Department of Statistics and Insurance Science
University of Piraeus

On Bayesian Model and Variable Selection Using MCMC

Ioannis Ntzoufras
e-mail:jbn@aueb.gr

A Joint Work with:

Petros Dellaportas,
Department of Statistics, Athens University of Economics and Business,

Jonathan J. Forster
Faculty of Mathematics, University of Southampton, UK

18 December 2000, Piraeus

Contents

Model Selection Strategies

What is Model Selection?

- Evaluation of performance of scientific scenarios and
- Selection of the 'best'.

'Best' Model?

- The 'best' performed model is totally subjective
- Different procedures (or scientists) support different scientific theories, scenarios and models.

Two **MAJOR** principles:

1. *Goodness of Fit*
   How close is theory [model] to reality [data].
2. *Parsimony*
   Simplicity of theory;
   In stats: Economy in parameters.

Available Methods

- Classical Model Selection: Significance Tests and Stepwise Methods:
  - Forward Strategy.
  - Backward Elimination.
  - Stepwise Procedures.
- Bayesian Model Selection
  - Posterior odds and posterior model probabilities.
  - Utility measures.
  - Predictive criteria.
- Model Selection Criteria
  - Akaike Information Criterion (AIC).
  - Bayes Information Criterion (BIC).
  - Other criteria.

Disadvantages of Classical Stepwise Procedures

- Large datasets we observe small p-values even if the hypothesized model is plausible.
- Exact significance level cannot be calculated since stepwise methods are sequential application of simple significance tests (Freedman, 1983).
- The maximum F-to-enter statistic 'is not even remotely like an F-distribution' (Miller, 1984).
- The selection of a single model ignores model uncertainty.
- We can compare only nested models.
- Different models are selected if we use different procedures or start from different models.

Bayesian Model Selection

Bayesian model selection is based on

1. Posterior odds of model m_0 versus model m_1 given by

$$PO_{01} = \frac{f(m_0|\underline{y})}{f(m_1|\underline{y})} = \underbrace{\frac{f(\underline{y}|m_0)}{f(\underline{y}|m_1)}}_{Bayes\ Factor} \times \underbrace{\frac{f(m_0)}{f(m_1)}}_{Prior\ Odds} .$$

2. Posterior probabilities given by

$$f(m|\underline{y}) = \frac{f(\underline{y}|m)f(m)}{\sum_{m_l \in \mathcal{M}} f(\underline{y}|m_l)f(m_l)} = \left(\sum_{m_l \in \mathcal{M}} PO_{m_l,m}\right)^{-1},$$

- $\mathcal{M}$: set of models under consideration,
- $\Sigma_{m' \in \mathcal{M}} f(m|\underline{y}) = 1$.

Markov Chain Monte Carlo Methodology

For illustration, consider the simple normal linear regression model

$$Y_i \sim N(\beta_1 X_{1i} + \beta_2 X_{2i}, \sigma^2).$$

A Bayesian model formulation consists from

1. The model likelihood $f(\underline{y}|\beta_1, \beta_2, \sigma^2)$
2. The prior distribution $f(\beta_1, \beta_2, \sigma^2)$

Then Bayesian methods are interested in estimating the posterior distributions:

- Joint posterior distribution

  $$f(\beta_1, \beta_2, \sigma^2|\underline{y}) \propto f(\underline{y}|\beta_1, \beta_2, \sigma^2) f(\beta_1, \beta_2, \sigma^2)$$

  and/or

- Marginal posterior distributions $f(\beta_1|\underline{y})$, $f(\beta_2|\underline{y})$ and $f(\sigma^2|\underline{y})$.

A...Gibbs Sampler

Gibbs sampler generates samples iteratively from each the full conditional posterior distributions.

- Generate β_1 from $f(\beta_1|\beta_2, \sigma, \underline{y})$
- Generate β_2 from $f(\beta_2|\beta_1, \sigma, \underline{y})$
- Generate σ^2 from $f(\sigma^2|\beta_1, \beta_2, \underline{y})$.

B...Metropolis Hastings Algorithm

In Metropolis-Hasting algorithm we

- Propose new candidate values $(\beta_1', \beta_2', \sigma^{2'})$ generated from a known easy-to-use distribution

$$q(\beta_1, \beta_2, \sigma^2, \beta_1', \beta_2', \sigma^{2'}).$$

- Accept the above proposed values with probability

$$\alpha = min\left(1, \frac{f(\beta_1', \beta_2', \sigma^{2'}|\underline{y})q(\beta_1', \beta_2', \sigma^{2'}, \beta_1, \beta_2, \sigma^2)}{f(\beta_1, \beta_2, \sigma^2|\underline{y})q(\beta_1, \beta_2, \sigma^2, \beta_1', \beta_2', \sigma^{2'})}\right)$$

Model Selection via Markov Chain Monte Carlo Methods

Problems in Bayesian model selection:

- Integrals involved in $f(m|\underline{y})$ and
- Size of $\mathcal{M}$.

Hence, MCMC methods become an extremely attractive alternative.

Descriprion:

- Generate sample $(m^{(t')}, \underline{\beta}^{(t')},\ t' = 1, \ldots, t)$
- Estimate posterior model probabilities by

$$\hat{f}(m) \quad = \quad \frac{1}{t} \sum_{t'=1}^{t} I(m^{(t')} = m) \qquad m \in \mathcal{M}$$

  $I(\cdot)$: Indicator function.

- Get samples from $f(\underline{\beta}_{(m)}|m, \underline{y})$ (automatically available).

Why use MCMC in Model selection:

- Automatic after defining the prior distribution,
- Cannot explore the model space otherwise,
- Integrals involved are intractable.
- Bayesian model averaging is straightforward.

- General Model Selection Methods
  - Reversible jump by Green (1995).
  - Carlin and Chib (1995) Gibbs sampler.
  - Markov chain Monte Carlo model composition (MC^3) by Madigan and York (1995).
- Variable selection samplers
  - Stochastic Search Variable Selection (SSVS) by George and McCulloch (1993).
  - Kuo and Mallick (1998) Gibbs sampler.
- Fast variable selection algorithms for normal models
  - Clyde *et al.* (1996).
  - Smith and Kohn (1996).
  - Clyde (1998).

Reversible Jump

The Procedure

Suppose that the current state of the Markov chain is $\left(\underline{\beta}_{(m)}, m\right)$, then

- Generate $\underline{\beta}_{(m)}$ from $f(\underline{\beta}_{(m)}|\underline{y}, m)$ (optional).
- Propose a new model m' with probability $j(m, m')$.
- Generate $\underline{u}$ from proposal $q(\underline{u}|\underline{\beta}_{(m)}, m, m')$.
- Set $(\underline{\beta}'_{(m')}, \underline{u}') = h_{m,m'}(\underline{\beta}_{(m)}, \underline{u})$.
  - $d(m) + d(\underline{u}) = d(m') + d(\underline{u}')$ and
  - $h_{m',m} = h^{-1}_{m,m'}$.
- Accept the proposed move to model m' with probability

  $\alpha = \min(1, A)$

$$A = \frac{f(\underline{y}|\underline{\beta}'_{(m')}, m')f(\underline{\beta}'_{(m')}|m')f(m')j(m', m)q(\underline{u}'|\underline{\beta}'_{(m')}, m', m)}{f(\underline{y}|\underline{\beta}_{(m)}, m)f(\underline{\beta}_{(m)}|m)f(m)j(m, m')q(\underline{u}|\underline{\beta}_{(m)}, m, m')}\left|\frac{\partial h(\underline{\beta}_{(m)}, \underline{u})}{\partial(\underline{\beta}_{(m)}, \underline{u})}\right|.$$

Carlin and Chib Gibbs Sampler

Characteristic:

Requires realisations of $\{\underline{\beta}_{(m_k)} : m_k \in \mathcal{M}, m\}$.

The Procedure

Suppose that the current state is $\left(\{\underline{\beta}_{(m_k)} : m_k \in \mathcal{M}, m\}\right)$, then

- Generate $\underline{\beta}_{(m)}$ from $f(\underline{\beta}_{(m)}|\underline{y}, m)$.
- Generate $\underline{\beta}_{(m_l)}$ from $f(\underline{\beta}_{(m_l)}|m_l \neq m)$.
  - Pseudo-parameters: $\underline{\beta}_{(m_l)}$ are called,
  - Pseudopriors or linking densities: $f(\underline{\beta}_{(m_l)}|m_l \neq m)$.
  - No need to specify different $f(\underline{\beta}_{(m_l)}|m_l \neq m)$ for different m.
- The model indicator m is generated by

$$f(m|\{\underline{\beta}_{(m_k)} : m_k \in \mathcal{M}\}, \underline{y}) = \frac{A_m}{\sum_{m_k \in \mathcal{M}} A_{m_k}}$$

$$A_m = f(\underline{y}|\underline{\beta}_{(m)}, m) \prod_{m_l \in \mathcal{M}} \left\{f(\underline{\beta}_{(m_l)}|m)\right\} f(m).$$

Drawback:

Specification and generation from many pseudopriors (at least $|\mathcal{M}| - 1$)

- computationally demanding (time, memory and hard disk limitations)
- procedure is impracticable for large problems.

Stochastic Search Variable Selection

<u>Characteristics</u>

- Variable Selection Method
- Originally for Normal models (1993) and then applied in other GLM type models.
- The dimension of the model is constant.
- The model likelihood is given by $f(\underline{y}|\underline{\beta})$ for all models.
- The model indicator m is substituted by $\underline{\gamma}^T = (\gamma_1, \ldots, \gamma_p)$.
- The indicator variables γ_j are involved in the model through the prior

$$\underline{\beta}_j|\gamma_j \sim \gamma_j N(0, \underline{\Sigma}_j) + (1-\gamma_j) N(0, k_j^{-2} \underline{\Sigma}_j)$$

  for specified k_j and $\underline{\Sigma}_j$.

- Generally SSVS results differ from usual model selection (tend to be close for large k_j).

The Procedure

Suppose that the current state is $(\underline{\beta}, \underline{\gamma})$, then

- Generate $\underline{\beta}_j$ from

$$f(\underline{\beta}_j | \underline{\beta}_{\backslash j}, \underline{\gamma}, \underline{y}) \propto f(\underline{y} | \underline{\beta}, \underline{\gamma}) f(\underline{\beta}_j | \gamma_j)$$

  for $j = 1, \ldots, p$.

  $\underline{\beta}_j$: vector of parameters of j term.

- Generate $\gamma_j \sim Bernoulli\left(\frac{O_j}{1+O_j}\right)$ with

$$O_j = \frac{f(\gamma_j = 1 | \underline{\beta}, \underline{\gamma}_{\backslash j}, \underline{y})}{f(\gamma_j = 0 | \underline{\beta}, \underline{\gamma}_{\backslash j}, \underline{y})} = \underbrace{\frac{f(\underline{\beta} | \gamma_j = 1, \underline{\gamma}_{\backslash j})}{f(\underline{\beta} | \gamma_j = 0, \underline{\gamma}_{\backslash j})}}_{Prior\ Ratio} \underbrace{\frac{f(\gamma_j = 1, \underline{\gamma}_{\backslash j})}{f(\gamma_j = 0, \underline{\gamma}_{\backslash j})}}_{Prior\ Odds}$$

  for $j = 1, \ldots, p$.

  $\underline{\gamma}_{\backslash j}$: all components of $\underline{\gamma}$ except γ_j.

Variable selection step does not (directly) depend on the model likelihood!

Kuo and Mallick Sampler

Characteristics:

- Variable Selection Method
- Originally for Normal models (1993) but can be applied in other GLM type models.
- Likelihood is given by $f(\underline{y}|\underline{\beta}, \underline{\gamma})$.
- Model indicator m is substituted by $\underline{\gamma}$.
- Indicator variables γ_j are involved in the model by substituting $\underline{\beta}_j$ by $\gamma_j \underline{\beta}_j$ in the linear predictor.
- Prior is given by $f(\underline{\beta})$ for all models.
- Generally KM results differ than other model selection methods due to the fact that the underlying priors are automatically defined by $f(\underline{\beta})$.

The Procedure

Suppose that the current state is $(\underline{\beta}, \underline{\gamma})$, then

- Generate $\underline{\beta}_j$ from
  - $f(\underline{y}|\underline{\beta}, \underline{\gamma})f(\underline{\beta}_j|\underline{\beta}_{\backslash j})$ if $\gamma_j = 1$
  - $f(\underline{\beta}_j|\underline{\beta}_{\backslash j})$ if $\gamma_j = 0$

  for all $j = 1, \ldots, p$.

- Generate $\gamma_j \sim Bernoulli\left(\frac{O_j}{1+O_j}\right)$ with

$$O_j = \underbrace{\frac{f(\underline{y}|\underline{\beta}, \gamma_j = 1, \underline{\gamma}_{\backslash j})}{f(\underline{y}|\underline{\beta}, \gamma_j = 0, \underline{\gamma}_{\backslash j})}}_{Likelihood\ Ratio} \underbrace{\frac{f(\gamma_j = 1, \underline{\gamma}_{\backslash j})}{f(\gamma_j = 0, \underline{\gamma}_{\backslash j})}}_{Prior\ Odds}.$$

Advantage: extremely straightforward.

Disadvantage: There is no flexibility to improve efficiency.

Developments of MCMC Model and Variable Selection Methods

- New Variable Selection Samplers
  - Gibbs variable selection.
  - Metropolised Carlin and Chib sampler.
- Connections between MCMC methods
- Discussion of selection of Proposal Distributions.
- Implementation in GLM

Gibbs Variable Selection

<u>Characteristics</u>

- Natural hybrid of SSVS and the Kuo and Mallick (1998) sampler.
- Same likelihood as in Kuo and Mallick Sampler.
- Specify prior as $f(\underline{\beta}|\underline{\gamma})f(\underline{\gamma})$.

  Consider the partition of $\underline{\beta} = \left\{\underline{\beta}_{(\underline{\gamma})}, \underline{\beta}_{(\backslash\underline{\gamma})}\right\}$ into

  - $\underline{\beta}_{(\underline{\gamma})}$: parameters in model ($\gamma_j = 1$)
  - $\underline{\beta}_{(\backslash\underline{\gamma})}$: parameters not in model ($\gamma_j = 0$)

  then $f(\underline{\beta}|\underline{\gamma})$ may be partitioned into

  - Prior: $f(\underline{\beta}_{(\underline{\gamma})}|\underline{\gamma})$ and
  - Pseudoprior: $f(\underline{\beta}_{(\backslash\underline{\gamma})}|\underline{\beta}_{(\underline{\gamma})}, \underline{\gamma})$.

The Procedure

Suppose that the current state is $(\underline{\beta}, \underline{\gamma})$, then

- Generate parameters $\underline{\beta}_{(\underline{\gamma})}$ from

$$f(\underline{\beta}_{(\underline{\gamma})}|\underline{\beta}_{(\backslash\underline{\gamma})}, \underline{\gamma}, \underline{y}) \propto f(\underline{y}|\underline{\beta}, \underline{\gamma}) f(\underline{\beta}_{(\underline{\gamma})}|\underline{\gamma}) f(\underline{\beta}_{(\backslash\underline{\gamma})}|\underline{\beta}_{(\underline{\gamma})}, \underline{\gamma})$$

- Generate pseudo-parameters $\underline{\beta}_{(\backslash\underline{\gamma})}$ from $f(\underline{\beta}_{(\backslash\underline{\gamma})}|\underline{\beta}_{(\underline{\gamma})}, \underline{\gamma})$

- Generate $\gamma_j \sim Bernoulli\left(\frac{O_j}{1+O_j}\right)$ with

$$O_j = \underbrace{\frac{f(\underline{y}|\underline{\beta}, \gamma_j = 1, \underline{\gamma}_{\backslash j})}{f(\underline{y}|\underline{\beta}, \gamma_j = 0, \underline{\gamma}_{\backslash j})}}_{Likelihood\ Ratio} \underbrace{\frac{f(\underline{\beta}|\gamma_j = 1, \underline{\gamma}_{\backslash j})}{f(\underline{\beta}|\gamma_j = 0, \underline{\gamma}_{\backslash j})}}_{Prior/Pseudoprior\ Ratio} \underbrace{\frac{f(\gamma_j = 1, \underline{\gamma}_{\backslash j})}{f(\gamma_j = 0, \underline{\gamma}_{\backslash j})}}_{Prior\ Odds}$$

Simpler Approach

- Assume prior

$$f(\underline{\beta}_j|\underline{\gamma}_j) = \gamma_j N(\mathbf{0}, \underline{\Sigma}_j) + (1-\gamma_j)N(\underline{\bar{\mu}}_j, \underline{S}_j),$$

$\underline{\bar{\mu}}_j$ and $\underline{S}_j$: are pseudoprior parameters (tuned to achieve optimal convergence).

- The full conditional posterior distribution is now given by

$$f(\underline{\beta}_j|\underline{\beta}_{\setminus j}, \underline{\gamma}, \underline{y}) \propto \begin{cases} f(\underline{y}|\underline{\beta}, \underline{\gamma})N(0, \underline{\Sigma}_j) & \gamma_j = 1 \\ N(\underline{\bar{\mu}}_j, \underline{S}_j) & \gamma_j = 0 \end{cases}$$

This approach is ...

  - Simple to apply
  - Efficient when covariates are not highly correlated.
  - Easy to specify pseudopriors

    Get $\underline{\bar{\mu}}_j$ and $\underline{S}_j$: from a pilot run of the full model; see Dellaportas and Forster (1999).

Comparison of Gibbs Based Methods

The variable selection step can be generally written as $\gamma_j \sim Bernoulli\left(\frac{O_j}{1+O_j}\right)$

$$O_j = LR_j \times PR_j \times PSR_j \times \underbrace{\frac{f(\gamma_j = 1, \underline{\gamma}_{\backslash j})}{f(\gamma_j = 0, \underline{\gamma}_{\backslash j})}}_{Prior\ Odds}$$

where

- LR_j is a likelihood ratio,
- PR_j is a prior ratio,
- PSR_j is a pseudoprior ratio.

In each method we have:

- SSVS: LR_j vanishes
- KM: PR_j and PSR_j disappear
- GVS: Trade off between SSVS and KM sampler.

  O_j depends on both LR and on PR and PSR.

Reversible Jump and 'Metropolised' Carlin and Chib

The Procedure

Suppose that the current state is $\left(\underline{\beta}_{(m)}, m\right)$, then

- Generate $\underline{\beta}_{(m)}$ from $f(\underline{\beta}_{(m)}|\underline{y}, m)$.
- Propose a new model m' with probability $j(m, m')$.
- Generate $\underline{\beta}'_{(m')}$ from the proposal $f(\underline{\beta}'_{(m')}|m' \neq m)$.
- Accept the proposed move with probability $\alpha = min(1, A)$

$$A = \frac{f(\underline{y}|\underline{\beta}'_{(m')}, m')f(\underline{\beta}'_{(m')}|m')f(\underline{\beta}_{(m)}|m')f(m')j(m', m)}{f(\underline{y}|\underline{\beta}_{(m)}, m)f(\underline{\beta}_{(m)}|m)f(\underline{\beta}'_{(m')}|m)f(m)j(m, m')}.$$

Important Features

- Requires (only) $\underline{\beta}_{(m)}$ and $\underline{\beta}'_{(m')}$ to calculate α.
- Model m' is proposed with probability $j(m, m')$, independently of the values of any model parameters.

Only need to sample from pseudoprior $f(\underline{\beta}'_{(m')}|m' \neq m)$!!!

RJ and MCC

- MCC is a reversible jump with ...
  - $\left(\underline{\beta}'_{(m')}, \underline{u}'\right) = \left(\underline{u}, \underline{\beta}_{(m)}\right)$,

    $\underline{u}' = \{\underline{\beta}_{(m_l)} : m_l \neq m'\}$, $\underline{u} = \{\underline{\beta}_{(m_l)} : m_l \neq m\}$,
  - proposal densities are replaced by

    $q(\underline{u}|\underline{\beta}_{(m)}, m, m') = \Pi_{m_l \in \mathcal{M} \setminus \{m'\}} \left\{ f(\underline{\beta}_{(m_l)}|m') \right\}$ and

    $q(\underline{u}'|\underline{\beta}'_{(m')}, m', m) = \Pi_{m_l \in \mathcal{M} \setminus \{m\}} \left\{ f(\underline{\beta}_{(m_l)}|m) \right\}$.
- MCC also coincides to the simpler RJ with:
  - $\left(\underline{\beta}'_{(m')}, \underline{u}'\right) = \left(\underline{u}, \underline{\beta}_{(m)}\right)$,

    $\underline{u}' = \underline{\beta}_{(m)}$ and $\underline{u} = \underline{\beta}'_{(m')}$,
  - proposal densities are replaced by

    $q(\underline{u}|\underline{\beta}_{(m)}, m, m') = f(\underline{\beta}'_{(m')}|m' \neq m)$ and

    $q(\underline{u}'|\underline{\beta}'_{(m')}, m, m') = f(\underline{\beta}_{(m)}|m \neq m')$

Reversible Jump and MC^3

The Procedure

- Suppose that ...
  - $f(\underline{\beta}_{(m)}|m, \underline{y})$ is available for all models $m \in \mathcal{M}$,
  - $f(\underline{y}|m)$ is also known.
- Consider MCC (or RJ) with

$$\boxed{q(\underline{\beta}_{(m)}|\underline{\beta}'_{(m')}, m, m') = f(\underline{\beta}_{(m)}|m, \underline{y}),}$$

- If $\left(\underline{\beta}_{(m)}, m\right)$ is the current state, then
  - Generate $\underline{\beta}_{(m)}$ from $f(\underline{\beta}_{(m)}|\underline{y}, m)$ (optional).
  - Propose a new model m' with probability $j(m, m')$.
  - Generate $\underline{\beta}'_{(m')}$ from the posterior $f(\underline{\beta}'_{(m')}|m, \underline{y})$.
  - Accept the proposed model m' with probability

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

Proposal Distributions

- Proposal Distributions for Model Parameters

  - Independent distributions for each term j: $N(\underline{\bar{\mu}}_j, \underline{S}_j)$.
  - SSVS type proposal: $N(\mathbf{0}_{d_j}, \underline{\Sigma}_j / k_j^2)$.
  - Maximum likelihood based: $N\left(\underline{\hat{\beta}}_{(m)}, \underline{\hat{\Sigma}}_{(m)}\right)$.
  - Alternative easy-to-use choice: $N(\underline{\hat{\beta}}_{(m)}, \underline{\Sigma}_{(m)}/k^2)$.
  - Using conditional maximised likelihood.
  - Giudici and Roberts (1998) automatic choice.

- Proposal Distributions on Model Space

  - Common proposal: Uniform distribution.
  - 'Local' and 'Global' proposals.
  - $j(m, m) = 0$ better than $j(m, m) > 0$ (Liu 1996a,b).
  - Set $j(m, m')$ using Laplace or BIC approximations.
  - Use an MC^3 When size of $\mathcal{M}$ is large.

General Variable Selection Frame for GLM

1. Generate $\underline{\beta}_{(\underline{\gamma})}$ from $f(\underline{\beta}_{(\underline{\gamma})}|\phi, \underline{\gamma}, \underline{y})$

2. If ϕ unknown then generate from $f(\phi|\underline{\beta}_{(\underline{\gamma})}, \underline{\gamma}, \underline{y})$.

3. Generate all pseudo-parameters from the pseudopriors or proposal distributions.

4. Generate γ_j using one of the following:

   (a) Metropolis within Gibbs approach.

   For all $j = 1, \ldots, p$

   - propose a change from γ_j to $1 - \gamma_j$ with probability $q_2(\gamma_j, 1 - \gamma_j)$ and
   - accept the proposed move with probability

$$\alpha = min\left(1, \frac{q_2(1-\gamma_j, \gamma_j)}{q_2(\gamma_j, 1-\gamma_j)} O_j^{1-2\gamma_j}\right).$$

(b) Random scan approach.

- Randomly select a new proposed model $\underline{\gamma}' \in nb(\underline{\gamma})$ with probability $j(\underline{\gamma}, \underline{\gamma}')$.
- accept the proposed move with probability

$$\alpha = min\left(1, \frac{j(\underline{\gamma}', \underline{\gamma})}{j(\underline{\gamma}, \underline{\gamma}')} O_j^{1-2\gamma_j}\right).$$

O_j is the full conditional posterior odds to include j term in the model:

$$O_j = \frac{f(\gamma_j = 1 | \underline{\beta}_{(\underline{\gamma})}, \underline{\gamma}_{\setminus j}, \underline{y})}{f(\gamma_j = 0 | \underline{\beta}_{(\underline{\gamma})}, \underline{\gamma}_{\setminus j}, \underline{y})}.$$

Can be analysed as a product of: LR_j, PR_j, PSR_j and the prior model odds.

All variable selection methods can be included in this frame.

Illustrated Examples

- Five (5) Simulated Regression Examples
- 2^6 Contingency Table Example
- $3 \times 2 \times 4$ Contingency Table
- Logistic Regression Example

Simulated Regression Examples

			Design	Generated Model		Supported Model	
Dataset	n	p	Structure	$\underline{\eta}$	σ	Backward/Forward	MCMC
1	50	15	1	$X_4 + X_5$	2.50	$X_4 + X_5 + X_{12}$	$X_4 + X_5$
2	50	15	2	$\sum_{j=1}^{5} X_j$	2.50	$\sum_{j=1}^{5} X_j + X_{12}/\ X_{14}$	X_{14}
3	50	15	1	0	2.50	*Empty*	*Empty*
4	50	15	2	0	2.50	$X_3 + X_{12}$	X_3
5	100	50	1	0	1.00	X_{19}	*Empty*
6	100	30	1	$0.5X_1$	0.87	X_1	X_1

Design structure 1: $X_0 = 1$, $X_j \sim N(0,1)$, for $j = 1, \ldots, p$,

Design structure 2: $X_0 = 1$, $X_j \sim N(0,1)$, for $j = 1, \ldots, 10$ and $X_j \sim N(0.3X_1 + 0.5X_2 + 0.7X_3 + 0.9X_4 + 1.1X_5, 1)$, for $j = 11, \ldots, 15$.

Simulated Regression Datasets:

Batch Standard Deviations of Highest Posterior Model Probability

	Dataset					
	1	2	3	4	5	6
GVS	0.017	0.077	0.024	0.016	0.041	0.027
KM	0.039	0.059	0.032	0.037	0.089	0.059
RJ	0.042	0.102	0.032	0.028	0.062	0.062
MCC(I)	0.044	—	0.043	0.026	0.143	0.078
MCC(M)	0.048	0.042	0.044	0.029	0.103	0.090
SSVS	0.138	0.122	0.065	0.111	0.109	0.138

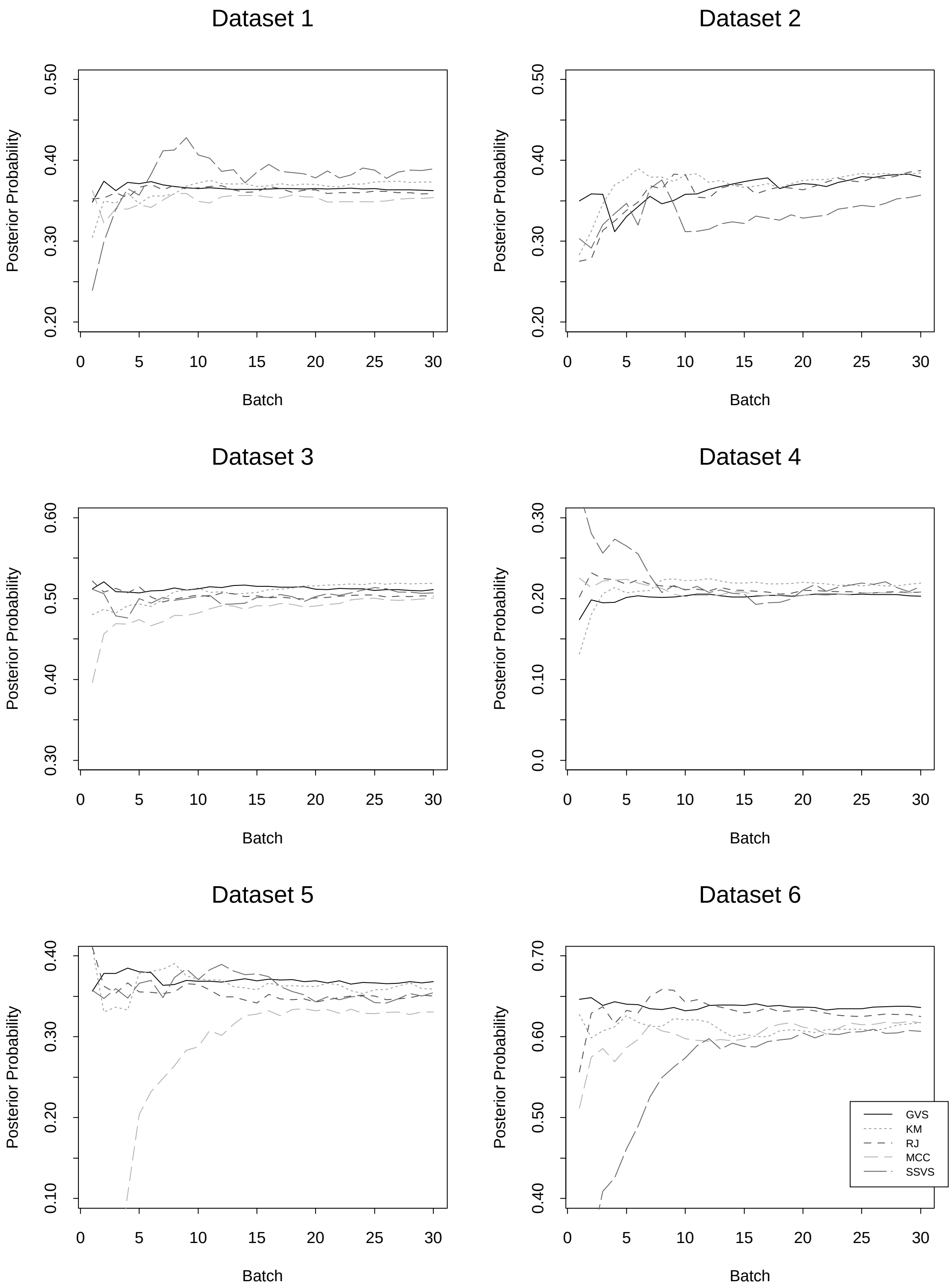

Figure 1: Simulated Regression Datasets: Ergodic Highest Posterior Model Probabilities.

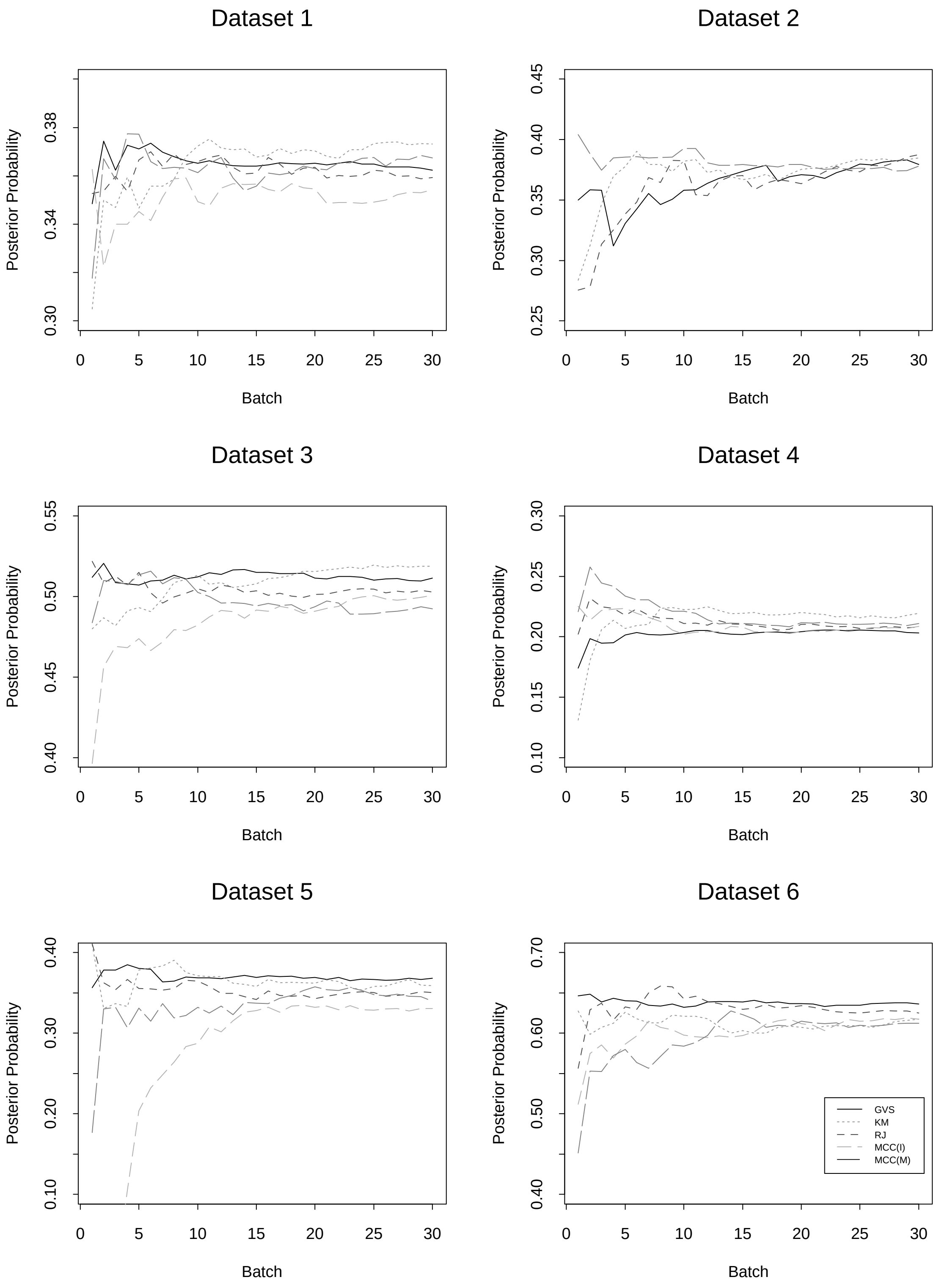

Figure 2: Ergodic Highest Posterior Model Probabilities for Simulated Datasets (GVS: Gibbs variable selection, KM: Kuo and Mallick's Gibbs sampler, RJ: Reversible jump as implemented by Dellaportas and Forster (1999), MCC(I): Metropolised Carlin and Chib with Independent Proposals, MCC(M): Metropolised Carlin and Chib with Multivariate Proposals).

A Logistic Regression Example

Dataset analysed by Healy (1988) describe the relationship between

- the number of survivals,
- the patient condition (more or less severe) and
- the received treatment (antitoxin medication or not).

Model	Deviance	AP	GVS	KM	RJ	MCC(I)	MCC(M)
μ	18.656	0.004	0.004	0.008	0.005	0.004	0.005
$\mu + a_j$	4.748	0.460	0.492	0.484	0.489	0.490	0.494
$\mu + b_l$	12.171	0.011	0.010	0.009	0.011	0.011	0.012
$\mu + a_j + b_l$	0.368	0.462	0.440	0.450	0.442	0.441	0.435
$\mu + a_j + b_l + (ab)_{jl}$	0.000	0.063	0.053	0.050	0.053	0.054	0.054

	Batch Standard Deviation	
	Model A	Model $A + B$
GVS	0.015	0.014
KM	0.045	0.041
RJ	0.033	0.027
MCC(I)	0.033	0.027
MCC(M)	0.026	0.020

Table 1: Logistic Regression Example: Batch Standard Deviation of Posterior Model Probabilities.

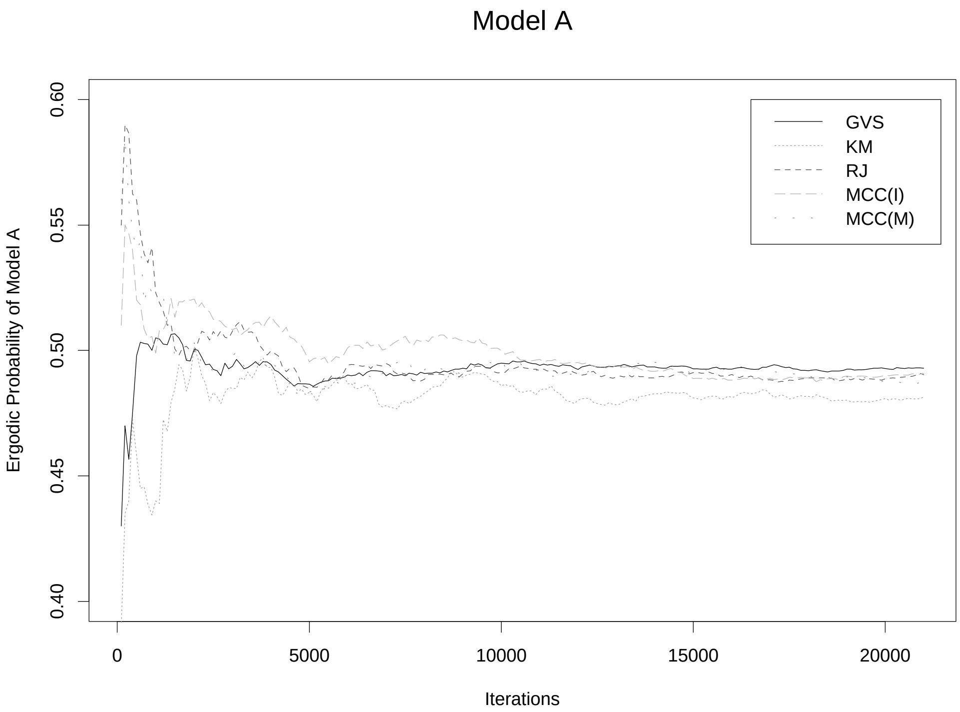

Figure 3: Ergodic Posterior Model Probability of Model A for Logistic Regression Example (GVS: Gibbs variable selection, KM: Kuo and Mallick's Gibbs sampler, RJ: Reversible jump as implemented by Dellaportas and Forster (1999), MCC(I): Metropolised Carlin and Chib with Independent Proposals, MCC(M): Metropolised Carlin and Chib with Multivariate Proposals).

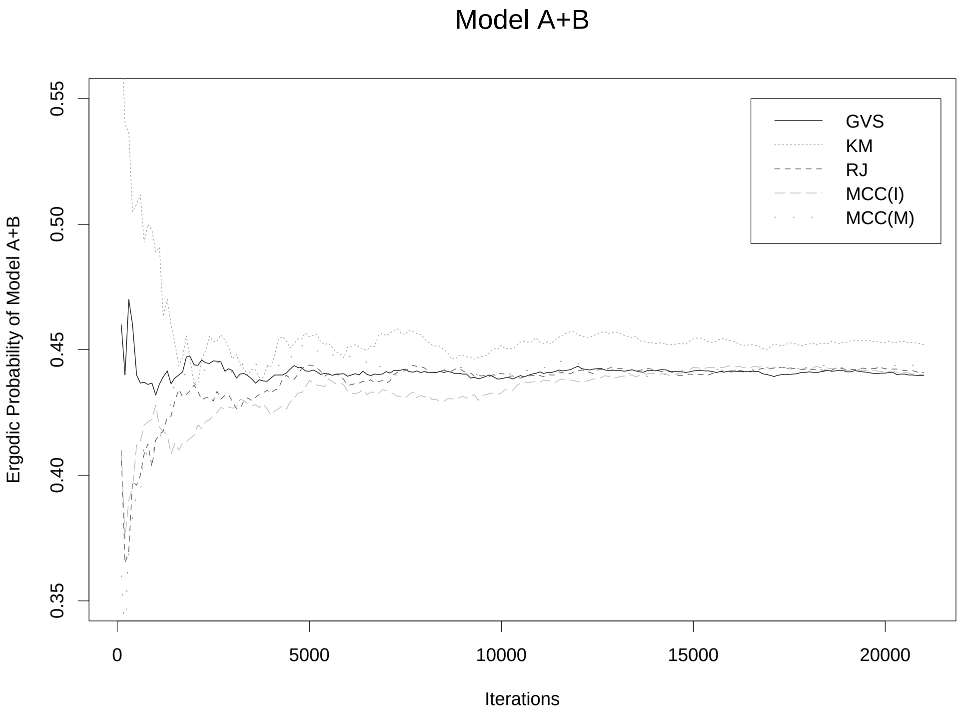

Figure 4: Ergodic Posterior Model Probability of Model $A+B$ for Logistic Regression Example (GVS: Gibbs variable selection, KM: Kuo and Mallick's Gibbs sampler, RJ: Reversible jump as implemented by Dellaportas and Forster (1999), MCC(I): Metropolised Carlin and Chib with Independent Proposals, MCC(M): Metropolised Carlin and Chib with Multivariate Proposals).

Simultaneous Covariate and Structural Identification in GLM

- Equivalent priors between different links.
- MCMC algorithms for variable and link selection
  - Using appropriate transformation.
  - Using Metropolised Carlin and Chib sampler
- Results are available for selection between traditional links (logit, probit, cloglog, loglog).
- Now working on selection between two general link families with additional parameters (t-ling vs. loggamma link).

On Prior Distributions for Model Selection?

That's a long painfull story! (Lindley's Paradox)

Research Work on Prior Distributions

- Information Criteria and imposed penalty.
- Normal models and independent priors
  - Dimensionality Penalty under Collinearity.
- Conditional prior odds at zero.
- Prior Specification via penalty determination.
- Posterior Odds at the limits of significance
  - Specification of Prior distributions Using p-values.
- Extensions to GLM and Interpretation of prior distributions Using imaginary data.

Discussion and Further Research

Future research will include

- Implementation of MCMC methods (especially of reversible jump) in specific complicated model selection issues (for example see Nobile and Green, 1997 or Vrontos *et al.* , 1998).
- Other general issues of MCMC need also to be investigated, such as
  - rates of convergence (Brooks and Giudici, 1998)
  - choices of proposals (Giudici and Roberts, 1998).
- Model selection between models with different quantity and quality of data.
- Calculation of Bayes factors variants (IBF,FBF,PBF etc.) using MCMC.
- Reference prior distributions.