

Package ‘RRHO’

October 8, 2014

Type Package

Title Inference on agreement between two lists

Version 1.2.0

Date 2013-06-21

Author Jonathan Rosenblatt and Jason Stein

Maintainer Jonathan Rosenblatt <john.ros.work@gmail.com>

Description

The package is aimed at inference on the amount of agreement in two sorted lists using the Rank-Rank Hypergeometric Overlap test.

Collate 'ExpressionAnalysis.R'

License GPL-2

Depends VennDiagram, grid

biocViews Genetics, SequenceMatching, Microarray, Transcription

R topics documented:

RRHO-package	2
pvalRRHO	2
RRHO	4

Index	6
--------------	----------

RRHO-package

Test overlap using the Rank-Rank Hypergeometric test

Description

The package is aimed at inference on the amount of agreement in two sorted lists using the Rank-Rank Hypergeometric Overlap test.

Details

Package: RRHO
Type: Package
Version: 0.3
Date: 2013-06-21
License: GPL-2

See [RRHO](#) to get started.

Author(s)

Jason Stein and Jonathan Rosenblatt Maintainer: Jonathan Rosenblatt <john.ros.work@gmail.com>

See Also

[RRHO](#)

pvalRRHO

Compute the significance of the overlap between two lists

Description

Computes the significance of the agreements between lists as returned by [RRHO](#) using resampling.

Usage

```
pvalRRHO(RRHO.obj, replications, stepsize=RRHO.obj$stepsize, FUN= max)
```

Arguments

RRHO.obj The output object of the [RRHO](#) function.
replications The number of samples to be taken from the distribution of the aggregated test statistic.

stepsize	Controls the resolution of the test: how many items between any two overlap tests (i.e., between any two i 's and two j 's.)
FUN	The function aggregating information from the whole overlap matrix into one summary statistic. Typically the <i>min</i> pvalue, or <i>max</i> on $-\log(pval)$ scale.

Details

The distribution of $FUN(-\log(pval))$ is computed using resampling.

The aggregating function will typically be the max function, corresponding to the maximal $-\log(pvalue)$, i.e., the most significant agreement over all sublists.

The distribution is computed by resampling pairs of null sequences, computing the significances of all the overlaps as done in the reference, applying the aggregating function supplied by the user, and returning the permutation based significance.

Value

pval	The FWER corrected significance of observed aggregated pvalue.
FUN.ecdf	The simulated sampling distribution of the aggregated pvalues.
FUN	The matrix aggregation function used. typically max for minimal p-value.
n.items	Length of lists.
stepsize	See RRHO
replications	The number of simulation replications.
call	The function call.

Note

Might take a long time to run. Depending on the number of replications, the item (gene) count and the stepsize.

Also note that the significance returned is a conservative value (by a constant of $1/\text{replications}$).

Author(s)

Jonathan Rosenblatt

See Also

[RRHO](#)

Examples

```
list.length <- 100
list.names <- paste(Gene,1:list.length, sep=)
gene.list1<- data.frame(list.names, sample(list.length))
gene.list2<- data.frame(list.names, sample(list.length))
RRHO.example <- RRHO(gene.list1, gene.list2)
pval.testing <- pvalRRHO(RRHO.example,50)
```

RRHO

*Rank-Rank Hypergeometric Overlap Test***Description**

The function tests for significant overlap between two sorted lists using the method in the reference.

Usage

```
RRHO(list1, list2, stepsize = defaultfstepSize(list1, list2), labels, plots = FALSE, outputdir = NULL, BY)
```

Arguments

<code>list1</code>	data.frame. First column is the element (possibly gene) identifier, and the second is its value.
<code>list2</code>	data.frame. First column is the element (possibly gene) identifier, and the second is its value.
<code>stepsize</code>	Controls the resolution of the test: how many items between any two overlap tests.
<code>labels</code>	Character vector with two elements: the labels of the two lists.
<code>plots</code>	Logical. Should output plots be returned?
<code>outputdir</code>	Path name where plots are returned.
<code>BY</code>	Logical. Should Benjamini-Yekutieli FDR corrected p-values be computed?

Details

Following the method in the reference, the function computes the number of overlapping elements in the first i and j elements of each list, and return the observed significance of this overlap using a hypergeometric test (see [fisher.test](#)). The output is returned as a list of matrices including: the overlap in the first i, j elements and the significance of this overlap. Optional outputs include plots of these matrices in .jpg format.

Value

<code>hypermat</code>	Matrix of $-\log(pvals)$ of the test for the first i, j elements of the lists.
<code>hypermat.counts</code>	Counts of the number of agreements in the first i, j elements of the lists.
<code>hypermat.by</code>	An optional output of the B-Y corrected p-values of <code>hypermat</code>

Warning

Unlike the reference, we output the p-values in natural log scale and not in log 10 scale.

Author(s)

Jonathan Rosenblatt and Jason Stein

References

Plaisier, Seema B., Richard Taschereau, Justin A. Wong, and Thomas G. Graeber. "Rank-rank Hypergeometric Overlap: Identification of Statistically Significant Overlap Between Gene-expression Signatures." *Nucleic Acids Research* 38, no. 17(September 1, 2010)

Benjamini, Y., and D. Yekutieli. 2001. "The Control of the False Discovery Rate in Multiple Testing Under Dependency." *ANNALS OF STATISTICS* 29 (4): 1165-1188.

See Also

[pvalRRHO](#)

Examples

```
list.length <- 100
list.names <- paste(Gene,1:list.length, sep="")
gene.list1<- data.frame(list.names, sample(100))
gene.list2<- data.frame(list.names, sample(100))
RRHO.example <- RRHO(gene.list1, gene.list2)
image(RRHO.example$hypermat)
```

Index

*Topic **htest**

pvalRRHO, 2

RRHO, 4

RRHO-package, 2

fisher.test, 4

pvalRRHO, 2, 5

RRHO, 2, 3, 4

RRHO-package, 2