On Bayesian Model and Variable Selection Using MCMC

Petros Dellaportas[1], Jonathan J. Forster[2] and Ioannis Ntzoufras[3]

6 January 1998

SUMMARY

Several MCMC methods have been proposed for estimating probabilities of models and associated 'model-averaged' posterior distributions in the presence of model uncertainty. We discuss, compare, develop and illustrate several of these methods, focussing on connections between them.

Keywords: Gibbs Sampler; Metropolis-Hastings; Reversible Jump

1 Introduction

A Bayesian approach to model selection proceeds as follows. Suppose that the data $\boldsymbol{y}$ are considered to have been generated by a model m, one of a set M of competing models. Each model specifies the distribution of $\boldsymbol{Y}$, $f(\boldsymbol{y}|m,\boldsymbol{\beta}_m)$ apart from an unknown parameter vector $\boldsymbol{\beta}_m \in B_m$, where B_m is the set of all possible values for the coefficients of model m. If $f(m)$ is the prior probability of model m, then the posterior probability is given by

$$f(m|\boldsymbol{y}) \;=\; \frac{f(m)f(\boldsymbol{y}|m)}{\sum\limits_{m\in\mathcal{M}} f(m)f(\boldsymbol{y}|m)}, \qquad m \in M \tag{1}$$

where $f(\boldsymbol{y}|m)$ is the marginal likelihood calculated using $f(\boldsymbol{y}|m) = \int f(\boldsymbol{y}|m,\boldsymbol{\beta}_m)f(\boldsymbol{\beta}_m|m)d\boldsymbol{\beta}_m$ and $f(\boldsymbol{\beta}_m|m)$ is the conditional prior distribution of $\boldsymbol{\beta}_m$, the model parameters for model m.

This integral is only analytically tractable in certain restricted examples. A further problem is that the size of the set of possible models M may be so great that calculation or approximation of $f(\boldsymbol{y}|m)$ for all $m \in M$ becomes infeasible. Therefore MCMC methods which generate observations from the joint posterior distribution of $(m,\boldsymbol{\beta}_m)$ have recently become popular for estimating $f(m)$ and $f(\boldsymbol{\beta}_m|m)$. We focus on the methods of Green ('reversible jump', 1995) and Carlin and Chib (1995), and describe a connection between them.

We also consider 'variable selection' problems where the models under consideration can be naturally represented by a set of binary indicator variables so that $M \subseteq \{0,1\}^p$, where p is the total possible number of variables. We introduce a modification of Carlin and Chib's method for variable selection problems, which can be more efficient in certain examples.

2 MCMC Model Selection Methods

2.1 Reversible Jump

Reversible jump MCMC (Green, 1995) is a flexible MCMC sampling strategy for generating observations from the joint posterior distribution $f(m,\boldsymbol{\beta}_m|\boldsymbol{y})$. The method is based on creating a Markov chain which

[1]Department of Statistics, Athens University of Economics and Business, Greece, email:petros@aueb.gr

[2]Department of Mathematics, University of Southampton, UK, email:jjf@maths.soton.ac.uk

[3]Department of Statistics, Athens University of Economics and Business, Greece, email:jbn@stat-athens.aueb.gr

can 'jump' between models with parameter spaces of different dimension, while retaining detailed balance which ensures the correct limiting distribution, provided the chain is irreducible and aperiodic.

Suppose that the current state of the Markov chain is $(m, \boldsymbol{\beta}_m)$, where $\boldsymbol{\beta}_m$ has dimension $d(\boldsymbol{\beta}_m)$, then one version of the procedure is as follows:

- Propose a new model m' with probability $j(m, m')$.
- Generate $\boldsymbol{u}$ from a specified proposal density $q(\boldsymbol{u}|\boldsymbol{\beta}_m, m, m')$.
- Set $(\boldsymbol{\beta}'_{m'}, \boldsymbol{u}') = g_{m,m'}(\boldsymbol{\beta}_m, \boldsymbol{u})$ where $g_{m,m'}$ is a specified invertible function. Hence $d(\boldsymbol{\beta}_m) + d(\boldsymbol{u}) = d(\boldsymbol{\beta}_{m'}) + d(\boldsymbol{u}')$. Note that $g_{m',m} = g^{-1}_{m,m'}$.
- Accept the proposed move to model m' with probability

$$\alpha = \min\left(1, \frac{f(\boldsymbol{y}|m', \boldsymbol{\beta}'_{m'})f(\boldsymbol{\beta}'_{m'}|m')f(m')j(m', m), q(\boldsymbol{u}'|\boldsymbol{\beta}_m, m', m)}{f(\boldsymbol{y}|m, \boldsymbol{\beta}_m)f(\boldsymbol{\beta}_m|m)f(m)j(m, m')q(\boldsymbol{u}|\boldsymbol{\beta}_m, m, m')} \left|\frac{\partial g(\boldsymbol{\beta}_m, \boldsymbol{u})}{\partial(\boldsymbol{\beta}_m, \boldsymbol{u})}\right|\right). \tag{2}$$

There are many variations or simpler versions of reversible jump that can be applied in specific model selection problems. In particular, if all parameters of the proposed model are generated from a proposal distribution, then $(\boldsymbol{\beta}'_{m'}, \boldsymbol{u}') = (\boldsymbol{u}, \boldsymbol{\beta}_m)$ with $d(\boldsymbol{\beta}_m) = d(\boldsymbol{u}')$ and $d(\boldsymbol{\beta}_{m'}) = d(\boldsymbol{u})$, and the jacobian term in (2) is one. This version of reversible jump could be used for jumping between models for which no appropriate parameter transformation exists. Where models m and m' may be described as nested, then there may be an extremely natural proposal distribution and transformation function $g_{m,m'}$ (may be the identity function) such that $d(\boldsymbol{u}') = 0$ and $\boldsymbol{\beta}_{m'} = g_{m,m'}(\boldsymbol{\beta}_m, \boldsymbol{u})$. See, for example, Dellaportas and Forster (1996). Finally, if $m' = m$, then the move is a standard Metropolis-Hastings step.

2.2 Carlin and Chib's Method

Carlin and Chib (1995) proposed using a Gibbs sampler to generate from the posterior distribution $f(m, \boldsymbol{\beta}_m)$. In order to do this, it is required to consider a Markov chain of realisations of $(m, \boldsymbol{\beta}_k : k \in M)$. Therefore, a prior distribution for $(m, \boldsymbol{\beta}_k : k \in M)$ is no longer completely specified by $f(m)$ and $f(\boldsymbol{\beta}_m|m)$, so Carlin and Chib proposed the use of pseudopriors or linking densities $f(\boldsymbol{\beta}_k|m \neq k)$, $k \in M$.

The full posterior conditional distributions are

$$f(\boldsymbol{\beta}_k|\boldsymbol{y},\ \{\boldsymbol{\beta}_l : l \neq k\}, m) \propto \begin{cases} f(\boldsymbol{y}|\boldsymbol{\beta}_m, m)f(\boldsymbol{\beta}_m|m) & k = m \\ f(\boldsymbol{\beta}_k|k \neq m) & k \neq m \end{cases} \tag{3}$$

and

$$f(m|\{\boldsymbol{\beta}_k : k \in M\}, \boldsymbol{y}) = \frac{A_m}{\sum_{k \in M} A_k} \tag{4}$$

where

$$A_m = f(\boldsymbol{y}|\boldsymbol{\beta}_m, m) \prod_{l \in M} [f(\boldsymbol{\beta}_l|m)]f(m).$$

When $k = m$, $\boldsymbol{\beta}_k$ are generated from the conditional posterior distribution $f(\boldsymbol{\beta}_m|\boldsymbol{y}, m)$, and when $k \neq m$ from the corresponding pseudoprior, $f(\boldsymbol{\beta}_k|m)$. The model indicator m is generated as a discrete random variable.

The main drawback of this method is the unavoidable specification of, and generation from, many pseudoprior distributionss. Carlin and Chib (1995) point out that, as the pseudopriors do not enter the marginal posterior distributions $f(m, \boldsymbol{\beta}_m)$ of interest, they should be chosen to make the method efficient. However, generation from $|M|-1$ pseudopriors at every cycle of the Gibbs sampler is still required, and this is computationally demanding.

2.3 Reversible Jump and 'Metropolised' Carlin and Chib

The Gibbs sampler proposed by Carlin and Chib (1995) requires the calculation of all A_k in the denominator of (4). An alternative approach is a hybrid Gibbs/Metropolis strategy, where the 'model selection' step is not based on the full conditional, but on a proposal for a move to model m', followed by acceptance or rejection of this proposal. If the current state is model m and we propose model m' with probability $j(m, m')$, then the acceptance probability is given by

$$\begin{aligned} \alpha &= \min\left(1, \frac{A_{m'} j(m', m)}{A_m j(m, m')}\right) \\ &= \min\left(1, \frac{f(\boldsymbol{y}|\boldsymbol{\beta}_{m'}, m') f(\boldsymbol{\beta}_{m'}|m') f(\boldsymbol{\beta}_m|m') f(m') j(m', m)}{f(\boldsymbol{y}|\boldsymbol{\beta}_m, m) f(\boldsymbol{\beta}_m|m) f(\boldsymbol{\beta}_{m'}|m) f(m) j(m, m')}\right) \end{aligned} \tag{5}$$

as all other pseudopriors cancel.

Note that when we are in model m and we propose model m', we require only values of $\boldsymbol{\beta}_m$ and $\boldsymbol{\beta}_{m'}$ to calculate α in (5). Furthermore, we are assuming that model m' is proposed with probability $j(m, m')$, independently of the values of any model parameters. Therefore if we reverse the order of sampling from j and the full conditional distributions for $\boldsymbol{\beta}_k$ in (3), there is no need to sample from any pseudopriors other than that for m'. The method now consists of the following three steps:

- Propose a new model m' with probability $j(m, m')$.
- Generate $\boldsymbol{\beta}_{m'}$ from the pseudoprior $f(\boldsymbol{\beta}_{m'}|m \neq m')$.
- Accept the proposed move to model m' with probability α given by (5).

It is straightforward to see that by a simple modification ('Metropolising' the model selection step), Carlin and Chib's method becomes a special case of reversible jump with $(\boldsymbol{\beta}'_{m'}, \boldsymbol{u}') = (\boldsymbol{u}, \boldsymbol{\beta}_m)$ and proposal $q(\boldsymbol{u}|\boldsymbol{\beta}_m, m, m')$ replaced by pseudoprior $f(\boldsymbol{\beta}_{m'}|m \neq m')$.

2.4 Using Posterior Distributions as Proposals

Suppose that, for each m, the posterior density $f(\boldsymbol{\beta}_m|m, \boldsymbol{y})$ is available, including the normalising constant which is the marginal likelihood $f(\boldsymbol{y}|m)$. If this distribution is used as a pseudoprior then the acceptance probability in (5) is given by

$$\begin{aligned} \alpha &= min\left(1, \frac{f(\boldsymbol{y}|m', \boldsymbol{\beta}'_{m'}) f(\boldsymbol{\beta}'_{m'}|m') f(m') j(m', m) f(\boldsymbol{\beta}_m|m, \boldsymbol{y})}{f(\boldsymbol{y}|m, \boldsymbol{\beta}_m) f(\boldsymbol{\beta}_m|m) f(m) j(m, m') f(\boldsymbol{\beta}'_{m'}|m', \boldsymbol{y})}\right) \\ &= min\left(1, B_{m'm} \frac{f(m') j(m', m)}{f(m) j(m, m')}\right) \end{aligned}$$

where $B_{m'm}$ is the Bayes factor of model m' against model m. In practice, we can not usually calculate $B_{m'm}$. In the special case where models are decomposable graphical models, Madigan and York (1995) used exactly this approach, which they called MC^3. Here there is no need to generate the model parameters $\boldsymbol{\beta}_m$ as part of the Markov chain. These can be generated separately from the known posterior distributions $f(\boldsymbol{\beta}_m|m, \boldsymbol{y})$ if required.

3 Variable Selection.

Many statistical models may be represented naturally as $(s, \boldsymbol{\gamma}) \in S \times \{0,1\}^p$, where the indicator vector $\boldsymbol{\gamma}$ represents which of the p possible sets of covariates are present in the model and s represents other structural properties of the model. For example, for a generalised linear model s may describe the distribution, link function and variance function, and the linear predictor may be writen as

$$\boldsymbol{\eta} = \sum_{i=1}^{p} \gamma_i \boldsymbol{X}_i \boldsymbol{\beta}_i \tag{6}$$

where $\boldsymbol{X}_i$ is the design matrix and $\boldsymbol{\beta}_i$ the parameter vector related to the ith term. In the following, we restrict consideration to variable selection aspects assuming that s is known, or dealt with in another way and therefore we substitute $\boldsymbol{\gamma}$ for model indicator m. For example, we can apply reversible jump to variable selection by substituting $\boldsymbol{\gamma}$ for m in (2).

Note that in some cases, it is sensible to set $f(\gamma_i|\boldsymbol{\gamma}_{\setminus i}) = f(\gamma_i)$ (where the subscript $\setminus i$ denotes all elements of a vector except the ith) , whereas in other cases (e.g. hierarchical or graphical log-linear models) it is required that $f(\gamma_i|\boldsymbol{\gamma}_{\setminus i})$ depends on $\gamma_{\setminus i}$; see Chipman (1995).

3.1 Gibbs Variable Selection

Here we introduce a modification of Carlin and Chib's method which is appropriate for variable selection, and can be implemented in Gibbs sampler software such as BUGS. Furthermore, it does not require unnecessary generation from pseudopriors.

We specify the prior for $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ as $f(\boldsymbol{\gamma}, \boldsymbol{\beta}) = f(\boldsymbol{\gamma}) f(\boldsymbol{\beta}|\boldsymbol{\gamma})$. If we consider a partition of $\boldsymbol{\beta}$ into $(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\beta}_{\setminus\boldsymbol{\gamma}})$ corresponding to those components of $\boldsymbol{\beta}$ which are included ($\gamma_i = 1$) or not included ($\gamma_i = 0$) in the model, then the prior $f(\boldsymbol{\beta}|\boldsymbol{\gamma})$ may be partitioned into model prior $f(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\boldsymbol{\gamma})$ and pseudoprior $f(\boldsymbol{\beta}_{\setminus\boldsymbol{\gamma}}|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma})$.

The full conditional posterior distributions are given by

$$f(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\boldsymbol{\beta}_{\setminus\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}) f(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\boldsymbol{\gamma}) f(\boldsymbol{\beta}_{\setminus\boldsymbol{\gamma}}|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}) \tag{7}$$

$$f(\boldsymbol{\beta}_{\setminus\boldsymbol{\gamma}}|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \boldsymbol{y}) \propto f(\boldsymbol{\beta}_{\setminus\boldsymbol{\gamma}}|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}) \tag{8}$$

and

$$\frac{f(\gamma_i = 1|\boldsymbol{\gamma}_{\setminus i}, \boldsymbol{\beta}, \boldsymbol{y})}{f(\gamma_i = 0|\boldsymbol{\gamma}_{\setminus i}, \boldsymbol{\beta}, \boldsymbol{y})} = \frac{f(\boldsymbol{y}|\boldsymbol{\beta}, \gamma_i = 1, \boldsymbol{\gamma}_{\setminus i})}{f(\boldsymbol{y}|\boldsymbol{\beta}, \gamma_i = 0, \boldsymbol{\gamma}_{\setminus i})} \, \frac{f(\boldsymbol{\beta}|\gamma_i = 1, \boldsymbol{\gamma}_{\setminus i})}{f(\boldsymbol{\beta}|\gamma_i = 0, \boldsymbol{\gamma}_{\setminus i})} \, \frac{f(\gamma_i = 1, \boldsymbol{\gamma}_{\setminus i})}{f(\gamma_i = 0, \boldsymbol{\gamma}_{\setminus i})} \tag{9}$$

Note that (7) seems less natural than (3) as $f(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\boldsymbol{\beta}_{\setminus\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \boldsymbol{y})$ may depend on $\boldsymbol{\beta}_{\setminus\boldsymbol{\gamma}}$. One way of avoiding this is to assume prior conditional independence of $\boldsymbol{\beta}_i$ terms given $\boldsymbol{\gamma}$, in which case $f(\boldsymbol{\beta}_{\setminus\boldsymbol{\gamma}}|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma})$ vanishes from (7). This is a restrictive assumption but may be realistic when priors are intended to be noninformative, particularly if the columns of different $\boldsymbol{X}_i$ in (6) are orthogonal to each other. Then, each prior for $\boldsymbol{\beta}_i|\boldsymbol{\gamma}$ consists of a mixture of two densities. The first, $f(\boldsymbol{\beta}_i|\gamma_i = 1, \boldsymbol{\gamma}_{\setminus i})$, is the true prior for the parameter whereas the second, $f(\boldsymbol{\beta}_i|\gamma_i = 0, \boldsymbol{\gamma}_{\setminus i})$, is a pseudoprior. Another way to make (7) usable is to define priors $f(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\boldsymbol{\beta}_{\setminus\boldsymbol{\gamma}}, \boldsymbol{\gamma}) = f(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\boldsymbol{\gamma})$ so that we need to calculate the true prior conditional density as $f(\boldsymbol{\beta}_i|\boldsymbol{\beta}_{\setminus i}, \gamma_i = 1, \boldsymbol{\gamma}_{\setminus i})$.

Expression (9) is similar to equivalent expressions in other proposed variable selection methods. George and McCulloch (1993) propose a 'Stochastic Search Variable Selection' strategy which assumes the maximal model throughout, but constrains $\boldsymbol{\beta}_i$ parameters to be close to zero when $\gamma_i = 0$. In this situation, $f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma})$

is independent of γ and so the first ratio on the right hand side of (9) vanishes. Kuo and Mallick (1997) propose a similar approach to the above but use a prior distribution for (γ, β) with β independent of γ. Then, the second term on the right hand side of (9) vanishes.

Carlin and Chib's method involves a single model indicator parameter. Therefore at each iteration of the Gibbs sampler, all parameters of all models are generated from either posterior distribution or pseudoprior, and the model selection step allows a simultaneous change of all γ_is. For Gibbs variable selection, an observation of γ is generated following generation of all β_k from posterior distributions or pseudopriors. This procedure will generally involve generating from p conditional distributions for β_k, a much smaller burden than required for Carlin and Chib's method. Furthermore, it would seem to be more efficient to generate pairs of (β_k, γ_k) succesively, possibly by a random scan, so that more local moves in model space are attempted.

Clearly, moves between models m (γ) and m' (γ') may also be based on a Metropolis step, as was suggested in Section 2.3. Then the pseudopriors may be thought of as part of the proposal density for parameters which are present in one model but not in the other. This highlights a drawback with the variable selection approaches discussed in this section, namely that parameters which are 'common' to both models remain unchanged, and therefore the procedure will not be efficient unless posterior distributions for such parameters are similar under both models.

4 A Logistic Regression Example

To illustrate and compare the different MCMC model choice methods, we consider a dataset analysed by Healy (1988). The data, presented in Table 1, reflect the relationship between the number of survivals, the patient condition (more or less severe) and the received treatment (antitoxin medication or not). We wish to

	Antitoxin	Death	Survivals
More Severe	Yes	15	6
	No	22	4
Less Severe	Yes	5	15
	No	7	5

Table 1: Logistic Regression Example Dataset.

select one of the 5 possible nested logistic regression models with response variable the number of survivals and explanatory factors the patient condition and the received treatment. The full model is given by

$$Y_{jl} \sim Bin(n_{jl}, p_{jl}), \quad , log(\frac{p_{jl}}{1-p_{jl}}) = \mu + a_j + b_l + (ab)_{jl}, \quad j, l = 1, 2$$

where Y_{jl}, n_{jl} and p_{jl} are the number of survivals, the total number of patients and the probability of survival under j level of severity and l treatment respectivelly; μ, a_j, b_l and $(ab)_{jl}$ are the model parameters corresponding to the constant term, j level of severity, l level of treatment, and interaction of j severity and l treatment.

We consider, for our illustration, the reversible jump, the metropolised version of Carlin and Chib's method presented in section 2.3, the Gibbs variable selection introduced in section 3.1, and the Kuo and

Mallick (1997) method. A rough guideline for our comparisons is an approximation to Bayes factor B_{10} of model m_1 against model m_0 based on Bayes information criterion (BIC)

$$-2logB_{10} \simeq -2log(\frac{f(\boldsymbol{y}|\hat{\boldsymbol{\beta}}_{m_1}, m_1)}{f(\boldsymbol{y}|\hat{\boldsymbol{\beta}}_{m_0}, m_0)}) + (d(\boldsymbol{\beta_{m_1}}) - d(\boldsymbol{\beta_{m_0}}))log(n) \tag{10}$$

where $\hat{\boldsymbol{\beta}}_m$ denotes the maximum likelihood estimator under model m and $n = \sum_{ij} n_{ij}$ for the logistic regression example; see Raftery (1996) for details. Calculation of all Bayes factors against the full model leads immediately to posterior model probabilities.

Assuming the usual sum-to-zero constraints, the parameter vector for the full model is given by $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3) = (\mu, a_2, b_2, (ab)_{22})$. We use the same $N(0, \sigma^2)$ prior distribution for each β_i, $i = 0, 3$, under all five models. Following the ideas of Dellaportas and Forster (1995) we chose $\sigma^2 = 4 \times 2$ as a variance which gives approximately non-informative priors.

The details of our numerical comparisons are as follows. All Markov chains were initiated at the full model with starting points $\beta_i = 0$ for all $i = 0, \ldots, 3$. In reversible jump and metropolised Carlin and Chib methods we always propose a new 'neighbour' model which differs only by one term and we set $j(m, m) = 0$. The proposal distributions for β_i in the above methods as well as the pseudopriors in Gibbs variable selection are $N(\bar{\beta}_i, \bar{\sigma}_i^2)$ where $\bar{\beta}_i$ and $\bar{\sigma}_i^2$ were estimated from a pilot run of 500 iterations in the full model, after discarding the first 100 as burn-in iterations. The resulting values of $\bar{\beta}_i$ and $\bar{\sigma}_i$ turned out to be $(-0.47, -0.87, 0.56, -0.17)$ and $(0.27, 0.27, 0.28, 0.27)$ for $i = 0, \ldots, 3$ respectively. Within each model, updating of the parameters β_i was performed via Gibbs sampling steps as described in Dellaportas and Smith (1993). Finally, each Markov chain ran for 110,000 iterations and the output summaries are based on ergodic averages taken on the last 100,000 iterations.

All MCMC methods give similar results; they favour two out of the five models with probability > 0.93. The results are given in Table 2. Although this is a small model choice problem, it is promising to see that Gibbs sampling variable selection can perform well when compared with Metropolis model choice methods.

Model	Deviance	AP	GVS	KM	RJ	MCC
μ	18.656	0.0039	0.0051	0.0052	0.0051	0.0053
$\mu + a_j$	4.748	0.4599	0.4933	0.4959	0.4914	0.4905
$\mu + b_l$	12.171	0.0112	0.0113	0.0095	0.0118	0.0116
$\mu + a_j + b_l$	0.368	0.4624	0.4387	0.4362	0.4394	0.4402
$\mu + a_j + b_l + (ab)_{jl}$	0.000	0.0625	0.0507	0.0532	0.0523	0.0524

Table 2: Posterior Model Probabilities: AP=approximate probabilities, GVS=Gibbs variable selection, KM=Kuo and Mallick Gibbs sampler, RJ=reversible jump, MCC=Metropolised Carlin and Chib's method.

References

[1] Carlin B.P. and Chib S. (1995). Bayesian Model Choice via Markov Chain Monte Carlo Methods. *Journal of Royal Statistical Society B*, **157**, 473–484.

[2] Chipman H. (1995). Bayesian Variable Selection with Related Predictors. *Canadian Journal of Statistics*, to appear.

[3] Dellaportas P. and Forster J.J. (1996). Markov Chain Monte Carlo Model Determination for Hierarchical and Graphical Models. *Technical Report*, Faculty of Mathematics, Southampton University, UK.

[4] Dellaportas P. and Smith A.F.M. (1993). Bayesian Inference for Generalised Linear and Proportional Hazards Models via Gibbs Sampling. *Applied Statistics*, **42**, 443–459.

[5] George E. and McCulloch R.E. (1993). Variable Selection via Gibbs Sampling. *Journal of the American Statistical Association*, **88**, 881–889.

[6] Green P. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, **82**, 711–732.

[7] Healy, M.J.R. (1988). Glim: An Introduction. *Claredon Press*, UK.

[8] Kuo L. and Mallick B. (1997). Variable Selection for Regression models. *Sankhya*, to appear

[9] Madigan D. and York J. (1995). Bayesian Graphical Models for Discrete Data. *International Statistical Review*, **63**, 215–232.

[10] Raftery A.E.(1996). Approximate Bayes Factors and Accounting for Model Uncertainty in Generalised Linear Models. *Biometrika*, **83**, 251–266.