Package 'LFDREmpiricalBayes'

Type Package

Title Estimating Local False Discovery Rates Using Empirical Bayes
Methods

Version 1.0

Date 2017-09-26

Author Ali Karimnezhad, Johnary Kim, Anna Akpawu, Justin Chitpin and David R Bickel

Maintainer Ali Karimnezhad <ali_karimnezhad@yahoo.com>

Description New empirical Bayes methods aiming at analyzing the association of single nu-
cleotide polymorphisms (SNPs) to some particular disease are implemented in this pack-
age. The package uses local false discovery rate (LFDR) estimates of SNPs within a sample pop-
ulation defined as a ``reference class'' and discovers if SNPs are associated with the correspond-
ing disease. Although SNPs are used throughout this document, other biologi-
cal data such as protein data and other gene data can be used. Karim-
nezhad, Ali and Bickel, D. R. (2016) <http://hdl.handle.net/10393/34889>.

Depends R(>= 2.14.2)

Imports matrixStats, stats, R6

Suggests LFDR.MLE, testthat

biocViews Bayesian, MathematicalBiology, MultipleComparison

URL https://davidbickel.com

License GPL-3

NeedsCompilation no

Repository CRAN

Date/Publication 2017-09-27 09:08:46 UTC

R topics documented:

LFDREmpiricalBayes-package

Estimating Local False Discovery Rates Using Empirical Bayes Methods

---

Description

New empirical Bayes methods aiming at analyzing the association of single nucleotide polymorphisms (SNPs) to some particular disease are implemented in this package. The package uses local false discovery rate (LFDR) estimates of SNPs within a sample population defined as a "reference class" and discovers if SNPs are associated with the corresponding disease. Although SNPs are used throughout this document, other biological data such as protein data and other gene data can be used. Karimnezhad, Ali and Bickel, D. R. (2016) <http://hdl.handle.net/10393/34889>.

Details

Package:	LFDREmpiricalBayes
Type:	Package
Version:	1.0
Date:	2017-09-26
License:	GPL-3
Depends:	R(>= 2.14.2)
Imports:	matrixStats, stats
Suggests:	LFDR.MLE
URL:	https://davidbickel.com

Author(s)

Ali Karimnezhad, Johnary Kim, Anna Akpawu, Justin Chitpin and David R Bickel

Maintainer: Ali Karimnezhad <ali_karimnezhad@yahoo.com>

References

Karimnezhad, A. and Bickel, D. R. (2016). Incorporating prior knowledge about genetic variants into the analysis of genetic association data: An empirical Bayes approach. Working paper. Retrieved from http://hdl.handle.net/10393/34889

See Also

For more information on how to interpret the outputs, look at the supplementary file in the vignette directory, "Using the LFDREmpiricalBayes Package."

caution.parameter.actions

Based on a Decision-Theoretic Approach, Performs a Multiple Hypothesis Testing Problem under a Zero-One Loss Function

Description

Assuming a zero-onr loss function, it provides three caution-type actions using estimated LFDRs computed based on both separate and combined reference classes.

Usage

```
caution.parameter.actions(x1,x2,l1=4,l2=1) # default values l1=4 and l2=1
# to obtain a threshold of 20%.
```

Arguments

x1	A vector of LFDRs in the combined reference class.
x2	A vector of LFDRs in the separate reference class.
l1	Loss value (Type-I error) for deriving the threshold of the Bayes action.
l2	Loss value (Type-II error) for deriving the threshold of the Bayes action.

Details

Accepts previously obtained LFDR estimates of SNPs falling inside the intersection of the separate and combined reference classes. The LFDR estimates of some biological feature (SNP or gene) within a sample population that we will refer to as 'reference class'. If a reference class, containing LFDR estimates is a subset of the other, it is referred to as 'separate class'. The entire set of LFDR estimates is called a 'combined' reference class. Then, a multiple hypothesis problem is conducted using three caution-type estimators. The threshold set for rejecting the null hypothesis is derived from pre-specified l1 and l2 values. Since having a type-I error is worse than a type-II error, l1 is recommende to be greater than l2.

In generating the output, there are two potential outputs for each index of the three caution-type actions. Check the **Value** section for the corresponding caution-type actions.

For each index of the output, one of two potential outputs based on Bayes action are shown:

0	Do not reject the null hypothesis
1	Reject the null hypothesis

For each corresponding index in the output, the decision on whether to reject or not reject the null hypothesis for biological feature can be based on CGM1, CGM0, and CGM0.5 decisions. Check **See Also** for more details on how to better interpret the outputs.

Value

Outputs three vectors of equal size as seen below:

CGM1   Decision values for the Conditional Gamma Minimax (CGMinimax).

CGM0   Decision values for the Conditional Gamma Minimin (CGMinimin).

CGM0.5   Decision values for the CG0.5 caution case (a balance between CGMinimax and
                        CGMinimin.

Note that the length of the input vectors x1 and x2 determines the number of null hypothesis values
seen in the output.

Note

A limitation to the code is that both reference classes: x1 and x2 must be of the same vector length.

Author(s)

Code: Ali Karimnezhad.
Documentation: Justin Chitpin, Anna Akpawu and Johnary Kim.

References

Karimnezhad, A. and Bickel, D. R. (2016). Incorporating prior knowledge about genetic variants
into the analysis of genetic association data: An empirical Bayes approach. Working paper. Re-
trieved from http://hdl.handle.net/10393/34889

See Also

For more information on how to interpret the outputs, look at the vignette, "Using LFDREmpiricalBayes".

Examples

```
#LFDR reference class values generated

#First reference class (separate class)
LFDR.Separate <- c(.14,.8,.251,.30)
#Second reference class (combined class)
LFDR.Combined <- c(.21,.61,.0888,.10)

# Default threshold at (20%).
output <- caution.parameter.actions(x1=LFDR.Separate, x2=LFDR.Combined)

# Three caution cases
output
```

ME.log	*Provides Reliable LFDR Estimates by Selecting an Appropriate Reference Class*

Description

Selects an appropriate reference class given two reference classes. Considers two vecotr of LFDR estimates computed based on the two alternative reference classes and provides a vector of more reliable LFDR estimates.

Usage

```
ME.log(stat,lfdr.C,p0.C,ncp.C,p0.S,ncp.S,a=3,lower.p0=0,upper.p0=1,
lower.ncp=0.1,upper.ncp=50,length.p0=200,length.ncp=200)
```

Arguments

stat	A vector of test statistics for SNPs falling inside the intersection of the separate and combined reference classes.
lfdr.C	A data frame of local false discovery rates of features falling inside the intersection of the separate and combined reference classes, computed based on all features belonging to the combined reference class.
p0.C	An estimate of the proportion of the non-associated features applied to the combined reference class.
ncp.C	A non-centrality parameter applied to the combined reference class.
p0.S	An estimate of the proportion of the non-associated features applied to the separate reference class.
ncp.S	A non-centrality parameter applied to the separate reference class.
a	Parameter used to define the grade of evidence that alternative reference class should be favoured instead of the separate reference class.
lower.p0	The lower bound for the proportion of unassociated features.
upper.p0	The upper bound for the proportion of unassociated features.
lower.ncp	The lower bound for the non-centrality parameter.
upper.ncp	The lower bound for the non-centrality parameter.
length.p0	Desired length of a sequence vector containing the proportion of non-associated features. The sequences starts from lower.p0 and ends at upper.p0.
length.ncp	Desired length of a sequence vector containing non-centrality parameters. The sequences starts from lower.ncp and ends up at upper.ncp.

Details

The terms 'separate' and 'combined' reference classes are used when one sample population (reference class) is a subset of the other. Detailed explanations can be found in the vignette "Using the LFDREmpiricalBayes Package".

Value

Returns the following values:

p0.hat	estimate of the proportion of non-associated SNPs
ncp.hat	estimate of the non-centrality parameter
LFDR.hat	A vector of LFDR estimates for features falling inside the intersection of the separate and combined reference classes, obtained by the Maximum Entropy method.

Note

The vector of test statistics: stat, need to be positive values in order for the function ME.log to work.

Author(s)

Code: Ali Karimnezhad.
Documentation: Johnary Kim and Anna Akpawu.

References

Karimnezhad, A. and Bickel, D. R. (2016). Incorporating prior knowledge about genetic variants into the analysis of genetic association data: An empirical Bayes approach. Working paper. Retrieved from http://hdl.handle.net/10393/34889

Examples

```
#import the function ``lfdr.mle'' from package``LFDR.MLE''
library(LFDR.MLE)

#Consider a separate reference class and a combined reference class below:

n.SNPs.S<-3  # number of SNPs in the separate reference class
n.SNPs.Sc<-2 # number of SNPs in the complement of the separate reference class.

#Create a series of test statistics for SNPs in the separate reference class.
stat.Small<-rchisq(n.SNPs.S,df=1,ncp=0)
ncp.Sc<-10

#Create a series of test statistics for SNPs in the combined reference class.
stat.Big<-c(stat.Small,rchisq(n.SNPs.Sc,df=1,ncp=ncp.Sc))

#Using lfdr.mle, a series of arguments are used.
dFUN=dchisq; lower.ncp = .1; upper.ncp = 50;
lower.p0 = 0; upper.p0 = 1;

#Maximum Likelihood estimates for the LFDRs of SNPs in the created
# separate reference class.
```

```
#Separate reference class.
estimates.S<-lfdr.mle(x=stat.Small,dFUN=dchisq,df=1,lower.ncp = lower.ncp,
upper.ncp = upper.ncp)
LFDR.Small<-estimates.S$LFDR
p0.Small<-estimates.S$p0.hat
ncp.Small<-estimates.S$ncp.hat

# Maximum Likelihood estimates for the LFDRs of SNPs in the created combined
# reference class.
estimates.C<-lfdr.mle(x=stat.Big,dFUN=dchisq,df=1,lower.ncp = lower.ncp,
upper.ncp = upper.ncp)
LFDR.Big<-estimates.C$LFDR
p0.Big<-estimates.C$p0.hat
ncp.Big<-estimates.C$ncp.hat


#The first three values of the combined reference class correspond to the
#separate reference class in this example
LFDR.SBig<-LFDR.Big[1:3]

LFDR.ME<-ME.log(stat=stat.Small,lfdr.C=LFDR.SBig,p0.C=p0.Big,ncp.C=ncp.Big,
p0.S=p0.Small,ncp.S=ncp.Small)

LFDR.ME
```

PRGM.action	*Based on the Robust Bayes Approach, Performs a Multiple Hyothesis Testing Problem under an Squared Error Loss Function*

Description

Assuming a squared error loss function, it provides Robust Bayes estimates of the LFDR estimates giving credit to both separate and combined reference classes.

Usage

```
PRGM.action(x1,x2)
```

Arguments

x1	Input numeric vector of LFDR estimates of the separate reference class.
x2	Input numeric vector of LFDR estimated of the combined reference class.

Value

The output is a vector of the LFDR estimates based on the two reference classes.

Author(s)

Code: Ali Karimnezhad.
Documentation: Johnary Kim and Anna Akpawu.

References

Karimnezhad, A. and Bickel, D. R. (2016). Incorporating prior knowledge about genetic variants into the analysis of genetic association data: An empirical Bayes approach. Working paper. Retrieved from http://hdl.handle.net/10393/34889

Examples

```
#LFDR reference class values generated

#First reference class
LFDR.Separate <- c(0.14, 0.8, 0.16, 0.30)
#Second reference class
LFDR.Combined <- c(0.21, 0.61, 0.12, 0.10)

output <- PRGM.action(LFDR.Separate, LFDR.Combined)

# Vector of the LFDR estimates
output
```

SEL.caution.parameter   *Based on a Decision-Theoretic Approach, Performs a Multiple Hyothesis Testing Problem under an Squared Error Loss Function*

Description

Assuming a squared error loss function, it provides three caution-type actions using estimated LFDRs computed based on both separate and combined reference classes.

Usage

```
SEL.caution.parameter(x1,x2)
```

Arguments

x1   Input numeric vector of LFDR estimates in the separate reference class.

x2   Input numeric vector of LFDR estimates in the combined reference class.

Value

Much like `caution.parameter.actions`, this function returns three vectors of equal size as seen below:

CGM1   Squared error loss value for the Conditional Gamma Minimax (CGMinimax).

CGM0   Squared error loss value for the Conditional Gamma Minimin (CGMinimin).

CGM0.5   Squared error loss value for the Action/Decision estimate (a balance between CGMinimax and CGMinimin.

For each index of the vectors, the squared error loss values are given.

Author(s)

Code: Ali Karimnezhad.
Documentation: Johnary Kim and Anna Akpawu.

References

Karimnezhad, A. and Bickel, D. R. (2016). Incorporating prior knowledge about genetic variants into the analysis of genetic association data: An empirical Bayes approach. Working paper. Retrieved from http://hdl.handle.net/10393/34889

Examples

```
#Similar to caution.parameter actions we have the following classes

#First reference class
LFDR.Separate <- c(0.14, 0.8, 0.16, 0.30)
#Second reference class
LFDR.Combined <- c(0.21, 0.61, 0.12, 0.10)

output <- SEL.caution.parameter(LFDR.Separate, LFDR.Combined)

# Three caution cases with SEL values.
output
```

Index