Package ‘RRHO’

March 29, 2017

Type Package
Title Inference on agreement between ordered lists
Version 1.14.0
Date 2014-06-26
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.
License GPL-2
Depends R (>= 2.10), grid
Imports VennDiagram
Suggests lattice
Enhances biocViews Genetics, SequenceMatching, Microarray, Transcription
NeedsCompilation no

R topics documented:

RRHO-package ............................................................... 1
HNP .................................................................................. 2
pvalRRHO ........................................................................ 3
RRHO ............................................................................. 4
RRHOComparison .............................................................. 6

Index

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
See **RRHO** to get started.

**Author(s)**

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

**See Also**

RRHO, RRHOComparison

---

**HNP**  
**RRHO comparison data sets.**

**Description**

RRHO comparison data sets. See references for details.

**Usage**

data(lists)

**Format**

Three data frames: HNP, My and Sestan. Each is a data.frame with gene identifiers and sorting values so that they can be used as inputs to RRHOComparison.

**References**


**Examples**

data(lists)  
str(HNP); str(Sestan); str(My)
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 $\min$ p-value, or $\max$ on $-\log(pval)$ scale.

**Details**

The distribution of $\text{FUN}(-\log(pval))$ 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
RRHO

Rank-Rank Hypergeometric Overlap Test

Description

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

Usage

```r
RRHO(
  list1, list2,
  stepsize = defaultStepSize(list1, list2),
  labels,
  alternative,
  plots = FALSE,
  outputdir = NULL,
  BY = FALSE,
  log10.ind=FALSE)
```

Arguments

- `list1`: data.frame. First column is the element (possibly gene) identifier, and the second is its value on which to sort. For differential gene expression, values are often -log10(P-value) * sign(effect).
- `list2`: data.frame. Same as `list1`.
- `stepsize`: Controls the resolution of the test: how many items between any two overlap tests.
- `labels`: Character vector with two elements: the labels of the two lists.
- `alternative`: Either "enrichment" for a one sided test, or "two.sided" for a two sided test. See Details section.
- `plots`: Logical. Should output plots be returned?
- `outputdir`: Path name where plots are returned.
- `BY`: Logical. Should Benjamini-Yekutieli FDR corrected pvalues be computed?
- `log10.ind`: Logical. Should pvalues be reported and plotted in -log10 scale and not -log scale?
Details

Following the method in the reference, the function computes the number of overlapping elements in the first \(i \ast \text{stepsize}\) and \(j \ast \text{stepsize}\) elements of each list, and return the observed significance of this overlap using a hypergeometric test (see \texttt{fisher.test}). The output is returned as a list of matrices including: the overlap in the first \(i \ast \text{stepsize}, j \ast \text{stepsize}\) elements and the significance of this overlap.

If \texttt{plots=TRUE} then plots of these matrices are stored in .jpg format. In the case of \texttt{alternative='two.sided'} the pvalue plots are signed, just like in [1], thus distinguishing between over and under enrichment.

Value

\begin{description}
\item[$\text{hypermat}$] Matrix of \(-\log(pvals)\) of the test for the first \(i, j\) elements of the lists.
\item[$\text{hypermat.counts}$] Counts of the number of agreements in the first \(i, j\) elements of the lists.
\item[$\text{hypermat.by}$] An optional output of the B-Y corrected p-values of \text{hypermat}
\item[$\text{hypermat.signed}$] Matrix of the type of deviation from the null. Negative for underenrichment and positive for overenrichment.
\end{description}

Notes

By default, pvalues are reported in (minus) the natural log scale and not in (minus) log 10 scale. This behaviour is governed by \texttt{log10.ind}.

The p-values of the two-sided hypothesis test differ from those in reference [1]. This is because the two-sided p-values suggested in [1], are based on taking either the upper or lower tail of the distribution without appropriately using both tails. This method does not correctly control the type I error rate. In the implementation here, for a two-sided test we sum the probabilities from both tails of the hypergeometric distribution. See the package vignette for a small simulation.

Author(s)

Jonathan Rosenblatt and Jason Stein

References


See Also

\texttt{pvalRRHO}; \texttt{RRHOComparison}
Examples

```r
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))

# Enrichment alternative
RRHO.example <- RRHO(gene.list1, gene.list2, alternative = 'enrichment')
image(RRHO.example$hypermat)

# Two sided alternative
RRHO.example <- RRHO(gene.list1, gene.list2, alternative = 'two.sided')
image(RRHO.example$hypermat)
```

Description

Comparing two RRHO maps where one of the lists is shared between the two maps as in {RRHO map 1: list1 vs list3} vs {RRHO map 2: list2 vs list3}.

Usage

```r
RRHOComparison(list1, list2, list3,
    stepsize, plots = FALSE,
    labels, outputdir = NULL,
    log10.ind)
```

Arguments

- `list1`: A data.frame from experiment 1 with two columns, column 1 is the ‘Gene Identifier’, column 2 is the signed ranking value (e.g. signed -log10 of p-value, or fold change).
- `list2`: Same as `list1`.
- `list3`: Same as `list1`.
- `stepsize`: Integer indicating how many genes to increase by in each algorithm iteration.
- `labels`: Character vector carrying the labels for the outputted plots.
- `plots`: Logical. Should comparisons be plotted?
- `outputdir`: Plot destination directory.
- `log10.ind`: Logical. Should p-values be reported and plotted in -log10 scale and not -log scale?

Details

The difference in [overlap between list1 and list3] compared to the [overlap between list2 and list3]. This is useful for determining if there is a statistically significant difference between two RRHO maps. In other words, this is useful for determining if the overlap between list1 and list3 is statistically different between the overlap between list2 and list3.
RRHO Difference maps are produced by calculating for each pixel the normal approximation of difference in log odds ratio and standard error of overlap between the two RRHO maps. This Z score is then converted to a P-value and corrected for multiple comparisons across pixels [3]. The function will return a RRHO of the significance of overlap between list1 and list3 and list2 and list3. A third RRHO gives the significance of the difference between these two overlap maps. Note that by default all pvalues are outputted in -log scale. This can be changed with the log10.ind argument.

**Value**

A object including:

- `hypermat1`: Pvalues of comparing list1 to list3.
- `hypermat2`: Pvalues of comparing list2 to list3.
- `Pdiff`: The pvalue of the test for a difference in difference between lists 1-3 and 2-3.
- `Pdiff.by`: Pvalues, corrected for the search over all of the list using Benjamini-Yekutieli.

**Author(s)**

Jason Stein and Jonathan Rosenblatt

**References**


**See Also**

RRHO

**Examples**

```r
size<- 500
data.frame(GeneIdentifier=paste('gen',1:size, sep=''),
RankingVal=-log(runif(size)))
RRHOComparison(list1,list2,list3,
    stepsize=10,
    labels=c("list1","list2","list3"),
    plots=TRUE,
    outputdir=temp.dir,
    log10.ind=FALSE)
```
Index

*Topic **datasets**
  HNP, 2
*Topic **htest**
  pvalRRHO, 3
  RRHO, 4
  RRHO-package, 1

  fisher.test, 5

  HNP, 2

  My (HNP), 2

  pvalRRHO, 3, 5

  RRHO, 2–4, 4, 7
  RRHO-package, 1
  RRHOCmparison, 2, 5, 6

  Sestan (HNP), 2