

# fastBVSr Manual

Quan Zhou (Advisor: Yongtao Guan)

May 26, 2017

This software package is a faster implementation of the Bayesian variable selection regression (BVSr) of Guan and Stephens [2011], which was developed for large-scale multi-SNP association analysis. Its usage is very similar to the program **piMASS**, which can be downloaded at <http://haplotype.org/software.html>.

A typical command for running fastBVSr is

```
./fastBVSr -g test.mg.txt -p test.ph -w 10000 -s 100000 -o try1
```

This command will read the mean genotype file “test.mg.txt” and the phenotype file “test.ph” and then run MCMC with 10000 burn-in iterations and 100000 sampling iterations. All the output files start with prefix “try1”. A test dataset is provided in the package.

## 1 Input File Formats

The mean genotype file format and the phenotype format are the same as in **piMASS**.

**Mean Genotype File** No header row. Each row is a SNP. Columns can be separated by tab, space or comma. The first column is the SNP name (NO whitespace or commas in SNP names); the second is the minor allele; the third is the major allele. The rest columns are the mean genotypes (numbers of copies of the minor allele) of every individual. “NA” or “?” represents a missing value. Example:

```
rs1  A  T  0  1  2
rs2  G  C  1  ?  1.05
```

**Phenotype File** No header row. The  $n$ -th line is the phenotypic value (a number) of the  $n$ -th individual. Missing values are not allowed. Example:

```
1
0.4
-0.05
```

## 2 Understanding the Output

For most users, the only output file of interest is the `***.beta.txt`. It contains 6 columns: the first column is the SNP name; the second is the  $\log_{10}$  BF computed from a single-SNP regression model with prior effect size  $\sigma = 0.2$ ; the third column is the posterior inclusion probability (PIP) estimate from MCMC, which can be expressed as

$$\mathbb{E}(\gamma_j \mid \mathbf{y}) = \mathbb{P}(\gamma_j = 1 \mid \mathbf{y})$$

where  $\gamma_j = 1$  if the  $j$ -th SNP is included in the model (thus associated with the SNP) and  $\gamma_j = 0$  otherwise; the fourth column is the Rao-Blackwellized estimate for PIP; the fifth column gives the estimate for the posterior mean of the effect size, which can be expressed as  $\mathbb{E}(\beta_j \mid \mathbf{y})$ ; the sixth column is the Rao-Blackwellized estimate of the posterior mean of  $\beta_j$ . Please refer to Guan and Stephens [2011] for details of the Rao-Blackwellization scheme.

NOTE: the 5th and 6th columns are estimates for  $\mathbb{E}(\beta_j \mid \mathbf{y})$ , not  $\mathbb{E}(\beta_j \mid \mathbf{y}, \gamma_j = 1)$  !

For users that are familiar with the BVSR method of Guan and Stephens [2011], the information in the `***.log.txt` and `***.path.txt` is probably self-evident.

## 3 List of Common Options

NOTE: all the listed arguments except `-R` have the same meaning as in **piMASS**. Required/recommended options:

- `-g`    *string*    the name of the input mean genotype file.
- `-p`    *string*    the name of the input phenotype file.
- `-o`    *string*    prefix of the output files.
- `-s`    *integer*    number of sampling iterations in MCMC.
- `-w`    *integer*    number of burn-in iterations in MCMC.

Advanced options:

- -r *integer* random seed.
- -nstart *integer* the initial model size.
- -hmax *double* maximum value for  $h$ .
- -hmin *double* minimum value for  $h$ .
- -pmax *integer*  $\frac{pmax}{N_{snp}}$  is the maximum value for  $\pi$ .
- -pmin *integer*  $\frac{pmin}{N_{snp}}$  is the minimum value for  $\pi$  ( $pmin \geq 1$ ).
- -smax *integer* maximum model size in MCMC.
- -smin *integer* minimum model size in MCMC ( $\geq 1$ ).
- -R *integer* (not to be confused with -r) do Rao-Blackwellization every  $R$  iterations (default value = 1000). If you run 1 million sampling iterations, consider using “-R 10000”.

## References

Yongtao Guan and Matthew Stephens. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *The Annals of Applied Statistics*, pages 1780–1815, 2011.