Package ‘twilight’

April 26, 2017

Version 1.52.0
Title Estimation of local false discovery rate
Author Stefanie Scheid <stefanie.scheid@gmx.de>

Description In a typical microarray setting with gene expression data observed under two conditions, the local false discovery rate describes the probability that a gene is not differentially expressed between the two conditions given its corresponding observed score or p-value level. The resulting curve of p-values versus local false discovery rate offers an insight into the twilight zone between clear differential and clear non-differential gene expression. Package ‘twilight’ contains two main functions: Function twilight.pval performs a two-condition test on differences in means for a given input matrix or expression set and computes permutation based p-values. Function twilight performs a stochastic downhill search to estimate local false discovery rates and effect size distributions. The package further provides means to filter for permutations that describe the null distribution correctly. Using filtered permutations, the influence of hidden confounders could be diminished.

Maintainer Stefanie Scheid <stefanie.scheid@gmx.de>
Depends R (>= 2.10), splines (>= 2.2.0), stats (>= 2.2.0), Biobase(>= 1.12.0)
Imports Biobase, graphics, grDevices, stats
Suggests golubEsets (>= 1.4.2), vsn (>= 1.7.2)
License GPL (>= 2)
URL http://compdiag.molgen.mpg.de/software/twilight.shtml
biocViews Microarray, DifferentialExpression, MultipleComparison
NeedsCompilation yes

R topics documented:
exfdr .................................................. 2
expval .................................................. 2
plot.twilight .......................................... 3
print.twilight ........................................ 5
Description

Application of function twilight on twilight object data(expval).

The function call was exfdr <- twilight(expval,B=1000).

Usage

data(exfdr)

Format

A twilight object.

References


Description

Application of function twilight.pval on leukemia data set of Golub et al. (1999), as given in data(Golub_Merge) in library(golubEsets).

First step was the variance-stabilizing normalization of Huber et al. (2002) in library(vsn):
golubNorm <- justvsn(Golub_Merge).

The function call was then expval <- twilight.pval(golubNorm,id) with id <- as.numeric(Golub_Merge$ALL.AML)
plot.twilight

Usage
data(expval)

Format
A twilight object.

References


plot.twilight  

Plot function for twilight objects

Description
Interface to plotting of twilight objects. Produces one of five possible plots.

Usage
## S3 method for class 'twilight'
plot(x, which = NULL, grayscale = FALSE, legend = TRUE, ...)

Arguments
x An object of class twilight.
which A character string specifying the plot to be made.
grayscale Logical value. Specifying whether plots should be colored or grayscale. Only necessary for "scores" and "fdr".
legend Logical value. Produces a legend for "scores" and "effectsize". A legend for "fdr" is only available if bootstrap estimates exist.
... Additional graphical arguments.
Details

Option `which="scores"` plots the expected versus the observed test statistics and draws confidence lines calculated from permutations. This plot is similar to plots in Tusher et al. (2001).

Option `which="qvalues"` plots \( q \)-values versus the number of rejected hypotheses.

Option `which="fdr"` plots \( p \)-values versus 1 - local false discovery rate, that is the conditional probability of being significant given the corresponding \( p \)-value, plus bootstrap estimates if computed. Bottom ticks are 1%-quantiles of \( p \)-values.

Option `which="volcano"` results in the volcano plot, that is observed score versus 1 - local false discovery rate. Bottom ticks are 1%-quantiles of scores.

Option `which="effectsize"` plots the observed fold change equivalent score distribution overlaid by the estimated effect size distribution, that is distribution of scores under the alternative. This plot is only available if function `twilight.pval` was applied with `method="fc"` as fold change equivalent scores are computed from log ratios.

Option `which="table"` tabulates histogram "effectsize".

Value

No value is returned except for "table".

Author(s)

Stefanie Scheid [http://www.molgen.mpg.de/~scheid](http://www.molgen.mpg.de/~scheid)

References


See Also

twilight.pval, twilight

Examples

```r
### contains a twilight object created by function twilight
data(exfdr)
plot(exfdr)
```
print.twilight

print.twilight  
Print function for twilight objects

Description

Extract and print information about a twilight object.

Usage

```r
## S3 method for class 'twilight'
print(x, ...)
```

Arguments

- `x`  
  Input object of class twilight.
- `...`  
  Additional printing arguments.

Value

No value is returned.

Author(s)

Stefanie Scheid [http://www.molgen.mpg.de/~scheid](http://www.molgen.mpg.de/~scheid)

References


See Also

- `plot.twilight`

Examples

```r
### contains a twilight object created by function twilight
data(exfdr)
print(exfdr)
```
twilight

*Estimation of the local false discovery rate*

**Description**

The function performs the successive exclusion procedure (SEP) as described in Scheid and Spang (2004).

**Usage**

```r
twilight(xin, lambda = NULL, B = 0, boot.ci = 0.95, clus = NULL, verbose = TRUE)
```

**Arguments**

- `xin` Numerical vector of \(p\)-values or a twilight object.
- `lambda` Numerical value denoting the penalty factor. If not specified, the function searches for an appropriate regularization parameter.
- `B` Numerical value specifying the number of bootstrap samples. If not specified, no bootstrap estimates are calculated.
- `boot.ci` Numerical value denoting the probability value for bootstrap confidence intervals of local false discovery rate and prior \(p_0\).
- `clus` A list or numerical value to be passed to `makeCluster(clus,...)` in library(snow). If specified, bootstrapping is performed in parallel. No checks on `clus` are performed. Please make sure that `makeCluster(clus)` works properly in your environment.
- `verbose` Logical value for message printing.

**Details**

In short, the successive exclusion procedure divides the set of \(p\)-values into two parts. The first part is chosen such that it resembles a uniform distribution while containing as many \(p\)-values as possible. This set of \(p\)-values represents the set of \(p\)-values derived from non-induced genes. The height of the uniform distribution is a natural estimate for the mixture parameter \(p_0\). The \(p\)-value not contained in the uniform part serve as representatives of \(p\)-values derived from induced genes. Their distribution is the basis of the final estimate of the local false discovery rate.

NOTE: Library(snow) has to be loaded manually. It is not loaded as part of ‘suggests’ or ‘depends’ because it is only available under UNIX. If twilight does not work with the current version of snow, please send a report.

**Value**

Returns a twilight object consisting of a data.frame named `result` with variables

- `pvalue` Sorted input vector.
- `qvalue` \(q\)-values computed as described in Storey and Tibshirani (2003) with new estimate \(p_0\).
- `index` Index of the original ordering.
- `fdr` Local false discovery rate averaged over 10 runs of SEP.
- `mean.fdr` Bootstrap estimate of local false discovery rate.
lower.fdr  Lower boot.ci-bootstrap confidence bound.
upper.fdr   Upper boot.ci-bootstrap confidence bound.

Values are sorted by p-value.

**Note**

Additional output consists of

- lambda Regularization parameter.
- pi0 SEP estimate of prior probability.
- boot.pi0 Bootstrap estimate and boot.ci-bootstrap confidence bounds.
- boot.ci Passes boot.ci for plotting.
- effect Histogram of effect size distributions averaged over 10 runs of SEP.

If `xin` is of class `twilight`, the remaining slots are filled with corresponding input values. If `xin` is not of class `twilight`, these slots remain free.

**Author(s)**

Stefanie Scheid [http://www.molgen.mpg.de/~scheid](http://www.molgen.mpg.de/~scheid)

**References**


**See Also**

`twilight.pval`, `plot.twilight`, `exfdr`

**Examples**

```r
### twilight object created with B=1000 bootstrap samples
data(exfdr)
print(exfdr)
plot(exfdr)
```
All permutations of a binary vector

Description

For a given binary input vector, the function completely enumerates all possible permutations.

Usage

twilight.combi(xin, pin, bin)

Arguments

xin
   Binary input vector, e.g. class labels.
pin
   Logical value. TRUE if samples are paired, FALSE if not.
bin
   Logical value. TRUE if permutations should be balanced, FALSE if not.

Details

Please note, that the resulting permutations are always as "balanced" as possible. The balancing
is done for the smaller subsample. If its sample size is odd, say 5, twilight.combi computes all
permutations with 2 or 3 samples unchanged. In the paired case, the output matrix contains only
one half of all permutations. The second half is simply 1-output which leads to the same absolute
test statistics in a paired test.

Value

Returns a matrix where each row contains one permuted vector. Note that even for balanced per-
mutations, the first row always contains the original vector. If the number of rows exceeds 10000,
NULL is returned.

Author(s)

Stefanie Scheid http://www.molgen.mpg.de/~scheid

References

Scheid S and Spang R (2004): A stochastic downhill search algorithm for estimating the local false
Scheid S and Spang R (2005): twilight; a Bioconductor package for estimating the local false
discovery rate, Bioinformatics 21(12), 2921–2922.
of large-scale genomic data, in: Apostolico A, Guerra C, Istrail S, Pevzner P, and Waterman M
(Eds.): Research in Computational Molecular Biology: 10th Annual International Conference,
Proceedings of RECOMB 2006, Venice, Italy, April 2-5, 2006. Lecture Notes in Computer Science

See Also

twilight.permute.pair, twilight.permute.unpair
Examples

```r
x <- c(rep(0,4),rep(1,3))
y <- twilight.combi(x, pin=FALSE, bin=FALSE)
```

Description

The function call invokes the filtering for permutations of class labels that produce a set of complete null scores. Depending on the test setting, the algorithm iteratively generates valid permutations of the class labels and computes scores. These are transformed to pooled p-values and each set of permutation p-values is tested for uniformity. Permutations with acceptable uniform p-value distributions are kept. The search stops if either the number `num.perm` of wanted permutations is reached or if the number of possible unique(!) permutations is smaller than `num.perm`. The default values are similar to function `twilight.pval` but please note the details below.

Usage

```r
twilight.filtering(xin, yin, method = "fc", paired = FALSE, s0 = 0, verbose = TRUE, num.perm = 1000)
```

Arguments

- `xin`: Either an expression set (ExpressionSet) or a data matrix with rows corresponding to features and columns corresponding to samples.
- `yin`: A numerical vector containing class labels. The higher label denotes the case, the lower label the control samples to test case vs. control. For correlation scores, `yin` can be any numerical vector of length equal to the number of samples.
- `method`: Character string: "fc" for fold change equivalent test (that is log ratio test), "t" for t-test, and "z" for Z-test. With "pearson" or "spearman", the test statistic is either Pearson’s correlation coefficient or Spearman’s rank correlation coefficient.
- `paired`: Logical value. Depends on whether the samples are paired. Ignored if `method="pearson"` or `method="spearman"`.
- `s0`: Fudge factor for variance correction in the Z-test. Takes effect only if `method="z"`. If `s0=0`: The fudge factor is set to the median of the pooled standard deviations.
- `verbose`: Logical value for message printing.
- `num.perm`: Number of permutations. Within `twilight.pval`, `num.perm` is set to 8.
- `num.take`: Number of permutations kept in each step of the iterative filtering. Within `twilight.pval`, `num.take` is set to the minimum of 50 and `ceiling(num.perm/20)`.

Details

See vignette.
Value

Returns a matrix with permuted input labels `yin` as rows. Please note that this matrix is already translated into binary labels for two-sample testing or to ranks if Spearman’s correlation was chosen. The resulting permutation matrix can be directly passed into function `twilight.pval`. Please note that the first row always contains the original input `yin` to be consistent with the other permutation functions in package `twilight`.

Author(s)

Stefanie Scheid [http://www.molgen.mpg.de/~scheid](http://www.molgen.mpg.de/~scheid)

References


See Also

twilight.pval

Examples

```r
## Not run:
### Leukemia data set of Golub et al. (1999).
library(golubEsets)
data(Golub_Train)

### Variance-stabilizing normalization of Huber et al. (2002).
library(vsn)
golubNorm <- justvsn(Golub_Train)

### A binary vector of class labels.
id <- as.numeric(Golub_Train$ALL.AML)

### Do an unpaired t-test.
### Let's have a quick example with 50 filtered permutations only.
### With num.take=10, we only need 5 iteration steps.
yperm <- twilight.filtering(golubNorm,id,method="t",num.perm=50,num.take=10)
dim(yperm)

### Let's check that the filtered permutations really produce uniform p-value distributions.
### The first row is the original labeling, so we try the second permutation.
yperm <- yperm[,-1]
b <- twilight.pval(golubNorm,yperm[,1],method="t",yperm=yperm)
hist(b$result$pvalue)
## End(Not run)
```
twilight.permute.pair  

**Permutation matrix of paired class labels**

**Description**

The function returns a matrix where each row is a (un)balanced permutation of the input twosample class labels.

**Usage**

```r
twilight.permute.pair(v, m, bal = TRUE)
```

**Arguments**

- `v`: A binary vector representing class labels in original order. Pairs must be in the same order.
- `m`: A numerical value giving the number of permutations.
- `bal`: Logical value. Results in balanced or unbalanced permutations.

**Value**

Returns a matrix where each row contains one permuted vector of class labels. Note that even for balanced permutations, the first row always contains the original vector.

**Author(s)**

Stefanie Scheid [http://www.molgen.mpg.de/~scheid](http://www.molgen.mpg.de/~scheid)

**References**


**See Also**

- `twilight.permute.unpair`
- `twilight.combi`
twilight.permute.unpair

Permutation matrix of unpaired class labels

Description

The function returns a matrix where each row is a (un)balanced permutation of the input twosample class labels.

Usage

twilight.permute.unpair(v, m, bal = TRUE)

Arguments

v    A binary vector representing class labels in original order.
m    A numerical value giving the number of permutations.
bal  Logical value. Results in balanced or unbalanced permutations.

Value

Returns a matrix where each row contains one permuted vector of class labels. Note that even for balanced permutations, the first row always contains the original vector.

Author(s)

Stefanie Scheid http://www.molgen.mpg.de/~scheid

References


See Also

twilight.permute.pair, twilight.combi
twilight.pval

Compute p-values from expression sets

Description
A function to compute two-sample t, Z and fold change equivalent test statistics (paired or unpaired) and correlation coefficients. Based on permutations, expected test statistics as given in Tusher et al. (2001) and empirical p-values are computed. Additional output are q-values computed as given in Storey and Tibshirani (2003). The resulting object is of class twilight and can be passed to functions twilight or plot.twilight.

Usage
```r
twilight.pval(xin, yin, method = "fc", paired = FALSE, B = 1000, yperm = NULL, balance = FALSE, quant.ci = 0.95, s0=NULL, verbose = TRUE, filtering = FALSE)
```

Arguments
- **xin**: Either an expression set (ExpressionSet) or a data matrix with rows corresponding to features and columns corresponding to samples.
- **yin**: A numerical vector containing class labels. The higher label denotes the case, the lower label the control samples to test case vs. control. For correlation scores, yin can be any numerical vector of length equal to the number of samples.
- **method**: Character string: "fc" for fold change equivalent test (that is log ratio test), "t" for t-test, and "z" for Z-test. With "pearson" or "spearman", the test statistic is either Pearson's correlation coefficient or Spearman's rank correlation coefficient.
- **paired**: Logical value. Depends on whether the samples are paired. Ignored if method="pearson" or method="spearman".
- **B**: Numerical value specifying the number of permutations.
- **yperm**: Optional matrix containing in each row a permutation of the class labels in binary(!) format for two-sample testing. For computation of correlation scores, the rows of yperm have to contain the appropriate values or ranks. Use this argument carefully! If yperm is specified, no other permutation will be done. Please note that the first row of yperm MUST be the input vector yin. Otherwise, the p-value calculation will be incorrect.
- **balance**: Logical value. Depends on whether balanced or unbalanced permutations should be done. Ignored if method="pearson" or method="spearman".
- **quant.ci**: Probability value for confidence lines. Lines are symmetric and denote the quant.ci-quantile of maximal absolute differences between each permutation and the expected scores.
- **s0**: Fudge factor for variance correction in the Z-test. Takes effect only if method="z". If s0=NULL: The fudge factor is set to the median of the pooled standard deviations.
- **verbose**: Logical value for message printing.
- **filtering**: Logical value for filtering for permutations of class labels that produce a set of complete null scores. Invokes function twilight.filtering. If yperm is specified, no filtering will be done. Note that the filtering is done on unbalanced permutations even if balance=TRUE.
Details

Please see vignette for detailed information.

Value

Returns a twilight object consisting of a data.frame named result with variables

- **observed**: Observed test statistics.
- **expected**: Mean of order statistics of the permutation statistics.
- **candidate**: Binary vector. "1" for observations exceeding the confidence lines.
- **pvalue**: Empirical p-values from two-sided hypothesis tests.
- **qvalue**: q-values computed as described in Storey and Tibshirani (2003).
- **index**: Index of the original ordering.

Values are sorted by absolute observed scores.

Note

Additional output consists of

- **ci.line**: Quantile corresponding to quant.ci, passed for plotting.
- **pi0**: Estimated prior probability.
- **s0**: Estimated fudge factor if method="z".
- **call**: Character string of function arguments.
- **quant.ci**: Passes quant.ci for plotting.

The remaining slots are left free for function twilight.

Author(s)

Stefanie Scheid [http://www.molgen.mpg.de/~scheid](http://www.molgen.mpg.de/~scheid)

References


See Also
twilight.plot.twilight, twilight.combi, twilight.filtering, expval, twilight.teststat

Examples

```r
### twilight object created from Golub data set
data(expval)
print(expval)
plot(expval)
```

twilight.teststat Interface to the test statistics provided within 'twilight'

Description

A function to compute two-sample t, Z and fold change equivalent test statistics (paired or unpaired) and correlation coefficients.

Usage

```r
twilight.teststat(xin, yin, method = "fc", paired = FALSE, s0 = NULL)
```

Arguments

- **xin**
  - Either an expression set (ExpressionSet) or a data matrix with rows corresponding to features and columns corresponding to samples.

- **yin**
  - A numerical vector containing class labels. The higher label denotes the case, the lower label the control samples to test case vs. control. For correlation scores, yin can be any numerical vector of length equal to the number of samples.

- **method**
  - Character string: "fc" for fold change equivalent test (that is log ratio test), "t" for t-test, and "z" for Z-test of Tusher et al. (2001). With "pearson" or "spearman", the test statistic is either Pearson’s correlation coefficient or Spearman’s rank correlation coefficient.

- **paired**
  - Logical value. Depends on whether the samples are paired. Ignored if method="pearson" or method="spearman".

- **s0**
  - Fudge factor for variance correction in the Z-test. Takes effect only if method="z". If s0=NULL: The fudge factor is set to the median of the pooled standard deviations.

Details

Please see vignette for detailed information.

Value

Returns a list with two components: a numerical vector of observed test statistics observed. Each entry corresponds to one row of the input data matrix. Also, the estimated fudge factor s0 is returned. In any other case except method="z", s0 is zero.
Author(s)

Stefanie Scheid http://www.molgen.mpg.de/~scheid

References


See Also

twilight.pval

Examples

### Z-test on random values

M <- matrix(rnorm(20000),nrow=1000)
id <- c(rep(1,10),rep(0,10))
stat <- twilight.teststat(M,id,method="z")

### Pearson correlation

id <- 1:20
stat <- twilight.teststat(M,id,method="pearson")
Index

*Topic datagen
  twilight.combi, 8
  twilight.filtering, 9
  twilight.permute.pair, 11
  twilight.permute.unpair, 12
*Topic datasets
  exfdr, 2
  expval, 2
*Topic hplot
  plot.twilight, 3
*Topic nonparametric
  twilight, 6
  twilight.pval, 13
  twilight.teststat, 15
*Topic print
  print.twilight, 5

exfdr, 2, 7
expval, 2, 15
plot.twilight, 3, 5, 7, 15
print.twilight, 5

twilight, 4, 6, 15
twilight.combi, 8, 11, 12, 15
twilight.filtering, 9, 15
twilight.permute.pair, 8, 11, 12
twilight.permute.unpair, 8, 11, 12
twilight.pval, 4, 7, 10, 13, 16
twilight.teststat, 15, 15