Package 'LGRF'

September 13, 2015

Type Package

Title Set-Based Tests for Genetic Association in Longitudinal Studies

Version 1.0

Date 2015-08-20

Author Zihuai He

Maintainer Zihuai He <zihuai@umich.edu>

Description Functions for the longitudinal genetic random field method (He et al., 2015, <doi:10.1111/biom.12310>) to test the association between a longitudinally measured quantitative outcome and a set of genetic variants in a gene/region.

License GPL-3

Depends CompQuadForm, SKAT, geepack

NeedsCompilation no

Repository CRAN

Date/Publication 2015-09-13 09:22:48

R topics documented:

IBS_pseudo *Generate IBS pseudo variables*

Description

If users want to calculate the IBS similarity, this function creates the IBS pseudo variables. This is in order to calculate the IBS similarity in an efficient way.

Usage

```
IBS_pseudo(x)
```

Arguments

x	An n by q matrix of genetic variants.

Value

It returns an n by 3p matrix of pseudo variables for efficiently calculating IBS similarity.

Examples

```
library(LGRF)

# Load data example
# Z: genotype matrix, n by q matrix

data(LGRF.example)
Z<-LGRF.example$Z
A<-IBS_pseudo(Z)

# Then the IBS matrix can be calculated by K.IBS<-AA^T.
```

LGRF.example *Data example for LGRF*

Description

The dataset contains outcome variable Y, covariate X, time and genotype data Z. The first column in time is the subject ID and the second column is the measured exam. Y, X and time are all in long form. Z is a genotype matrix where each row corresponds to one subject.

Usage

```
data(LGRF.example)
```

Examples

```
data(LGRF.example)
```

LGRF.SSD.All *LGRF tests for multiple regions/genes using SSD format files*

Description

Test the association between an outcome variable and multiple regions/genes using SSD format files.

Usage

```
LGRF.SSD.All(SSD.INFO, result.null, Gsub.id=NULL, interGXT=FALSE, similarity='GR',
impute.method='fixed', MinP.compare=FALSE, ...)
```

Arguments

SSD.INFO	SSD format information file, output of function “Open_SSD". The sets are defined by this file.
result.null	Output of function “null.LGRF".
Gsub.id	The subject id corresponding to the genotype matrix, an m dimensional vector. This is in order to match the phenotype and genotype matrix. The default is NULL, where the order is assumed to be matched with Y, X and time.
interGXT	Whether to incorperate the gene-time interaction effect. Incorperating this effect can improve power if there is any gene-time interaction, but has slight power loss otherwise. The default is FALSE. *Please note that the second column of time should be included as a covairate when interGXT is TRUE.
similarity	Choose the similarity measurement for the genetic variants. Can be either "GR" for genetic relationship or "IBS" for identity by state. The default is "GR" for better computational efficiency.
impute.method	Choose the imputation method when there is missing genotype. Can be "random", "fixed" or "bestguess". Given the estimated allele frequency, "random" simulates the genotype from binomial distribution; "fixed" uses the genotype expectation; "Best guess" uses the genotype with highest probability.
MinP.compare	Whether to compare with the GEE based minimum p-value (MinP) test. The default is FALSE. Please note that implementing the GEE based MinP test is time consuming.
...	Other options of the GEE based MinP test. Defined same as in function “test.MinP".

Value

results	First column contains the set ID; Second column contains the p-values; Third column contains the number of tested SNPs.

Examples

```
# * Since the Plink data files used here are hard to be included in a R package,
# The usage is marked by "#" to pass the package check.

#library(LGRF)

###############################################

# Plink data files: File.Bed, File.Bim, File.Fam
# Files defining the sets: File.SetID, File.SSD, File.Info
# For longitudinal data, outcome and covariates are saved in a separate file: Y, time, X.
# Null model was fitted using function null.LGRF.

# Create the MW File
# File.Bed<-"./example.bed"
# File.Bim<-"./example.bim"
# File.Fam<-"./example.fam"
# File.SetID<-"./example.SetID"
# File.SSD<-"./example.SSD"
# File.Info<-"./example.SSD.info"

# Generate SSD file
# To use binary ped files, you have to generate SSD file first.
# If you already have a SSD file, you do not need to call this function.
# Generate_SSD_SetID(File.Bed, File.Bim, File.Fam, File.SetID, File.SSD, File.Info)

# SSD.INFO<-Open_SSD(File.SSD, File.Info)
# Number of samples
# SSD.INFO$nSample
# Number of Sets
# SSD.INFO$nSets

## Fit the null model
# Y: outcomes, n by 1 matrix where n is the total number of observations
# X: covariates, n by p matrix
# time: describe longitudinal structure, n by 2 matrix
# result.null<-null.LGRF(Y,time,X=cbind(X,time[,2]))

# *Please note that the second column of time should be included as a covairate if
# the gene by time interaction effect will be incorperated.

## Test all regions
# out_all<-LGRF.SSD.All(SSD.INFO, result.null)

# Example result
# out.all$results
#      SetID   P.value N.Marker
# 1  GENE_01 0.6568851       94
# 2  GENE_02 0.1822183       84
# 3  GENE_03 0.3836986      108
# 4  GENE_04 0.1265337      101
```

```
# 5  GENE_05 0.3236089      103
# 6  GENE_06 0.9401741       94
# 7  GENE_07 0.1043820      104
# 8  GENE_08 0.6093275       96
# 9  GENE_09 0.6351147      100
# 10 GENE_10 0.5631549      100

## Test all regions, and compare with GEE based MinP test
# out_all<-LGRF.SSD.All(SSD.INFO, result.null,MinP.compare=T)

# Example result
# out.all$results
#      SetID P.value P.value.MinP N.Marker
# 1  GENE_01 0.62842       1.0000       94
# 2  GENE_02 0.06558       0.2718       84
# 3  GENE_03 0.61795       1.0000      108
# 4  GENE_04 0.39667       0.7052      101
# 5  GENE_05 0.17371       0.5214      103
# 6  GENE_06 0.90104       1.0000       94
# 7  GENE_07 0.10143       0.1188      104
# 8  GENE_08 0.78082       0.3835       96
# 9  GENE_09 0.21966       0.5364      100
# 10 GENE_10 0.25468       0.3527      100
```

`LGRF.SSD.OneSet_SetIndex`

LGRF tests for a single region/gene using SSD format files

Description

Test the association between an outcome variable and one region/gene using SSD format files.

Usage

```
LGRF.SSD.OneSet_SetIndex(SSD.INFO, SetIndex, result.null, Gsub.id=NULL, interGXT=FALSE,
similarity='GR', impute.method='fixed', MinP.compare=FALSE, ...)
```

Arguments

`SSD.INFO`	SSD format information file, output of function “Open_SSD". The sets are defined by this file.
`SetIndex`	Set index. From 1 to the total number of sets.
`result.null`	Output of function “null.LGRF".
`Gsub.id`	The subject id corresponding to the genotype matrix, an m dimensional vector. This is in order to match the phenotype and genotype matrix. The default is NULL, where the order is assumed to be matched with Y, X and time.

interGXT	Whether to incorperate the gene-time interaction effect. Incorperating this effect can improve power if there is any gene-time interaction, but has slight power loss otherwise. The default is FALSE. *Please note that the second column of time should be included as a covairate when interGXT is TRUE.
similarity	Choose the similarity measurement for the genetic variants. Can be either "GR" for genetic relationship or "IBS" for identity by state. The default is "GR" for better computational efficiency.
impute.method	Choose the imputation method when there is missing genotype. Can be "random", "fixed" or "bestguess". Given the estimated allele frequency, "random" simulates the genotype from binomial distribution; "fixed" uses the genotype expectation; "Best guess" uses the genotype with highest probability.
MinP.compare	Whether to compare with the GEE based minimum p-value (MinP) test. The default is FALSE. Please note that implementing the GEE based MinP test is time consuming.
...	Other options of the GEE based MinP test. Defined same as in function “test.MinP".

Value

p.value	p-value of the LGRF test.
n.marker	number of tested SNPs in the SNP set.

Examples

```
# * Since the Plink data files used here are hard to be included in a R package,
# The usage is marked by "#" to pass the package check.

#library(LGRF)

##############################################

# Plink data files: File.Bed, File.Bim, File.Fam
# Files defining the sets: File.SetID, File.SSD, File.Info
# For longitudinal data, outcome and covariates are saved in a separate file: Y, time, X.
# Null model was fitted using function null.LGRF.

# Create the MW File
# File.Bed<-"./example.bed"
# File.Bim<-"./example.bim"
# File.Fam<-"./example.fam"
# File.SetID<-"./example.SetID"
# File.SSD<-"./example.SSD"
# File.Info<-"./example.SSD.info"

# Generate SSD file
# To use binary ped files, you have to generate SSD file first.
# If you already have a SSD file, you do not need to call this function.
# Generate_SSD_SetID(File.Bed, File.Bim, File.Fam, File.SetID, File.SSD, File.Info)

# SSD.INFO<-Open_SSD(File.SSD, File.Info)
```

```
# Number of samples
# SSD.INFO$nSample
# Number of Sets
# SSD.INFO$nSets

## Fit the null model
# Y: outcomes, n by 1 matrix where n is the total number of observations
# X: covariates, n by p matrix
# time: describe longitudinal structure, n by 2 matrix
# result.null<-null.LGRF(Y,time,X=cbind(X,time[,2]))

# *Please note that the second column of time should be included as a covairate if
# the gene by time interaction effect will be incorperated.

## Test a single region
# out_single<-LGRF.SSD.OneSet_SetIndex(SSD.INFO=SSD.INFO, SetIndex=1,
# result.null=result.null, MinP.compare=F)

# Example result
# $p.value
# [1] 0.6284

# $n.marker
# [1] 94

## Test a single region, and compare with GEE based MinP test
# out_single<-LGRF.SSD.OneSet_SetIndex(SSD.INFO=SSD.INFO, SetIndex=1,
# result.null=result.null,MinP.compare=T)

# $p.value
#       LGRF MinP
# [1,] 0.6284    1

# $n.marker
# [1] 94
```

`null.LGRF` *Fit the null model for longitudinal genetic random field model*

Description

Before testing a specific region using a score test, this function fits the longitudinal genetic random field model under the null hypothesis.

Usage

```
null.LGRF(Y, time, X = NULL)
```

Arguments

Y	The outcome variable, an n*1 matrix where n is the total number of observations
time	An n*2 matrix describing how the observations are measured. The first column is the subject id. The second column is the measured exam (1,2,3,etc.).
X	An n*p covariates matrix where p is the total number of covariates.

Value

It returns a list used for function test.LGRF().

Examples

```
library(LGRF)

# Load data example
# Y: outcomes, n by 1 matrix where n is the total number of observations
# X: covariates, n by p matrix
# time: describe longitudinal structure, n by 2 matrix
# Z: genotype matrix, m by q matrix where m is the total number of subjects

data(LGRF.example)
Y<-LGRF.example$Y;time<-LGRF.example$time;X<-LGRF.example$X;Z<-LGRF.example$Z

# Fit the null model
result.null<-null.LGRF(Y,time,X=cbind(X,time[,2]))

# *Please note that the second column of time should be included as a covairate if
# the gene by time interaction effect will be incorperated.
```

---

test.LGRF *Test the association between an outcome variable and a region/gene by LGRF*

Description

Once the model under the null model is fitted using "null.LGRF()", this function tests a specifc region/gene.

Usage

```
test.LGRF(Z, result.null, Gsub.id=NULL, interGXT = FALSE, similarity = "GR",
impute.method="fixed")
```

Arguments

Z	Genetic variants in the target region/gene, an m*q matrix where m is the subject ID and q is the total number of genetic variables. Note that the number of rows in Z should be same as the number of subjects.
result.null	The output of function "null.LGRF()"
Gsub.id	The subject id corresponding to the genotype matrix, an m dimensional vector. This is in order to match the phenotype and genotype matrix. The default is NULL, where the order is assumed to be matched with Y, X and time.
interGXT	Whether to incorperate the gene-time interaction effect. Incorperating this effect can improve power if there is any gene-time interaction, but has slight power loss otherwise. The default is FALSE. *Please note that the second column of time should be included as a covairate when interGXT is TRUE.
similarity	Choose the similarity measurement for the genetic variants. Can be either "GR" for genetic relationship or "IBS" for identity by state. The default is "GR" for better computational efficiency.
impute.method	Choose the imputation method when there is missing genotype. Can be "random", "fixed" or "bestguess". Given the estimated allele frequency, "random" simulates the genotype from binomial distribution; "fixed" uses the genotype expectation; "Best guess" uses the genotype with highest probability.

Value

p.value	p-value of the LGRF test.
n.marker	number of tested SNPs in the SNP set.

Examples

```
## null.LGRF fits the null model.
# Input: Y, time, X (covariates)
## test.LGRF tests a region and give p-value.
# Input: Z (genetic variants) and result of null.longGRF

library(LGRF)

# Load data example
# Y: outcomes, n by 1 matrix where n is the total number of observations
# X: covariates, n by p matrix
# time: describe longitudinal structure, n by 2 matrix
# Z: genotype matrix, m by q matrix where m is the total number of subjects

data(LGRF.example)
Y<-LGRF.example$Y;time<-LGRF.example$time;X<-LGRF.example$X;Z<-LGRF.example$Z

# Fit the null model
result.null<-null.LGRF(Y,time,X=cbind(X,time[,2]))

# *Please note that the second column of time should be included as a covairate if
# the gene by time interaction effect will be incorperated.
```

```
# The LGRF-G test
pLGRF_G<-test.LGRF(Z,result.null)

# The LGRF-GT test
pLGRF_GT<-test.LGRF(Z,result.null,interGXT=TRUE)

# The LGRF-G test using the IBS similarity
pLGRF_G_IBS<-test.LGRF(Z,result.null,similarity="IBS")

# The LGRF-GT test, main effect is modeled using the IBS similarity
pLGRF_GT_IBS<-test.LGRF(Z,result.null,interGXT=TRUE,similarity="IBS")
```

`test.MinP` *Test the association between an outcome variable and a region/gene by MinP*

Description

If users want to compare LGRF with the minimum p-value (MinP) test, this function tests a specifc region/gene by a GEE based minimum p-value test after fitting "null.LGRF()".

Usage

```
test.MinP(Z, result.null, Gsub.id=NULL, corstr="exchangeable", MinP.adjust=0.95,
impute.method="fixed")
```

Arguments

`Z`	Genetic variants in the target region/gene, an m*q matrix where m is the subject ID and q is the total number of genetic variables. Note that the number of rows in Z should be same as the number of subject.
`result.null`	The output of function "null.LGRF()".
`Gsub.id`	The subject id corresponding to the genotype matrix, an m dimensional vector. This is in order to match the phenotype and genotype matrix. The default is NULL, where the order is assumed to be matched with Y, X and time.
`corstr`	The working correlation as specified in 'geeglm'. The following are permitted: '"independence"', '"exchangeable"', '"ar1"', '"unstructured"' and '"userdefined".
`MinP.adjust`	The minimum p-value is adjusted by the number of independent tests. Choose the adjustment thereshold as specified in Gao, et al. (2008) "A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms". Values from 0 to 1 are permitted.
`impute.method`	Choose the imputation method when there is missing genotype. Can be "random", "fixed" or "bestguess". Given the estimated allele frequency, "random" simulates the genotype from binomial distribution; "fixed" uses the genotype expectation; "Best guess" uses the genotype with highest probability.

Value

p.value	p-value of the MinP test.
n.marker	number of tested SNPs in the SNP set.

Examples

```
## null.LGRF fits the null model.
# Input: Y, time, X (covariates)
## test.MinP tests a region and give p-value.
# Input: Z (genetic variants) and result of null.longGRF

library(LGRF)

# Load data example
# Y: outcomes, n by 1 matrix where n is the total number of observations
# X: covariates, n by p matrix
# time: describe longitudinal structure, n by 2 matrix
# Z: genotype matrix, m by q matrix where m is the total number of subjects

data(LGRF.example)
Y<-LGRF.example$Y;time<-LGRF.example$time;X<-LGRF.example$X;Z<-LGRF.example$Z

# Fit the null model
result.null<-null.LGRF(Y,time,X=X)

# The minimum p-value test based on GEE
pMinP<-test.MinP(Z,result.null,corstr="exchangeable",MinP.adjust=0.95)
```

Index