Package 'ScoreEB'

October 12, 2022

Type Package

Title Score Test Integrated with Empirical Bayes for Association Study

Version 0.1.1

Author Wenlong Ren

Maintainer Wenlong Ren <wenlongren@ntu.edu.cn>

Description Perform association test within linear mixed model framework using score test integrated with Empirical Bayes for genome-wide association study. Firstly, score test was conducted for each marker under linear mixed model framework, taking into account the genetic relatedness and population structure. And then all the potentially associated markers were selected with a less stringent criterion. Finally, all the selected markers were placed into a multilocus model to identify the true quantitative trait nucleotide.

License GPL-3

Imports data.table

Encoding UTF-8

LazyData true

NeedsCompilation yes

Depends R (>= 3.5.0)

RoxygenNote 7.1.1

Repository CRAN

Date/Publication 2021-09-15 21:10:12 UTC

R topics documented:

ebayes_EM *Empirical Bayes for multi-locus selection*

Description

Empirical Bayes using expectation–maximization algorithm.

Usage

```
ebayes_EM(x,z,y,EMB.tau,EMB.omega)
```

Arguments

x	fixed effect vector or matrix.
z	genotype data.
y	phenotype data.
EMB.tau	one of hyperparameters in inverse chi-square distribution.
EMB.omega	one of hyperparameters in inverse chi-square distribution.

Value

u	The effect values of markers, and their absolute values are used as the basis for further screening.

Examples

```
data(geno)
data(pheno)
EMB.tau <- 0
EMB.omega <- 0
z <- t(geno[,-c(1:4)])
y <- as.matrix(pheno)
nsample <- dim(z)[1]
x <- as.matrix(rep(1,nsample))
ebayes_EM(x,z,y,EMB.tau,EMB.omega)
```

geno *Genotype of example data*

Description

Genotype dataset with SNP chromosome, position and etc.

Usage

```
data(geno)
```

Details

Dataset input of genotype in ScoreEB function.

`likelihood` *Carry out likelihood ratio test*

Description

Snps selected via EM-Bayes to further identified by likelihood ratio test.

Usage

```
likelihood(xxn,xxx,yn,bbo)
```

Arguments

`xxn`	fixed effect vector or matrix.
`xxx`	snp matrix which are selected by EM-Bayes.
`yn`	phenotype data.
`bbo`	effect value of snp estimated by EM-Bayes.

Value

`lod`	Odds of logarithm vector of markers.

Examples

```
data(geno)
data(pheno)
z <- t(geno[,-c(1:4)])
y <- as.matrix(pheno)
n.sample <- dim(z)[1]
m.marker <- dim(z)[2]
x <- as.matrix(rep(1,n.sample))
beta <- rnorm(m.marker)
likelihood(x,z,y,beta)
```

multinormal *Multivariate normal distribution*

Description

Obtain P value with multivariate normal distribution.

Usage

```
multinormal(y,mean,sigma)
```

Arguments

y	column vector.
mean	arithmetic mean.
sigma	standard deviation.

Value

pdf_value	A vector of multivariate normal distribution density function.

Examples

```
data(pheno)
y <- pheno
mean <- 2.0
sigma <- 1.5
multinormal(y,mean,sigma)
```

PCG *Preconditioned Conjugate Gradient*

Description

Conduct preconditioned conjugate gradient method to accelerate.

Usage

```
PCG(G,b,m.marker,sigma.k2,sigma.e2,tol,miter)
```

Arguments

G	genotype data.
b	column vector.
m.marker	the number of markers.
sigma.k2	variance of polygenic.
sigma.e2	variance of residual error.
tol	convergence threshold.
miter	the maximum number of iterations.

Value

x	x is approximate solution of linear equations.

Examples

```
data(geno)
G <- t(geno[,-c(1:4)])
n.sample <- dim(G)[1]
m.marker <- dim(G)[2]
b <- rnorm(n.sample)
sigma.k2 <- 6.0
sigma.e2 <- 10.0
tol <- 5e-4
miter <- 20
PCG(G,b,m.marker,sigma.k2,sigma.e2,tol,miter)
```

---

pheno *Phenotype of example data*

Description

Phenotype dataset of multiple traits.

Usage

```
data(pheno)
```

Details

Dataset input of phenotype in ScoreEB function.

ScoreEB *Score Test Integrated with Empirical Bayes for Association Study*

Description

Perform association test within linear mixed model framework using score test integrated with Empirical Bayes for genome-wide association study. Firstly, score test was conducted for each marker under linear mixed model framework, taking into account the genetic relatedness and population structure. And then all the potentially associated markers were selected with a less stringent criterion. Finally, all the selected markers were placed into a multi-locus model to identify the true quantitative trait nucleotide.

Usage

```
ScoreEB(genofile, phenofile, popfile = NULL, trait.num = 1, EMB.tau = 0,
EMB.omega = 0, B.Moment = 20, tol.pcg = 1e-4, iter.pcg = 100, bin = 100,
lod.cutoff = 3.0, seed.num = 10000, dir_out)
```

Arguments

genofile	Genotype file name, change the file path where it is located, i.e.,"D:/Genotype_Example.csv".
phenofile	Phenotype file name, change the file path where it is located, i.e.,"D:/Phenotype_Example.csv".
popfile	Population structure file name, change the file path where it is located,i.e.,"D:/Population.csv".
trait.num	trait.num stands for computing trait from the 1st to the "trait.num"
EMB.tau	EMB.tau and EMB.omega are two values of hyperparameters in empirical Bayes step, which are set to 0 by default.
EMB.omega	As describe in EMB.tau
B.Moment	B.Moment is a parameter to obtain trace of NxN matrix approximately using method of moment. B.Moment is set to 20 by default.
tol.pcg	tol.pcg and iter.pcg are tolerance and maximum iteration number in preconditioned conjugate gradient algorithm.
iter.pcg	As describe in tol.pcg
bin	bin is to choose the maximum score within a certain range.
lod.cutoff	lod.cutoff is the threshold to determine identified QTNs.
seed.num	Set a random number.
dir_out	Give the path where it will be saved,i.e.,"D:/Result"

Value

result.total	A data frame of identified markers, including "Trait", "Id", "Chr", "Pos", "Score", "Beta", "Lod" and "Pvalue" of markers.

Note

1. genofile and phenofile are the required input file, while popfile is the optional input file.
2. In the "tempdir()" folder, there are two results files "ScoreEB.Result.csv" and "ScoreEB.time.csv" generated and saved after the run.
3. The results file "ScoreEB.Result.csv" has 8 columns, including "Trait", "Id", "Chr", "Pos", "Score", "Beta", "Lod" and "Pvalue".
4. The time file "ScoreEB.time.csv" includes 3 rows, which are "User", "System", "Elapse" time, respectively.

Author(s)

Wenlong Ren
Wenlong Ren <wenlongren@ntu.edu.cn>

Examples

```
genofile <- system.file("extdata", "Genotype_Example.csv", package="ScoreEB")
phenofile <-system.file("extdata", "Phenotype_Example.csv", package="ScoreEB")
dir_out <- tempdir()
ScoreEB(genofile, phenofile, popfile = NULL, trait.num = 1, EMB.tau = 0,
EMB.omega = 0, B.Moment = 20, tol.pcg = 1e-4, iter.pcg = 100, bin = 100,
lod.cutoff = 3.0, seed.num = 10000, dir_out)
```

Index