G x E based on genomic information Practical

S. Hong Lee

(Jan/17)

Preparation

- Individual laptop

  Download mtg2 (v 2.04), example0, example1 and example 5 from

  https://sites.google.com/site/honglee0707/mtg2

Or

- Login the server (hong1.une.edu.au)

  mkdir <your directory>

  cd <your directory>

  cp ../honglee/mtg2 mtg2

  cp -r ../honglee/example0 example0

  …

Preparation

- Literature
  - Lee, SH and van der Werf, JHJ. MTG2: An efficient algorithm for multivariate linear mixed model analysis based on genomic information. *Bioinformatics* 32: 1420 - 1422 (2016)
  - Maier, R., et al. Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder and major depression disorder. *The American Journal of Human Genetics* 96, 283-294 (2015)
  - Lee, S. H.; Van der Werf, J. H. J. An efficient variance component approach implementing an average information REML suitable for combined LD and linkage mapping with a general complex pedigree. *Genetics Selection Evolution* 38: 25-43 (2006)

Multi-trait GREML/GBLUP

- mtg2

  ./mtg2 -p {plink fam file name} -d {phenotype file name} –g {grm file name} -cc {class covariate file name} –qc {continuous covariate file name} -out {output file name} -sv {starting value file name} –mod {number of traits}

  Go to exapmple0/

  ./mtg2 -p toy.fam -d toy.phen -g toy.rtmx -mod 1

  Or

  ./mtg2 -fam toy.fam -pheno toy.phen -grm toy.rtmx -mod 1

Input files for MTG2

<fam file for -p>

The PLINK fam file is your *.fam file that used in estimating the grm.

<grm file for -g>

For the grm file, you should unzip the .gz file from GCTA, delete the third column.

Then, it looks like

1 1 0.999

2 1 0.011

2 2 1.031

3 1 0.02

…..

Input files for MTG2

<gz format grm file for -zg>

MTG2 can read gz format GRM files generated by GCTA or PLINK1.9.
E.g., ./mtg2 -p test.fam -d test.dat -zg test.grm.gz -cc test.cov -qc test.pc -out test.out -mod 5

<binary grm file for -bg>

MTG2 can read binary GRM files generated by GCTA or PLINK1.9.

E.g., ./mtg2 -p test.fam -d test.dat -bg test.grm.bin -cc test.cov -qc test.pc -out test.out -mod 5

Phenotype file for MTG2

<phenotype file for -d>

With 5 traits model, the columns have FID, IID, t1, t2, t3, t4 and t5 (phenotypes for trait 1 ~ 5). It looks like

1 1 0.02 0.71 -0.02 0.04 -0.62

1 2 0.12 0.31 -0.27 NA -0.35

2 1 0.22 0.25 -0.28 0.63 -0.15

……

Missing values should be coded as NA.

Covariate files for MTG2

<files for –cc (class covariate) and –qc (continuous covariate)>

The FID and IID order for phenotype file, covariate files (cc, qc) should be the same.

Missing values should be coded as NA.

Principal components using MTG2

In order to get eigenvalues and eigenvectors with the prefix “test.grm”, i.e. test.grm.eval and test.grm.evec, the following command can be used

./mtg2 -p test.fam -g test.grm -pca n (n is the number of the first n PCs).

MTG2 extra options

-cove 1: parameterising residual covariance

-thread k: k paralleled computation

-sv {file name}

-mg {file name} instead of -g {file name}: multiple random effects model

-bv {file name}: BLUP estimation

-inv 1: inverting matrix

-bend 1: bending NPD matrix making it PD

-nit {value}: maximum number of iterations (default is 200)

-conv {value}: convergence criteria for log likelihood (default is 0.001)

-frq 1: estimating allele frequency given plink files

MTG2 extra options

-trf_h2: transform between observed and liability scale for h2

-sbv: converting between individual BLUP and SNP BLUP

-delta: Delta method to get SE of the ratio for target VC

-rtmx: estimating GRM

-Me: effective # chromosome segments given effective size

-pred_acc: accuracy of genomic prediction based on theory
(and power of genomic prediction)

-simcoal: coalescence simulation (SNP genotype)

-simreal: phenotype simulation based on given genotypes

-fix: constrain some parameters during REML estimation

Multivariate linear mixed model (5 traits model)

- Go to example1/

./mtg2 -p example1.fam -d example1.dat -g example1.grm -mod 5 (without residual covariance)

./mtg2 -p example1.fam -d example1.dat -g example1.grm -mod 5 -cove 1 (with residual covariance)

this will give estimated parameters in ascm.out (with -out <file_name>, the outputs will be in <file_name>)

Multivariate linear mixed model (5 traits model)

- To speed up, using eigen-decomposition technique

./mtg2 -p example1.fam -g example1.grm -pca 1908
this will give example1.grm.eval and example1.grm.evec

Then, using -eig substantially reduce the computing time

./mtg2 -p example1.fam -d example1.dat -eig example1.grm -mod 5 -cove 1 -out ascm.out_eig

Multivariate linear mixed model

DIY

- ✓ Can you obtain the full eigenvector matrix from example1.grm? (see prac page 9)
  i. What should be the dimension of the eigenvector matrix?
  ii. What is the maximum value of the eigenvalues?
  iii. What is the minimum value of the eigenvalues?
  iv. What does a negative eigenvalue mean?
  v. Should you make NPD to PD? (Lecture page 46)

Multivariate linear mixed model

DIY

- ✓ Can you estimate genetic variances from 5 trait model? (see prac page 12)
  - i. Using –g, what is computing time for inverting V and for obtaining derivatives?
  - ii. Can you try the same model with –eig (prac page 13)? What is computing time?
  - iii. Can you compare the estimates with –g and –eig? Are they the same or not?
  - iv. What is the estimated h2 for the third trait?
  - v. What is the estimated genetic correlation between trait 1 and 4?
  - vi. With and without –cove 1, how does the estimate change?
  - vii. When do you not need to estimate residual covariance?

Random regression model (lecture page 38)

It needs a parameter file having # order for each random effect and value for each environment (see rrm.par3)

```
3 3          ! # order for the first and second random effects
0 15 30 75   ! environments variable (time at the measurements)
```

Random regression model (lecture page 38)

./mtg2 -p example1.fam -d example1.dat -mg joint.rtmx -rrm rrm.par3 -mod 4 -out rrm.out

Joint.rtmx is the file specifying genetic covariance matrices for the random effects, e.g.

example1.idm

example1.grm

if one gives example1.idm as n x n identity matrix, residual covariances can be modelled.

With –eig, this can be done using with –rrme 1.

./mtg2 -p example1.fam -d example1.dat -eig example1.rtmx -rrm rrm.par3 -rrme 1 -mod 4 -out rrm.out_eig

Random regression model (lecture page 38)

DIY

- ✓ Can you run RRM with –mg? (see prac page 17)
- ✓ Can you run RRM with –eig? (see prac page 17)
  - i. Compare the speed and results with and without eigen-decomposition (i.e. rrm.out vs. rrm.out_eig)
  - ii. Can you transform the RR coefficients to variance covariance matrix of phenotypic observation? (lecture page 38 and 42, and see the r code)

$$\mathrm{var}\begin{pmatrix} g_1 & \cdots & g_{1,T} \\ \vdots & \ddots & \vdots \\ g_{8,T} & \cdots & g_T \end{pmatrix} = \mathbf{\Phi K \Phi}'$$

Random regression model (lecture page 38)

To find the best model, change # order in the model (i.e. rrm.par3) and run it.

```
1 1            ! # order for the first and second random effects
0 15 30 75     ! environments variable (time at the measurements)

1 2            ! # order for the first and second random effects
0 15 30 75     ! environments variable (time at the measurements)

1 3            ! # order for the first and second random effects
0 15 30 75     ! environments variable (time at the measurements)
…

4 3            ! # order for the first and second random effects
0 15 30 75     ! environments variable (time at the measurements)
```

Random regression model (lecture page 38)

- ✓ Can you make a table like that in lecture note page 41 with the maximum likelihood values?
- ✓ And, AIC or BIC values ?

  AIC = 2p – 2ln(Likelihood)

  BIC = ln(n)p – 2ln(Likelihood) where n is # records and p is # parameters

		# Order for residual variance		
		1	2	3
# Order for genetic variance	1			
	2			
	3			
	4			

Spline function (manual page 8)

Smooth piecewise cubic polynomials (e.g. ASReml, also see Kachman (2011) course note for more detail)

The same as the usual RRM except

-spl instead of –rrm

<files for –spl>

The first line is for the number of knots for spline function for each random effect, the second line is for corresponding value for each environment (e.g. time at which repeated measurements were taken) and the third line is the location of the knots.

Spline function (manual page 8)

E.g. if you have spline function with # knots k=3 (of which the location is 0, 37.5 and 75) and glucose level as phenotypes taken at 0, 15, 30 and 75 minutes after intraperitoneal glucose injection, it looks like

3

0 15 30 75

0 37.5 75

Note: The location of the first knot should be equal or less than the value of the first environment and the last knot should be equal or larger than the value of the last environment.

E.g. ./mtg2 -p example1.fam -d example1.dat -eig example1.grm -mod 4 -spl spl.par3 -out spl.out -rrme 1

Spline function (manual page 8)

DIY

✓ How many parameters should you fit when using spline function?

$$\boldsymbol{G} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ \vdots & \\ 1 & 12 \end{pmatrix} \begin{pmatrix} \sigma^2_{a1} & \sigma_{a12} \\ \sigma_{a21} & \sigma^2_{a2} \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ \vdots & \\ 1 & 12 \end{pmatrix}' + \boldsymbol{Z} \begin{pmatrix} \sigma^2_{as} & & & \\ & \sigma^2_{as} & & \\ & & \ddots & \\ & & & \sigma^2_{as} \end{pmatrix} \boldsymbol{Z}'$$

$$\boldsymbol{G} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ \vdots & \\ 1 & 12 \end{pmatrix} \begin{pmatrix} \sigma^2_{a1} & \sigma_{a12} \\ \sigma_{a21} & \sigma^2_{a2} \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ \vdots & \\ 1 & 12 \end{pmatrix}' + \boldsymbol{Z}\boldsymbol{Z}'\sigma^2_{as}$$

Steve Kachman asreml course note (2011)

Spline function (manual page 8)

DIY

- ✓ Can you fit spine function with # knots = 3 (at 0, 37.5 and 75) for the example1 data? (see prac page 22)
- ✓ Can you compare the best model from Legendre polynomial (usual RRM) to see it gives a better fit? (you can use likelihood ratio test, AIC or BIC)
- ✓ When should spline be more useful than Legendre polynomial?
- ✓ You can try with different # knots and locations to see how the likelihood changes

Multivariate RRM (manual page 9)

- ./mtg2 -p {plink fam file name} -d {phenotype file name} -g {grm file name} -out {output file name} -sv {starting value file name} -mod {number of sites} -mrrm {number of traits} -rrm {parameter required for fitting random regression}

  E.g. for bivariate RRM,

  ./mtg2 -p example5.fam -d example5.dat -eig example5.rtmx -out ascm.out -mod 6 -mrrm 2 -rrm mrrm.par -rrme 1 -cove 1

Multivariate RRM (manual page 9)

<files for -rrm>

The first line is for the number of site for each trait. So the sum of the sites for the traits should be the same as the total number of sites (i.e. -mod {number of sites}).

The second line is for the number of order for Legendre polynomial for each random effect, and the third line is for corresponding value for each environment, which is for the first trait. The fourth and fifth lines are the same as the second and third, but they are for the second trait, and so on for the next traits.

E.g. if you have Legendre polynomial order k=2 and glucose level (first trait) and insulin level (second trait) as phenotypes taken at 0, 15, 30 and 75 minutes after intraperitoneal glucose injection, it looks like

4 4

2

0 15 30 75

2

0 15 30 75

Multivariate RRM (manual page 9)

To speed up, using -eig,

./mtg2 -p example5.fam -d example5.dat -eig example5.rtmx -out ascm.out -mod 6 -mrrm 2 -rrm mrrm.par

+ residual covariance across environments

./mtg2 -p example5.fam -d example5.dat -eig example5.rtmx -out ascm.out -mod 6 -mrrm 2 -rrm mrrm.par -rrme 1

+ residual covariance between traits

./mtg2 -p example5.fam -d example5.dat -mg joint.rtmx -out ascm.out -mod 6 -mrrm 2 -rrm mrrm.par -cove 1

Multivariate RRM (manual page 9)

DIY

- ✓ Go to example5/ and have a look at the phenotype file (example5.dat)
- ✓ Phenotypes were measured at three different time points (0, 15 and 30min) for two traits. Can you fit bivariate RRM for this data set? (see prac page 27)
- ✓ Can you also estimate residual covariance across environments in the bivariate RRM?
- ✓ Can you also additionally estimate residual covariance between traits in the bivariate RRM?
- ✓ Can you interpret the output file?