Package ‘Shrinkage’

October 23, 2015

Type Package
Title Several Shrinkage Effect-Size Estimators
Version 1.0
Date 2015-10-20
Author Corey M. Yanofsky, Zahra Montazeri, Marta Padilla, Alaa Ali and David R. Bickel
Maintainer M. Padilla <padilla.mpf@gmail.com>
Description Several shrinkage estimators of the effect-size of a parameter of interest.
License GPL-3
Depends R (>= 2.14.0), methods, PsiHat, multtest, limma
NeedsCompilation no
Repository CRAN
Date/Publication 2015-10-23 18:26:09

R topics documented:

Shrinkage-package .......................................................... 2
nhard.threshold.est ......................................................... 3
nIC.est ................................................................. 5
nlofdr.est ............................................................... 6
nQs.est ................................................................. 9
nscottberger.est ......................................................... 10
other.est ............................................................ 11

Index 14
Several Shrinkage Estimators.

Description

Several shrinkage estimators of the effect-size of a parameter of interest based on different criteria.

Details

Package: Shrinkage
Type: Package
Version: 1.0
Date: 2015-10-20
License: GPL-3
Depends: methods, PsiHat, multtest, limma

Author(s)

Code: Corey M. Yanofsky, Zahra Montazeri, David R. Bickel and Marta Padilla
Documentation: Alaa Ali and Marta Padilla
Maintainer: M. Padilla <padilla.mpf@gmail.com>

References


See Also

PsiHat, multtest and locfdr packages.

Examples

# simulate some data sets: matrices of log-abundance levels
nsam<-5  # number of individuals
nfeat<-6  # number of features (metabolites, genes,...)
diffs<-c(1,4)  # features with differential log-abundance levels
lfcc<-5  # differential quantity

# create data sets:
nhard.threshold.est

Hard-threshold estimators.

Description

Hard-threshold estimators based on the raw estimates of the log fold change and their p-values.

Usage

```r
nhard.threshold.est(x, y = NULL, opt = c("fold.change", "twer", "fwer", "lfdr", "lfdr0", "lfdr1"), alpha = 0.05, pval.fun = t.test, arglis.pvalfun = list(), alternative = "two.sided", ...)
nfc.threshold.est(x, y = NULL, threshold = 0.5, ...)
```

Arguments

- **x**: Input data matrix: features(rows) x samples (columns). See examples.
- **y**: Optional input data matrix.
- **opt**: Option for selecting the type of estimator, it is a character:
  - "fold.change" estimator based on the log of the raw estimated fold change (can use function nfc.threshold.est instead)
  - "twer" estimator based on the raw p-value that controls test-wise error rate (TWER).
  - "fwer" estimator based on the adjusted p-value that controls family-wise error rate (FWER) from multtest package.
  - "lfdr" estimator based on the Efrons local false discovery rate (LFDR) from locfdr package.
- **alpha**: Hard threshold value for the general function nhard.threshold.est.
- **threshold**: Hard threshold value for function nfc.threshold.est.
pval.fun Function to compute p-values from the input data. Usually: "t.test", "wilcox.test", etc.

alternative Argument for input function pval.fun, type of p-values to be computed: "less", "greater", "two-sided" (see stats R package).

arglis.pvalfun Further arguments for the input function pval.fun (see stats R package).

Arguments to pass to function mt.rawp2adjp (for opt = FWER), locfdr (for opt = LFDR) or to another internal function.

Value
A vector of length equal to the total number of features (i.e. proteins, genes,...).

Author(s)
Code: Zahra Montazeri, Corey M. Yanofsky, David R. Bickel and Marta Padilla (modifications)
Documentation: Alaa Ali and Marta Padilla

References

See Also
multtest and locfdr packages.

Examples
# simulate some data sets: matrices of log-abundance levels
nsam<-10 # number of individuals
nfeat<40 # number of features (metabolites, genes,...)
diffs<1:7 # features with differential log-abundance levels
lfc<-5 # differential quantity

# create data sets:
x <- matrix(runif(nfeat*nsam), nrow = nfeat, ncol = nsam) # case
y <- matrix(runif(nfeat*nsam), nrow = nfeat, ncol = nsam) # control
x[diffs,] <- x[diffs,] + lfc

# hard threshold estimator on fold change: ---------
z1 <- nhard.threshold.est(x,y,alpha=0.3, opt="fold.change")
z2 <- nfc.threshold.est(x,y,threshold=0.5)

# other options: ------
z4 <- nhard.threshold.est(x=x,y=y,pval.fun=t.test, opt="TWER")
z5 <- nhard.threshold.est(x=x,y=y,alpha=0.05, opt="FWER")
z6 <- nhard.threshold.est(x=x,y=y,pval.fun=wilcox.test, opt="LFDR")
Estimators based on model selection criteria.

Usage

\[
\text{nIC.est}(x, y = \text{NULL}, \text{opt} = \text{c}(\text{"BF"}, \text{"AIC"}, \text{"BIC"}), \text{param} = \text{NULL}, \text{logx} = \text{TRUE}, \ldots)
\]

\[
\text{nBF_estimator}(x, y = \text{NULL}, \text{param} = \text{NULL}, \text{logx} = \text{TRUE}, \ldots)
\]

\[
\text{nAICc_estimator}(x, y = \text{NULL}, \text{param} = \text{NULL}, \text{logx} = \text{TRUE}, \ldots)
\]

\[
\text{nBIC_estimator}(x, y = \text{NULL}, \text{param} = \text{NULL}, \text{logx} = \text{TRUE}, \ldots)
\]

Arguments

- **x**: Input data matrix: features(rows) x samples (columns). See examples.
- **y**: Optional input data matrix.
- **opt**: Option for selecting the type of estimator, it is a character:
  - "BIC" estimator based on the Bayesian information criterion (BIC). Equivalent function: nBIC_estimator
  - "AIC" estimator based on the Akaike information criterion corrected for small samples (AICc). Equivalent function: nAICc_estimator
  - "BF" estimator based on the Bayes factor (BF). Equivalent function: nBF_estimator.
- **param**: Numeric vector, the effect-size of the parameter of interest. If input param = NULL, it is internally computed from the input matrices x and y if they are given.
- **param0**: Value of the effect-size of the parameter of interest corresponding to the null hypothesis (null value)(i.e. log fold change corresponding to no change, usually 0). If input param0 = NULL, it is internally set.
- **logx**: If logx = TRUE (and param = NULL and param0 = NULL), param is computed internally considering that input matrices x and y are logarithms and thus param0 is set to 0.
- ... Further arguments to pass to an internal function.

Value

A vector of length equal to the total number of features (i.e. proteins, genes,...).
Note

When inputs param and/or param0 are not given, they are computed internally from matrices x and y. If logx = TRUE then param = \bar{x} - \bar{y} and param0 is set to 0, while if logx = FALSE then param = \bar{x}/\bar{y} and param0 is set to 1.

Author(s)

Code: Zahra Montazeri, Corey M. Yanofsky, David R. Bickel and Marta Padilla (modifications)
Documentation: Alaa Ali and Marta Padilla

References


Examples

```r
# simulate some data sets: matrices of log-abundance levels
nsam<5 # number of individuals
nfeat<6 # number of features (metabolites, genes,...)
diffs<4 # features with differential log-abundance levels
lfc<5 # differential quantity

# create xprnSet, xprnSetPair and numeric objects:
x <- matrix(rnorm(nfeat*nsam), nrow = nfeat, ncol = nsam) # case
y <- matrix(rnorm(nfeat*nsam), nrow = nfeat, ncol = nsam) # control
x[diffs,] <- x[diffs,] + lfc

# examples: ---------
z1 <- nIC.est(x=x, opt='BIC')
z2 <- nIC.est(x=x, opt='BF')
z3 <- nIC.est(x=x, opt='AIC')
z4 <- nIC.est(x=x, y=y, opt='BIC')
z5 <- nIC.est(x=x, y=y, opt='BF')
z6 <- nIC.est(x=x, y=y, opt='AIC')
```

nlocfdr.est

LFDR based shrinkage estimator.

Description

Shrinkage estimator based on Efrons local false discovery rate (LFDR).
Usage

nlocfdr.est(x.stat = NULL, y.pvalue = NULL, pval.fun = t.test, arglis.pvalfun = list(),
alternative = "greater", param = NULL, param0 = NULL, logx = TRUE, nulltype = 0,
q.norm = T, ...)

nlocfdr.x(x = NULL, y = NULL, pval.fun = t.test, arglis.pvalfun = list(),
alternative = "greater", param = NULL, param0 = NULL, logx = TRUE, nulltype = 0,
q.norm = T, ...)

nlocfdr.stat(stat = NULL, pvalue = NULL, param = NULL, param0 = 0, nulltype = 0,
q.norm = T, ...)

Arguments

x.stat Input data matrix (features (rows) x samples (columns)) or numeric vector of
statistics for the general function nlocfdr.est.
y.pvalue Optional input data matrix or numeric vector of pvalues for the general function
nlocfdr.est.
x Input data matrix: features (rows) x samples (columns) for function nlocfdr.x. See examples.
y Optional input data matrix for function nlocfdr.x.
stat Input (numeric) vector of statistics for function nlocfdr.stat. Pvalues (pvalue)
or statistics (stat) must be non null.
pvalue Input (numeric) vector of pvalues for function nlocfdr.stat (see Note).
pval.fun Function to compute p-values from the input data. Usually: "t.test", "wilcox.test",
etc.
alternative Argument for input function pval.fun, type of p-values to be computed: "less",
"greater", "two-sided" (see stats R package).
arglis.pvalfun Further arguments for the input function pval.fun (see stats R package).
param Numeric vector as the parameter of interest (effect-size). If input param = NULL,
it is internally computed from the input matrices x and y if they are given.
param0 Null value of the parameter of interest (i.e. log fold change corresponding to no
change, usually 0). If input param0 = NULL, it is internally set.
logx If logx = TRUE (and param = NULL and param0 = NULL), param is computed
internally considering that input matrices x and y are logarithms, and param0 is
set to 0.
nulltype Parameter for selection of the type of null hypothesis distribution in Efrons
method (package locfdr):
0: theoretical null hypothesis distribution,
1-3: empirical null hypothesis distribution (See function locfdr in locfdr pack-
age).
q.norm If q.norm = TRUE, stat is internally computed from pvalue (see Note).
... Further arguments to pass to function locfdr (see package locfdr).
Value

A vector of length equal to the total number of features (i.e. proteins, genes,...).

Note

when stat is empty (or q.norm = TRUE), the vector stat is internally computed by qnorm(pvalue)
if pvalue is available.

When inputs param and/or param0 are not given, they are computed internally from matrices x and y.
If logx = TRUE then param = \( \bar{x} - \bar{y} \) and param0 is set to 0, while if logx = FALSE then param
= \( \bar{x}/\bar{y} \) and param0 is set to 1.

Function nlocfdr.stat needs inputs param and param0.

Author(s)

Code: Corey M. Yanofsky, Zahra Montazeri, David R. Bickel and Marta Padilla (modifications)
Documentation: Alaa Ali and Marta Padilla

References

Yanofsky, C. M., & Bickel, D. R. (2010). Validation of differential gene expression algorithms:
Application comparing fold-change estimation to hypothesis testing. BMC Bioinformatics, 11, 63.
Montazeri, Z., Yanofsky, C. M., & Bickel, D. R. (2010). Shrinkage estimation of effect sizes as an
alternative to hypothesis testing followed by estimation in high-dimensional biology: Applications
to differential gene expression. Statistical Applications in Genetics and Molecular Biology, 9, 23.

See Also

Function and package locfdr.

Examples

# simulate some data sets: matrices of log-abundance levels
nsam<20 # number of individuals
nfeat<40 # number of features (metabolites, genes,...)
diffs<1:4 # features with differential log-abundance levels
lfc<5 # differential quantity

# create data sets:
x <- matrix(rnorm(nfeat*nsam), nrow = nfeat, ncol = nsam) # case
y <- matrix(rnorm(nfeat*nsam), nrow = nfeat, ncol = nsam) # control
x[diffs,] <- x[diffs,] + lfc

stat<-rnorm(nfeat) # a vector of statistics
stat[diffs]<-stat[diffs]+lfc

# shrinkage estimator ---------
z1 <- nlocfdr.stat(stat=stat,param=rowMeans(x))
z2 <- nlocfdr.x(x=x,y=y,pval.fun="wilcox.test",nulltype=1,df=3)
Description

Frequentist Q1 and Q2 estimators.

Usage

nqs.est(x, y = NULL, opt = "Q1", mu0 = 0, c = 0.5, a = 0.4, b = 0.01, h = 1)
nQ1.est(x, y=NULL, mu0=0, c=0.5, h=1)
nQ2.est(x, y=NULL, mu0=0, c=0.5, a=0.4, b=0.01)

Arguments

x Input data matrix: features (rows) x samples (columns). See examples.
y Optional input data matrix.
opt Option for selecting the type of estimator, it is a character:
"Q1" Estimator based on the log of the raw estimated fold change. Equivalent function: nQ1.est
"Q2" Estimator based on the raw p-value that controls test-wise error rate (TWER). Equivalent function: nQ2.est
h Tuning parameters for Q1 estimator.
a, b Tuning parameters for Q2 estimator.
c Tuning parameter for Q1 and Q2 estimator.
mu0 Effect size corresponding to the null hypothesis (i.e. log fold change corresponding to no change, usually 0).

Value

A vector of length equal to the total number of features (i.e. proteins, genes,...).

Author(s)

Code: Zahra Montazeri, Corey M. Yanofsky, David R. Bickel and Marta Padilla (modifications)
Documentation: Alaa Ali and Marta Padilla
References


Examples

```r
# simulate some data sets: matrices of log-abundance levels
nsam<5  # number of individuals
nfeat<6  # number of features (metabolites, genes,...)
diffs<-c(1,4)  # features with differential log-abundance levels
lfc<5  # differential quantity

# create data matrices; features x samples:
x <- matrix(rnorm(nfeat*nsam), nrow = nfeat, ncol = nsam)  # case
y <- matrix(rnorm(nfeat*nsam), nrow = nfeat, ncol = nsam)  # control
x[diffs,] <- x[diffs,] + lfc

# Q1: ---------
out <- nQs.est(x=x, opt='Q1')
out <- nQ1.est(x=x, y=y, h=0.9)
out <- nQ1.est(x=x, y=y)
out <- nQ1.est(x=x, mu=0.1, c=0.4)

# Q2: ---------
z1 <- nQs.est(x=x, y=y, opt='Q2', mu=0.2)
z2 <- nQ2.est(x=x, y=y, c=0.4)
z3 <- nQ2.est(x=x, a=0.4, b=0.02)
z4 <- nQ2.est(x=x)
```

nscottberger.est  

Scott-Berger estimator.

Description

Estimator that implements a slightly altered version of the model of Scott and Berger (see references).

Usage

nscottberger.est(x, y = NULL, scaled = F, ...)

Arguments

- **x**: Input data matrix: features(rows) x samples (columns). See examples.
- **y**: Optional input data matrix.
- **scaled**: Logical.
- **...**: Further arguments to pass to an internal function.
**Value**

A vector of length equal to the total number of features (i.e. proteins, genes,...).

**Author(s)**

Code: Corey M. Yanofsky, Zahra Montazeri, David R. Bickel and Marta Padilla (modifications)
Documentation: Alaa Ali and Marta Padilla

**References**


**Examples**

```r
# simulate some data sets: matrices of log-abundance levels
nsam<-10       # number of individuals
nfeat<-40      # number of features (metabolites, genes,...)
diffs<-c(1:4)  # features with differential log-abundance levels
lfc<-5         # differential quantity

# create data sets:
x <- matrix(runif(nfeat*nsam), nrow = nfeat, ncol = nsam) # case
y <- matrix(runif(nfeat*nsam), nrow = nfeat, ncol = nsam) # control
x[diffs,] <- x[diffs,] + lfc

# scottberger estimator: -------
z1 <- nscottberger.est(x=x)
z2 <- nscottberger.est(x=x, y=y)
```

**Other shrinkage estimators.**

**Description**

Other shrinkage estimators.

**Usage**

```r
other.est(x, y = NULL, opt = c("limma","pseudo","lfdr0","lfdr1"), pval.fun = t.test, alternative = "greater", arglis.pvalfun = list(), ...)
npseudo.est(x, y = NULL, ...)
```
nlimma.est(x, y = NULL, ...)
nlfdr0.est(x, y = NULL, pval.fun = t.test, alternative = "greater",
arglis.pvalfun = list(), ...)
nlfdr1.est(x, y = NULL, pval.fun = t.test, alternative = "greater",
arglis.pvalfun = list(), ...)

Arguments

x
Input data matrix: features(rows) x samples (columns). See examples.

y
Optional input data matrix.

opt
Option for selecting the type of estimator, it is a character:
"locfdr0" estimator based on the local false discovery rate (LFDR) with theoretical null (null hypothesis distribution follows N(0,1)). Equivalent function: nlfdr0.est.
"locfdr1" estimator based on the local false discovery rate (LFDR) with empirical null (null hypothesis distribution estimated from data). Equivalent function: nlfdr1.est.
"limma" estimator based on the raw p-value that controls test-wise error rate (TWER). Equivalent function: nlimma.est
"pseudo" estimator based on the adjusted p-value that controls family-wise error rate (FWER). Equivalent function: npseudo.est

pval.fun
Function to compute p-values from the input data. Usually: "t.test", "wilcox.test", etc.

alternative
Argument for input function pval.fun, type of p-values to be computed: "less", "greater", "two-sided" (see stats R package).

arglis.pvalfun
Further arguments to pass to input function pval.fun (see stats R package).

...
Further arguments to pass to internal an function.

Value
A vector of length equal to the total number of features (i.e. proteins, genes,...).

Author(s)
Code: Corey M. Yanofsky, Zahra Montazeri, David R. Bickel and Marta Padilla (modifications)
Documentation: Alaa Ali and Marta Padilla

References
Examples

```r
# simulate some data sets: matrices of log-abundance levels
nsam<-25  # number of individuals
nfeat<-50  # number of features (metabolites, genes,...)
diffs<-c(1,4)  # features with differential log-abundance levels
lfc<-5  # differential quantity

# create data sets:
x <- matrix(runif(nfeat*nsam), nrow = nfeat, ncol = nsam) # case
y <- matrix(runif(nfeat*nsam), nrow = nfeat, ncol = nsam) # control
x[diffs,] <- x[diffs,] + lfc

# moderated t-stat estimators: --------

z1 <- other.est (x=x,y=y,opt="limma")
z2 <- other.est (x=x,y=y,opt="pseudo")
z3 <- other.est (x=x,y=y,opt="lfdr0",pval.fun="t.test")
```
Index

*Topic AIC
nIC.est, 5

*Topic Akaike information criterion
nIC.est, 5

*Topic BIC
nIC.est, 5

*Topic Bayes factor
nIC.est, 5

*Topic Bayesian information criterion
nIC.est, 5

*Topic FWER
nhard.threshold.est, 3

*Topic Hard-threshold
nhard.threshold.est, 3

*Topic LFDR
nhard.threshold.est, 3

*Topic Q1
nQs.est, 9

*Topic Q2
nQs.est, 9

*Topic Scottberger estimator
nscottberger.est, 10

*Topic TWER
nhard.threshold.est, 3

*Topic effect size estimator
Shrinkage-package, 2

*Topic effect size
Shrinkage-package, 2

*Topic family-wise error rate
nhard.threshold.est, 3

*Topic frequentist shrinkage estimator
nQs.est, 9

*Topic limma
other.est, 11

*Topic local false discovery rate
nhard.threshold.est, 3

nlocfdr.est, 6

*Topic locfdr
other.est, 11

*Topic pseudo
other.est, 11

*Topic shrinkage effect size estimator
Shrinkage-package, 2

*Topic shrinkage estimator
Shrinkage-package, 2

*Topic shrinkage
Shrinkage-package, 2

*Topic test-wise error rate
nhard.threshold.est, 3

nAICc_estimator (nIC.est), 5
nBF_estimator (nIC.est), 5
nBIC_estimator (nIC.est), 5
nfic.threshold.est

(nhard.threshold.est), 3

nhard.threshold.est, 3
nIC.est, 5

nlfdr0.est (other.est), 11
nlfdr1.est (other.est), 11
nlimma.est (other.est), 11
nlocfdr.est, 6

nlocfdr.stat (nlocfdr.est), 6

nlocfdr.x (nlocfdr.est), 6

npseudo.est (other.est), 11

nQ1.est (nQs.est), 9
nQ2.est (nQs.est), 9
nQs.est, 9

nscottberger.est, 10

other.est, 11

PsiHat, 2

Shrinkage (Shrinkage-package), 2
Shrinkage-package, 2