Package 'gTestsMulti'

October 13, 2022

Type Package

Title New Graph-Based Multi-Sample Tests

Version 0.1.0

Suggests ade4

Description New multi-sample tests for testing whether multiple samples are from the same distribution. They work well particularly for high-dimensional data.
Song, H. and Chen, H. (2022)
<arXiv:2205.13787>.

Author Hoseung Song [aut, cre],
Hao Chen [aut]

Maintainer Hoseung Song <hosong@ucdavis.edu>

License GPL (>= 2)

Imports Matrix, MASS

Encoding UTF-8

NeedsCompilation no

Repository CRAN

Date/Publication 2022-06-09 08:10:21 UTC

R topics documented:

gTestsMulti *New graph-based multi-sample tests*

Description

This package can be used to determine whether multiple samples are from the same distribution.

Author(s)

Hoseung Song and Hao Chen

Maintainer: Hoseung Song (hosong@ucdavis.edu)

References

Song, H. and Chen, H. (2022). New graph-based multi-sample tests for high-dimensional and non-Euclidean data. arXiv:2205.13787

See Also

gtestsmulti

Examples

```
## Mean difference in Gaussian distribution.
d = 50
mu = 0.2
sam = 50

set.seed(500)
X1 = matrix(rnorm(d*sam), sam)
X2 = matrix(rnorm(d*sam,mu), sam)
X3 = matrix(rnorm(d*sam,2*mu), sam)

data_list = list(X1, X2, X3)

# We use 'mstree' in 'ade4' package to construct the minimum spanning tree.
require(ade4)
x = rbind(X1, X2, X3)
E = mstree(dist(x))

a = gtestsmulti(E, data_list, perm = 1000)
# output results based on the permutation and the asymptotic results
# the test statistic values can be found in a$teststat
# p-values can be found in a$pval
```

gtestsmulti *New graph-based multi-sample tests*

Description

This function provides graph-based multi-sample tests.

Usage

```
gtestsmulti(E, data_list, perm=0)
```

Arguments

E	The edge matrix for the similarity graph. Each row contains the node indices of an edge.
data_list	The list of multivariate matrices corresponding to the K different classes. The length of the list is K. Each element of the list is a matrix containing observations as the rows and features as the columns.
perm	The number of permutations performed to calculate the p-value of the test. The default value is 0, which means the permutation is not performed and only approximated p-value based on the asymptotic theory is provided. Doing permutation could be time consuming, so be cautious if you want to set this value to be larger than 10,000.

Value

Returns a list `teststat` with each test statistic value and a list `pval` with p-values of the tests. See below for more details.

S	The value of the test statistic S.
S_A	The value of the test statistic S^A.
S_appr	The approximated p-value of S based on asymptotic theory with a Bonferroni procedure.
S_A_appr	The approximated p-value of S^A based on asymptotic theory.
S_perm	The permutation p-value of S when argument 'perm' is positive.
S_A_perm	The permutation p-value of S^A when argument 'perm' is positive.

See Also

gTestsMulti

Examples

```
## Mean difference in Gaussian distribution.
d = 50
mu = 0.2
sam = 50

set.seed(500)
X1 = matrix(rnorm(d*sam), sam)
X2 = matrix(rnorm(d*sam,mu), sam)
X3 = matrix(rnorm(d*sam,2*mu), sam)

data_list = list(X1, X2, X3)

# We use 'mstree' in 'ade4' package to construct the minimum spanning tree.
require(ade4)
x = rbind(X1, X2, X3)
E = mstree(dist(x))


a = gtestsmulti(E, data_list, perm = 1000)
# output results based on the permutation and the asymptotic results
# the test statistic values can be found in a$teststat
# p-values can be found in a$pval
```

Index