Package: PAGWAS (via r-universe)

August 28, 2024

Type Package

Title Pathway Analysis Methods for Genomewide Association Data

Version 2.0

Date 2015-12-07

Author Marina Evangelou

Maintainer Marina Evangelou <m.evangelou@ic.ac.uk>

Description Bayesian hierarchical methods for pathway analysis of genomewide association data: Normal/Bayes factors and Sparse Normal/Adaptive lasso. The Frequentist Fisher's product method is included as well.

License GPL (>= 2)

Depends lars, foreach, mnormt

NeedsCompilation no

Date/Publication 2015-12-08 14:53:45

Repository https://mevangelou.r-universe.dev

RemoteUrl https://github.com/cran/PAGWAS

RemoteRef HEAD

RemoteSha f72f5ebcca81da66e5d2b0e1a8156001995d3373

Contents

PAGWAS-package *Pathway Analysis Methods for Genomewide Association Data*

Description

Bayesian hierarchical methods for pathway analysis of genomewide association data: Normal/Bayes factors and Sparse Normal/Adaptive lasso. The Frequentist Fisher's method is included as well.

Details

Package:	PAGWAS
Type:	Package
Version:	2.0
Date:	2015-12-02
License:	GPL (>=2)
LazyLoad:	yes

Author(s)

Marina Evangelou Maintainer: Marina Evangelou <m.evangelou@ic.ac.ukt>

References

Evangelou, M., Dudbridge, F., Wernisch, L. (2014). Two novel pathway analysis methods based on a hierarchical model. Bioinformatics 30(5), 690 - 697

Evangelou, M., Rendon, A., Ouhewand, W. H., Wernisch, L., Dudbridge, F. (2012) Comparison of methods for competitive tests of pathway analysis. Plos One 7(7): e41018

create.pathway.df *Creates a pathway data frame*

Description

Returns a data frame with L rows and M columns. L is the number of SNPs in the genotypes data frame and M is the number of tested pathways.

Usage

```
create.pathway.df(genotypes,snps.paths)
```

Arguments

`genotypes`	Genotype matrix, with L SNPs (columns) and N individuals (rows)
`snps.paths`	A list with entries the SNP members of each pathway. The size of the list is M

Value

A data frame with columns equal to the number of pathways in the pathway.snps list and rows equal to the number of tested SNPs

See Also

`SNPs`, `genes`, `snps.to.pathways` `snps.to.genes`

Examples

```
data(SNPs)
data(genes)
data(pathways)
data(genotypes)
snps.genes <- snps.to.genes(snp.info=SNPs,gene.info=genes, distance=0)
pathway.snps <- snps.to.pathways(pathways,snps.genes)
P <- create.pathway.df(genotypes=genotypes,snps.paths=pathway.snps)
```

create.pathway.matrix *Creates a pathway matrix*

Description

Creates a pathway matrix, with rows the SNPs assigned to each pathway

Usage

```
create.pathway.matrix(genotypes,pathway.snps)
```

Arguments

`genotypes` Genotype matrix, with L SNPs (columns) and N individuals (rows)

`pathway.snps` A list of the SNPs members of each pathway

Value

A matrix with columns equal to the number of pathways in the pathway.snps list and rows equal to the number of SNPs in the genotypes data-frame

See Also

`SNPs`, `genes`, `snps.to.pathways`, `snps.to.genes`

Examples

```
## Not run:
data(SNPs)
data(genes)
data(pathways)
data(genotypes)
snps.genes <- snps.to.genes(snp.info=SNPs,gene.info=genes, distance=50)
pathway.snps <- snps.to.pathways(pathways,snps.genes)
P <- create.pathway.matrix(genotypes,pathway.snps)

## End(Not run)
```

`FM.chi.pvalue` *Calculates the Fisher's method p-value for each tested pathway*

Description

Calculates the Fisher's method *p*-value for a set of *p*-values. It returns both the *p*-value and the test statistic value of the Fisher's product method.

Usage

```
FM.chi.pvalue(x)
```

Arguments

`x` A vector of *p*-values. These *p*-values can be either gene or SNP *p*-values of a tested pathway

Value

`FMstatistic` Fisher's product method test statistic

`FMpvalue` Fisher's method *p*-value, computed using the exact distribution of the Fisher's method test statistic which is a *Chi*^2 distribution with degrees of freedom twice the size of vector x

References

Evangelou M, Rendon A, Ouwehand WH, Wernisch L, Dudbridge F (2012) Comparison of Methods for Competitive Tests of Pathway Analysis. PLoS ONE 7(7): e41018. doi:10.1371/journal.pone.0041018

See Also

`pathways`, `snps.to.pathways`

Examples

```
FM.chi.pvalue(x=c(0.05,0.1))
```

---

genes *A data frame of 20 artificial genes with their chromosomes and positions on the genome*

Description

A data frame with 20 rows and 4 columns.

Usage

```
data(genes)
```

Format

Column names:

`Name` Name of gene

`Start` Start position of gene on the genome

`End` End position of gene on the genome

`Chr` Chromosome of gene

See Also

`SNPs`

Examples

```
data(genes)
print(genes[1:5,])
```

genotypes *Genotypes for 100 SNPs and 75 individuals*

Description

A data frame with 75 rows (individuals) and 100 columns (SNPs). The entries of the genotype matrix are 0, 1 and 2. There are no missing values.

Usage

```
data(genotypes)
```

See Also

SNPs

Examples

```
data(genotypes)
```

list.of.parameters *A list with possible hyper-parameters for NBF*

Description

A list with 8 different combinations of the hyper-parameters of NBF. Possible values for the hyper-parameters a, b, s2_0 and nu_0 are given in each entry of the list

Usage

```
data(list.of.parameters)
```

Examples

```
data(list.of.parameters)
a=list.of.parameters[[1]][1]
b=list.of.parameters[[1]][2]
s2_0=list.of.parameters[[1]][3]
nu_0=list.of.parameters[[1]][4]
```

NBF *Normal/Bayes factors method for finding associated pathways*

Description

A vector of the computed Bayes factors for the tested pathways.

Usage

```
NBF(y, G, P, a, b, s2, nu)
```

Arguments

`y`	Response vector of length N
`G`	Genotype matrix, with N rows and L columns (number of tested SNPs)
`P`	Pathway matrix, with L columns and M columns (number of tested pathways)
`a`	Hyper-parameter of the variance assumed for the integrated out SNP effects
`b`	Hyper-parameter of the variance assumed for the pathway effects
`s2`	Hyper-parameter of the Inverse-Chi-squared distribution assumed for the variance of the response vector
`nu`	Hyper-parameter of the Inverse-Chi-squared distribution assumed for the variance of the response vector

Value

A vector of the computed Bayes factors of the same length as the number of tested pathways

References

Evangelou, M., Dudbridge, F., Wernisch, L. (2014). Two novel pathway analysis methods based on a hierarchical model. Bioinformatics, 30(5), 690 - 697.

Examples

```
## Not run:
 data(genotypes)
 G=genotypes
 data(pathways)
 data(SNPs)
 data(genes)
 snps.genes=snps.to.genes(SNPs,genes,distance=0)
 snps.paths=snps.to.pathways(pathways,snps.genes)
 P=create.pathway.df(G,snps.paths)
 y=rnorm(nrow(G),mean=0,sd=10)
 NBF(y,G,P,a,b,s2,nu)
## End(Not run)
```

pathways *A list of 2 pathways with their gene members*

Description

A list of two pathways. The gene members of each pathway are given.

Usage

```
data(pathways)
```

See Also

genes, SNPs

Examples

```
data(pathways)
```

return.a.SNAL *Returns the hyper-parameter a of SNAL*

Description

This function returns the appropriate hyper-parameter a for the integrated out SNP effects of SNAL

Usage

```
return.a.SNAL(y, G, P, no.simulations, no.paths, no.snps)
```

Arguments

y	Response vector of length N
G	Genotype matrix, with N rows and L columns (L is the number of tested SNPs)
P	Pathway matrix, with L columns and M columns (M is the number of tested pathways)
no.simulations	Number of simulations to run
no.paths	Number of pathways assumed to be causal for the simulations
no.snps	Number of SNPs assumed to be causal for the simulations

Value

Returns the value of the hyper-parameter a that gives the highest the power of SNAL

References

Evangelou, M., Dudbridge, F., Wernisch, L. (2013). Two novel pathway analysis methods based on a hierarchical model. Bioinformatics (to appear)

See Also

`SNAL`, `return.s2.SNAL`

`return.bf.NBF` *Returns the Bayes factor threshold of NBF*

Description

This function returns a data-frame of true positive rates and false positive rates of NBF for the Bayes factors thresholds chosen by the user

Usage

```
return.bf.NBF(y, G, P, a, b, s2, nu,
 no.simulations, no.paths, no.snps, BF.cutoffs)
```

Arguments

`y`	Response vector of length N
`G`	Genotype matrix, with N rows and L columns (L is the number of tested SNPs)
`P`	Pathway matrix, with L columns and M columns (M is the number of tested pathways)
`a`	Hyper-parameter a of the variance assumed for the integrated out SNP effects
`b`	Hyper-parameter b of the variance assumed for the pathway effects
`s2`	Hyper-parameter s2_0 of the Inverse-Chi-squared distribution assumed for the variance of the response vector
`nu`	Hyper-parameter nu_0 of the Inverse-Chi-squared distribution assumed for the variance of the response vector
`no.simulations`	Number of simulations to run
`no.paths`	Number of pathways assumed to be causal for the simulations
`no.snps`	Number of SNPs assumed to be causal for the simulations
`BF.cutoffs`	A vector of Bayes factors (BFs) thresholds to test

Value

Returns a data-frame of true positive and false positive rates for the Bayes factors (BFs) thresholds given by the user

References

Evangelou, M., Dudbridge, F., Wernisch, L. (2013). Two novel pathway analysis methods based on a hierarchical model. Bioinformatics (to appear)

See Also

NBF, return.hyperparameters.NBF

Examples

```
## Not run: return.bf.NBF(y,G,P,a=1e-4,b=1e-2,s2=0.25,nu=200,
no.simulations=100,no.paths=10,no.snps=20,BF.cutoffs=c(0.5,0.75,0.9,1))
## End(Not run)
```

---

return.hyperparameters.NBF

Returns the four hyper-parameters of NBF

Description

This function returns the four hyper-parameters a, b, s2_0, nu_0 of NBF

Usage

```
return.hyperparameters.NBF(y, G, P, no.simulations,
no.paths, no.snps, list.parameters)
```

Arguments

`y`	Response vector of length N
`G`	Genotype matrix, with N rows and L columns (L is the number of tested SNPs)
`P`	Pathway matrix, with L columns and M columns (M is the number of tested pathways)
`no.simulations`	Number of simulations to run
`no.paths`	Number of pathways assumed to be causal for the simulations
`no.snps`	Number of SNPs assumed to be causal for the simulations
`list.parameters`	A list of various combinations of the four hyper-parameters, see list.of.parameters

Value

Returns a list with the four hyper-parameters of NBF

`a` Hyper-parameter of the variance assumed for the integrated out SNP effects

`b` Hyper-parameter of the variance assumed for the pathway effects

`s2_0` Hyper-parameter of the Inverse-Chi-squared distribution assumed for the variance of the response vector

`nu_0` Hyper-parameter of the Inverse-Chi-squared distribution assumed for the variance of the response vector

References

Evangelou, M., Dudbridge, F., Wernisch, L. (2013). Two novel pathway analysis methods based on a hierarchical model. Bioinformatics (to appear)

See Also

`NBF`, `return.bf.NBF`

Examples

```
## Not run: hyper=return.hyperparameters.NBF(y,G,P,no.simulations=100,
no.paths=10,no.snps=20,
list.parameters=list(c(1e-3,1e-3,0.25,200),c(1e-2,1e-3,0.25,100)))

## End(Not run)
```

`return.s2.SNAL` *Returns the s2 tuning parameter of SNAL*

Description

This function returns a data-frame of true positive rates and false positive rates of SNAL for the s2 thresholds chosen by the user

Usage

```
return.s2.SNAL(y, G, P, a, no.simulations, no.paths, no.snps, s2.cutoffs)
```

Arguments

`y` Response vector of length N

`G` Genotype matrix, with N rows and L columns (L is the number of tested SNPs)

`P` Pathway matrix, with L columns and M columns (M is the number of tested pathways)

`a` Hyper-parameter of the variance assumed for the integrated out SNP effects `no.simulations` Number of simulations to run

`no.paths` Number of pathways assumed to be causal for the simulations

`no.snps` Number of SNPs assumed to be causal for the simulations

`s2.cutoffs` A vector of s2 thresholds to test

Value

Returns a data-frame of true positive and false positive rates for the s2 thresholds given by the user

References

Evangelou, M., Dudbridge, F., Wernisch, L. (2013). Two novel pathway analysis methods based on a hierarchical model. Bioinformatics (to appear)

See Also

`SNAL`,`return.a.SNAL`

Examples

```
## Not run: return.s2.SNAL(y,G,P,a=1e-2,no.simulations=100,
  no.paths=10,no.snps=20,s2.cutoffs=c(0.5,1,2))

## End(Not run)
```

`roc.convex` *Computes the area under a ROC curve*

Description

Computes the area under a ROC curve using the convex hull of the curve

Usage

```
roc.convex(sens, spec)
```

Arguments

`sens` Vector with the values of the recorded sensitivity (true positive rate)

`spec` Vector with the values of the recorded specificity (1-false positive rate)

Value

Returns the computed area under the ROC curve

Author(s)

Marina Evangelou, Lorenz Wernisch

References

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27

Evangelou, M., Dudbridge, F., Wernisch, L. (2013). Two novel pathway analysis methods based on a hierarchical model. Bioinformatics (to appear)

Examples

```
roc.convex(sens=c(0.1,0.5,1),spec=c(0.1,0.3,0.7))
```

SNAL *Sparse Normal/Adaptive lasso method for finding associated pathways*

Description

Sparse Normal/Adaptive lasso method applied for finding the associated pathways. The iterative algorithm suggested by Wipf and Nagarajan (2008) is applied. A vector equal to the number of tested pathways is returned, the zero entries of the vector correspond to the pathways that are not associated. The posterior estimates of the beta coefficients are also returned as they are described by Wipf and Nagarajan (2008).

Usage

```
SNAL(y, G, P, a, s2)
```

Arguments

`y` Response vector of length N

`G` Genotype matrix, with N rows and L columns (number of tested SNPs)

`P` Pathway matrix, with L columns and M columns (number of tested pathways)

`a` Hyper-parameter of the variance assumed for the integrated out SNP effects

`s2` Variance assumed for the response variable, the tuning parameter of adaptive lasso

Value

`gamma.star` Estimates of gamma hyper-parameters

`ARD` Posterior estimates of beta coefficients

References

Evangelou, M., Dudbridge, F., Wernisch, L. (2014). Two novel pathway analysis methods based on a hierarchical model. Bioinformatics, 30(5), 690 - 697.

Wipf, D. and Nagarajan, S. (2008). A new view of automatic relevance determination. Advances in Neural Information Processing Systems, 20

See Also

SNAL.calculation

Examples

```
## Not run:
 data(genotypes)
 G=genotypes
 data(pathways)
 data(SNPs)
 data(genes)
 snps.genes=snps.to.genes(SNPs,genes,distance=0)
 snps.paths=snps.to.pathways(pathways,snps.genes)
 P=create.pathway.df(G,snps.paths)
 y=rnorm(nrow(G),mean=0,sd=10)
 SNAL(y,G,P,a,s2)
## End(Not run)
```

SNAL.calculation *Sparse Normal/Adaptive lasso method for finding associated variables. The SNAL method is applied to the linear regression Y= Phi beta + epsilon*

Description

For more details please read SNAL.

Usage

```
SNAL.calculation(Y, Phi, s2)
```

Arguments

`Y` Response vector of length N

`Phi` Design matrix, with N rows and M columns (number of tested variables)

`s2` Variance assumed for the response variable, the tuning parameter of the adaptive lasso problem

Value

`gamma.star` Estimates of gamma hyper-parameters

`ARD` Posterior estimates of beta coefficients

References

Evangelou, M., Dudbridge, F., Wernisch, L. (2014). Two novel pathway analysis methods based on a hierarchical model. Bioinformatics, 30(5), 690 - 697

Wipf, D. and Nagarajan, S. (2008). A new view of automatic relevance determination. Advances in Neural Information Processing Systems, 20

See Also

SNAL

Examples

```
## Not run: SNAL.calculation(Y,Phi,s2=0.5)
```

SNPs *A data frame of 100 artificial SNPs with their chromosomes and positions on the genome*

Description

A data frame with 100 rows and 3 columns.

Usage

```
data(SNPs)
```

Format

Column names:

Name SNP name

Position Position of SNP on the genome

Chr Chromosome of the SNP

See Also

genes, genotypes

Examples

```
data(SNPs)
print(SNPs[1:5,])
```

snps.to.genes *Assigns SNPs to genes*

Description

Assigns SNPs to genes based on their physical distance.

Usage

```
snps.to.genes(snp.info, gene.info, distance)
```

Arguments

snp.info	A data frame with 3 columns with names: Name, Position and Chr that correspond to the SNP name, its position on the genome and its chromosome, respectively
gene.info	A data frame with 4 columns with names: Name, Start, End and Chr that correspond to the gene name, start and end positions on the genome and its chromosome, respectively
distance	A number that corresponds to the distance below and above the Start and End positions of the gene that all SNPs in that region should be assigned to the gene

Value

A list of the same size as the number of genes of the gene.info data frame. The names of the SNPs assigned to each gene are returned

See Also

SNPs, genes, snps.to.pathways

Examples

```
data(SNPs)
data(genes)
snps.to.genes(snp.info=SNPs,gene.info=genes,distance=50)
```

snps.to.pathways *Assigns SNPs to pathways*

Description

Assigns SNPs to pathways, using the pathway gene members and the SNPs assigned to each gene.

Usage

```
snps.to.pathways(pathways,gene.snps)
```

Arguments

`pathways`	A list of pathways with their gene members
`gene.snps`	A list of genes with the SNPs assigned to them according to their physical distance on the genome

Value

A list of the same size as the number of pathways in the pathway list. The names of the SNPs assigned to each pathway are returned. Empty pathways are also returned.

See Also

`SNPs`, `genes`, `snps.to.genes`

Examples

```
data(SNPs)
data(genes)
data(pathways)
snps.genes <- snps.to.genes(snp.info=SNPs,gene.info=genes, distance=50)
pathway.snps <- snps.to.pathways(pathways,snps.genes)
```

Index