Code for *Individual adaptation: an adaptive MCMC scheme for variable selection problems*

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

Abstract

Individual adaptation is an algorithm with adaptive proposals for variable selection in linear regression models. Instruction for an MCMC individual adaptation algorithm with parallel tempering and an SMC individual adaptation algorithm with parallel tempering are described.

1 Introduction

The model that is fitted is

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

where $\mathbf{1}$ is an $(n \times 1)$-dimensional vector of 1's, X_γ is the sub-matrix of X where the i-th column is included if $\gamma_i = 1$, β_γ is a $(p_\gamma \times 1)$-dimensional vector and $\epsilon \sim \mathrm{N}(0, \sigma^2 I_n)$. It will be useful to define the notation $\theta_\gamma = (\alpha, \beta_\gamma)$. The priors are

$$p(\alpha, \sigma^2, \beta_\gamma, \gamma) \propto \sigma^{-2} p(\beta_\gamma | \sigma^2, \gamma) p(\gamma) \tag{1}$$

with

$$\beta_\gamma | \sigma^2, \gamma \sim \mathrm{N}(0, \sigma^2 V_\gamma) \qquad \text{and} \qquad p(\gamma) = h^{p_\gamma} (1-h)^{p - p_\gamma}.$$

*School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, CT2 7NF, U.K. and † Department of Statistics, University of Warwick, Coventry, CV4 7AL, U.K.

2 Functions

```
[output] = IA_PT_sampler(data, target, g, mode,
           RAPA, hparam, tau, nu, burnin, numbofits, thin,
           numbofreps, heat)
```

- `data` – An $(n \times p)$-dimensional data matrix X.
- `target` – An $(n \times 1)$-dimensional response Y.
- `g` – The hyperparameter g.
- `gprior` – If `gprior` is 1, a g-prior is used for β_γ with $V_\gamma = g(X_\gamma^T X_\gamma)^{-1}$. Otherwise, an independence prior for β_γ with $V_\gamma = gI$.
- `mode` – If `mode` is 2, samples of the model are stored. Otherwise, their values are not stored.
- `RAPA` – If the `RAPA` is 1, the RAPA-IA variant of the algorithm is run. Otherwise, the IA variant is run.
- `hparam` – If the length of `hparam` is 1, this is the value of h. If the length of `hparam` is 2, $h \sim$ Be(a, b) where `hparam=[a b]`.
- `tau` – The target acceptance rate τ.
- `nu` – Parameter controlling 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 thining level. Every `thin`-th iteration is recorded after the burn-in period.
- `numbofreps` – The number of multiple chains used in the MCA-IA, MCA-RAPA-IA, MCA-IA-PT and MCA-RAPA-IA-PT. The single chain versions of the these algorithms IA, RAPA-IA, IA-PT and RAPA-IA-PT can be run by setting `numbofreps=1`.
- `heat` – The initial temperature schedule. Note: this is a vector of the form $(t_m, t_{m-1}, \ldots, t_1)$. The version without parallel tempering occurs if we set `heat=1`.
- `output` is a Matlab structure containing
  - `prob_inclusion` – Estimated posterior inclusion probabilities.
  - `logpost` – The log posterior value at every recorded iteration.

- `modelsize` – A sample of model sizes.
- `accept` – The acceptance probability 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 acceptance probability of swaps between consecutive tempered chains.
- `heat` – The final temperature scale.

```
[output] = IA_SMC_sampler(data, target, g,
           RAPA, hparam, tau, c_ESS, nu, burnin, numbofits, thin,
           numbofreps, heat)
```

- `data` – An $(n \times p)$-dimensional data matrix X.
- `target` – An $(n \times 1)$-dimensional response Y.
- `g` – The hyperparameter g.
- `gprior` – If `gprior` is 1, a g-prior is used for β_γ with $V_\gamma = g(X_\gamma^T X_\gamma)^{-1}$. Otherwise, an independence prior for β_γ with $V_\gamma = gI$.
- `RAPA` – If the `RAPA` is 1, the RAPA-IA variant of the algorithm is run. Otherwise, the IA variant is run.
- `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.
- `c_ESS` – This parameter controls the ESS at the next temperature in the adaptive scheme. The ESS will be `c_ESS`$\times N$ where N is the number of particles.
- `nu` – Parameter controlling starting values of A and D (see paper for further details).
- `numbofparts` – The number of particles.
- `numbofits` – The number of iterations run at each step the parallel tempering scheme.
- `output` is a Matlab structure containing
  - `prob_inclusion` – Estimated posterior inclusion probabilities.
  - `logpost` – The log posterior value at every recorded iteration.
  - `modelsize` – A sample of model sizes.
  - `gamma` – A sample of models, γ.
  - `zetaAS` – The final value of A.
  - `zetaDS` – The final value of D.