

# Package ‘HDMT’

December 5, 2019

**Type** Package

**Title** A Multiple Testing Procedure for High-Dimensional Mediation Hypotheses

**Version** 1.0.1

**Date** 2019-12-4

**Author** James Dai [aut, cre],  
Xiaoyu Wang [aut]

**Maintainer** James Dai <jdai@fredhutch.org>

**Description** A multiple-testing procedure for high-dimensional mediation hypotheses. Mediation analysis is of rising interest in epidemiology and clinical trials. Among existing methods for mediation analyses, the popular joint significance (JS) test yields an overly conservative type I error rate and therefore low power. In the R package 'HDMT' we implement a multiple-testing procedure that accurately controls the family-wise error rate (FWER) and the false discovery rate (FDR) when using JS for testing high-dimensional mediation hypotheses. The core of our procedure is based on estimating the proportions of three component null hypotheses and deriving the corresponding mixture distribution of null p-values. Results of the data examples include better-behaved quantile-quantile plots and improved detection of novel mediation relationships on the role of DNA methylation in genetic regulation of gene expression. With increasing interest in mediation by molecular intermediaries such as gene expression and epigenetic markers, the proposed method addresses an unmet methodological challenge.

**Depends** R (>= 3.4.0)

**Imports** cp4p,fdrtool

**LazyLoad** no

**License** GPL (>= 2)

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2019-12-05 05:50:02 UTC

## R topics documented:

adjust\_quantile . . . . . 2

correct_qqplot . . . . .	3
exercise_input . . . . .	4
fdrest . . . . .	5
fwerest . . . . .	6
nullestimation . . . . .	7
snp_input . . . . .	8

## Index 10

---

adjust_quantile	<i>A function to compute the quantiles of pmax using either approximation or exact method</i>
-----------------	---

---

### Description

A function to generate the quantiles of pmax (the maximum of the two p-values using the mixture null distribution)

### Usage

```
adjust_quantile(alpha00, alpha01, alpha10, alpha1, alpha2,
input_pvalues, method = 0)
```

### Arguments

alpha00	A numeric number represents the proportion of null $H_{00}$
alpha01	A numeric number represents the proportion of null $H_{01}$
alpha10	A numeric number represents the proportion of null $H_{10}$
alpha1	A numeric number represents the proportion of null $\alpha=0$ (association between exposure and mediator)
alpha2	A numeric number represents the proportion of null $\beta=0$ (association between mediator and outcome adjusted for exposure)
input_pvalues	A matrix contains two columns of p-values for candidate mediators. Column 1 is the p-value of testing if a exposure is associated with the mediator ( $\alpha \neq 0$ ). Column 2 is the p-value of testing if a mediator is associated with the outcome adjusted for exposure ( $\beta \neq 0$ )
method	Use the option to choose from two methods. method=0: approximation method without estimating the CDFs; method=1: estimate the CDFs nonparametrically

### Details

A function used to generate the quantiles for the observed p-max values using the mixture null distribution.

### Value

A vector contains the p-values based on generated null distribution

**Author(s)**

James Y. Dai and X. Wang

**References**

James Y. Dai, Janet L. Stanford, Michael LeBlanc. A multiple-testing procedure for high-dimensional mediation hypotheses, *Journal of the American Statistical Association*, 2019, submitted.

**Examples**

```
data(snp_input)
input_pvalues=snp_input
#To save time for illustration, we use 10 percent of rows
input_pvalues=input_pvalues[sample(1:nrow(input_pvalues),
size=ceiling(nrow(input_pvalues)/10)),]

nullprop <- nullestimation(input_pvalues,lambda=0.5)
pnull <- adjust_quantile(nullprop$alpha00,nullprop$alpha01,nullprop$alpha10,
nullprop$alpha1,nullprop$alpha2,input_pvalues,method=0)
```

---

correct\_qqplot

*A function to draw the quantile-quantile plot*

---

**Description**

A function to draw the corrected quantile-quantile (Q-Q) plots, generated by the mixture null distribution (green dots) and by the uniform distribution (red dots)

**Usage**

```
correct_qqplot(pmax, pnull, opt="all")
```

**Arguments**

pmax	The vector for maximum p-values
pnull	The quantiles of p-values based on the estimated mixture null distribution
opt	Option to draw the plot. opt="all":use all the data points, opt="subset": use a subset of the data points, in case there are too many points in a genome-wide setting, to avoid overcrowded Q-Q plot.

**Author(s)**

James Y. Dai and X. Wang

**References**

James Y. Dai, Janet L. Stanford, Michael LeBlanc. A multiple-testing procedure for high-dimensional mediation hypotheses, *Journal of the American Statistical Association*, 2019, submitted.

**Examples**

```

data(snp_input)
input_pvalues=snp_input
#To save time for illustration, we use 10 percent of rows
input_pvalues=input_pvalues[sample(1:nrow(input_pvalues),
size=ceiling(nrow(input_pvalues)/10)),]

pmax <- apply(input_pvalues,1,max)
nullprop <- nullestimation(input_pvalues,lambda=0.5)
pnull1 <- adjust_quantile(nullprop$alpha10,nullprop$alpha01,nullprop$alpha00,
nullprop$alpha1,nullprop$alpha2,input_pvalues,method=1)
correct_qqplot(pmax,pnull1)

```

---

exercise\_input

*An example dataset to demonstrate the usage of 'HDMT'*


---

**Description**

The dataset was included to show the mediation role of DNA methylation in the study of regulation of prostate cancer progression in a Seattle-based cohort of patients diagnosed with clinically localized PCa with exercise as the exposure. Due to space limit, a subset (10 percent) of the full dataset used in the paper is included here.

The dataset is a matrix containing two columns of p-values for candidate mediators. Column 1 is the p-value of testing if a mediator is associated with the outcome conditional on the exposure ( $\beta \neq 0$ ). Column 2 is the p-value of testing if a mediator is associated with the outcome ( $\alpha \neq 0$ ).

**Usage**

```
data("exercise_input")
```

**Format**

The format of exercise\_input is: num [1:47900, 1:2] 0.4966344 0.1048730 0.1005355 0.4946623 ...

**References**

James Y. Dai, Janet L. Stanford, Michael LeBlanc. A multiple-testing procedure for high-dimensional mediation hypotheses, *Journal of the American Statistical Association*, 2019, submitted.

**Examples**

```

data(exercise_input)
dim(exercise_input)

```

---

fdrest *A function to compute the pointwise FDR for every observed p-max*

---

### Description

A function to compute the pointwise FDR based on the proposed JS-mixture method

### Usage

```
fdrest(alpha00, alpha01, alpha10, alpha1, alpha2, input_pvalues, method = 0)
```

### Arguments

alpha00	A numeric number represents the proportion of null $H_{00}$
alpha01	A numeric number represents the proportion of null $H_{01}$
alpha10	A numeric number represents the proportion of null $H_{10}$
alpha1	A numeric number represents the proportion of null $\beta=0$
alpha2	A numeric number represents the proportion of null $\alpha=0$
input_pvalues	A matrix contains two columns of p-values for candidate mediators. Column 1 is the p-value of testing if an exposure is associated with the mediator ( $\alpha \neq 0$ ). Column 2 is the p-value of testing if a mediator is associated with the outcome adjusted for the exposure ( $\beta \neq 0$ )
method	The option to choose from two methods. method=0: approximation without estimating the CDFs; method=1: estimate the CDFs nonparametrically

### Details

A function to compute the pointwise FDR based on the proposed method to estimate the mixture null distribution

### Value

The computed pointwise FDR

### Author(s)

James Y. Dai and X. Wang

### References

James Y. Dai, Janet L. Stanford, Michael LeBlanc. A multiple-testing procedure for high-dimensional mediation hypotheses, Journal of the American Statistical Association, 2019, submitted.

**Examples**

```

data(snp_input)
input_pvalues=snp_input
#To save time for illustration, we use 10 percent of rows
input_pvalues=input_pvalues[sample(1:nrow(input_pvalues),
size=ceiling(nrow(input_pvalues)/10)),]

nullprop <- nullestimation(input_pvalues,lambda=0.5)
fdr <- fdrtest(nullprop$alpha00,nullprop$alpha01,nullprop$alpha10,nullprop$alpha1,
nullprop$alpha2,input_pvalues,method=0)

```

---

fwerest	<i>A function used to compute Family wise error rate (FWER) cutoff for p-max at a designated level</i>
---------	--

---

**Description**

A function to compute FWER cutoff based on the proposed method using the estimated mixture null distribution

**Usage**

```

fwerest(alpha10, alpha01, alpha00, alpha1, alpha2, input_pvalues,
alpha = 0.05, method = 0)

```

**Arguments**

alpha00	A numeric number represents the proportion of null $H_{00}$
alpha01	A numeric number represents the proportion of null $H_{01}$
alpha10	A numeric number represents the proportion of null $H_{10}$
alpha1	A numeric number represents the proportion of null $\beta=0$
alpha2	A numeric number represents the proportion of null $\alpha=0$
input_pvalues	A matrix contains two columns of p-values for candidate mediators. Column 1 is the p-value of testing if an exposure is associated with the mediator ( $\alpha \neq 0$ ). Column 2 is the p-value of testing if a mediator is associated with the outcome adjusted for the exposure ( $\beta \neq 0$ )
alpha	The designated significance level for FWER
method	The option to choose from two methods. method=0: approximation without estimating the CDFs; method=1: estimate the CDFs nonparametrically

**Details**

A function to compute FWER cutoff for p-max accounting for the mixture null distribution.

**Value**

A numeric number represents the output FWER cutoff

**Author(s)**

James Y. Dai and X. Wang

**References**

James Y. Dai, Janet L. Stanford, Michael LeBlanc. A multiple-testing procedure for high-dimensional mediation hypotheses, *Journal of the American Statistical Association*, 2019, submitted.

**Examples**

```
data(snp_input)
input_pvalues=snp_input
#To save time for illustration, we use 10 percent of rows
input_pvalues=input_pvalues[sample(1:nrow(input_pvalues),
size=ceiling(nrow(input_pvalues)/10)),]

nullprop <- nullestimation(input_pvalues,lambda=0.5)
fwercut <- fwerest(nullprop$alpha00,nullprop$alpha01,nullprop$alpha10,
nullprop$alpha1,nullprop$alpha2,input_pvalues,alpha=0.05,method=0)
```

---

nullestimation

*A function to estimate the proportion of three component nulls*


---

**Description**

A function to estimate the proportion of three component nulls. We developed a JS (joint significance)-mixture method based the estimated nulls which can provide much more accurate control of the family-wise error rate (FWER) and the false discovery rate (FDR) compared to the JS-uniform method.

**Usage**

```
nullestimation(input_pvalues, lambda = 0.5)
```

**Arguments**

**input\_pvalues** A matrix contains two columns of p-values for candidate mediators. Column 1 is the p-value of testing if an exposure is associated with the mediator ( $\alpha \neq 0$ ). Column 2 is the p-value of testing if a mediator is associated with the outcome adjusted for the exposure ( $\beta \neq 0$ )

**lambda** A tuning parameter between 0 and 1, the default value is 0.5.

**Details**

A function used to estimate the proportions of the three types of component null hypotheses:

$H_{00}$ :  $\alpha=0$  and  $\beta=0$

$H_{01}$ :  $\alpha=0$  and  $\beta \neq 0$

$H_{10}$ :  $\alpha \neq 0$  and  $\beta=0$

**Value**

A list contains five elements.

alpha00	A numeric number represents the proportion of null $H_{00}$
alpha01	A numeric number represents the proportion of null $H_{01}$
alpha10	A numeric number represents the proportion of null $H_{10}$
alpha1	A numeric number represents the proportion of null $\beta=0$
alpha2	A numeric number represents the proportion of null $\alpha=0$

**Author(s)**

James Y. Dai and X. Wang

**References**

James Y. Dai, Janet L. Stanford, Michael LeBlanc. A multiple-testing procedure for high-dimensional mediation hypotheses, *Journal of the American Statistical Association*, 2019, submitted.

**Examples**

```
data(snp_input)
input_pvalues=snp_input
#To save time for illustration, we use 10 percent of rows
input_pvalues=input_pvalues[sample(1:nrow(input_pvalues),
size=ceiling(nrow(input_pvalues)/10)),]

nullprop <- nullestimation(input_pvalues,lambda=0.5)
```

---

snp\_input

*An example dataset to demonstrate the usage of 'HDMT'*

---

**Description**

The dataset was included to show the mediation role of DNA methylation in the study of genetic regulation of gene expression in primary prostate cancer (PCa) samples from The Cancer Genome Atlas (TCGA) with risk SNPs as the exposure.

The dataset is a matrix containing two columns of p-values for candidate mediators. Column 1 is the p-value of testing if a mediator is associated with the outcome conditional on the exposure ( $\beta \neq 0$ ). Column 2 is the p-value of testing if a mediator is associated with the outcome ( $\alpha \neq 0$ ).



**Usage**

```
data("snp_input")
```

**Format**

The format of `snp_input` is: num [1:69602, 1:2] 0.106 0.999 0.101 0.173 0.89 ...

**References**

James Y. Dai, Janet L. Stanford, Michael LeBlanc. A multiple-testing procedure for high-dimensional mediation hypotheses, *Journal of the American Statistical Association*, 2019, submitted.

**Examples**

```
data(snp_input)  
dim(snp_input)
```

# Index

- \*Topic **composite null**
  - adjust\_quantile, 2
  - fdrest, 5
  - fwertest, 6
  - nullestimation, 7
- \*Topic **dataset**
  - exercise\_input, 4
  - snp\_input, 8
- \*Topic **joint significance**
  - adjust\_quantile, 2
  - fdrest, 5
  - fwertest, 6
  - nullestimation, 7
- \*Topic **mediation**
  - adjust\_quantile, 2
  - fdrest, 5
  - fwertest, 6
  - nullestimation, 7

adjust\_quantile, 2

correct\_qqplot, 3

exercise\_input, 4

fdrest, 5

fwertest, 6

nullestimation, 7

snp\_input, 8