Package ‘TreeBH’

April 28, 2022

Type Package

Title Error Control for Tree-Structured Hypotheses

Version 1.0

Date 2018-03-02

Author Christine B. Peterson

Maintainer Christine B. Peterson <cbpeterson@gmail.com>

Description This package implements a multiple testing procedure which addresses the challenge of controlling error rates at multiple levels of resolution. We frame this problem as the selection of hypotheses which are organized in a tree structure. The procedure controls relevant error rates (in particular, the selective FDR in each level of the tree) given certain assumptions on the dependence among the hypotheses.

Depends qvalue (>= 2.0.0)

License LGPL

R topics documented:

`TreeBH-package` *Error Control for Tree-Structured Hypotheses*

Description

This package implements a multiple testing procedure which addresses the challenge of controlling error rates at multiple levels of resolution. We frame this problem as the selection of hypotheses which are organized in a tree structure. The procedure controls relevant error rates (in particular, the selective FDR in each level of the tree) given certain assumptions on the dependence among the hypotheses. For code designed specifically for error control in the context of single and multi-tissue eQTL analysis, see the TreeQTL package.

Details

Package:	TreeBH
Type:	Package
Version:	1.0
Date:	2013-03-02
License:	LGPL

Author(s)

Christine B. Peterson

Maintainer: Christine B. Peterson <cbpeterson@gmail.com>

References

Marina Bogomolov*, Christine B. Peterson*, Yoav Benjamini, and Chiara Sabatti (2021). Hypotheses on a tree: new error rates and testing strategies. Biometrika. 108(3): 575-590. *Authors contributed equally.

Examples

```
# This example follows the simulation set-up in Section 5.2 of Bogomolov et al. (2021)
# [https://doi.org/10.1093/biomet/asaa086] and Section 6.2 of Barber and Ramdas (2017)
# [http://dx.doi.org/10.1111/rssb.12218]. Specifically, there are a total of 100 x 100
# hypotheses, with the non-null hypotheses (true signals) arranged into 2 blocks of size
# 15 x 15 with 15 additional non-null hypotheses along the diagonal.

# Set up matrix with 0 for true null hypotheses and 1 for non-nulls
truth <- matrix(0, nrow = 100, ncol = 100)
truth[1:15, 1:15] <- 1
truth[16:30, 16:30] <- 1
for (i in 31:45) {
  truth[i, i] <- 1
}

# Generate p-values following known structure
mu <- truth * 3
X <- mu + matrix(rnorm(100 * 100), nrow = 100, ncol = 100)
pval_table <- 1 - pnorm(X)

# Define a three-level hierarchy, with hypotheses grouped
# 1. By row
# 2. In blocks of size 15 (with a final block of size 10) within each row
# 3. Individually within each block
level1_grouping <- as.vector(matrix(rep(1:100, 100), ncol = 100, byrow = FALSE))
level2_grouping <- as.vector(matrix(c(rep(1, 15), rep(2, 15), rep(3, 15),
                                      rep(4, 15), rep(5, 15), rep(6, 15),
                                    rep(7, 10)), ncol = 100, nrow = 100, byrow = TRUE) +
                              t(matrix(0:99 * 7, ncol = 100, nrow = 100, byrow = TRUE)))
level3_grouping <- 1:10000
groups <- cbind(level1_grouping, level2_grouping, level3_grouping)
```

```
# Apply TreeBH procedure, targeting a selective FDR of 0.2 in each level
target_level <- 0.2
treeBH_results <- get_TreeBH_selections(pvals = as.vector(pval_table),
                                        groups = groups,
                                        q = rep(target_level, 3))
```

`get_fisher_p` *Compute Fisher p-value given a list of p-values*

Description

Compute Fisher p-value given a list of p-values, assuming that NAs do not count toward the total number of tests

Usage

```
get_fisher_p(pvals)
```

Arguments

`pvals`	Vector of p-values (which may include NAs)

Value

Value of the Fisher p-value

Author(s)

Christine B. Peterson

References

Fisher, R.A. (1925). *Statistical Methods for Research Workers*.

`get_simes_p` *Compute Simes p-value given a list of p-values*

Description

Compute Simes p-value given a list of p-values, assuming that NAs do not count toward the total number of tests

Usage

```
get_simes_p(pvals)
```

Arguments

`pvals`	Vector of p-values (which may include NAs)

Value

Value of the Simes p-value

Author(s)

Christine B. Peterson

References

R. J. Simes (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*. **73**(3): 751–754.

`get_TreeBH_selections` *Performs multi-level hierarchical selection*

Description

This function identifies hypotheses of interest, where the hypotheses are organized within a hierarchical tree structure. Given p-values for the leaf nodes and a definition of the tree structure, this procedure identifies selected hypotheses, controlling the targeted selective false discovery rate at each level in the tree.

Usage

```
get_TreeBH_selections(pvals, groups, q, test = "simes")
```

Arguments

`pvals`	Vector of `N` p-values (some of which may be `NA`) corresponding to the individual hypotheseses in the most granular level of the tree
`groups`	Matrix defining how hypotheses are grouped within each level of the tree. Specifically, an `N x L` matrix where if hypothesis `n` belongs to the `k`th group in level `l`, `groups[n,l]` should be set to `k`. The final column should just be `1:N` since the most granular grouping is by individual hypothesis
`q`	Vector of `L` error rates (e.g. 0.05) to be targeted for each level in the tree
`test`	Vector of `L-1` combination methods to obtain level `l-1` p-values from level `l` p-values. Options are "simes" and "fisher". Default is "simes" at each level.

Value

`N x L` binary matrix. This matrix matches up with the input argument `groups`. If the `k`th group in level `l` was selected, the first matrix element corresponding to this group will be set to 1 to indicate the group was selected. The remaining entries will be 0s.

Author(s)

Christine B. Peterson

References

Marina Bogomolov*, Christine B. Peterson*, Yoav Benjamini, and Chiara Sabatti (2021). Hypotheses on a tree: new error rates and testing strategies. Biometrika. 108(3): 575-590. *Authors contributed equally.

Index