Package ‘HIMA’

March 7, 2018

Type Package
Title High-Dimensional Mediation Analysis
Version 1.0.7
Date 2018-03-06
License GPL-3
Depends R (>= 3.3), ncvreg
Imports stats, iterators, parallel, foreach, doParallel
LazyData false
URL https://github.com/YinanZheng/HIMA
BugReports https://github.com/YinanZheng/HIMA/issues
RoxygenNote 6.0.1
NeedsCompilation no
Author Yinan Zheng [aut, cre],
Haixiang Zhang [aut],
Lifang Hou [aut],
Lei Liu [aut, cph]
Maintainer Yinan Zheng <y-zheng@northwestern.edu>
Repository CRAN
Date/Publication 2018-03-06 23:19:18 UTC
R topics documented:

HIMA-package ................................................................. 2
hima ................................................................. 2
simHIMA ................................................................. 4

Index 6

HIMA-package  

High-dimensional Mediation Analysis

Description

HIMA is an R package for estimating and testing high-dimensional mediation effects in genomic/epigenomic studies.

Package: HIMA
Type: Package
Version: 1.0.7
Date: 2018-03-06
License: GPL-3

Author(s)

Yinan Zheng <y.zheng@northwestern.edu>, Haixiang Zhang <haixiang.zhang@tju.edu.cn>, Lifang Hou <l.hou@northwestern.edu> Lei Liu <lei.liu@wustl.edu>
Maintainer: Yinan Zheng <y.zheng@northwestern.edu>

References


hima

High-dimensional Mediation Analysis

Description

hima is used to estimate and test high-dimensional mediation effects.

Usage

hima(X, Y, M, COV.XM = NULL, COV.MY = COV.XM, family = c("gaussian", "binomial"), penalty = c("MCP", "SCAD", "lasso"), topN = NULL, parallel = FALSE, ncore = 1, verbose = FALSE, ...)
Arguments

- **X**: a vector of exposure.
- **Y**: a vector of outcome. Can be either continuous or binary (0-1).
- **M**: a data.frame or matrix of high-dimensional mediators. Rows represent samples, columns represent variables.
- **COV.XM**: a data.frame or matrix of covariates dataset for testing the association $M \sim X$. Covariates specified here will not participate penalization. Default = NULL. If the covariates contain mixed data types, please make sure all categorical variables are properly formatted as factor type.
- **COV.MY**: a data.frame or matrix of covariates dataset for testing the association $Y \sim M$. Covariates specified here will not participate penalization. If not specified, the same set of covariates for $M \sim X$ will be applied. Using different sets of covariates is allowed but this needs to be handled carefully.
- **family**: either 'gaussian' or 'binomial', depending on the data type of outcome ($Y$). See ncvreg
- **penalty**: the penalty to be applied to the model. Either 'MCP' (the default), 'SCAD', or 'lasso'. See ncvreg.
- **topN**: an integer specifying the number of top markers from sure independent screening. Default = NULL. If NULL, topN will be either ceiling($n/\log(n)$) if family = 'gaussian', or ceiling($n/(2+\log(n))$) if family = 'binomial', where $n$ is the sample size. If the sample size is greater than topN (pre-specified or calculated), all mediators will be included in the test (i.e. low-dimensional scenario).
- **parallel**: logical. Enable parallel computing feature? Default = TRUE.
- **ncore**: number of cores to run parallel computing Valid when parallel == TRUE. By default max number of cores available in the machine will be utilized.
- **verbose**: logical. Should the function be verbose? Default = FALSE.
- **...**: other arguments passed to ncvreg.

Value

A data.frame containing mediation testing results of selected mediators.

- **alpha**: coefficient estimates of exposure ($X$) $\rightarrow$ mediators ($M$).
- **beta**: coefficient estimates of mediators ($M$) $\rightarrow$ outcome ($Y$) (adjusted for exposure).
- **gamma**: coefficient estimates of exposure ($X$) $\rightarrow$ outcome ($Y$) (total effect).
- **alpha*beta**: mediation effect.
- **% total effect**: alpha*beta / gamma. Percentage of the mediation effect out of the total effect.
- **adjusted.p**: statistical significance of the mediator (Bonferroni procedure).
- **BH.FDR**: statistical significance of the mediator (Benjamini-Hochberg procedure).
Examples

n <- 100  # sample size
p <- 500  # the dimension of covariates

# the regression coefficients alpha (exposure --> mediators)
alpha <- rep(0, p)

# the regression coefficients beta (mediators --> outcome)
beta1 <- rep(0, p)  # for continuous outcome
beta2 <- rep(0, p)  # for binary outcome

# the first four markers are true mediators
alpha[1:4] <- c(0.45, 0.5, 0.6, 0.7)
beta1[1:4] <- c(0.55, 0.6, 0.65, 0.7)
beta2[1:4] <- c(1.45, 1.5, 1.55, 1.6)

# these are not true mediators
alpha[7:8] <- 0.5
beta1[5:6] <- 0.8
beta2[5:6] <- 1.7

# Generate simulation data
simdat_cont = simHIMA(n, p, alpha, beta1, seed=1029)
simdat_bin = simHIMA(n, p, alpha, beta2, binaryOutcome = TRUE, seed=1029)

# Run HIMA with MCP penalty by default
# When Y is continuous (default)
hima.fit <- hima(simdat_cont$X, simdat_cont$Y, simdat_cont$M, verbose = TRUE)
head(hima.fit)

# When Y is binary (should specify family)
hima.logistic.fit <- hima(simdat_bin$X, simdat_bin$Y, simdat_bin$M,
  family = "binomial", verbose = TRUE)
head(hima.logistic.fit)

---

simHIMA  
Simulation Data Generator for High-dimensional Mediation Analysis

Description

simHIMA is used to generate simulation data for high-dimensional mediation analysis.

Usage

simHIMA(n, p, alpha, beta, binaryOutcome = FALSE, seed)
simHIMA

Arguments

n  an integer specifying sample size.
\(p\)  an integer specifying the dimension of mediators.
alpha  a numeric vector specifying the regression coefficients alpha (exposure -> mediators).
beta  a numeric vector specifying the regression coefficients beta (mediators -> outcome).
binaryOutcome  logical. Should the simulated outcome variable be binary?
seed  an integer specifying a seed for random number generation.

See Also

see hima to run HIMA.

Examples

n <- 100  # sample size
p <- 500  # the dimension of covariates

# the regression coefficients alpha (exposure -> mediators)
alpha <- rep(0, p)

# the regression coefficients beta (mediators -> outcome)
beta <- rep(0, p)

# the first four markers are true mediators.
alpha[1:4] <- c(0.45, 0.5, 0.55, 0.6)
beta[1:4] <- c(0.5, 0.45, 0.4, 0.35)

alpha[7:8] <- 0.5
beta[5:6] <- 0.5

# Generate simulation data
simdat = simHIMA(n, p, alpha, beta, seed=1029)
Index

*Topic package
  HIMA-package, 2

HIMA (HIMA-package), 2
hima, 2, 5
HIMA-package, 2

ncvreg, 3

simHIMA, 4