

SparseLMM: User Manual

April 16, 2019

`fitNULLGLMM`

Fit the null logistic mixed model and estimate the variance ratio by a set of randomly selected variants

Description

Fit the null logistic mixed model and estimate the variance ratio by a set of randomly selected variants

Usage

```
fitNULLGLMM(plinkFile = "", phenoFile = "", phenoCol = "",
  traitType = "binary", invNormalize = FALSE, covarColList = NULL,
  qCovarCol = NULL, sampleIDColinphenoFile = "", tol = 0.02,
  maxiter = 20, tolPCG = 1e-05, maxiterPCG = 500, nThreads = 1,
  Cutoff = 2, numMarkers = 30, skipModelFitting = FALSE,
  memoryChunk = 2, tauInit = c(0, 0), LOCO = FALSE,
  traceCVcutoff = 1, ratioCVcutoff = 1, outputPrefix = "",
  kinFile = "", sparse_kin = FALSE, nThreadsOMP = 1)
```

Arguments

<code>plinkFile</code>	character. Path to plink file to be used for calculating elements of the genetic relationship matrix (GRM)
<code>phenoFile</code>	character. Path to the phenotype file
<code>phenoCol</code>	character. Column name for the trait e.g. "CAD"
<code>traitType</code>	character. e.g. "binary" or "quantitative". By default, "binary"
<code>invNormalize</code>	logical. Whether to perform the inverse normalization of the trait or not. E.g. TRUE or FALSE. By default, FALSE
<code>covarColList</code>	vector of characters. Covariates to be used in the glm model e.g c("Sex", "Age")
<code>qCovarCol</code>	vector of characters. Categorical covariates to be used in the glm model (NOT work yet)

sampleIDColinphenoFile	character. Column name for the sample IDs in the phenotype file e.g. "IID".
nThreads	integer. Number of threads to be used. By default, 1
numMarkers	integer (>0). Number of markers to be used for estimating the variance ratio. By default, 30
skipModelFitting	logical. Whether to skip fitting the null model and only calculating the variance ratio, By default, FALSE. If TRUE, the model file ".rda" is needed
memoryChunk	integer or float. The size (Gb) for each memory chunk. By default, 4
tauInit	vector of numbers. e.g. c(1,1), Unitial values for tau. For binary traits, the first element will be always be set to 1. If the tauInit is not specified, the second element will be 0.5 for binary traits.
LOCO	logical. Whether to apply the leave-one-chromosome-out (LOCO) approach. By default, FALSE. This option has not been extensively tested.
traceCVcutoff	float. The threshold for coefficient of variation (CV) for the trace estimator. Number of runs for trace estimation will be increased until the CV is below the threshold. By default 1. suggested: 0.0025. This option has not been extensively tested.
ratioCVcutoff	float. The threshold for coefficient of variation (CV) for estimating the variance ratio. The number of randomly selected markers will be increased until the CV is below the threshold. By default 1. suggested 0.001. This option has not been extensively tested.
outputPrefix	character. Path to the output files with prefix.
kinFile	(new)character. Path to the kinship matrix file
sparse_kin	(new)Logical, whether kinship matrix is stored as sparse form (MatrixMarket format), default is FALSE
nThreadsOMP	(new)integer, number of threads for OpenMP with external sparse GRM, default is 1

Value

a file ended with .rda that contains the glmm model information, a file ended with .varianceRatio.txt that contains the variance ratio value, and a file ended with #markers.SAIGE.results.txt that contains the SPAGMMAT tests results for the markers used for estimating the variance ratio.

Description

SparseLMM is an R package integrated with C++ Armadillo library and OpenMP API with a modified version of SAIGE's 'fitNULLGLMM()' function, which enables the flexibility to use non-genetic variance component information while yielding feasible computation and memory efficiency.

Details

The DESCRIPTION file: This package was not yet installed at build time.

Index: This package was not yet installed at build time.

SparseLMM is an R package integrated with C++ Armadillo library and OpenMP API with a modified version of SAIGE's 'fitNULLGLMM()' function, which enables the flexibility to use non-genetic variance component information while yielding feasible computation and memory efficiency.

Author(s)

Yongwen Zhuang

Maintainer: Yongwen Zhuang <zyongwen@umich.edu>

References

- [1] Chen, Han, Chaolong Wang, Matthew P Conomos, Adrienne M Stilp, Zilin Li, Tamar Sofer, Adam A Szpiro, et al. 2016. "Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models." *The American Journal of Human Genetics* 98 (4). Elsevier: 653–66.
- [2] OpenMP Architecture Review Board. 2013. "OpenMP Application Program Interface Version 4.0." <https://www.openmp.org/wp-content/uploads/OpenMP4.0.0.pdf>.
- [3] Zhou, Wei, Jonas B Nielsen, Lars G Fritsche, Rounak Dey, Maiken E Gabrielsen, Brooke N Woldford, Jonathon LeFaive, et al. 2018. "Efficiently Controlling for Case-Control Imbalance and Sample Relatedness in Large-Scale Genetic Association Studies." *Nature Genetics* 50 (9). Nature Publishing Group: 1335.

Examples

```
#### set params for test ####
plinkFile="./example/sampledData"
phenoFile="./example/sampledData_pheno.txt"
phenoCol="PheCode1"
covarColList="x1,sex"
covarColList <- strsplit(covarColList,",")[[1]]
sampleIDColInphenoFile="IID"
traitType="binary"
outputPrefix="./example/testoutput"
nThreads=1
LOCO=FALSE
memoryChunk=0.2
invNormalize = FALSE
```

```

qCovarCol = NULL
tol=0.02
maxiter=20
tolPCG=1e-5
maxiterPCG=500
Cutoff = 2
numMarkers = 100
skipModelFitting = FALSE
tauInit = c(0,0)
traceCVcutoff = 1
ratioCVcutoff = 1
kinFile = "./example/sampleddata.kmat.sp"
nThreadsOMP = 2

# run main function
fitNULLGLMM(plinkFile=plinkFile,
             phenoFile = phenoFile,
             phenoCol = phenoCol,
             traitType = traitType,
             covarColList = covarColList,
             sampleIDColinphenoFile = sampleIDColinphenoFile,
             nThreads = nThreads,
             LOCO = LOCO,
             outputPrefix = outputPrefix,
             memoryChunk = memoryChunk,
             invNormalize =invNormalize,
             qCovarCol = qCovarCol,
             tol = tol,
             maxiter = maxiter,
             Cutoff = Cutoff,
             numMarkers = numMarkers,
             skipModelFitting = skipModelFitting,
             tauInit = tauInit,
             traceCVcutoff = traceCVcutoff,
             ratioCVcutoff = ratioCVcutoff,
             kinFile = kinFile,
             sparse_kin = TRUE,
             nThreadsOMP = nThreadsOMP
)

```

Index

*Topic **package**

sparseLMM-package, [2](#)

fitNULLGLMM, [1](#)

sparseLMM(*sparseLMM-package*), [2](#)

sparseLMM-package, [2](#)