

Documentation for Software Package SBAMS

Xiaoquan Wen

June 20, 2013

1 Introduction

Software package SBAMS (**S**tructured **B**ayesian **M**odel **S**election) implements the Bayesian statistical methods discussed in [1], and provides the functionality to perform structured Bayesian model selections in the Multivariate Linear Regression (MVLRL) models. The relevant technical details are discussed in the paper and its supplementary materials. The purpose of this documentation is to explain the software implementation and corresponding input/output formats.

2 Approximate Bayes Factor Computation

2.1 R implementation

The file `mvlr.R` implements the approximate Bayes factor calculation for the MVLRL model. The code is self-explanatory and it is convenient to use model comparisons in a set of candidate MVLRL models.

2.2 C++ implementation

We also implement the C++ class MVLRL, which can be conveniently re-used in the C++ implementations of other software packages. The public and private interfaces of the MVLRL class are documented in the header file `MVLRL.h`. The compilation of the object file requires GNU Scientific Library (GSL).

3 MCMC Algorithm for Bayesian Model Selection

We also implement the MCMC algorithm described in [1] to perform Bayesian model selection for MVLRL model using the C++ programming language. The program, *sbams-mvrl*, can be used for some

rather general purposes. Because the usual multiple linear regression model can be viewed a special case of MVLR, *sbams_mvlr* applies in the setting where a multiple linear regression suffices.

3.1 Input Files

There are two required input files containing data and prior information respectively.

The data file contains a data matrix for response and covariate variables used in the MVLR. Each row of the data matrix has the following format,

```
header  variable_id  value_1  value_2 ... value_n
```

In particular, the “header” column takes values either “response” or “covariate”, indicating the role of a particular variable in the MVLR; the “variable_id” column contains a string identifier, and the following columns contain the numerical values of the variable for all the subjects. The following data section is excerpted from a simulated data set:

```
response F -0.059   0.168 ...
response L  1.072  -2.986 ...
response T  1.955  -5.209 ...
covariate rs1  0.76 -0.24 ...
covariate rs2  0.47 -0.53 ...
covariate rs3 -0.17  0.83 ...
...
```

3.2 Running Program

3.3 Program Parameters

3.4 Input/Output Options

3.5 Program Options

3.6 Output from Program

References

- [1] Wen, X. “Bayesian Model Selection in Complex Linear Systems, as Illustrated in Genetic Association Studies”, submit to Biometrics.