Package 'GroupTest'

January 20, 2025

Type Package

Title Multiple Testing Procedure for Grouped Hypotheses

Version 1.0.1

Date 2015-11-20

Author Zhigen Zhao

Maintainer Zhigen Zhao <zhaozhg@temple.edu>

Description
Contains functions for a two-stage multiple testing procedure for grouped hypothesis, aiming at controlling both the total posterior false discovery rate and within-group false discovery rate.

License GPL-3

NeedsCompilation no

Repository CRAN

Date/Publication 2015-11-26 14:25:49

Contents

GroupTest-package	*Multiple Hypothesis Testing Procedure for the Grouped Hypotheses*

Description

This package provides functions for the multiple hypotheses testing when there exists group structures.

Details

Package:	GroupTest
Type:	Package
Version:	1.0
Date:	2015-11-20
License:	GPL-3

This package provides functions for multiple testing for the grouped hypotheses. The data is an array of list with G list where G is the total number of groups. Each list within this array corresponds to a group, with the test statistic and the group size as its two elements. Under the null hypotheses, the test statistic follows a standard normal distribution.

The main function is GT.wrapper(). One example is provided under this function, explaining the data structure and how to use the package.

Author(s)

Zhigen Zhao <zhaozhg@temple.edu>

Maintainer: Zhigen Zhao <zhaozhg@temple.edu>

References

Liu, Y., Sarkar, S. K., and Zhao, Z. (2015) A New Approach to Multiple Testing of Grouped Hypotheses

He, L., Sarkar, S. K. and Zhao, Z. (2015) Capturing the severity of Type II errors in high-dimensional multiple testing. Journal of Multivariate Analysis. Vol. 142, 106-116.

AYP   *AYP of California, 2013*

Description

This data set is adequate yearly progress (AYP) study of California elementary schools in 2013 comparing the academic performance for socioeconomically advantaged (SEA) against socioeconomically disadvantaged (SED) students in the elementary schools. What is compared are the success rates of SEA students and SED students. The z-test statistic based two sample proportions test is cacluated for each schools. After removing schools with extremely small or large z-values, there are 4118 schools within 701 qualified school districts.

Usage

```
data("AYP")
```

Format

An array of lists.

Details

AYP data set is an array of lists, with each list corresponding to one school district. In each list, three variables are stored:

X: the test statistic for each individual schools within this school district.

md: the number of schools within this school district.

School.District: the name of the school district.

Source

<http://www.cde.ca.gov/ta/ac/ay/aypdatafiles.asp>

References

Liu, Y., Sarkar, S. K., and Zhao, Z. (2015) A New Approach to Multiple Testing of Grouped Hypotheses

Efron, B. (2008) Microarrays, empirical bayes and the two-groups model. Statisitcal Science, 23, 1-22.

Examples

```
data(AYP)

AYP.result <- GT.wrapper( AYP, alpha=0.1, eta=alpha, pi1.ini=0.5,
pi2.1.ini=0.05, L=2, muL.ini=c(3,-2), sigmaL.ini=c(1,1),
cL.ini=c(0.5,0.5), DELTA=0.0001, sigma.KNOWN=TRUE )
```

GroupTest_simulate   *Simulated data set to demonstrate the package*

Description

Simulated data set to demonstrate the package. In this data set, there are three groups. There are 3, 4, and 5 hypotheses respectively among the groups.

Usage

```
data("GroupTest_simulate")
```

Format

An array of lists.

Examples

```
data(GroupTest_simulate)

GT.test <- GT.wrapper( GroupTest_simulate, alpha=0.05, eta=alpha,
pi1.ini=0.7, pi2.1.ini=0.4, L=2, muL.ini=c(-1,1), sigmaL.ini=c(1,2),
cL.ini=c(0.4,0.6), DELTA=0.001, sigma.KNOWN=FALSE )
```

GT.decision   *Between- and within-group decisions*

Description

Based on the α-level and the the local fdr scores, this function provides the decision on between- and within-group levels.

Usage

```
GT.decision(TestStatistic, alpha = 0.05, eta = alpha)
```

Arguments

TestStatistic	An array of list. Each list of the array corresponds to one group, containing the test statistic, stored as X, and the group size, stored as mg.
alpha	the targeted FDR level.
eta	the targeted FDR level within each group. The default and recommended choice is alpha.

Value

TestStatistic An array of list. Each list of the array corresponds to one group, two additional
varialbes: within.group.rej and between.group.rej are stored in each list.

Examples

```
data(GroupTest_simulate)
GroupTest_simulate <- GT.localfdr( GroupTest_simulate, L=2, pi1=0.5, pi2.1=0.5,
muL=c(-1, 1), sigmaL=c(1,2), cL=c(0.4,0.6) )

GroupTest.decision <- GT.decision(GroupTest_simulate, alpha=0.05)
```

GT.em   *EM Algorithm*

Description

This function estimates all the parameters using the EM algorithm. The iteration is termined when
the sum of squared difference of the current updated values and the previous values of the parame-
ters is less than DELTA. A list consisting of all the estimated values of the parameters is returned.

Usage

```
GT.em(TestStatistic, pi1.ini, pi2.1.ini, L, muL.ini, sigmaL.ini, cL.ini,
DELTA, sigma.KNOWN)
```

Arguments

TestStatistic	An array of list. Each list of the array corresponds to one group, containing the test statistic, stored as X, and the group size, stored as mg.
L	The number of Gaussian component under the alternative hypothesis.
pi1.ini	Initial value: the probability that a group is significant.
pi2.1.ini	Initial value: the probability that an individual null hypothesis is false given that the group is significant.
muL.ini	Initial value: a vector of means for all the components of the Gaussian mixture.
sigmaL.ini	Initial value: a vector of standard deviation of all the components of the Gaussian mixture.
cL.ini	Initial value: a vector of the probability for all the components of the Gaussian mixture.
DELTA	The criteria to stop the EM algorithm.
sigma.KNOWN	The boolean variable, indicating whether the variance is known.

Value

This function return a list, consisting of the estimated values of all the parameters. The variables within this list are shown as following:

pi1	estimated value of π_1, the proportion of a group being significant
pi2.1	estimated value of $\pi_{2|1}$, the proportion of a null hypothesis being false within a significant group.
muL	a vector of estimated means for all the components of the Gaussian mixture
sigmaL	a vector of estimated standard deviation of all the components of the Gaussian mixture
cL	a vector of the probability for all the components of the Gaussian mixture
L	the number of components in the Gaussian mixture

Examples

```
data(GroupTest_simulate)
em.estimate <- GT.em( GroupTest_simulate, L=2, pi1.ini=0.7, pi2.1.ini=0.4,
muL.ini=c(-1,1), sigmaL.ini=c(1,2), cL.ini=c(0.4,0.6), DELTA=0.001,
sigma.KNOWN=FALSE )
```

GT.localfdr	*Between and within group local fdr scores*

Description

This function calculate the between-group and within-group local fdr scores for a given set of all the parameters.

Usage

```
GT.localfdr(TestStatistic, pi1, pi2.1, L, muL, sigmaL, cL)
```

Arguments

TestStatistic	An array of list. Each element of the array corresponds to one group, containing the test statistic, stored as X, and the group size, stored as mg.
L	The number of Gaussian component under the alternative hypothesis.
pi1	π_1, the probability that a group is significant.
pi2.1	$\pi_{2|1}$, the probability that an individual null hypothesis is false given that the group is significant.
muL	a vector of means for all the components of the Gaussian mixture.
sigmaL	a vector of standard deviation of all the components of the Gaussian mixture.
cL	a vector of the probability for all the components of the Gaussian mixture.

Value

This function returns an array of G lists where G is the number of groups.

TSGroupTest[[g]]
in each element, the individual conditional local fdr score ($P(\theta_{gj} = 0|x, \theta_g = 1)$), the group-wise local fdr score ($P(\theta_g = 0|x)$), are stored.

Examples

```
data(GroupTest_simulate)
GroupTest_simulate <- GT.localfdr( GroupTest_simulate, L=2, pi1=0.5,
    pi2.1=0.5, muL=c(-1, 1), sigmaL=c(1,2), cL=c(0.4,0.6) )
```

GT.wrapper   *Multiple testing procedure for the grouped hypothesis*

Description

This function is the main function to perform the two-stage testing for the grouped hypotheses.

Usage

```
GT.wrapper(TestStatistic, alpha = 0.05, eta = alpha, pi1.ini = 0.7,
pi2.1.ini = 0.4, L = 2, muL.ini = c(-1, 1), sigmaL.ini = c(1, 1),
cL.ini = c(0.5, 0.5), DELTA = 0.001, sigma.KNOWN=FALSE)
```

Arguments

TestStatistic	An array of list. Each list of the array corresponds to one group, containing the test statistic, stored as X, and the group size, stored as mg.
alpha	the targeted FDR level. By default, it is chosen as 0.05.
eta	the targeted FDR level within each group. The default and recommended choice is alpha. By default, it is chosen as α.
pi1.ini	Initial value: the probability that a group is significant. By default, it is chosen as 0.7
pi2.1.ini	Initial value: the probability that an individual null hypothesis is false given that the group is significant. By default, it is chosen as 0.4.
L	The number of Gaussian component under the alternative hypothesis. By default, it is chosen as 2.
muL.ini	Initial value: a vector of means for all the components of the Gaussian mixture. By default, is is chosen as -1 and 1.
sigmaL.ini	Initial value: a vector of standard deviation of all the components of the Gaussian mixture. By default, it is chosen as 1 and 1.
cL.ini	Initial value: a vector of the probability for all the components of the Gaussian mixture. By default, it is chosen as 50% and 50%.

DELTA	The criteria to stop the EM algorithm. In this algorithm, we calcualte the maximum of absolution difference of the current estiamted value and its previous value for the parameters. By default, it is chosen as 0.0001.
sigma.KNOWN	The boolean variable, indicating whether the variance is known. Be default, it is chosen as FALSE.

Value

The function returns a TSGroupTest object. It contains

parameter	this is a list, consisting of estimated parameters based on the EM algorithm. The elements are π_1, $\pi_{2|1}$, c_l, μ_l, σ_l.
TSGroupTest[[g]]	
	all the quntities regarding the g-th group, including the test statistic within this group, the individual conditional local fdr score ($P(\theta_{gj} = 0|x, \theta_g = 1)$), the group-wise local fdr score ($P(\theta_g = 0|x)$), between-group decision, within-group decision

Examples

```
data(GroupTest_simulate)

GT.Test <- GT.wrapper( GroupTest_simulate, alpha=0.05, eta=alpha,
pi1.ini=0.7, pi2.1.ini=0.4, L=2, muL.ini=c(-1,1), sigmaL.ini=c(1,2),
cL.ini=c(0.4,0.6), DELTA=0.001, sigma.KNOWN=FALSE )
```

Index