

User's Guide for Gmatrix Ver2.

A program for computing Genomic relationship matrix

Guosheng Su and Per Madsen

Department of Genetics and Biotechnology, Aarhus University

Gmatrix is a program to setup the genomic relationship matrix based on marker (SNP) information. Gmatrix can store the genomic relationship matrix and/or its inverse on disk.

The following gives a description of: parameter file, input files, and output files

1. parameter file: par.dat

Gmatrix has many options, which are specified in a file with the name "par.dat". Each option consists of two lines, one line for a key word, followed by one line for a value. Some of the options have default values. The following is a list of the options with default values in **bold**:

\$MINMAF
0.01

\$FREQMETH
1

\$SCALEMETHOD
1

\$DIAG_ADD
0.01

\$G_ADD
0

\$DIAG_ONE
1

\$PROP_A_to_G
0

\$CAL_DET
1

\$OUT_GMATRIX
0

\$OUT_IGMATRIX

1

\$MINMAF: minimum minor allele frequency.

The marker lower than the given value will be deleted

\$FREQMETHOD: method to get marker allele frequency

1: calculated from data.

2: estimated base p. (not implemented yet)

\$SCALEMETHOD: method to scale Gmatrix.

1: divide by sum of $2pq$.

2: Mmatrix divide by $\sqrt{2pq}$ for each locus.

3: divide a beta function (not implement yet)

\$PROP_A_TO_G: percentage of A matrix to modify G matrix.

If no Amatrix to be used, set 0.

If Amatrix will be used, give a value, usually 0.05. Pedigree file (pedigree.dat) must be provided.

\$DIAG_ADD: A value to be added to diagonal of Gmatrix in order to make Gmatrix positive(semipositive) definite. Usually 0 0.02.

\$G_ADD: A value to be added to all elements of Gmatrix.

It is said that adding a very small values can improve the relationship matrix but it has not been confirmed.

\$DIAG_ONE: Scale for rescaling GMatrix by dividing the mean of diagonals and then times this scale

For example, given the scale being 1, the mean of diagonals of resulting Gmatrix is 1.

0: No rescale

>0: Rescale so that the mean of diagonals equal to this value

\$CAL_DET: require for calculating determinant of Gmatrix

1 to calculate determinant of Gmatrix.

0 without calculation

\$OUT_GMATRIX: require for output of Gmatrix

1 output of Gmatrix

0 not

\$OUT_IGMATRIX: require for output of inverse Gmatrix

1 output of inverse Gmatrix

0 not

It is only necessary to specify a option for, which a default value is defined, if a different value should be used. The options can be specified in any order.

2. Input file

- 1) **map.dat**: Map file. Just used for calculate number of marker loci. Each line for each locus.
You can use only one column with `event=1`.
- 2) **marker.dat**: Marker file. One line for each individual.
first column is id number,
the 2, 3 column are alleles of locus 1,
the 4, 5 column are alleles of locus 2,
and so on
- 3) **ped.dat**: Pedigree file. This file is only needed in the case where the A matrix should be used (`$PROP_A_TO_G > 0`).
column 1: ID
column 2: sire
column 3: dam

3. outputfiles

- 1) **gmatrix.dat** (if `$OUT_GMATRIX=1`): Gmatrix file which include 3 columns.
column 1: ID of the i animal
column 2: ID of the j animal
column 3: relation coefficient between the i animal and the j animal

This file can be used as the genomic relationship matrix for the variance structure “PGMIX” in DMU

- 2) **igmatrix.dat** (if `$OUT_IGMATRIX=1`): Inverse Gmatrix file which include 3 columns.
column 1: ID of the i animal
column 2: ID of the j animal
column 3: relation coefficient between the i animal and the j animal
However, an extra line (the first line) is added, that is:
0 0 log(determinant of Gmatrix)

This file can be used as the inverse genomic relationship matrix for GBLUP by the variance structure “COR” in DMU

3. Other files

Script file: If you want to run the program by submitting it as a job in Linux/Unix system or in PC, A script file or a batch file is required.