

Adaptive MC³ and Gibbs Algorithms for Bayesian Model Averaging in Linear Regression Models

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

Abstract

The MC³ (Madigan and York, 1995) and Gibbs (George and McCulloch, 1997) samplers are the most widely implemented algorithms for Bayesian Model Averaging (BMA) in linear regression models. These samplers draw a variable at random in each iteration using uniform selection probabilities and then propose to update that variable. This may be computationally inefficient if the number of variables is large and many variables are redundant. In this work, we introduce adaptive versions of these samplers that retain their simplicity in implementation and reduce the selection probabilities of the many redundant variables. The improvements in efficiency for the adaptive samplers are indicated in real and simulated datasets.

Keywords: Linear Regression; Bayesian Model Averaging; Gibbs sampler; Adaptive MCMC

Supplementary Material

All the supplemental files are contained in a single zipped archive and can be obtained via a single download.

*Cyprus University of Technology, University of Kent and University of Warwick

1. Data Sets

`FLSgrowth.mat` The FLS data set contains economic growth data for 72 countries.

The vector Y is the average per capita GDP growth over the period 1960-1992 while the matrix X is the centred design matrix containing measurements for 41 determinants of economic growth.

`SDMgrowth.mat` The SDM data set contains economic growth data for 88 countries.

The vector Y is the average per capita GDP growth over the period 1960-1996 while the matrix X is the centred design matrix containing measurements for 67 determinants of economic growth.

2. Computer Code

`ADMC3.m` This MATLAB file implements the adaptive MC^3 algorithms for Bayesian Model Averaging in linear regression. The user is responsible for setting the response variable Y , the design matrix X and the prior setting on the single parameter g `PRIOR_G`. There are two available prior settings on g , the g -BRIC prior of Fernández et al. (2001) (g -BRIC') and the Hyper- g/n prior of Liang et al. (2008) ('Hyper'). The user is also responsible for choosing whether to use ('Yes') or not ('No') a prior on the mean of model size `PRIOR_W`. In the case of choosing 'Yes' then a Beta prior is used on w and W is the prior choice for the mean of model size. The `METHOD` to be used is also set by the user. There are three methods, the 'MC3' implementing the MC^3 algorithm, the 'Sigma' implementing the adaptive MC^3 algorithm that use the sample variances to update the variable selection probabilities and the 'PIP' implementing the adaptive MC^3 algorithm that use the inclusion frequencies to update the variable selection probabilities. The other user's input are the number of MCMC iterations `NUM_ITER`, the burn-in period `NUM_BURN` and the thinning of the chain `NUM_THIN`. The output are the posterior inclusion probabilities `PROB_INCLUSION`, their effective sample sizes

ESS, the posterior sample of visited models and the between-model acceptance rate ACCEPTANCE.

`ADGibbs.m` This MATLAB file implements the adaptive Gibbs algorithm for Bayesian Model Averaging in linear regression. The user guidelines for this program are analogous to those for `ADMC3.m`.

References

- Fernández, C., E. Ley, and M. F. J. Steel (2001). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics* 16, 563–76.
- George, E. I. and R. E. McCulloch (1997). Approaches for Bayesian variable selection. *Statistica Sinica* 7, 339–373.
- Liang, F., R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger (2008). Mixture of g -priors for Bayesian variable selection. *Journal of the American Statistical Association* 103, 410–423.
- Madigan, D. and J. York (1995). Bayesian graphical models for discrete data. *International Statistical Review* 63, 215–232.