Package 'MRBEEX'

October 9, 2024

Type Package

Title Mendelian Randomization using Bias-Correction Estimating Equations with Extended Functions

Version 0.1.0

Author Yihe Yang

Maintainer Yihe Yang <yxy1234@case.edu>

Description MRBEEX extends the functionality of the MRBEE package by incorporating advanced methods for Mendelian Randomization (MR) analysis. It introduces the MRBEEX function, which uses bias-corrected estimating equations to mitigate weak instrument bias due to estimation errors in GWAS effect estimates for exposures and outcomes. For addressing horizontal pleiotropy, the package employs the IPOD algorithm to identify uncorrelated horizontal pleiotropy (UHP) and a two-mixture regression model for correlated horizontal pleiotropy (CHP). Additionally, it integrates SuSiE for exposure selection, enhancing interpretability (UseSuSiE=T). The package also includes MRBEEX.UV for univariable MR analysis, offering similar methods for managing UHP and CHP. Both functions support the inclusion of correlated instrumental variables using an LD matrix and provide advanced options for exposure selection and horizontal pleiotropy correction. Furthermore, it offers CisMRBEEX for performing multivariable cis-Mendelian randomization.

License MIT LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.3.1

Imports MASS,
Rcpp (>= 1.0.0),
RcppArmadillo,
FDRestimation,
mvtnorm,
mixtools,
utils,
data.table,
varbvs,
susieR,
CppMatrix,
Matrix

Remotes harryyiheyang/CppMatrix,
stephenslab/susieR

ORCID 0000-0001-6563-3579

R topics documented:

`CisMRBEEX` *Cis Multivariable Mendelian Randomization using Bias-corrected Estimating Equations*

Description

This function performs multivariable cis-Mendelian randomization that removes weak instrument bias using Bias-corrected Estimating Equations and identifies uncorrelated horizontal pleiotropy (UHP). Additionally, it integrates SuSiE for exposure selection, enhancing interpretability (use.susie=T).

Usage

```
CisMRBEEX(
  by,
  bX,
  byse,
  bXse,
  LD,
  Rxy,
  cluster.index = c(1:length(by)),
  block.rho = 0,
  use.susie = T,
  estimate_residual_variance = T,
  residual_variance = 1,
  reliability.thres = 0.9,
  Lvec = c(1:5),
  pip.thres = 0.2,
  xQTL.max.L = 10,
  xQTL.sampling = 500,
  xQTL.pip.thres = 0.2,
  xQTL.Nvec,
  post.selection = T,
  tauvec = seq(3, 30, by = 3),
  rho = 2,
  ridge = 0.05,
  max.iter = 100,
```

```
  max.eps = 0.001,
  susie.iter = 100,
  ebic.theta = 1,
  ebic.gamma = 2,
  maxdiff = 3,
  theta.ini = F,
  gamma.ini = F
)
```

Arguments

by	A vector of effect estimates from the outcome GWAS.
bX	A matrix of effect estimates from the exposure GWAS.
byse	A vector of standard errors of effect estimates from the outcome GWAS.
bXse	A matrix of standard errors of effect estimates from the exposure GWAS.
LD	The linkage disequilibrium (LD) matrix.
Rxy	The correlation matrix of estimation errors of exposures and outcome GWAS. The last column corresponds to the outcome.
cluster.index	A vector indicating the LD block indices each IV belongs to. The length is equal to the number of IVs, and values are the LD block indices.
block.rho	A parameter used in AR1-like blockwise thresholding of residual covariance matrix. Defaults to 0.
use.susie	An indicator of whether using SuSiE to select causal exposures. Defaults to T.
estimate_residual_variance	An indicator of whether estimating the residual_variance in SuSiE. Defaults to T.
residual_variance	When estimate_residual_variance = T, the initial value of the residual variance. When estimate_residual_variance = F, the fixed value of the residual variance. Defaults to 1.
reliability.thres	A threshold for the minimum value of the reliability ratio. If the original reliability ratio is less than this threshold, only part of the estimation error is removed so that the working reliability ratio equals this threshold.
Lvec	When SuSiE is used, the candidate vector for the number of single effects. Default is c(1:min(10, nrow(bX))).
pip.thres	A threshold of minimum posterior inclusion probability. Default is 0.2.
xQTL.max.L	The maximum number of L in estimating the xQTL effects. Defaults to 10.
xQTL.sampling	The number of subsampling times in estimating the standard errors of xQTL effects. Defaults to 100.
xQTL.pip.thres	The minimum empirical posterior inclustion probability (PIP) used in subsampling procedure.
xQTL.Nvec	The vector of sample sizes of exposures.
post.selection	An indicator of whether using OLS to refit xQTL effect. Defaults to T.
tauvec	When choosing "IPOD", the candidate vector of tuning parameters for the MCP penalty function. Default is seq(3, 30, by=3).
rho	When choosing "IPOD", the tuning parameter in the nested ADMM algorithm. Default is 2.

ridge	A ridge.parameter on causal effect estimate. Defaults to 0.05.
max.iter	Maximum number of iterations for causal effect estimation. Defaults to 100.
max.eps	Tolerance for stopping criteria. Defaults to 0.001.
susie.iter	Number of iterations in SuSiE per iteration. Default is 100.
ebic.theta	EBIC factor on causal effect. Default is 1.
ebic.gamma	EBIC factor on horizontal pleiotropy Default is 2.
maxdiff	The maximum difference between the MRBEE causal estimate and the initial estimator. Defaults to 1.5.
theta.ini	Initial value of theta. If FALSE, the default method is used to estimate. Default is FALSE.
gamma.ini	Initial value of gamma. Default is FALSE.

Value

A list that contains the results of the MRBEEX with respect to different methods applied:

theta Causal effect estimate.

theta.se Standard error of the causal effect estimate.

theta.cov Covariance matrix of the causal effect estimate.

theta.pip Empirical posterior inclusion probability (PIP) of the causal effect in the subsampling procedure.

theta.pratt Pratt index estimate of exposure.

gamma Estimate of horizontal pleiotropy.

gamma.pratt Pratt index estimate of horizontal pleiotropy.

Bic A vector or matrix recording the Bayesian Information Criterion (BIC) values.

theta.ini Initial value of theta used in the estimation procedure.

gamma.ini Initial value of gamma used in the estimation procedure.

reliability.adjust Estimated reliability-adjusted values.

thetalist List of theta estimates recorded during each iteration in the subsampling procedure.

gammalist List of gamma estimates recorded during each iteration in the subsampling procedure.

Examples

```
## Not run:
# Example usage of MRBEEX function
result <- CisMRBEEX(by, bX, byse, bXse)

## End(Not run)
```

clump_cluster *Clustering second data frame based on closest SNP centers from first data frame*

Description

This function performs clustering of SNPs in a second data.frame based on the closest SNP centers defined in a first data.frame. Both data.frames should include SNP, BP, and CHR columns. This function scales CHR and BP to ensure distinctiveness across chromosomes and employs Euclidean distance to find the nearest cluster centers from the first data.frame for each SNP in the second data.frame.

Usage

```
clump_cluster(df1, df2)
```

Arguments

df1	A data.frame representing the output of a plink clump with parameters r2=0.01. It contains columns for SNP, BP (base pair position), CHR (chromosome), and P (p-value).
df2	A data.frame similar to df1, representing a plink output with a less stringent r2 value, typically r2=0.5, including columns for SNP, BP, CHR, and P.

Details

The function first standardizes the CHR and BP columns by multiplying CHR by 10000 and dividing BP by 1e6. This standardization helps to manage the scale differences between chromosome numbers and base pair positions. After standardization, it calculates the Euclidean distances between each SNP in df2 to all SNP centers in df1, assigns each SNP in df2 to the nearest center from df1, and adds a new column 'cluster' to df2 to reflect this assignment.

Value

A modified version of df2 where each SNP is annotated with a 'cluster' index corresponding to the closest SNP center from df1 based on scaled CHR and BP values.

Examples

```
df1 <- data.frame(SNP=c("rs1", "rs2"), CHR=c(1, 1), BP=c(150000, 250000), P=c(0.001, 0.002))
df2 <- data.frame(SNP=c("rs1", "rs3", "rs2", "rs4"), CHR=c(1,1,1,1),
                  BP=c(150000,160000,250000,260000),
                  P=c(0.001,0.003,0.002, 0.004))
clustered_df2 <- clump_cluster(df1, df2)
```

`cluster_snps` *Clustering SNPs based on p-value and proximity with a PLINK C+T file.*

Description

This function clusters SNPs within a given window size based on their P-value and proximity. It iterates through each chromosome, finds the SNP with the smallest P-value, and groups all SNPs within the specified window size around this SNP into a cluster.

Usage

```
cluster_snps(df, window_size = 1e+06)
```

Arguments

`df`	A data.frame containing SNP data with columns for SNP (SNP ID), CHR (chromosome), BP (base pair position), and P (p-value).
`window_size`	An integer specifying the window size around each SNP (in base pairs) within which other SNPs are considered for clustering. Default to 1e6.

Details

The function processes each chromosome independently. It orders the SNPs by their base pair positions, identifies the SNP with the smallest P-value, and clusters all SNPs within the specified window size around this SNP. The process is repeated until all SNPs are assigned to a cluster.

Value

A data.frame containing the clustered SNPs with an additional column 'ClusterSize' indicating the number of SNPs in each cluster.

Examples

```
df <- data.frame(SNP=c("rs1", "rs2", "rs3", "rs4", "rs5"),
                 CHR=c(1, 1, 1, 1, 2),
                 BP=c(100000, 150000, 200000, 250000, 300000),
                 P=c(0.01, 0.02, 0.03, 0.04, 0.05))
window_size <- 50000
clustered_snps <- cluster_snps(df, window_size)
```

errorCov *Estimate Error Covariance Matrix Using GWAS Insignificant Effects*

Description

This function estimates the error covariance matrix by subsampling a proportion of insignificant GWAS effects and calculating their correlation coefficients.

Usage

```
errorCov(
  ZMatrix,
  Zscore.cutoff = 2,
  subsampling.ratio = 0.1,
  subsampling.time = 1000
)
```

Arguments

ZMatrix	A matrix of Z-scores for exposure and outcome, with the outcome GWAS in the last column.
Zscore.cutoff	The cutoff for significance. Defaults to 2.
subsampling.ratio	The proportion of effects to subsample for each iteration. Defaults to 0.1.
subsampling.time	The number of subsampling iterations. Defaults to 1000.

Value

A matrix representing the estimated error covariance.

filter_align *Filter and Align GWAS Data to a Reference Panel*

Description

The filter_align function processes a list of GWAS summary statistics data frames, harmonizes alleles according to a reference panel, removes duplicates, and aligns data to common SNPs. It's used to prepare data for further analysis such as LDSC.

Usage

```
filter_align(gwas_data_list, ref_panel, allele_match = T)
```

Arguments

gwas_data_list	A list of data.frames where each data.frame contains GWAS summary statistics for a trait. Each data.frame should include columns for SNP identifiers, Z-scores of effect size estimates, sample sizes (N), effect allele (A1), and reference allele (A2).
ref_panel	A data.frame containing the reference panel data. It must include columns for SNP, A1, and A2.
allele_match	An indicator of whether matching the effect alleles of GWAS files to the reference panel.

Details

The function performs several key steps: adjusting alleles according to a reference panel, removing duplicate SNPs, and aligning all GWAS data frames to a set of common SNPs. This is often a necessary preprocessing step before performing genetic correlation and heritability analyses.

Value

A list of data.frames, each corresponding to an input GWAS summary statistics data frame, but filtered, harmonized, and aligned to the common SNPs found across all data frames.

GWPT *Genome-Wide Pleiotropy Test*

Description

This function performs a genome-wide pleiotropy test (GWPT) after Mendelian randomization. It offers an option for a two-mixture model, where the residual is chosen as the smaller one resulting from the two causal effect estimates from two mixtures.

Usage

```
GWPT(by, byse, bX, bXse, Rxy, theta, theta.cov)
```

Arguments

by	A vector of effect estimates from the outcome GWAS.
byse	A vector of standard errors of effect estimates from the outcome GWAS.
bX	A matrix of effect estimates from the exposure GWAS.
bXse	A matrix of standard errors of effect estimates from the exposure GWAS.
Rxy	The correlation matrix of estimation errors of exposures and outcome GWAS. The last column corresponds to the outcome.
theta	The causal effect estimate.
theta.cov	The covariance matrix of the causal effect estimate.

Value

A list with two components:

BETA	The estimated residual values.
SE	The standard errors of the residual estimates.

GWPT_Mixture *Genome-Wide Pleiotropy Test for mixture model*

Description

This function performs a genome-wide pleiotropy test (GWPT) after Mendelian randomization. It offers an option for a two-mixture model, where the residual is chosen as the smaller one resulting from the two causal effect estimates from two mixtures.

Usage

```
GWPT_Mixture(
  by,
  byse,
  bX,
  bXse,
  Rxy,
  theta1,
  theta.cov1,
  theta2,
  theta.cov2,
  LD.blcok
)
```

Arguments

by	A vector of effect estimates from the outcome GWAS.
byse	A vector of standard errors of effect estimates from the outcome GWAS.
bX	A matrix of effect estimates from the exposure GWAS.
bXse	A matrix of standard errors of effect estimates from the exposure GWAS.
Rxy	The correlation matrix of estimation errors of exposures and outcome GWAS. The last column corresponds to the outcome.
theta1	The causal effect estimate of the first mixture.
theta.cov1	The covariance matrix of the causal effect estimate of the first mixture.
theta2	The causal effect estimate of the second mixture.
theta.cov2	The covariance matrix of the causal effect estimate of the second mixture.
LD.blcok	A vector of indices of LD blocks.

Value

A list with two components:

BETA	The estimated residual values.
SE	The standard errors of the residual estimates.

MRBEEX *Multivariable Mendelian Randomization using Bias-corrected Estimating Equations*

Description

This function removes weak instrument bias using Bias-corrected Estimating Equations and identifies uncorrelated horizontal pleiotropy (UHP) and correlated horizontal pleiotropy (CHP) through two distinct methods. UHP is detected using the IPOD algorithm, where outliers are interpreted as UHP. For CHP, a two-mixture regression model (Mixture) is implemented by the mixtools R package. Additionally, it integrates SuSiE for exposure selection, enhancing interpretability (use.susie=T). Both the IPOD algorithm and the Mixture method support the inclusion of correlated instrumental variables using an LD matrix and provide advanced options for exposure selection and horizontal pleiotropy correction.

Usage

```
MRBEEX(
  by,
  bX,
  byse,
  bXse,
  LD = "identity",
  Rxy,
  cluster.index = c(1:length(by)),
  reliability.thres = 0.8,
  Method = c("IPOD", "Mixture"),
  use.susie = T,
  tauvec = seq(3, 30, by = 3),
  rho = 2,
  main.cluster.thres = 0.48,
  min.cluster.size = 10,
  robust.se = T,
  Lvec = c(1:min(10, nrow(bX))),
  pip.thres = 0.3,
  max.iter = 100,
  max.eps = 0.001,
  susie.iter = 100,
  ebic.theta = 1,
  ebic.gamma = 2,
  maxdiff = 3,
  sampling.time = 100,
  sampling.iter = 10,
  theta.ini = F,
  gamma.ini = F
)
```

Arguments

by	A vector of effect estimates from the outcome GWAS.
bX	A matrix of effect estimates from the exposure GWAS.

byse A vector of standard errors of effect estimates from the outcome GWAS.

bXse A matrix of standard errors of effect estimates from the exposure GWAS.

LD The linkage disequilibrium (LD) matrix. Default is the identity matrix, assuming independent instrumental variables (IVs).

Rxy The correlation matrix of estimation errors of exposures and outcome GWAS. The last column corresponds to the outcome.

cluster.index A vector indicating the LD block indices each IV belongs to. The length is equal to the number of IVs, and values are the LD block indices.

reliability.thres A threshold for the minimum value of the reliability ratio. If the original reliability ratio is less than this threshold, only part of the estimation error is removed so that the working reliability ratio equals this threshold.

Method Method for handling horizontal pleiotropy. Options are "IPOD" and "Mixture".

use.susie An indicator of whether using SuSiE to select causal exposures. Defaults to T.

tauvec When choosing "IPOD", the candidate vector of tuning parameters for the MCP penalty function. Default is seq(3, 30, by=3).

rho When choosing "IPOD", the tuning parameter in the nested ADMM algorithm. Default is 2.

main.cluster.thres When choosing "Mixture", a threshold for weights belonging to the first category. To prevent instability caused by small-effect IVs falling into both categories, we slightly lower the voting threshold for the first category to below 0.5, ensuring it remains dominant. Default is 0.48.

min.cluster.size When choosing "Mixture", a threshold for the minimum number of IVs in the second cluster. If our initial check reveals that the number is below this threshold, the IPOD algorithm will be applied. Default is 10.

robust.se When choosing "Mixture", an indicator of whether the robust covariance estimate is applied to calculate the empirical covariance matrix of causal effect estimates from subsampling results. Default is T.

Lvec When SuSiE is used, the candidate vector for the number of single effects. Default is c(1:min(10, nrow(bX))).

pip.thres Posterior inclusion probability (PIP) threshold. Individual PIPs less than this value will be shrunk to zero. Default is 0.3.

max.iter Maximum number of iterations for causal effect estimation. Defaults to 100.

max.eps Tolerance for stopping criteria. Defaults to 0.001.

susie.iter Number of iterations in SuSiE per iteration. Default is 100.

ebic.theta EBIC factor on causal effect. Default is 1.

ebic.gamma EBIC factor on horizontal pleiotropy Default is 2.

maxdiff The maximum difference between the MRBEE causal estimate and the initial estimator. Defaults to 1.5.

sampling.time Number of blockwise bootstrapping times. Default is 100.

sampling.iter Number of iterations per blockwise bootstrapping procedure. Default is 5.

theta.ini Initial value of theta. If FALSE, the default method is used to estimate. Default is FALSE.

gamma.ini Initial value of gamma. Default is FALSE.

Value

A list that contains the results of the MRBEEX with respect to different methods applied:

`theta` Causal effect estimate.

`theta.se` Standard error of the causal effect estimate.

`theta.cov` Covariance matrix of the causal effect estimate.

`theta.pip` Empirical posterior inclusion probability (PIP) of the causal effect in the subsampling procedure.

`theta.pratt` Pratt index estimate of exposure.

`gamma` Estimate of horizontal pleiotropy.

`gamma.pratt` Pratt index estimate of horizontal pleiotropy.

`Bic` A vector or matrix recording the Bayesian Information Criterion (BIC) values.

`theta.ini` Initial value of `theta` used in the estimation procedure.

`gamma.ini` Initial value of `gamma` used in the estimation procedure.

`reliability.adjust` Estimated reliability-adjusted values.

`thetalist` List of `theta` estimates recorded during each iteration in the subsampling procedure.

`gammalist` List of `gamma` estimates recorded during each iteration in the subsampling procedure.

`theta1` Causal effect estimate for the first mixture component (when `Method="Mixture"`).

`theta2` Causal effect estimate for the second mixture component (when `Method="Mixture"`).

`theta.se1` Standard error of `theta1`.

`theta.se2` Standard error of `theta2`.

`theta.cov1` Covariance matrix of `theta1`.

`theta.cov2` Covariance matrix of `theta2`.

`theta.pratt1` Pratt index estimates of exposures in the first mixture.

`theta.pratt2` Pratt index estimates of exposures in the second mixture.

`theta.pip1` Empirical PIP of `theta1` in the subsampling procedure.

`theta.pip2` Empirical PIP of `theta2` in the subsampling procedure.

`thetalist1` List of `theta1` estimates recorded during each iteration in the subsampling procedure.

`thetalist2` List of `theta2` estimates recorded during each iteration in the subsampling procedure.

`cluster1` Indices of individual IVs in the first mixture component.

`cluster2` Indices of individual IVs in the second mixture component.

`fit.mixture` Return of fit.mixture in the lastest iteration.

`IsMixture` If this component is provided, our initial check found that the number was below a threshold of minimum cluster size of the second cluster, the IPOD algorithm was applied.

Examples

```
## Not run:
# Example usage of MRBEEX function
result <- MRBEEX(by, bX, byse, bXse)

## End(Not run)
```

MRBEEX_UV *Univariable Mendelian Randomization using Bias-corrected Estimating Equations*

Description

This function removes weak instrument bias using Bias-corrected Estimating Equations and identifies uncorrelated horizontal pleiotropy (UHP) and correlated horizontal pleiotropy (CHP) through two distinct methods. UHP is detected using the IPOD algorithm, where outliers are interpreted as UHP. For CHP, a two-mixture regression model is applied. Both the IPOD algorithm and the Mixture method support the inclusion of correlated instrumental variables using an LD matrix and provide advanced options for exposure selection and horizontal pleiotropy correction.

Usage

```
MRBEEX_UV(
  by,
  bX,
  byse,
  bXse,
  LD = "identity",
  Rxy,
  cluster.index = c(1:length(by)),
  reliability.thres = 0.8,
  Method = c("IPOD", "Mixture"),
  ebic.theta = 0,
  tauvec = seq(3, 30, by = 3),
  rho = 2,
  ebic.gamma = 2,
  main.cluster.thres = 0.48,
  min.cluster.size = 10,
  robust.se = T,
  max.iter = 100,
  max.eps = 0.001,
  maxdiff = 3,
  sampling.time = 100,
  sampling.iter = 10,
  theta.ini = F,
  gamma.ini = F
)
```

Arguments

by	A vector of effect estimates from the outcome GWAS.
bX	A vector of effect estimates from the exposure GWAS.
byse	A vector of standard errors of effect estimates from the outcome GWAS.
bXse	A vector of standard errors of effect estimates from the exposure GWAS.
LD	The linkage disequilibrium (LD) matrix. Default is the identity matrix, assuming independent instrumental variables (IVs).

Rxy	The correlation matrix of estimation errors of exposures and outcome GWAS. The last element corresponds to the outcome.
cluster.index	A vector indicating the LD block indices each IV belongs to. The length is equal to the number of IVs, and values are the LD block indices.
reliability.thres	A threshold for the minimum value of the reliability ratio. If the original reliability ratio is less than this threshold, only part of the estimation error is removed so that the working reliability ratio equals this threshold.
Method	Method for handling horizontal pleiotropy. Options are "IPOD" and "Mixture".
ebic.theta	EBIC factor on causal effect. Default is 0.
tauvec	When choosing "IPOD", the candidate vector of tuning parameters for the MCP penalty function. Default is seq(3, 30, by=3).
rho	When choosing "IPOD", the tuning parameter in the nested ADMM algorithm. Default is 2.
ebic.gamma	EBIC factor on horizontal pleiotropy Default is 2.
main.cluster.thres	When choosing "Mixture", a threshold for weights belonging to the first category. To prevent instability caused by small-effect IVs falling into both categories, we slightly lower the voting threshold for the first category to below 0.5, ensuring it remains dominant. Default is 0.4.
min.cluster.size	When choosing "Mixture", a threshold for the minimum number of IVs in the second cluster. If our initial check reveals that the number is below this threshold, the IPOD algorithm will be applied. Default is 10.
robust.se	When choosing "Mixture", an indicator of whether the robust covariance estimate is applied to calculate the empirical covariance matrix of causal effect estimates from subsampling results. Default is T.
max.iter	Maximum number of iterations for causal effect estimation. Defaults to 100.
max.eps	Tolerance for stopping criteria. Defaults to 0.001.
maxdiff	The maximum difference between the MRBEE causal estimate and the initial estimator. Defaults to 1.5.
sampling.time	Number of resampling times. Default is 100.
sampling.iter	Number of iterations per resampling. Default is 5.
theta.ini	Initial value of theta. If FALSE, the default method is used to estimate. Default is FALSE.
gamma.ini	Initial value of gamma. Default is FALSE.

Value

A list containing the results of the MRBEEX.UV analysis:

theta Causal effect estimate.

theta.se Standard error of the causal effect estimate.

theta.cov Covariance matrix of the causal effect estimate.

theta.pratt Pratt index estimate of exposure.

gamma Estimate of horizontal pleiotropy.

gamma.pratt Pratt index estimate of horizontal pleiotropy.

`Bic` A vector or matrix recording the Bayesian Information Criterion (BIC) values.

`theta.ini` Initial value of `theta` used in the estimation procedure.

`gamma.ini` Initial value of `gamma` used in the estimation procedure.

`reliability.adjust` Estimated reliability-adjusted values.

`thetalist` List of `theta` estimates recorded during each iteration in the subsampling procedure.

`gammalist` List of `gamma` estimates recorded during each iteration in the subsampling procedure.

`theta1` Causal effect estimate for the first mixture component (when `Method="Mixture"`).

`theta2` Causal effect estimate for the second mixture component (when `Method="Mixture"`).

`theta.se1` Standard error of `theta1`.

`theta.se2` Standard error of `theta2`.

`theta.cov1` Covariance matrix of `theta1`.

`theta.cov2` Covariance matrix of `theta2`.

`theta.pratt1` Pratt index estimates of exposures in the first mixture.

`theta.pratt2` Pratt index estimates of exposures in the second mixture.

`thetalist1` List of `theta1` estimates recorded during each iteration in the subsampling procedure.

`thetalist2` List of `theta2` estimates recorded during each iteration in the subsampling procedure.

`cluster1` Indices of individual IVs in the first mixture component.

`cluster2` Indices of individual IVs in the second mixture component.

`fit.mixture` Return of fit.mixture in the lastest iteration.

Examples

```
## Not run:
# Example usage of MRBEEX.UV function
result <- MRBEEX_UV(by, bX, byse, bXse)

## End(Not run)
```

`MRBEE_IMRP` *Mendelian randomization with bias-correction estimating equation: detecting horizontal pleiotropy via hypothesis test.*

Description

This function estimates the causal effect using a bias-correction estimating equation, considering potential pleiotropy and measurement errors.

Usage

```
MRBEE_IMRP(
  by,
  bX,
  byse,
  bXse,
  Rxy,
  max.iter = 30,
```

```
  max.eps = 1e-04,
  pv.thres = 0.05,
  var.est = "variance",
  FDR = T,
  adjust.method = "Sidak",
  maxdiff = 1.5
)
```

Arguments

by	A vector (n x 1) of the GWAS effect size of outcome.
bX	A matrix (n x p) of the GWAS effect sizes of p exposures.
byse	A vector (n x 1) of the GWAS effect size SE of outcome.
bXse	A matrix (n x p) of the GWAS effect size SEs of p exposures.
Rxy	A matrix (p+1 x p+1) of the correlation matrix of the p exposures and outcome. The last one should be the outcome.
max.iter	Maximum number of iterations for causal effect estimation. Defaults to 30.
max.eps	Tolerance for stopping criteria. Defaults to 1e-4.
pv.thres	P-value threshold in pleiotropy detection. Defaults to 0.05.
var.est	Method for estimating the variance of residual in pleiotropy test. Can be "robust", "variance", or "ordinal". Defaults is "variance" that estimates the variance of residual using median absolute deviation (MAD).
FDR	Logical. Whether to apply the FDR to convert the p-value to q-value. Defaults to TRUE.
adjust.method	Method for estimating q-value. Defaults to "Sidak".
maxdiff	The maximum difference between the MRBEE causal estimate and the initial estimator. Defaults to 1.5.

Value

A list containing the estimated causal effect, its covariance, and pleiotropy

---

summary_generation *Generating simulated data for Mendelian randomization simulation*

Description

This function generates simulated data for Mendelian Randomization (MR) analysis, considering genetic effects, estimation errors, and horizontal pleiotropy. It allows for different distributions of genetic effects and pleiotropy, and accommodates both independent and correlated instrumental variables (IVs).

Usage

```
summary_generation(
  theta,
  m,
  Rbb,
  Ruv,
  Rnn,
  Nxy,
  Hxy,
  LD = "identity",
  non.zero.frac,
  UHP.frac = 0,
  CHP.frac = 0,
  UHP.var = 0.5,
  UHP.dis = "uniform",
  CHP.effect = c(1, rep(0, length(theta) - 1)),
  effect.dis = "normal"
)
```

Arguments

theta	An (px1) vector of causal effects.
m	The number of instrumental variables (IVs).
Rbb	An (pxp) correlation matrix of genetic effects.
Ruv	An ((p+1)x(p+1)) correlation matrix of residuals in outcome and exposures; the outcome is the first one.
Rnn	An ((p+1)x(p+1)) correlation matrix of sample overlap; the outcome is the first one.
Nxy	An ((p+1)x1) vector of GWAS sample sizes; the outcome is the first one.
Hxy	An ((p+1)x1) vector of heritabilities; the outcome is the first one.
LD	An (mxm) correlation matrix of the IVs or "identity" indicating independent IVs.
non.zero.frac	An (px1) vector with all entries in (0,1]; each entry is the probability of deltaj such that betaj=betaj'*deltaj.
UHP.frac	A number indicating the fraction of IVs affected by UHP.
CHP.frac	A number indicating the fraction of IVs affected by CHP.
UHP.var	A number indicating the variance attributed to UHP.
UHP.dis	Distribution of pleiotropy effects: "normal" (default), "uniform", "t" distribution (with degree of freedom 5).
CHP.effect	A vector of effects corresponding to the variables correlated with the correlated horizontal pleiotropy.
effect.dis	Distribution of genetic effects: "normal" (default), "uniform", "t" distribution (with degree of freedom 5).

Value

A list containing simulated GWAS effect sizes for exposures (bX), their standard errors (bXse), the GWAS effect size for the outcome (by), its standard error (byse), the pleiotropy effects (pleiotropy), and the true effects.

Index