Adaptive Monte Carlo for Bayesian Variable Selection in Regression Models

Demetris Lamnisos, Jim E. Griffin and Mark F.J. Steel

March 9, 2011

Abstract

This article describes a method for efficient posterior simulation for Bayesian variable selection in Generalized Linear Models with many regressors but few observations. A proposal on model space is described which contains a tuneable parameter. An adaptive approach to choosing this tuning parameter is described which allows automatic, efficient computation in these models. The method is applied to examples from normal linear and probit regression.

Keywords: Probit Regression; Linear Regression; Metropolis-within-Gibbs

1 Introduction

The availability of datasets with large numbers of variables has lead to interest in the use of variable selection methods for regression models with large numbers of potential regressors. In this paper, we will concentrate on Bayesian variable selection methods applied to datasets with hundreds of regressors where Markov chain Monte Carlo methods can be effectively used. We will work in the context of Generalized Linear Models (GLM) (McCullagh and Nelder, 1989) where it is assumed that \( y_1, y_2, \ldots, y_n \) are observed dependent variables and the \( i \)-th observation is associated with a set of \( p \) regressors \( \mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip}) \) and that

\[
f(y_i | \mathbf{x}_i, \beta, \tau) = \exp \left\{ a^{-1}(\tau) [y_i g_i - b(g_i) + c(y_i, \tau)] \right\}
\]

where \( g_i = g(\eta_i) \) is the link function and \( \eta_i = \alpha + \mathbf{x}_i \beta \) is the linear predictor, with intercept \( \alpha \in \mathbb{R} \) and regression coefficient \( \beta \in \mathbb{R}^p \). Different choices of \( a, b \) and \( c \) lead to different conditional models for \( y_i \). Linear regression and probit regression have been the most thoroughly explored models with large datasets and this paper
will concentrate on these examples (although the methods are applicable to all GLMs). For example, Lee et al. (2003) and Sha et al. (2003, 2004) consider using MCMC methods for Bayesian variable selection in a probit regression model to find gene expression levels related to a binary response (such as diseased or non-diseased). In linear regression, large numbers of variables are common in spectroscopy, chemometrics or proteomics and can also occur in economics. In both regression models, we would usually assume that only a subset of the regressors are needed to predict $y_i$ and the vector of indicator variables $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_p)$ is introduced to represent inclusion ($\gamma_i = 1$) or exclusion ($\gamma_i = 0$) of the $i$-th regressor leading to model size $p_\gamma = \sum_{i=1}^p \gamma_i$ for $\gamma$. Then the model $\gamma$ uses the linear predictor $\eta_i = \alpha + x_{i\gamma}^T \beta_\gamma$ in equation (1), where $x_{i\gamma}$ and $\beta_\gamma$ group the $p_\gamma$ elements for which $\gamma_i = 1$. Bayesian inference proceeds by placing a prior on $\gamma$ as well as $\alpha$, $\beta_\gamma$, and $\tau$.

Summaries of the posterior distribution are typically computed using either efficient methods to select a small subset of interesting models combined with an approximation for $\pi(y|\gamma)$ (Yeung et al., 2005; Hans et al., 2007; Clyde et al., 2011) or Markov chain Monte Carlo methods that produce samples from the posterior distribution using a Gibbs sampler. This paper follows the second approach. The model $\gamma$ is usually updated by a Metropolis-Hastings step which proposes new models by either including an excluded variable, excluding an included variables or swapping one included variable with one excluded variable. This is similar in spirit to a Random Walk Metropolis (RWM) sampler, which generates a Markov chain by proposing a new value as a perturbed version of the current value with the difference that here the state space is a lattice. In the RWM sampler, the variance of the perturbations can be controlled to obtain an optimal RWM sampler. Lamnisos et al. (2009) find that acceptance rates for Metropolis-Hastings samplers that include, exclude or swap a single variable usually have high acceptance rates for variable selection in probit regression models when there are many regressors and few observations. They describe a tuneable proposal for variable selection problems with a single parameter defined on $(0, 1)$ controlling the difference between the current and proposed model. They find that in probit modelling with many regressors, the optimal sampler occurs for a proposal which leads to an average acceptance rate close to 0.234, which is the value for the optimal RWM sampler in many continuous problems (Roberts and Rosenthal, 2001; Sherlock and Roberts, 2009).

Metropolis-Hastings samplers with tuneable proposal can lead to efficient algorithms but tuning can be time-consuming. Recently, there has been interest in adaptive Monte Carlo methods where the distribution of the proposal is adjusted during the MCMC run. These methods are difficult to implement in general since the Markov property is violated and standard theory for convergence of the chain to the target distribution does not apply. However, convergence to the target dis-
tribution can be verified for particular forms of adjustment. The first adaptive algorithm that could be shown to converge to the target distribution was introduced by Haario et al. (2001) who used methods from Stochastic Approximation. This important idea and other methods are reviewed by Andrieu and Thoms (2008).

Our goal is to automatically produce a Markov chain with good mixing properties which gives accurate answers with the smallest possible run length. The posterior distribution will be complicated and vast (there will be \(2^p\) potential models if there are \(p\) potential regressors). Adaptive methods are important because MCMC methods often mix slowly (and so proposals that encourage good mixing are important) and the running of many pilot runs is unsatisfactory due to the large number of iterations needed to give good estimates of posterior summaries. Such methods have been previously applied to variable selection problems by Nott and Kohn (2005). They allow the probability that a particular variable is proposed to be included in or removed from the model to adapt over the chain. This is rather different from the method developed here where the expected number of variables to be updated, rather than the variable-specific update probability, is adapted over the chain. The method uses a form of proposal for variable selection described by Lamnisos et al. (2009) which can be tuned in a similar way to a RWM sampler.

The paper is organised as follows: Section 2 describes the Bayesian approach to variable selection and some MCMC algorithms for posterior exploration, as well as a tuneable proposal on model space for variable selection problems, Section 3 describes adaptive versions of the algorithms, Section 4 discusses some numerical examples that illustrate the utility of the approach and Section 5 contains some concluding comments.

## 2 MCMC algorithms for Bayesian Variable Selection

The normal linear regression model has greater analytical tractability than other GLMs and so we start by discussing that model here. It assumes that observations \(y = (y_1, y_2, \ldots, y_n)^T\) are generated by

\[
y \sim N(\alpha 1 + X_\gamma \beta_\gamma, \tau I_n).
\]

Within a Bayesian analysis, \(\alpha, \beta_\gamma, \tau\) and \(\gamma\) are given prior distributions. The intercept \(\alpha\) is given the commonly used noninformative improper prior for location parameters

\[
\pi(\alpha) \propto 1, \quad (2)
\]

the regression coefficients have the multivariate normal prior

\[
\beta_\gamma | \tau, \gamma \sim N_p(0, \tau V_\gamma), \quad (3)
\]
and the error variance $\tau$ is assigned the usual noninformative improper prior for scale parameters
\[ \pi(\tau) \propto \frac{1}{\tau}. \] (4)

Finally, the prior on $\gamma$ assumes that $\pi(\gamma_i = 1) = w_i$, independently for $i = 1, \ldots, p$.

The priors (2), (3) and (4) result in an analytical expression for the marginal likelihood $\pi(y|\gamma)$ of model $\gamma$ given by
\[ \pi(y|\gamma) \propto |X_\gamma^T X_\gamma + V_\gamma^{-1}|^{-1/2} |V_\gamma|^{-1/2} (\tilde{y}^T \tilde{y} - y^T X_\gamma (X_\gamma^T X_\gamma + V_\gamma^{-1})^{-1} X_\gamma^T y)^{-(n-1)/2}, \]
where $\tilde{y} = y - \bar{y}1$ and $\bar{y}$ is the mean of the response $y$.

This analytical expression for the marginal likelihood $\pi(y|\gamma)$ facilitates the development of MCMC methods that are used to estimate posterior model characteristics. A Metropolis-Hastings algorithm with the posterior model distribution $\pi(\gamma|y)$ as stationary distribution can be implemented. This algorithm proceeds as follows:

**Algorithm 1 (MH-LR)** Let $\gamma$ be the current state of the chain.

1. Select model $\gamma'$ with probability $q(\gamma'|\gamma)$.

2. Jump to the model $\gamma'$ with probability
\[ \alpha(\gamma, \gamma') = \min \left\{ 1, \frac{\pi(y|\gamma') \pi(\gamma') q(\gamma'|\gamma)}{\pi(y|\gamma) \pi(\gamma) q(\gamma'|\gamma)} \right\}. \]

The availability of a closed form expression for $\pi(y|\gamma)$ allows a Metropolis-Hastings algorithm to be defined directly on $\gamma$. However, this expression is not analytically available for other GLMs and a sampler must be defined on the joint posterior of $\gamma$, $\alpha$, $\beta_\gamma$, and $\tau$. Let us focus on the particular case of the probit model. If the GLM is a probit regression model then Albert and Chib (1993) show that the model can be written in the following way
\[ z_i \sim \text{N}(\alpha + x_i^T \beta_\gamma, 1) \]
where $y_i = 0$ if $z_i < 0$ and $y_i = 1$ if $z_i > 1$ and $x_i^T$ denotes the $i$th row of $X_\gamma$.

Latent variables $z = (z_1, z_2, \ldots, z_n)^T$ are introduced which allow the model to be expressed as a linear regression with known error variance in $z_i$. The prior on $\beta_\gamma$ will be similar to (3), namely
\[ \beta_\gamma | \gamma \sim \text{N}_{p_\gamma}(0, V_\gamma), \]
and for $\alpha$ we now take a vague Normal prior

$$\alpha \sim N(0, h),$$

with some large value for $h$ (we adopt $h = 100$ in the sequel). In probit models, the use of the improper prior in (2) can lead to identification issues between $\alpha$ and the scale of $V_\gamma$, as documented in Lamnisos (2010). We shall define $\theta_\gamma = (\alpha, \beta_\gamma^T)^T$ in what follows. Various MCMC methods for dealing with these models are described by Lamnisos et al. (2009). Some methods propose a joint update of $(\gamma, z)$ by updating $\gamma$ given $\theta_\gamma$ and $z$ given $(\gamma, \theta_\gamma)$. One of these methods is the Automatic Generic sampler described by Green (2003) and extended by Lamnisos et al. (2009). The MLE of $\theta_\gamma$ is asymptotically normal and so the full conditional of $\theta_\gamma$ can be approximated by a normal distribution. The automatic generic method exploits the fact that a normally distributed random variable, $x$, with mean $\mu$ and variance-covariance matrix $\Sigma$ can be written as $x = \mu + \Sigma^{1/2} \epsilon$ where $\Sigma^{1/2}$ is the Cholesky decomposition of $\Sigma$ and $\epsilon$ is standard normal with the same dimension as $x$. In what follows we shall use the notation $\tilde{X}_\gamma = (1 : X_\gamma)$ and $\tilde{V}_\gamma = \begin{pmatrix} h' & 0' \\ 0 & V_\gamma \end{pmatrix}$. The Automatic Generic algorithm is then:

**Algorithm 2 (AG)** Let $(\gamma, \theta_\gamma)$ be the current state of the chain.

1. Generate $z_1, z_2, \ldots, z_n$ where $z_i \sim N(\alpha + x_i^T \beta_\gamma, 1)$ and $z_i > 0$ if $y_i = 1$ or $z_i < 0$ if $y_i = 0$.

2. Select model $\gamma'$ with probability $q(\gamma'|\gamma)$.

3. Propose $\theta_{\gamma'}$ in the following way. Let $\mu_\gamma$ and $\Sigma_\gamma$ be an approximation of the mean and variance of the posterior distribution of $\theta_\gamma$ and let $B_\gamma$ be the Cholesky decomposition of $\Sigma_\gamma$ and $v = B_\gamma^{-1}(\theta_\gamma - \mu_\gamma)$. Then we propose $\theta_{\gamma'} = \mu_{\gamma'} + B_{\gamma'} v'$ where

$$v' = \begin{cases} (v_1, \ldots, v_{p_{\gamma'}})^T & \text{if } p_{\gamma'} < p_\gamma \\ v & \text{if } p_{\gamma'} = p_\gamma \\ (v^T, \epsilon^T)^T & \text{if } p_{\gamma'} > p_\gamma \end{cases}$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_{p_{\gamma'} - p_\gamma})^T$ has i.i.d. $N(0, 1)$ elements.

4. Jump to the model $\gamma'$ and parameter $\theta_{\gamma'}$ with probability $\alpha(\gamma, \gamma', \theta_\gamma, \theta_{\gamma'}) =$

$$\min \left\{ 1, \frac{\pi(\gamma', \theta_{\gamma'}) q(\gamma'|\gamma') \pi(y|\theta_{\gamma'}) |B_{\gamma'}|}{\pi(\gamma, \theta_\gamma) q(\gamma'|\gamma) \pi(y|\theta_\gamma) |B_{\gamma}| \times K} \right\}$$
where

$$K = \begin{cases} 
(2\pi)^{-\frac{1}{2}(p_{\gamma'} - p_{\gamma})} \exp \left\{ -\frac{1}{2}(e')^T(e') \right\} & \text{if } p_{\gamma'} < p_{\gamma} \\
1 & \text{if } p_{\gamma'} = p_{\gamma} \\
(2\pi)^{\frac{1}{2}(p_{\gamma'} - p_{\gamma})} \exp \left\{ \frac{1}{2}e^T e \right\} & \text{if } p_{\gamma'} > p_{\gamma},
\end{cases}$$

and $e'$ is the obvious counterpart of $e$.

5. Draw a sample $\theta_{\gamma} \sim \mathcal{N}(\bar{X}_\gamma^T \bar{X}_\gamma + \bar{V}_\gamma^{-1})^{-1} \bar{X}_\gamma^T z, (\bar{X}_\gamma^T \bar{X}_\gamma + \bar{V}_\gamma^{-1})^{-1}$.

There are several methods of finding $\mu_\gamma$ and $\Sigma_\gamma$. Two are considered in this paper: the Laplace approximation and the Iterated Weighted Least Squares (IWLS) approximation for one iteration (see Lamnisos et al. (2009) for more details).

Alternative automatic methods for moving between models are described by Brooks et al. (2003) which are applied to probit regression by Lamnisos et al. (2009). In the case of proposals that increase the model size, the coefficient vector is completed with $u_{\gamma}(\epsilon) = \mu + \sigma \epsilon$ where $\epsilon$ has a standard normal distribution.

The Conditional Maximization method chooses $\mu$ to maximize the posterior distribution $\pi(\theta_{\gamma}, u_{\gamma} | y)$ with respect to $u_{\gamma}$. The variance, $\sigma^2$, is chosen to ensure that the acceptance probability of the move with $u_{\gamma}(\epsilon) = \mu$ is 1 and is given by

$$\sigma = \left( \frac{\pi(\gamma, \theta_{\gamma}) q(\gamma | \gamma) \pi(y | \theta_{\gamma})}{(2\pi)^{(p_{\gamma'} - p_{\gamma})/2} \pi(\gamma', \theta_{\gamma'}) q(\gamma | \gamma') \pi(y | (\theta_{\gamma}, \mu))} \right)^{\frac{1}{2}}.$$

Here

$$\alpha(\gamma, \gamma', \theta_{\gamma}, \theta_{\gamma'}) = \min \left\{ 1, \frac{\pi(\gamma', \theta_{\gamma'}) q(\gamma' | \gamma') \pi(y | \theta_{\gamma'}) \left( \sigma^2 \right)^{\frac{1}{2}(p_{\gamma'} - p_{\gamma})}}{\pi(\gamma, \theta_{\gamma}) q(\gamma | \gamma) \pi(y | \theta_{\gamma})} \right\} K \right\} \tag{5}$$

where

$$K = (2\pi)^{\frac{1}{2}(p_{\gamma'} - p_{\gamma})} \exp \left\{ \frac{1}{2}e^T e \right\}.$$

The pseudocode representation of the Conditional Maximization method is as follows:

**Algorithm 3 (C-M)** If current state is $(\gamma, \theta_{\gamma})$, then

1. Generate $z_1, z_2, \ldots, z_n$ where $z_i \sim \mathcal{N}(\alpha + x_i^T \beta_{\gamma}, 1)$ and $z_i > 0$ if $y_i = 1$ or $z_i < 0$ if $y_i = 0$.
2. Select model $\gamma'$ with probability $q(\gamma' | \gamma)$. 

6
3. Determine the location $\mu$ and the scale $\sigma$ of the proposal random variable $u_\gamma$ as described above.

4. Generate $\epsilon \sim N_{p_\gamma - p_\gamma}(0, I_{p_\gamma - p_\gamma})$.

5. Set $u_\gamma(\epsilon) = \mu + \sigma \epsilon$ and $\theta_{\gamma'} = (\theta_\gamma, u_\gamma)$.

6. Jump to the model $\gamma'$ and $\theta_{\gamma'}$ with probability given by (5). Otherwise, the proposal is rejected.

7. Draw a sample $\theta_{\gamma} \sim N((\bar{X}_\gamma^T \bar{X}_\gamma + \bar{V}_\gamma^{-1})^{-1} \bar{X}_\gamma^T \bar{V}_\gamma^{-1} \bar{X}_\gamma + \bar{V}_\gamma^{-1})^{-1}$.

Alternatively, Holmes and Held (2006) describe an algorithm which updates $\gamma$ and $\theta_\gamma$ jointly and can be extended to multinomial regression models.

**Algorithm 4 (H-H)** Suppose that $(\gamma, \theta_\gamma)$ are the current values of the chain

1. Generate $z_1, z_2, \ldots, z_n$ where $z_i \sim N(\alpha + x_i^2 \beta_\gamma, 1)$ and $z_i > 0$ if $y_i = 1$ or $z_i < 0$ if $y_i = 0$.

2. Select model $\gamma'$ with probability $q(\gamma'|\gamma)$.

3. Jump to the model $\gamma'$ with probability

$$
\alpha(\gamma, \gamma') = \min \left\{ 1, \frac{\pi(\gamma')q(\gamma'|\gamma')\pi(z|\gamma')}{\pi(\gamma)q(\gamma'|\gamma)\pi(z|\gamma)} \right\}.
$$

4. If the jump is accepted, draw $\theta_{\gamma'} \sim N((\bar{X}_{\gamma'}^T \bar{X}_{\gamma'} + \bar{V}_{\gamma'}^{-1})^{-1} \bar{X}_{\gamma'}^T \bar{V}_{\gamma'}^{-1} \bar{X}_{\gamma'} + \bar{V}_{\gamma'}^{-1})^{-1}$.

### 2.1 A Tuneable Proposal On Model Space

Lamnisos et al. (2009) propose a new general model proposal $q_\zeta(\gamma'|\gamma)$ which draws a new model in the following way:

1. A value $N^{(t)}$ is generated from a Binomial distribution with $N - 1$ trials and success probability $\zeta$.

2. One of three possible moves: “Add”, “Delete” and “Swap” is chosen uniformly at random. If Add is selected then $N^{(t)} + 1$ regressors are chosen to be added to those included in $\gamma$ to form $\gamma'$, if Delete is selected then $N^{(t)} + 1$ regressors are chosen to be removed from the model and if Swap is selected then $N^{(t)} + 1$ regressors are swapped without changing the model size (provided $p_\gamma \geq N^{(t)} + 1$; if not, the Add step is chosen).
This model proposal combines local moves with more global ones by changing simultaneously a block of variables. Two parameters determine this proposal: \( N \) is the maximum number of variables that can be changed from the current model and \( \zeta \) determines the degree of “localness” since the mean number of variables changed is \( 1 - \zeta + N \zeta \). The value of \( N \) will usually be fixed and the parameter \( \zeta \) chosen to control the mixing of the chain. The application of this proposal to microarray data by Lamnisos et al. (2009) suggests that the optimum effective sample size is obtained when the average acceptance rate falls in the range 0.25 to 0.40. This is true for a wide-range of sampling schemes. Rather like RWM samplers, this optimal choice of acceptance rate can be achieved by carefully tuning the parameter \( \zeta \) of the model proposal using a series of pilot runs. In each pilot run, the sampler is run for a chosen value of \( \zeta \) and the average acceptance rate calculated. If the acceptance rate is too high then \( \zeta \) is increased in the next run and if the acceptance rate is too low then \( \zeta \) is decreased in the next run. However, this tuning process is typically a computationally expensive task since trial and error is required.

As an alternative solution, we consider adaptive MCMC algorithms which can automatically handle this parameter tuning. The problem is similar to adaptation in RWM samplers since there is a tuneable proposal and an optimal acceptance rate to be achieved. We will extend the Adaptive Random Walk Metropolis (ARWM) algorithm proposed by Atchadé and Rosenthal (2005) to the variable selection problem. The ARWM algorithm automatically finds the optimal scale parameter in the RWM algorithm that results in the optimal acceptance rate \( \bar{\tau} = 0.234 \). Similarly, our adaptive MCMC algorithms automatically find \( \zeta \) such that the resulting acceptance rate falls in the range 0.25 to 0.4. A comprehensive survey of recent advances in adaptive MCMC methodology and their applications can be found in Rosenthal (2008).

3 Some Adaptive Algorithms

The ARWM algorithm of Atchadé and Rosenthal (2005) sequentially adapts the scale parameter \( \sigma \) of the RWM with a \( d \)-variate normal proposal density centred on the current value and with variance \( \sigma^2 \mathbf{I}_d \) and for which the stationary distribution is the positive continuous density \( \pi(x) \). The entire past of the stochastic process is used to adapt the scale parameter \( \sigma \). The resulting sequence of scale parameters \( \{\sigma^{(t)} : t \in \mathbb{N}\} \) converges to an optimal value that leads to the optimal acceptance rate \( \bar{\tau} = 0.234 \). Atchadé and Rosenthal (2005) fix values \( \delta_L \) and \( \delta_U \) such that \( 0 < \delta_L < \delta_U \) and defined the set \( \Delta = \{\sigma : \delta_L \leq \sigma \leq \delta_U\} \). They also assume that \( \Delta \) contains a unique value which results in the optimal acceptance rate \( \bar{\tau} \). Then,
they define the following function of $\sigma$

$$\rho(\sigma) = \begin{cases} 
\delta_L & \text{if } \sigma < \delta_L \\
\sigma & \text{if } \sigma \in \Delta \\
\delta_U & \text{if } \sigma > \delta_U.
\end{cases}$$

The aim of this function is to contain the adaptive algorithm inside $\Delta$. Finally, they define a discount factor $s(t)$, which is a positive sequence of real numbers such that $s(t) = O(t^{-\lambda})$ for some constant $1/2 < \lambda \leq 1$. This assumption ensures the ergodicity of the ARWM algorithm. The simple choice $s(t) = a t^{-1}$ for some $a > 0$ will meet this condition. The pseudocode of the ARWM algorithm proceeds as follows:

**Algorithm 5 (ARWM)** If at iteration $t$ the current state is $x(t) \in \mathbb{R}^d$ and the scale parameter of the proposal density is $\sigma(t) \in \Delta$, then

1. **Generate** $y \sim N(x(t), (\sigma(t))^2 I_d)$.

2. Set $x(t+1) = y$ with probability

   $$\alpha(x(t), y) = \min \left\{ 1, \frac{\pi(y)}{\pi(x(t))} \right\},$$

   otherwise take $x(t+1) = x(t)$.

3. **Compute**

   $$\sigma(t+1) = \rho(\sigma(t)) + s(t)(\alpha(x(t), y) - \bar{\tau})). \quad (6)$$

The acceptance rate is monitored by (6). The algorithm decreases the scale parameter $\sigma(t+1)$ when the acceptance rate is small and increases $\sigma(t+1)$ when the acceptance rate is high. Atchadé and Rosenthal (2005) showed under certain assumptions that the generated stochastic process is ergodic with stationary distribution the positive continuous density $\pi(x)$.

Turning to the MCMC algorithms of Section 2 with model proposal $q_\zeta(\gamma' | \gamma)$, the parameter $\zeta$ behaves like the scale parameter $\sigma$ of the RWM because values of $\zeta$ close to 0 yield more local moves and high acceptance rate and values of $\zeta$ close to 1 more global moves and small acceptance rate. Moreover, our applications to some gene expression datasets suggest an optimal acceptance rate $\bar{\tau}$ between 0.25 and 0.4. Therefore, we adopt ideas of the ARWM to develop adaptive version of each transdimensional MCMC sampler described in Section 2. In our case $\zeta \in [0, 1]$, thus values of $\delta_L$ and $\delta_U$ are chosen close to 0 and 1, respectively. In our applications of the adaptive algorithms, $\delta_L$ and $\delta_U$ are set to 0.01 and 0.99,
respectively. The parameter $\zeta$ can be made adaptive by updating it at the $t$-th iteration in the following way, analogous to the RWM,

$$\zeta^{(t+1)} = \rho(\zeta^{(t)} + s^{(t)}(\alpha^{(t)} - \bar{\tau}))$$

(7)

where $\alpha^{(t)}$ is the acceptance probability at the $t$-th iteration of the chain.

All the algorithms of Section 2 can be made adaptive by updating $\zeta$ at each iteration using the recursion in (7). The pseudocode representation of the adaptive MH-LR algorithm (denoted by ADMH-LR) adjusts the model proposal step and adds an extra step (step 3 below) to the corresponding non-adaptive algorithm (Algorithm 1).

Algorithm 6 (ADMH-LR) Let $\gamma$ be the current state of the chain and $\zeta^{(t)} \in \Delta$.

1. Select model $\gamma'$ with probability $q_{\zeta^{(t)}}(\gamma'|\gamma)$.

2. Jump to the model $\gamma'$ with probability

$$\alpha(\gamma, \gamma') = \min \left\{ 1, \frac{\pi(\gamma'|\gamma') \pi(\gamma') q(\gamma'|\gamma)}{\pi(\gamma|\gamma) \pi(\gamma) q(\gamma'|\gamma)} \right\}.$$

3. Compute

$$\zeta^{(t+1)} = \rho(\zeta^{(t)} + s^{(t)}(\alpha(\gamma, \gamma') - \bar{\tau})).$$

Similarly, the pseudocode for the adaptive version of the Holmes and Held algorithm for the probit regression model has the form

Algorithm 7 (ADH-H) If at iteration $t$ the current state is $(\gamma, \theta_\gamma)$ and $\zeta^{(t)} \in \Delta$, then

1. Generate $z_1, z_2, \ldots, z_n$ where $z_i \sim \mathcal{N}(\alpha + \beta^T \gamma, 1)$ and $z_i > 0$ if $y_i = 1$ or $z_i < 0$ if $y_i = 0$.

2. Select model $\gamma'$ with probability $q_{\zeta^{(t)}}(\gamma'|\gamma)$.

3. Jump to the model $\gamma'$ with probability

$$\alpha(\gamma, \gamma') = \min \left\{ 1, \frac{\pi(\gamma'|\gamma') \pi(\gamma') q(\gamma'|\gamma)}{\pi(\gamma|\gamma) \pi(\gamma) q(\gamma'|\gamma)} \right\}.$$

4. Compute

$$\zeta^{(t+1)} = \rho(\zeta^{(t)} + s^{(t)}(\alpha(\gamma, \gamma') - \bar{\tau})).$$

5. If $\gamma'$ is accepted, draw $\theta_{\gamma'} \sim \mathcal{N}((\tilde{X}_{\gamma'}^T \tilde{X}_{\gamma'} + \tilde{V}_{\gamma'}^{-1})^{-1} \tilde{X}_{\gamma'}^T z, (\tilde{X}_{\gamma'}^T \tilde{X}_{\gamma'} + \tilde{V}_{\gamma'}^{-1})^{-1})$.

The other algorithms of Section 2 can be made adaptive in the same way by changing $q(\gamma'|\gamma)$ to $q_{\zeta^{(t)}}(\gamma'|\gamma)$ and updating $\zeta$ after calculating the acceptance probability of the Metropolis-Hastings step.
3.1 Ergodicity of the algorithms

General conditions for ergodicity of the adaptive algorithms are discussed by Roberts and Rosenthal (2007) who establish two conditions which are sufficient:

- Simultaneous uniform ergodicity
- Diminishing adaptation

For the adaptive version of the algorithm for the linear regression setting, we can immediately state

**Theorem 1** Algorithm ADMH-LR is ergodic

**Proof**

The MH-LR algorithm with model proposal $q_\zeta(\gamma'|\gamma)$ has finite state space $\mathcal{X} = \{0, 1\}^p$ and is irreducible and aperiodic. According to Theorem 12.3.1 in Liu (2001) the MH-LR algorithm is geometrically ergodic for each $\zeta \in \Delta = [\delta_L, \delta_U]$. The proposal density of the ADMH-LR algorithm is for $k = 1, \ldots, N$

$$ q_\zeta(\gamma'|\gamma) = \begin{cases} 
\frac{1}{3 |\gamma^+|} \left( \binom{N - 1}{k - 1} \right) \zeta^{k-1}(1 - \zeta)^{N-k} & \text{if } \sum_{i=1}^p |\gamma'_i - \gamma_i| = k, \text{addition} \\
\frac{1}{3 |\gamma^0|} \left( \binom{N - 1}{k - 1} \right) \zeta^{k-1}(1 - \zeta)^{N-k} & \text{if } \sum_{i=1}^p |\gamma'_i - \gamma_i| = 2k, \text{swap} \\
\frac{1}{3 |\gamma^-|} \left( \binom{N - 1}{k - 1} \right) \zeta^{k-1}(1 - \zeta)^{N-k} & \text{if } \sum_{i=1}^p |\gamma'_i - \gamma_i| = k, \text{deletion} \\
0, \text{otherwise}, 
\end{cases} $$

where $|\gamma^+| = \#$ of neighboring models of $\gamma$ with dimension $p_\gamma + k$, $|\gamma^0| = \#$ of neighboring models of $\gamma$ with dimension $p_\gamma$ and $|\gamma^-| = \#$ of neighboring models of $\gamma$ with dimension $p_\gamma - k$. This proposal density is continuous and bounded in the compact interval $\Delta$. Therefore, the probability of moving to $\gamma'$

$$ p_\zeta(\gamma; \gamma') = \min \left\{ q_\zeta(\gamma'|\gamma), q_\zeta(\gamma|\gamma') \frac{\pi(\gamma'|y)}{\pi(\gamma|y)} \right\} $$

and the probability of accepting a proposal from $\gamma$

$$ \alpha_\zeta(\gamma) = \sum_{\gamma' \in \mathcal{X}} \min \left\{ q_\zeta(\gamma'|\gamma), q_\zeta(\gamma|\gamma') \frac{\pi(\gamma'|y)}{\pi(\gamma|y)} \right\} $$

11
are continuous functions in \( \Delta \). Consequently, the transition kernel
\[
P_{\zeta}(\gamma, \gamma') = (1 - \alpha_{\zeta}(\gamma))\delta_{\gamma} + p_{\zeta}(\gamma, \gamma')
\]
is a continuous functions of \( \zeta \in \Delta \), where \( \delta_{\gamma} \) is a point-mass at \( \gamma \). Then, from Lemma 1 of Roberts and Rosenthal (2007), the function
\[
f(\zeta) = ||P_{\zeta}^n(\gamma, \cdot) - \pi(\cdot)||
\]
is continuous in \( \Delta \) for each fixed \( n \in \mathbb{N} \), where \( ||P_{\zeta}^n(\gamma, \cdot) - \pi(\cdot)|| = \sup_{\Gamma \in \mathcal{X}} |P_{\zeta}^n(\gamma, \Gamma) - \pi(\Gamma)| \) is the usual total variation distance. Therefore, from the proof of Corollary 3 of Roberts and Rosenthal (2007), there is \( N = N(\varepsilon, \gamma) \in \mathbb{N} \) such that
\[
||P_{\zeta}^n(\gamma, \cdot) - \pi(\cdot)|| \leq \varepsilon \quad \text{for all} \quad \zeta \in \Delta.
\]
Letting \( N = \max_{\gamma \in \mathcal{X}} N(\varepsilon, \gamma) \) then the simultaneous uniform ergodicity condition is satisfied.

The condition of diminishing adaptation follows from the following argument: the transition kernel \( P_{\zeta}(\gamma, \gamma') \) exhibits diminishing adaptation because it is continuous with respect to \( \zeta \) and \( |\zeta^{(t+1)} - \zeta^{(t)}| \to 0 \) as \( t \to \infty \). \( \square \)

The use of Adaptive Monte Carlo methods in Metropolis-within-Gibbs sampler is discussed by Haario et al. (2005), Roberts and Rosenthal (2009) and Latuszyński et al. (2011). Our algorithms are simpler than the methods considered by these authors since there is only adaptation in one step of the sampler. Therefore, the chain will be ergodic if the update involving the adaptation is ergodic. For the adaptive versions of the algorithms for the probit regression setting, we can state

**Theorem 2** The algorithms ADH-H, ADAG and ADC-M are ergodic

**Proof**
The H-H, AG and C-M algorithms are \( \phi \)--irreducible and aperiodic and this implies that they are ergodic. The ADH-H is ergodic because the update involving the adaptation is the same as for the ADMH-LR.

In the case of ADAG and ADC-M, the probability of moving to \( \gamma' \) with \( p_{\gamma'} > p_{\gamma} \) is
\[
p_{\zeta}(\gamma, \gamma') = \int_{\epsilon} \alpha_{\zeta}(\gamma, \gamma', \theta_{\gamma}, \theta_{\gamma'}) q_{\zeta}(\gamma' | \gamma) q(\epsilon) d\epsilon,
\]
where \( q(\epsilon) \) is the density function of a standard normal. The probability (8) is a continuous functions of \( \zeta \in \Delta \) from the dominated convergence theorem. The probability of moving to \( \gamma' \) with \( p_{\gamma'} \leq p_{\gamma} \)
\[
p_{\zeta}(\gamma, \gamma') = \min \left\{ q_{\zeta}(\gamma' | \gamma), q_{\zeta}(\gamma' | \gamma') \frac{\pi(\gamma', \theta_{\gamma'}) \pi(y | \theta_{\gamma'})}{\pi(\gamma, \theta_{\gamma}) \pi(y | \theta_{\gamma})} \times K' \right\},
\]
is also a continuous function of $\zeta \in \Delta$, where

$$
K' = \begin{cases} 
\frac{|B_{\gamma'}|}{|B_{\gamma}|} \times K & \text{for ADAG} \\
\sigma^{(p_{\gamma'} - p_{\gamma})} \times K & \text{for ADC-M}.
\end{cases}
$$

Therefore, the probability of accepting a proposal from $\gamma$

$$
\alpha_{\zeta}(\gamma) = \sum_{\gamma' \in X} p_{\zeta}(\gamma, \gamma')
$$

and the transition kernel

$$
P_{\zeta}(\gamma, \gamma') = (1 - \alpha_{\zeta}(\gamma)) \delta_{\gamma} + p_{\zeta}(\gamma, \gamma')
$$

are continuous functions of $\zeta \in \Delta$. A similar argument to the proof of Theorem 1 shows that the update involving the adaptation satisfies the simultaneous uniform ergodicity condition.

The transition kernel $P_{\zeta}(\gamma, \gamma')$ exhibits diminishing adaptation because it is continuous with respect to $\zeta$ and $|\zeta^{(t+1)} - \zeta^{(t)}| \to 0$ as $t \to \infty$. □

4 Illustrations

The performance of the adaptive MCMC algorithms is evaluated using examples from probit regression and linear regression. All the adaptive MCMC samplers start with initial parameter value $\zeta_0 = 0.5$ and use the positive sequence of real numbers $\{s^{(t)} = \zeta_0/t : t \in \mathbb{N}\}$. The number of iterations was 2,000,000, the burn-in period 100,000 and the thinning 10 resulting in an MCMC sample size $T$ of 190,000. Finally, we specify the value 0.3 as an optimal acceptance rate $\bar{\tau}$ because the optimum effective sample size of the MCMC algorithms that explore the model and parameter space of our problem is obtained when acceptance rates are between 0.25 and 0.4. Adopting $\bar{\tau} = 0.234$ instead makes very little difference to our results. We compare the adaptive version to its non-adaptive counterpart using the parameter settings $\zeta = 0, 0.25, 0.5, 0.75, 0.95$ and $N = 4$. In both applications, we assume that $V_{\gamma}$ is a diagonal matrix $cI_{p_{\gamma}}$. This implies that the coefficients are independent a priori and we choose $c = 5$ which is the value chosen by Sha et al. (2004). We also use mean prior model size $pw$ equal to five.

The efficiency of an MCMC sampler can be measured using the Effective Sample Size (ESS) which is $T/(1 + 2 \sum_{j=1}^{\infty} \rho_j)$ for an MCMC run of length $T$ with lag $j$ autocorrelation $\rho_j$ (see e.g., Liu (2001). The interpretation is that the
MCMC sampler has the same accuracy of estimates as a Monte Carlo sampler (where all the draws are independent) run for ESS iterations. In this paper, the MCMC output monitored consists of the components $\gamma_i$ of $\gamma$ since the posterior inclusion probabilities are the main quantities of interest in variable selection. An estimate of the integrated autocorrelation time $\tau_i = 1 + 2 \sum_{j=1}^{\infty} \rho_j$ for each $\gamma_i$ was computed using the Initial Positive Sequence Estimator (Geyer, 1992). We calculate the mean $m$ of $\tau_i$’s for each algorithm and estimate the Effective Sample Size by $\text{ESS} = \frac{T}{m}$. The algorithms have different running times and so we define the efficiency ratio for a sampler to be

$$\text{ER}(\text{Sampler}) = \frac{\text{ESS}(\text{Sampler})}{\text{CPU}(\text{Sampler})},$$

which standardizes the effective sample size by CPU run time and so penalizes computationally inefficient algorithms. We are interested in the performance of each adaptive algorithm to the non-adaptive algorithm with $\zeta = 0$ (which is the standard MCMC proposal for these types of models and represents a baseline) and with the optimal value of $\zeta$ among five candidates ($\zeta = 0, 0.25, 0.5, 0.75, 0.95$), which is the value resulting in the highest ER. The relative efficiency of the non-adaptive over the adaptive algorithm is defined by

$$\text{R.E} = \frac{\text{ER}(\text{Non-Adaptive})}{\text{ER}(\text{Adaptive})}.$$  

### 4.1 Probit Regression

The performance of the adaptive MCMC algorithms in probit regression is evaluated using two microarray datasets. These are the Arthritis data (Sha et al., 2003) for which we have $n = 31$ observations and $p = 755$ possible regressors and the Colon Tumour data (Alon et al., 1999) with $n = 62$ and $p = 1224$. Adaptive versions of all algorithms in Section 2 were tested and are denoted as follows:

1. ADH-H : Adaptive Holmes and Held algorithm
2. ADAG-LA : Adaptive automatic generic sampler with Laplace approximation
3. ADAG-IWLS : Adaptive automatic generic sampler with Iterated Weighted Least Squares approximation

Non-adaptive versions of the algorithms are indicated by dropping the first two letters “AD”.
Table 1: The effective sample size ESS, the CPU time in seconds of the adaptive and non-adaptive algorithms with relative efficiencies of the non-adaptive algorithm over the adaptive algorithm for the Arthritis and Colon Tumour datasets.

**Arthritis data**

<table>
<thead>
<tr>
<th>Method</th>
<th>ESS</th>
<th>CPU</th>
<th>R.E</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-H ((\zeta = 0))</td>
<td>35144</td>
<td>9795</td>
<td>0.99</td>
</tr>
<tr>
<td>H-H ((\zeta = 0.25))</td>
<td>35489</td>
<td>9880</td>
<td>1.00</td>
</tr>
<tr>
<td>ADH-H</td>
<td>34902</td>
<td>9668</td>
<td></td>
</tr>
<tr>
<td>AG-LA ((\zeta = 0))</td>
<td>58001</td>
<td>40135</td>
<td>0.75</td>
</tr>
<tr>
<td>AG-LA ((\zeta = 0.5))</td>
<td>80701</td>
<td>40346</td>
<td>1.04</td>
</tr>
<tr>
<td>ADAG-LA</td>
<td>80853</td>
<td>41966</td>
<td></td>
</tr>
<tr>
<td>AG-IWLS ((\zeta = 0))</td>
<td>50942</td>
<td>9781</td>
<td>0.74</td>
</tr>
<tr>
<td>AG-IWLS ((\zeta = 0.5))</td>
<td>68233</td>
<td>9870</td>
<td>0.98</td>
</tr>
<tr>
<td>ADAG-IWLS</td>
<td>69126</td>
<td>9822</td>
<td></td>
</tr>
<tr>
<td>C-M ((\zeta = 0))</td>
<td>55641</td>
<td>49173</td>
<td>0.95</td>
</tr>
<tr>
<td>C-M ((\zeta = 0.5))</td>
<td>66801</td>
<td>49993</td>
<td>1.12</td>
</tr>
<tr>
<td>ADC-M</td>
<td>64243</td>
<td>53944</td>
<td></td>
</tr>
</tbody>
</table>

**Colon Tumour data**

<table>
<thead>
<tr>
<th>Method</th>
<th>ESS</th>
<th>CPU</th>
<th>R.E</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-H ((\zeta = 0))</td>
<td>32752</td>
<td>13968</td>
<td>0.99</td>
</tr>
<tr>
<td>ADH-H</td>
<td>32228</td>
<td>13587</td>
<td></td>
</tr>
<tr>
<td>AG-LA ((\zeta = 0))</td>
<td>47354</td>
<td>41355</td>
<td>0.97</td>
</tr>
<tr>
<td>AG-LA ((\zeta = 0.25))</td>
<td>53653</td>
<td>42345</td>
<td>1.08</td>
</tr>
<tr>
<td>ADAG-LA</td>
<td>54777</td>
<td>45716</td>
<td></td>
</tr>
<tr>
<td>AG-IWLS ((\zeta = 0))</td>
<td>45494</td>
<td>14165</td>
<td>0.88</td>
</tr>
<tr>
<td>AG-IWLS ((\zeta = 0.25))</td>
<td>51971</td>
<td>14246</td>
<td>1.00</td>
</tr>
<tr>
<td>ADAG-IWLS</td>
<td>51231</td>
<td>14066</td>
<td></td>
</tr>
<tr>
<td>C-M ((\zeta = 0))</td>
<td>39596</td>
<td>48159</td>
<td>0.94</td>
</tr>
<tr>
<td>C-M ((\zeta = 0.25))</td>
<td>42370</td>
<td>48313</td>
<td>1.00</td>
</tr>
<tr>
<td>ADC-M</td>
<td>41861</td>
<td>47789</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 presents results of the adaptive algorithms and various non-adaptive algorithms with fixed values of \(\zeta\) (the standard choice of \(\zeta = 0\) and the optimal value among the five values mentioned above) for the Arthritis and Colon Tumour datasets. The relative efficiency for all sampling methods against the standard...
proposal ($\zeta = 0$) is always less than 1 indicating that the adaptive method is superior. The increase in performance depends on the particular method and the form of the posterior. However, the effect can be large in some cases. For example, the standard method only obtains 75% of the efficiency of the adaptive method with the Arthritis data and AG-LA and AG-IWLS algorithms. It is also clear that the Automatic Generic methods gain the most benefit. In fact, the effective sample size of the adaptive algorithms are very similar to the optimal non-adaptive algorithms in terms of mixing. Furthermore, the increase in CPU time of the adaptive algorithms is small. This leads to relative efficiencies quite close to 1 and therefore the adaptive algorithms achieve essentially the same efficiency as the optimal non-adaptive algorithms. Crucially, however, the adaptive algorithms avoid the pilot runs needed to tune the model proposal parameter $\zeta$.

Figure 1 and Figure 2 show the trace plots of both the model proposal parameter $\zeta$ (left panels) and the empirical acceptance rate (right panels) of the adaptive algorithms for the Arthritis and Colon Tumour datasets, respectively. The param-
Figure 2: Trace plots of the model proposal parameter $\zeta$ and the empirical acceptance rate of the adaptive algorithms for the Colon Tumour dataset; MCMC iterations are represented on a log scale.

Parameter $\zeta$ of each adaptive algorithm converges to a value close to the optimal one obtained by manual tuning. Furthermore, the empirical acceptance rates converge to values quite close to the target acceptance rate 0.3. These results illustrate that the adaptive MCMC algorithms automatically find model proposal parameters $\zeta$ that give asymptotically the optimal acceptance rate $\overline{\tau} = 0.3$.

Figure 3 displays the scatter-plots of the log estimated posterior gene inclusion probabilities of the adaptive and optimal non-adaptive algorithms for the Arthritis and Colon Tumour datasets. The log posterior gene inclusion probabilities are very similar indicating empirically that the stationary distribution of the stochastic process generated by the adaptive MCMC algorithms is the target posterior distribution $\pi(\gamma, \theta, y)$. 
Figure 3: Scatter-plots of the log estimated posterior gene inclusion probabilities of the adaptive and optimal non-adaptive algorithms for the Arthritis and Colon Tumour datasets

4.2 Linear Regression

An adaptive version of the MH-LR algorithm was applied to the Tecator dataset \((n = 172, p = 100)\) which is discussed in Griffin and Brown (2010). Figure 4 displays the estimated posterior inclusion probabilities of the MH-LR algorithm with \(\zeta = 0\). The posterior inclusion probabilities are very similar for different values of \(\zeta\). Figure 5 shows how the new general model proposal improves the
Figure 4: Estimated posterior inclusion probabilities and scatter-plot of the log estimated posterior inclusion probabilities with the Tecator data for different values of $\zeta$

ESS, even though it decreases the between-model acceptance rate. The MH-LR algorithm has maximum ESS for $\zeta = 0.25$ which gives an acceptance rate of 0.33. This acceptance rate falls in the range 0.25 to 0.4 and this result is consistent with that found in probit regression.

Figure 5: Effective sample size and acceptance rate of the MH-LR algorithms for five different model proposal parameters using the Tecator data

Table 2 presents results of the adaptive and non-adaptive algorithms with fixed values of $\zeta$. The relative efficiency of the sampling method with standard proposal ($\zeta = 0$) is less than 1 indicating that the adaptive method is superior. The relative efficiency of the optimal non-adaptive algorithm is almost 1 and therefore the adaptive algorithm achieves essentially the same efficiency as the optimal non-adaptive algorithms.

Figure 6 shows the trace plots of both the model proposal parameter $\zeta$ (left panel) and the empirical acceptance rate (right panel) of the adaptive algorithm for
Table 2: The effective sample size ESS, the CPU time in seconds of the adaptive and non-adaptive algorithms with relative efficiencies of the non-adaptive algorithm over the adaptive algorithm for the Tecator dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>ESS</th>
<th>CPU</th>
<th>R.E</th>
</tr>
</thead>
<tbody>
<tr>
<td>MH-LR ($\zeta = 0$)</td>
<td>27222</td>
<td>25941</td>
<td>0.91</td>
</tr>
<tr>
<td>MH-LR ($\zeta = 0.25$)</td>
<td>30134</td>
<td>26132</td>
<td>1.01</td>
</tr>
<tr>
<td>ADMH-LR</td>
<td>30351</td>
<td>26241</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6: Trace plots of the model proposal parameter $\zeta$ and the empirical acceptance rate of the adaptive MH-LR algorithm for the Tecator dataset; MCMC iterations are represented on a log scale

the Tecator dataset. The parameter $\zeta$ of the adaptive algorithm converges to a value close to the optimal one obtained by manual tuning and the empirical acceptance rates converge to a value quite close to the target acceptance rate 0.3. This result illustrates that the adaptive MH-LR algorithm automatically finds model proposal parameters $\zeta$ that give asymptotically the optimal acceptance rate $\bar{\tau} = 0.3$.

Figure 7 displays the scatter-plot of the log estimated posterior inclusion probabilities of the adaptive and optimal non-adaptive algorithms for the Tecator dataset. These log posterior inclusion probabilities are very similar for both algorithms.

5 Discussion

This paper describes an adaptive Monte Carlo algorithm for posterior simulation for variable selection in generalized linear models with many regressors, where acceptance rates for standard MCMC algorithms tend to be fairly large. The algorithm leads to Markov chains with good mixing properties without the need for pilot runs. In fact, the effective sample sizes for the adaptive algorithms are almost identical to those for the algorithms run at an optimized value of the pro-
Figure 7: Scatter-plot of the log estimated posterior inclusion probabilities of the adaptive and optimal non-adaptive algorithms for the Tecator dataset

proposal parameter $\zeta$ (found using trial-and-error). The methods are useful when there are a large number of variables that could potentially be included in the model, which leads to high average acceptance rates for standard algorithms. If the number of regressors is not large, then acceptance rates will not be high and an average acceptance rate of 0.3 may not be achievable. In this case, $\zeta$ should be close to zero and the value of $\zeta^{(t)}$ will converge to a value close to zero showing the robustness of the algorithm. Therefore, we suggest the use of adaptive MCMC algorithms to efficiently explore the model space using Bayesian variable selection in linear and probit regression problems with many covariates. Extensions to other types of GLMs can also be accommodated. For example, Frühwirth-Schnatter and Frühwirth (2006, 2007) propose an auxiliary mixture sampling approach which uses approximations through mixtures of normal distributions to develop easy Gibbs sampling schemes for Poisson and logistic regression models, respectively. The MCMC algorithms for such models could be made adaptive using the same ideas as developed in this paper.

References


