Package ‘SMUT’

September 24, 2019

Type Package
Title Multi-SNP Mediation Intersection-Union Test
Version 1.1
Author Wujuan Zhong
Maintainer Wujuan Zhong <zhongwujuan@gmail.com>
Description Testing the mediation effect of multiple SNPs on an outcome through a mediator.
LazyData true
License GPL (>= 2)
Depends R (>= 2.10)
Imports Rcpp (>= 0.12.14), SKAT, MASS
LinkingTo Rcpp, RcppEigen
NeedsCompilation yes
Repository CRAN
Date/Publication 2019-09-24 04:40:08 UTC

R topics documented:

eigenMapMatMult .................................................. 2
Generalized_Testing_coefficient_of_mediator .................. 2
Genotype_data ................................................. 3
GSMUT .......................................................... 4
SMUT .......................................................... 6
Testing_coefficient_of_mediator ................................. 8

Index 10
Description
Matrix multiplication using RcppEigen.

Usage
\[
eigenMapMatMult(A, B)
\]

Arguments
- **A, B**: numeric (double) complex matrices or vectors.

Value
The matrix product. The value is the same as \( A \times B \).

Examples
```
library(SMUT)
A=matrix(1:9,3,3)
A=A+0
B=as.matrix(c(5.0, 2.0, 0.0))
eigenMapMatMult(A,B)
```

# Thanks for using our R package SMUT

---

Generalized_Testing_coefficient_of_mediator

**Description**
Testing coefficient of mediator, namely theta, in the outcome model. The outcome model is the following.
\[
\text{outcome} \sim \text{intercept} + G \times \gamma + \text{mediator} \times \theta + \text{error}
\]

**Usage**
\[
\text{Generalized\_Testing\_coefficient\_of\_mediator}(G, \text{mediator}, \text{outcome}, \text{covariates}=\text{NULL}, \text{outcome\_type}, \text{approxi}=\text{TRUE}, \text{verbose}=\text{FALSE})
\]
**Arguments**

- **G**
  - n by p matrix (n rows and p columns). Each row is one individual; each column is one SNP.
- **mediator**
  - a vector length of n. It is the mediator variable.
- **outcome**
  - a vector length of n. It is the outcome variable.
- **covariates**
  - n by r matrix (n rows and r columns). Each row is one individual; each column is one covariate.
- **outcome_type**
  - Type of the outcome variable. "continuous" for a continuous outcome; "binary" for a binary outcome; "count" for a count outcome; "survival" for a survival outcome.
- **approxi**
  - a boolean value. This is an indicator whether the approximation of computing derivatives is applied to save computing time. Default is TRUE.
- **verbose**
  - a boolean value. If TRUE a lot of computing details is printed. Default is FALSE.

**Value**

- **p_value**
  - P value for testing the coefficient of mediator in the outcome model.
- **theta_hat**
  - The point estimate of theta (coefficient of mediator) in the outcome model.

**Author(s)**

- Wujuan Zhong

---

**Genotype_data**

*Example genotype data for SMUT*

---

**Description**

Example genotype data for SMUT. It is a matrix with 100 rows and 200 columns. Each row is an individual; each column is a SNP.

**Format**

It is a matrix with 100 rows and 200 columns. Each row is an individual; each column is a SNP.
Generalized Multi-SNP Mediation Intersection-Union Test

Description
Testing the mediation effect of multiple SNPs on an outcome following an exponential family distribution or a survival outcome through a continuous mediator.

Usage
GSMUT(G, mediator, outcome, covariates=NULL, outcome_type, approxi=TRUE, verbose=FALSE)

Arguments
- **G**: n by p matrix (n rows and p columns). Each row is one individual; each column is one SNP.
- **mediator**: a vector length of n. It is the mediator variable.
- **outcome**: a vector length of n. It is the outcome variable.
- **covariates**: n by r matrix (n rows and r columns). Each row is one individual; each column is one covariate.
- **outcome_type**: Type of the outcome variable. "continuous" for a continuous outcome; "binary" for a binary outcome; "count" for a count outcome; "survival" for a survival outcome.
- **approxi**: a boolean value. This is an indicator whether the approximation of computing derivatives is applied to save computing time. Default is TRUE.
- **verbose**: a boolean value. If TRUE a lot of computing details is printed. Default is FALSE.

Value
- **p_value_IUT**: The p value for testing the mediation effect (theta*beta) based on intersection-union test.
- **p_value_theta**: The p value for testing theta in the outcome model. The outcome model is the following.
  outcome ~ intercept + covariates*iota + G*gamma + mediator*theta
- **theta_hat**: The point estimate of theta (coefficient of mediator) in the outcome model.
- **p_value_beta**: The p value for testing beta in the mediator model. The mediator model is the following.
  mediator ~ intercept + covariates*iota + G*beta + error

Author(s)
Wujuan Zhong
Examples

library(SMUT)
# load the Genotype data included in this R package
data("Genotype_data")

##### for a binary outcome #####
set.seed(1)

# generate two covariates
covariate_1=rnorm(nrow(Genotype_data),0,1)
covariate_2=sample(c(0,1),size=nrow(Genotype_data),replace = TRUE)
covariates=cbind(covariate_1,covariate_2)

# generate a mediator
beta=rnorm(ncol(Genotype_data),0,0.5)
tau_M=c(-0.3,0.2)
e1 = rnorm(nrow(Genotype_data), 0, 1)
mediator = 1 + eigenMapMatMult(Genotype_data,beta) +
eigenMapMatMult(covariates, tau_M) + e1

##### generate a binary outcome #####
theta=1
gamma=rnorm(ncol(Genotype_data),0,0.5)
tau=c(-0.2,0.2)
etta=1 + eigenMapMatMult(Genotype_data, gamma) +
eigenMapMatMult(covariates, tau) + theta * mediator
pi=1/(1+exp( -(etta ) ))
outcome=rbinom(length(pi),size=1,prob=pi)
result=GSMUT(G=Genotype_data,mediator=mediator,outcome=outcome,
covariates=covariates,outcome_type="binary")
print(result)
# p_value_IUT is the p value for the mediation effect.

## Not run:

##### generate a count outcome #####
theta=1
gamma=rnorm(ncol(Genotype_data),0,0.5)
tau=c(-0.2,0.2)
etta=1 + eigenMapMatMult(Genotype_data, gamma) +
eigenMapMatMult(covariates, tau) + theta * mediator
mu_param=exp(etta) # the mean parameter
phi_param=10 # the shape parameter
outcome=rnbinom(length(mu_param),size=phi_param,mu=mu_param)
result=GSMUT(G=Genotype_data,mediator=mediator,outcome=outcome,
covariates=covariates,outcome_type="count")
print(result)
# p_value_IUT is the p value for the mediation effect.

##### generate a survival outcome #####
theta=2
\[
\begin{align*}
\text{gamma} &= \text{rnorm(ncol(Genotype\_data)),0,0.5) \\
\text{tau} &= c(-0.2,0.2) \\
\text{eta} &= 1 + \text{eigenMapMatMult(Genotype\_data, gamma)} + \\
&\quad \text{eigenMapMatMult(covariates, tau)} + \theta \times \text{mediator} \\
\text{v} &= \text{runif(nrow(Genotype\_data))} \\
\lambda &= 0.01; \rho = 1; \text{rateC} = 0.001 \\
\text{Tlat} &= (-\log(v) / (\lambda \times \exp(\text{eta})))^{1/\rho} \\
\text{C} &= \text{rexp(nrow(Genotype\_data)}, \text{rate}=\text{rateC}) \\
\text{time} &= \text{pmin(Tlat, C)} \\
\text{status} &= \text{as.numeric(Tlat} <= \text{C}) \\
\text{outcome} &= \text{cbind(time, status)} \\
\text{result} &= \text{GSMUT(G=Genotype\_data, mediator=mediator, outcome=outcome,}
\text{ covariates=covariates, outcome_type="survival")}
\end{align*}
\]

## End(Not run)

---

**SMUT**

**Multi-SNP Mediation Intersection-Union Test**

### Description

Testing the mediation effect of multiple SNPs on an outcome through a mediator.

### Usage

```
SMUT(G, mediator, outcome, 
outcome_type="continuous", method="score", approxi=TRUE, debug=FALSE)
```

### Arguments

- **G** n by p matrix (n rows and p columns). Each row is one individual; each column is one SNP.
- **mediator** a vector length of n. It is the mediator variable.
- **outcome** a vector length of n. It is the outcome variable.
- **outcome_type** Type of the outcome variable. For now, this package only deals with continuous outcome. Default is "continuous".
- **method** The method of testing coefficient of mediator in the outcome model. The score test is used. Default is "score".
- **approxi** a boolean value. This is an indicator whether the approximation of the score statistic is applied to save computing time. Default is TRUE.
debug

A boolean value. If TRUE a lot of computing details is printed; otherwise the function is completely silent. Default is FALSE.

**Value**

- **p_value_IUT** The p value for testing the mediation effect (theta*beta) based on intersection-union test.
- **p_value_theta** The p value for testing theta in the outcome model. The outcome model is the following.
  \[
  \text{outcome} \sim \text{intercept} + G^*\text{gamma} + \text{mediator}*\text{theta} + \text{error}.
  \]
- **p_value_beta** The p value for testing beta in the mediator model. The mediator model is the following.
  \[
  \text{mediator} \sim \text{intercept} + G^*\text{beta} + \text{error}
  \]

**Author(s)**

Wujuan Zhong

**References**


**Examples**

```r
library(SMUT)
# load the Genotype data included in this R package
data("Genotype_data")

# generate one mediator and one outcome

# first example, the mediation effect is significant
set.seed(1)
beta = rnorm(ncol(Genotype_data),1,2)
e1 = rnorm(nrow(Genotype_data), 0, 1)
mediator = 1 + eigenMapMatMult(Genotype_data,beta) + e1
theta=0.8
gamma=rnorm(ncol(Genotype_data),0.5,2)
e2 = rnorm(nrow(Genotype_data), 0, 1)
outcome = 2 + eigenMapMatMult(Genotype_data,gamma) + theta*mediator + e2

p_value=SMUT(G=Genotype_data,mediator=mediator,outcome=outcome)
print(p_value)

# p_value_IUT is the p value for the mediation effect.
# we have significant(at alpha level 0.05) mediation effects (p_value_IUT = 0.001655787).

# second example, the mediation effect is non-significant
set.seed(1)
beta = rnorm(ncol(Genotype_data),1,2)
```
Testing coefficient of mediator, namely theta, in the outcome model. The outcome model is the following:

\[
\text{outcome} \sim \text{intercept} + G \cdot \gamma + \text{mediator} \cdot \theta + \text{error}
\]

Usage

```r
Testing_coefficient_of_mediator(G, mediator, outcome, outcome_type="continuous", method="score", approxi=TRUE, debug=FALSE)
```
Arguments

**G**
- n by p matrix (n rows and p columns). Each row is one individual; each column is one SNP.

**mediator**
- a vector length of n. It is the mediator variable.

**outcome**
- a vector length of n. It is the outcome variable.

**outcome_type**
- Type of the outcome variable. For now, this package only deals with continuous outcome. Default is "continuous".

**method**
- The method of testing coefficient of mediator in the outcome model. The score test is used. Default is "score".

**approxi**
- a boolean value. This is an indicator whether the approximation of the score statistic is applied to save computing time. Default is TRUE.

**debug**
- a boolean value. If TRUE a lot of computing details is printed; otherwise the function is completely silent. Default is FALSE.

Value

P value for testing the coefficient of mediator in the outcome model.

Author(s)

Wujuan Zhong

Examples

```r
library(SMUT)
# load the Genotype data included in this R package
data("Genotype_data")

# generate one mediator and one outcome
set.seed(1)
beta=rnorm(ncol(Genotype_data),1,2)
e1 = rnorm(nrow(Genotype_data), 0, 1)
mediator = 1 + eigenMapMatMult(Genotype_data,beta) + e1
ttheta=0.8
gamma=rnorm(ncol(Genotype_data),0.5,2)
e2 = rnorm(nrow(Genotype_data), 0, 1)
ooutcome = 2 + eigenMapMatMult(Genotype_data,gamma) + theta*mediator + e2

p_value=Testing_coefficient_of_mediator(G=Genotype_data,mediator=mediator,outcome=outcome)
print(p_value)

# Thanks for using our R package SMUT
```
Index

eigenMapMatMult, 2

Generalized Testing_coefficient_of_mediator, 2

Genotype_data, 3

GSMUT, 4

SMUT, 6

Testing_coefficient_of_mediator, 8