M. (1997) Automatic Bayesian curve fitting. Working Paper. Imperial College.

DiCiccio, T.J., Kass, R.A., Raftery, A.E. and Wasserman, L. (1997) Computing Bayes' factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association*, **92,** 903–915.

Gamerman, D. and Lopes, H. F. (2006) *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference (2nd edition).* Chapman and Hall.

Gelfand, A.E. and Dey, D. (1994) Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society, Ser.* B, **56,** 501–514.

Gilks, W. R., Richardson, S. and Spiegelhalter, D.J. (1996) *Markov Chain Monte Carlo in Practice.* Chapman and Hall.

Godsill, S.J. (1998) On the relationship between MCMC model uncertainty methods. Technical Report CUED/F-INFENG/TR.305. Cambridge University Engineering Department.

Green, P. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82,** 711–732.

Hastings, W.K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57,** 97–109.

Huerta, G. and West, M. (1999) Priors and component structures in autoregressive time series models. *Journal of the Royal Statistical Society-Series* B, **61,** 881–899.

Insua, D.R. and Müller, P. (1998) Feedforward neural networks for nonpametric regression. Discussion Paper 98–02. Institute of Statistics and Decision Sciences, Duke University.

Kass, R.A. and Raftery, A.E. (1995) Bayes' factor. *Journal of the American Statistical Association*, **90,** 773–795.

Lewis, S.M. and Raftery, A.E. (1997) Estimating Bayes' factors via posterior simulation with the Laplace-Metropolis estimator. *Journal of the American Statistical Association*, **92,** 648–655.

Meng, X.L. and Wong, W.H. (1996) Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica*, **6,** 831–860.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N. and Teller, A.H. (1995) Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21,** 1087–1092.

Newton, M.A. and Raftery, A.E. (1994) Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society, Ser.* B, **56,** 3–48.

Peskun, P.H. (1973) Optimum Monte-Carlo sampling using Markov chains. *Biometrika*, **60,** 607–612.

Richardson, S. and Green, P. (1997) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination (with discussion). *Journal of the Royal Statistical Society - Series* B, **59,** 731–758.

Troughton, P.T. and Godsill, S.J. (1997) Reversible jump sampler for autoregressive time series, employing full conditionals to achieve efficient model space moves. Technical Report CUED/F-INFENG/TR.304. Cambridge University Engineering Department.