Code for *Adaptive MCMC schemes for Bayesian variable selection problems*

J. E. Griffin, K. Łatuszyński and M. F. J. Steel*

Abstract

Instructions for Matlab code implementing two adaptive MCMC algorithms for Bayesain variable selection, Individual adaptation and Adaptive Scaled Independence, are given.

Keywords: Bayesian variable selection; spike-and-slab priors; high-dimensional daa; large p, small n problems; linear regression

1 Introduction

The model fitted is

$$y = \alpha \mathbf{1} + X_\gamma \beta_\gamma + \epsilon, \qquad \epsilon \sim \mathrm{N}(0, \sigma^2 I_n) \tag{1}$$

*Jim Griffin is Professor, School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, CT2 7NF, U.K. (Email: J.E.Griffin-28@kent.ac.uk), Krys Łatuszyński is Royal Society University Research Fellow and Associate Professor (Email: K.G.Latuszynski@warwick.ac.uk) and Mark Steel is Professor, Department of Statistics, University of Warwick, Coventry, CV4 7AL, U.K. (Email: m.steel@warwick.ac.uk). The authors are grateful to Błażej Miasojedow for helpful comments.

where y is an $(n \times 1)$-dimensional vector of responses, $X = (x_1, \ldots, x_p)$ is an $(n \times p)$-dimensional data matrix and $\gamma = (\gamma_1, \ldots, \gamma_p)$ is a vector of indicator variables in which γ_i denotes whether the i-th variable is included in the model (when $\gamma_i = 1$). The matrix X_γ is formed by those columns of X corresponding to included variables. The prior is

$$p(\alpha, \sigma^2, \beta_\gamma, \gamma) \propto \sigma^{-2} p(\beta_\gamma | \sigma^2, \gamma) p(\gamma) \tag{2}$$

with $\beta_\gamma | \sigma^2, \gamma \sim \mathrm{N}(0, \sigma^2 V_\gamma)$ and $p(\gamma) = h^{p_\gamma}(1-h)^{p-p_\gamma}$ where $p_\gamma = \sum_{j=1}^{p} \gamma_j$.

2 Functions

```
[output] = EIA_sampler(data, target, g, gprior, mode, RB, hparam,
 nu, burnin, numbofits, thin, numbofreps, heat, adap_type)
```

- `data` – An $(n \times p)$-dimensional data matrix X
- `target` – An $(n \times 1)$-dimensional response y
- `g` – The value of the hyperparameter g
- `gprior` – if `gprior` is 1, a g-prior is used for β_γ with $V_\gamma = g\left(X_\gamma^T X_\gamma\right)^{-1}$. Otherwise, an independence prior for β_γ with $V_\gamma = g\, I$ is used.
- `mode` – if `mode` is 2, samples of the model are stored. Otherwise, their values are not stored.
- `RB` – if `RB` is 1, a Rao-Blackwellised estimates of the posterior inclusion probabilities is calculated.
- `hparam` – If the length of `hparam` is 1, this is the value of h. If the length of `hparam` is 2, $h \sim \mathrm{Be}(a, b)$ where `hparam=[a b]`.
- `nu` – Parameter controlling the starting values of A and D (see paper for further details).

- `burnin` – The number of iterations used in the burn-in period.
- `numbofits` – The total number of samples recorded for each multiple chain.
- `thin` – The thinning level. Every `thin`-th iteration is recorded after the burn-in period.
- `numbofreps` – The number of multiple chains
- `heat` – The initial temperature schedule for a parallel tempering version of the algorithm. The version without parallel tempering occurs if `heat=1`.
- `adap_type` – if `adap_type=1`, there is infinite adaptation. If `adap_type=2`, adaptation is restricted to the burn-in period.
- `output` is a Matlab structure containing
  - `prob_inclusion` – Estimated posterior inclusion probabilities which is the Rao-Blackwellised estimate if `RB=1` or an empirical average otherwise.
  - `prob_inclusion2` – Estimated posterior inclusion probabilities (without using Rao-Blackwellisation)
  - `logpost` – The log posterior value at every recorded iteration
  - `modelsize` – A sample of model sizes
  - `accept` – The average acceptance probability calculated at every recorded iteration.
  - `gamma` – A sample of models, γ (if `mode=2`)
  - `zetaAS` – The final value of A
  - `zetaDS` – The final value of D
  - `accept2` –The average acceptance probability of swaps between consecutive tempered chains
  - `heat` –The final temperature schedule

`[output] = ASI_sampler(data, target, g, gprior, mode, RB, hparam, tau, nu, burnin, numbofits, thin, numbofreps, heat, adap_type)`

- `data` – An $(n \times p)$-dimensional data matrix X
- `target` – An $(n \times 1)$-dimensional response y
- `g` – The value of the hyperparameter g
- `gprior` – if `gprior` is 1, a g-prior is used for β_γ with $V_\gamma = g\left(X_\gamma^T X_\gamma\right)^{-1}$. Otherwise, an independence prior for β_γ with $V_\gamma = g\, I$ is used.
- `mode` – if `mode` is 2, samples of the model are stored. Otherwise, their values are not stored.
- `RB_adap` – if `RB_adap` is 1, the PIP's used in the adaptive proposal are updated using the Rao-Blackwellisation after the burn-in, otherwise the PIP's used in the adaptive proposal are updated using the current values of γ only.
- `RB` – if `RB` is 1, a Rao-Blackwellised estimates of the posterior inclusion probabilities is calculated.
- `hparam` – If the length of `hparam` is 1, this is the value of h. If the length of `hparam` is 2, $h \sim \text{Be}(a, b)$ where `hparam=[a b]`.
- `tau` – The target acceptance rate τ
- `nu` – Parameter controlling the starting values of A and D (see paper for further details).
- `burnin` – The number of iterations used in the burn-in period.
- `numbofits` – The total number of samples recorded for each multiple chain.
- `thin` – The thinning level. Every `thin`-th iteration is recorded after the burn-in period.
- `numbofreps` – The number of multiple chains

- `heat` – The initial temperature schedule for a parallel tempering version of the algorithm. The version without parallel tempering occurs if `heat=1`.
- `adap_type` – if `adap_type=1`, there is infinite adaptation. If `adap_type=2`, adaptation is restricted to the burn-in period.
- `output` is a Matlab structure containing
  - `prob_inclusion` – Estimated posterior inclusion probabilities which is the Rao-Blackwellised estimate if `RB=1` or an empirical average otherwise.
  - `prob_inclusion2` – Estimated posterior inclusion probabilities (without using Rao-Blackwellisation)
  - `logpost` – The log posterior value at every recorded iteration
  - `modelsize` – A sample of model sizes
  - `accept` – The average acceptance probability calculated at every recorded iteration.
  - `gamma` – A sample of models, γ (if `mode=2`)
  - `zeta` – The final value of ζ
  - `zetaAS` – The final value of A
  - `zetaDS` – The final value of D
  - `accept2` –The average acceptance probability of swaps between consecutive tempered chains
  - `heat` –The final temperature schedule