Overview
About
Credits
Questions and Help Requests
Citation
Download
Basic options
Bayesian alphabet

Overview

About

GCTB is a software tool that comprises a family of Bayesian linear mixed models for complex trait analyses using genome-wide SNPs. It was developed to simultaneously estimate the joint effects of all SNPs and the genetic architecture parameters for a complex trait, including SNP-based heritability, polygenicity and the joint distribution of effect sizes and minor allele frequencies.

Credits

Jian Zeng developed the software with supports from Jian Yang, Futao Zhang and Zhili Zheng. Part of the code are adopted from *GCTA* and *GenSel*. Luke Lloyd-Jones contributed to the BayesR module with supports from Jian Zeng and Mike Goddard.

Questions and Help Requests

If you have any bug reports or questions, please send an email to Jian Zeng (j.zeng@uq.edu.au) or Jian Yang (jian.yang@uq.edu.au).

Citation

Zeng et al. (2018) Signatures of negative selection in the genetic architecture of human complex traits. *Nature Genetics*, doi: 10.1038/s41588-018-0101-4.

Download

Executable Files

gctb_1.0_Linux.zip

gctb_1.0_Mac.zip

Source code

Standard version

MPI version

The MPI version implements a distributed computing strategy that scales the analysis to very large sample sizes. A significant improvement in computing time is expected for a sample size > 10,000. The MPI version needs to be compiled on user's machine. See README.html in the tarball for instructions of compilation and usage. A testing dataset is also included in each tarball.

Update log

1. 1 Dec, 2017: first release.

Basic options

Input and output

--bfile test

Input PLINK binary PED files, e.g. test.fam, test.bim and test.bed (see PLINK user manual for details).

--pheno test.phen

Input phenotype data from a plain text file, e.g. test.phen.

--out test

Specify output root filename.

Data management

--keep test.indi.list

Specify a list of individuals to be included in the analysis.

--chr 1

Include SNPs on a specific chromosome in the analysis, e.g. chromosome 1.

--extract test.snplist

Specify a list of SNPs to be included in the analysis.

--exclude test.snplist

Specify a list of SNPs to be excluded from the analysis.

--mpheno 2

If the phenotype file contains more than one trait, by default, GCTB takes the first trait for analysis (the third column of the file) unless this option is specified. For example, **--mpheno** 2 tells GCTB to take the second trait for analysis (the fourth column of the file).

--covar test.qcovar

Input quantitative covariates from a plain text file, e.g. test.qcovar. Each quantitative covariate is recognized as a continuous variable.

MCMC settings

--seed 123

Specify the seed for random number generation, e.g. 123. Note that giving the same seed value would result in exactly the same results between two runs.

--chain-length 21000

Specify the total number of iterations in MCMC, e.g. 21000 (default).

--burn-in 1000

Specify the number of iterations to be discarded, e.g. 1000 (default).

--out-freq 100

Display the intermediate results for every 100 iterations (default).

--thin 10

Output the sampled values for SNP effects and genetic architecture parameters for every 10 iterations (default). Only non-zero sampled values of SNP effects are written into a binary file.

--no-mcmc-bin

Suppress the output of MCMC samples of SNP effects.

Bayesian alphabet

--bayes S

Specify the Bayesian alphabet for the analysis, e.g. S. Different alphabet launch different models, which differ in the prior specification for the SNP effects. The available alphabet include

- B: Each SNP effect is assumed to have an i.i.d. mixture prior of a t-distribution $t(0, \tau^2, \nu)$ with a probability π and a point mass at zero with a probability $1 - \pi$.
- C: Each SNP effect is assumed to have an i.i.d. mixture prior of a normal distribution $N(0, \sigma^2)$ with a probability π and a point mass at zero with a probability $1 - \pi$.
- S: Similar to C but the variance of SNP effects is related to minor allele frequency (p) through a parameter S, i.e. $\sigma_j^2 = \left[2p_j(1 - p_j)\right]^S \sigma^2$.
- N: nested BayesC. SNPs within a 0.2 Mb non-overlapping genomic region are collectively considered as a window (specify the distance by **--wind** 0.2). This nested approach speeds up the analysis by skipping over windows with zero effect.
- NS: nested BayesS.

- R: BayesR. Each SNP effect is assumed to have an i.i.d. mixture prior of multiple normal distributions $N(0, \gamma_k \sigma_k^2)$ with a probability π_k and a point mass at zero with a probability $1 - \sum_k \pi_k$, where γ_k is a given constant.

--fix-pi

An option to fix π to a constant (the value is specified by the option --pi below). The default setting is to treat π as random and estimate it from the data.

--pi 0.05

A starting value for the sampling of π when it is estimated from the data, or a given value for π when it is fixed. The default value is 0.05. When BayesR is used, it is a string seperated by comma where the number of values defines the number of mixture components and each value defines the starting value for each component (the first value is reserved for the zero component); the default values are 0.95,0.03,0.01,0.01.

--gamma 0,0.01,0.1,1

When BayesR is used, this speficies the gamma values seperated by comma, each representing the scaling factor for the variance of a mixture component. Note that the number of values should match that in --pi.

--hsq 0.5

A starting value for the sampling of SNP-based heritability, which may improve the mixing of MCMC algorithm if it starts with a good estimate. The default value is 0.5.

--S 0

A starting value for the sampling of the parameter S (relationship between MAF and variance of SNP effects) in BayesS, which may improve the mixing of MCMC algorithm if it starts with a good estimate. The default value is 0.

--wind 0.2

Specify the window width in Mb for the non-overlapping windows in the nested models, e.g. 0.2 Mb. The default value is 1 Mb.

Examples

Standard version of gctb:

```
gctb --bfile test --pheno test.phen --bayes S --pi 0.1 --hsq 0.
5 --chain-length 25000 --burn-in 5000 --out test > test.log 2>&
1
```

MPI version of gctb (when using intelMPI libraries and two nodes):

```
mpirun -f $PBS_NODEFILE -np 2 gctb_mpi --bfile test --pheno tes
t.phen --bayes S --pi 0.1 --hsq 0.5 --chain-length 25000 --burn
-in 5000 --out test > test.log 2>&1
```

The output files include:

test.log: a text file of running status, intermediate output and final results;

test.snpRes: a text file of posterior statistics of SNP effects;

test.covRes: a text file of posterior statistics of covariates;

test.parRes: a text file of posterior statistics of key model parameters;

test.mcmcsamples.CovEffects: a text file of MCMC samples for the covariates fitted in the model;

test.mcmcsamples.SnpEffects: a binary file of MCMC samples for the SNP effects;

test.mcmcsamples.Par: a text file of MCMC samples for the key model parameters;

Citations

GCTB software and BayesS method
Zeng et al. (2018) Signatures of negative selection in the genetic architecture of human complex traits. *Nature Genetics*, doi: 10.1038/s41588-018-0101-4.

BayesR
Moser et al. (2015) Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model. *PLoS Genetics*, 11: e1004969.

BayesN
Zeng et al. (2018) A nested mixture model for genomic prediction using whole-genome SNP genotypes. *PLoS One*, 13: e0194683.

BayesCπ
Habier et al. (2011) Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics*, 12: 186.

BayesB
Meuwissen et al. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157: 1819-1829.