Package 'MRBEEX'

September 30, 2025

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 7.3.2

Imports MASS,
FDRestimation,
mvtnorm,
mixtools,
utils,
data.table,
varbvs,
susieR,
CppMatrix,
Matrix,
MRBEE

Suggests CARMA

Remotes harryyiheyang/CppMatrix,
noahlorinczcomi/MRBEE,
stephenslab/susieR

ORCID 0000-0001-6563-3579

Contents

`block_cutoff` *Cluster pruning by Local Pratt Index (retain up to 5 per eligible cluster)*

Description

Given `cluster_info` produced by `clump_cluster` and two row-aligned effect-size vectors `marginal_effect` and `direct_effect`, this function prunes rows within clusters whose size exceeds `cutoff` by selecting, per such cluster, at most the top 5 variants ranked by the local Pratt index accumulating their normalized contributions until reaching `contribution`; if the threshold is not reached, the top 5 are kept. Clusters with size *less than or equal to* `cutoff` are fully retained.

Usage

```
block_cutoff(
  cluster_info,
  marginal_effect,
  direct_effect,
  cutoff = 5,
  contribution = 0.9
)
```

Arguments

`cluster_info`	A data frame from `clump_cluster`, containing at least a numeric column `cluster` giving the cluster index for each row; must be row-aligned with the effect vectors.
`marginal_effect`	Numeric vector of GWAS marginal effect sizes, row-aligned with `cluster_info`. Must be standardized to a comparable scale (e.g., z-scores) before calling this function.

direct_effect	Numeric vector of direct effect sizes (from fine-mapping or PRS), row-aligned with cluster_info. Must be standardized to a comparable scale before calling. If estimated by SBayesRC, a common standardization is $\beta * \sqrt{2 * f * (1 - f)}$, where f is the allele frequency.
cutoff	Integer. Maximum cluster size threshold for pruning: clusters with size <= cutoff are entirely kept; clusters with size > cutoff are pruned using the Local Pratt rule.
contribution	Numeric in $(0, 1]$. Target cumulative contribution (based on normalized Local Pratt weights) used to decide how many top variants to retain per pruned cluster; no more than 5 will be kept.

Value

A list with:

- cluster_info_updated: the pruned subset of cluster_info
- selected_global_idx: global row indices retained
- dropped_global_idx: global row indices removed

---

build_blockdiag_ld *Build block-diagonal LD matrix by SNP clustering with VIF-based pruning (no shrinkage)*

Description

Build block-diagonal LD matrix by SNP clustering with VIF-based pruning (no shrinkage)

Usage

```
build_blockdiag_ld(
  SNP_DF,
  window_size = 1500000,
  R,
  kappa_thr = 30,
  vif_thr = 5
)
```

Arguments

SNP_DF	data.frame with columns: SNP, CHR, BP, P.
window_size	numeric, window (bp) for defining independent SNP centers (default 1.5e6).
R	numeric LD/correlation matrix; row/colnames must be SNP IDs; square.
kappa_thr	numeric, condition-number threshold to trigger VIF pruning (default 30).
vif_thr	numeric, VIF threshold on diag(solve(R_block)) to mark variants high-VIF (default 5).

Value

list with: - R1: block-diagonal matrix (no shrinkage) after pruning - cluster.index: data.frame with SNP, CHR, BP, cluster (after pruning) - removed_snp: character vector of removed SNP IDs

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.

Usage

```
CisMRBEEX(
  by,
  bX,
  byse,
  bXse,
  LD,
  Rxy,
  model.infinitesimal = F,
  reliability.thres = 0.75,
  Lvec = c(1:5),
  causal.pip.thres = 0.2,
  xQTL.method = "SuSiE",
  xQTL.selection.rule = "top_K",
  top_K = 1,
  xQTL.pip.min = 0.2,
  xQTL.max.L = 10,
  xQTL.cred.thres = 0.95,
  xQTL.pip.thres = 0.5,
  xQTL.Nvec,
  tauvec = seq(3, 30, by = 3),
  xQTL.weight = NULL,
  outlier.switch = T,
  Annotation = NULL,
  output.labels = NULL,
  carma.iter = 5,
  carma.inner.iter = 5,
  xQTL.max.num = 10,
  carma.epsilon.threshold = 0.001,
  admm.rho = 2,
  ridge.diff = 1000,
  max.iter = 100,
  max.eps = 0.001,
  susie.iter = 500,
  coverage.xQTL = 0.95,
  coverage.causal = 0.95,
  ebic.theta = 0,
  ebic.gamma = 1,
  theta.ini = F,
  gamma.ini = F,
```

```
  xQTLfitList = NULL,
  verbose = T
)
```

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.
model.infinitesimal	An indicator of whether using REML to model infinitesimal effects. Defaults to F.
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))).
causal.pip.thres	A threshold of minimum posterior inclusion probability. Default is 0.2.
xQTL.method	The method used in purifying the xQTLs. SuSiE or CARMA can be used here, where the latter can be more accurate but much most computationally costly. Defaults is SuSiE.
xQTL.selection.rule	The method for purifying informative xQTLs within each credible set. Options include "minimum_pip", which selects all variables with PIPs exceeding a specified threshold, and "top_K", which ensures at least K variables are selected based on their PIP ranking. Defaults to "top_K".
top_K	The maximum number of variables selected in each credible sets. Defaults to 1.
xQTL.pip.min	The minimum empirical PIP used in purifying variables in each credible set. Defaults to 0.2.
xQTL.max.L	When choosing "SuSiE", the maximum number of L in estimating the xQTL effects. Defaults to 10.
xQTL.cred.thres	When choosing "SuSiE", the minimum empirical posterior inclusion probability (PIP) used in getting credible sets of xQTL selection. Defaults to 0.95.
xQTL.pip.thres	The threshold of individual PIP when selecting xQTL. Defaults to 0.5.
xQTL.Nvec	When choosing "SuSiE", the vector of sample sizes of exposures.
tauvec	The candidate vector of tuning parameters for the MCP penalty function. Default is seq(3, 30, by=3).
xQTL.weight	When choosing "SuSiE", the vector of weights used in specifying the prior weights of SuSiE. Defaults to NULL.
outlier.switch	When choosing "CARMA", an indicator of whether turning on outlier detection. Defaults to F.

Annotation	When choosing "CARMA", the annotation matrix of SNP. Default is NULL.
output.labels	When choosing "CARMA", output directory where output will be written while CARMA is running. Defaults to NULL, meaning that a temporary folder will be created and automatically deleted upon completion of the computation.
carma.iter	When choosing "CARMA", the maximum iterations for EM algorithm to run. Defaults to 5.
carma.inner.iter	When choosing "CARMA", the maximum iterations for Shotgun algorithm to run per iteration within EM algorithm. Defaults to 5.
xQTL.max.num	When choosing "CARMA", the maximum number of causal variants assumed per locus, which is similar to the number of single effects in SuSiE. Defaults to 10.
carma.epsilon.threshold	When choosing "CARMA", the convergence threshold measured by average of Bayes factors. Defaults to 1e-3.
admm.rho	The tuning parameter in the nested ADMM algorithm. Default is 2.
ridge.diff	A ridge.parameter on the differences of causal effect estimate in one credible set. Defaults to 10.
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 500.
coverage.xQTL	The coverage of defining a credible set in xQTL selection. Defaults to 0.95.
coverage.causal	The coverage of defining a credible set in cis-MRBEE. Defaults to 0.95.
ebic.theta	EBIC factor on causal effect. Default is 1.
ebic.gamma	EBIC factor on horizontal pleiotropy Default is 2.
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.
xQTLfitList	Initial fits of xQTLs for exposures. This should be a list. Each component corresponds to the susie.fit of each exposure when xQTL.method = "SuSiE". When xQTL.method = "CARMA", this should be the list of results from a CARMA analysis. Users can customize additional SuSiE or CARMA parameters to improve performance. Default is NULL.
verbose	A logical indicator of whether to display the execution time of the method. Default is T.

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.

`susie.theta` The fit of causal effect resulted from susie.

`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.

var.error The variance of residuals.

var.error The variance of infinitesimal effect.

estimated.reliability.ratio The estimated reliability ratios of exposures.

xQTLfitList The results of sparse predictions of exposures, yielded by SuSiE or CARMA.

`CisMRBEE_UV` *CisMRBEE_UV: Cis Univariable Mendelian Randomization using Bias-corrected Estimating Equations*

Description

This function performs univariable cis-Mendelian randomization that removes weak instrument bias using Bias-corrected Estimating Equations and identifies uncorrelated horizontal pleiotropy (UHP).

Usage

```
CisMRBEE_UV(
  by,
  bX,
  byse,
  bXse,
  LD,
  Rxy,
  xQTL.N,
  xQTL.selection.rule = "top_K",
  top_K = 1,
  xQTL.pip.min = 0.2,
  xQTL.max.L = 10,
  xQTL.cred.thres = 0.95,
  xQTL.pip.thres = 0.5,
  reliability.thres = 0.75,
  tauvec = seq(3, 30, by = 1.5),
  admm.rho = 2,
  coverage.xQTL = 0.95,
  coverage.causal = 0.95,
  max.iter = 100,
  max.eps = 0.001,
  ebic.gamma = 2,
  xQTLfit = NULL
)
```

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 LD matrix of variants.
Rxy	The correlation matrix of estimation errors of exposures and outcome GWAS. The last column corresponds to the outcome.
xQTL.N	The sample sizes of exposure.
xQTL.selection.rule	The method for purifying informative xQTLs within each credible set. Options include "minimum_pip", which selects all variables with PIPs exceeding a specified threshold, and "top_K", which ensures at least K variables are selected based on their PIP ranking. Defaults to "top_K".
top_K	The maximum number of variables selected in each credible sets. Defaults to 1.
xQTL.pip.min	The minimum empirical PIP used in purifying variables in each credible set. Defaults to 0.2.
xQTL.max.L	The maximum number of L in estimating the xQTL effects. Defaults to 10.
xQTL.cred.thres	The minimum empirical posterior inclusion probability (PIP) used in getting credible sets of xQTL selection. Defaults to 0.95.
xQTL.pip.thres	If SuSiE fails to find any credible set, the threshold of individual PIP when selecting xQTL. Defaults to 0.5.
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.
tauvec	A vector of tuning parameters used in penalizing the direct causal effect. Default is 'seq(3,10,by=1)'.
admm.rho	A parameter set in the ADMM algorithm. Default is 2.
coverage.xQTL	The coverage of defining a credible set in xQTL selection. Defaults to 0.95.
coverage.causal	The coverage of defining a credible set in cis-MRBEE. Defaults to 0.95.
max.iter	The maximum number of iterations for the ADMM algorithm. Default is 15.
max.eps	The convergence tolerance for the ADMM algorithm. Default is 0.005.
ebic.gamma	The extended BIC factor for model selection. Default is 2.
xQTLfit	Initial fits of xQTLs for exposures. This should only be yielded by SuSiE, as CARMA is not allowed for cis-UVMR analysis currently. Default is NULL.

Value

A list containing:

theta The estimated effect size of the tissue-gene pair.

gamma The estimated effect sizes of the direct causal variants.

theta.cov The variance of the estimated effect size 'theta'.

`theta.se` The standard error of the estimated effect size 'theta'.

`theta.z` The z-score of the estimated effect size 'theta'.

`Bic` The BIC values for each tuning parameter.

`eQTL.fit` The SuSiE result of xQLT selection of exposure.

`var.error` The variance of residuals.

`var.error` The variance of infinitesimal effect.

`causal.fit` The SuSiE result of causal effect calibration of exposure.

`reliability.adjust` Estimated reliability-adjusted values.

`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.block
)
```

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.block	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)),
  method = c("IPOD", "Mixture"),
  use.susie = T,
  group.penalize = F,
  group.index = c(1:ncol(bX)[1]),
  group.diff = 100,
  main.cluster.thres = 0.48,
  min.cluster.size = 5,
  tauvec = seq(2.5, 40, by = 2.5),
  admm.rho = 2,
  Lvec = c(1:min(10, ncol(bX))),
  pip.thres = 0.5,
  pip.min = 0.1,
  cred.pip.thres = 0.95,
  max.iter = 100,
  max.eps = 0.001,
  susie.iter = 100,
  ebic.theta = 0,
  ebic.gamma = 1,
  ridge.diff = 1000,
  sampling.time = 100,
  sampling.iter = 10,
  maxdiff = 3,
  reliability.thres = 0.75,
  coverage.causal = 0.95,
  theta.ini = F,
  gamma.ini = F,
```

```
  verbose = T,
  gcov = NULL,
  ldsc = NULL
)
```

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.
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.
group.penalize	An indicator of whether using SuSiE to penalize highly correlated exposures. Defaults to F.
group.index	A vector of the group index of exposure. Defaults to NULL.
group.diff	The tuning penalizing difference of highly correlated exposure prediction. Defaults to 100.
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 minimum sample size of the second mixture. Default is 5.
tauvec	When choosing "IPOD", the candidate vector of tuning parameters for the MCP penalty function. Default is seq(3, 30, by=3).
admm.rho	When choosing "IPOD", the tuning parameter in the nested ADMM algorithm. Default is 2.
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 0.5.
pip.min	The minimum empirical PIP used in purifying variables in each credible set. Defaults to 0.1.
cred.pip.thres	The threshold of PIP of each credible set. Defaults to 0.95.
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 `0`.

`ebic.gamma` EBIC factor on horizontal pleiotropy. Default is `1`.

`ridge.diff` A ridge.parameter on the differences of causal effect estimate in one credible set. Defaults to `1e3`.

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

`sampling.iter` Number of iterations per blockwise bootstrapping procedure. Default is `10`.

`maxdiff` The maximum difference between the MRBEE causal estimate and the initial estimator. Defaults to `3`.

`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.

`coverage.causal` The coverage of defining a credible set in MRBEEX when `use.susie = T`. Defaults to `0.95`.

`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`.

`verbose` A logical indicator of whether to display the execution time of the method. Default is `T`.

`gcov` A matrix (p+1 x p+1) of the per-snp genetic covariance matrix of the p exposures and outcome. The last one should be the outcome.

`ldsc` A vector (n x 1) of the LDSCs of the IVs.

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.

`Voting` A list that contains (1) the weights of two mixtures and (2) the voting results of two mixture based on main.cluster.thres.

`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 = "IPOD",
  ebic.theta = 0,
  tauvec = seq(5, 30, by = 2),
  rho = 2,
  ebic.gamma = 2,
  max.iter = 100,
  max.eps = 0.001,
  maxdiff = 3,
  sampling.time = 1000,
  sampling.iter = 30,
  theta.ini = F,
  gamma.ini = F,
  ldsc = NULL,
  gcov = NULL
)
```

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.
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.
ldsc	A vector (n x 1) of the LDSCs of the IVs.
gcov	A matrix (p+1 x p+1) of the per-snp genetic covariance matrix of the p exposures and outcome. The last one should be the outcome.

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.bootstrap Resampled causal effect estimates in bootstrap.

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.

`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,
  group.penalize = F,
  group.index = NULL,
  group.diff = 1000
)
```

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.
group.penalize	An indicator of whether using SuSiE to penalize highly correlated exposures. Defaults to F.
group.index	A vector of the group index of exposure. Defaults to c(1:ncol(bX)).
group.diff	The tuning penalizing difference of highly correlated exposure prediction. Defaults to 10.

Value

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

---

MRBEE_TL *Estimate Non-Transferable Causal Effect with MRBEE and SuSiE*

Description

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

Usage

```
MRBEE_TL(
  by,
  bX,
  byse,
  bXse,
  Rxy,
  LD = "identity",
```

```
  cluster.index = c(1:length(by)),
  group.penalize = F,
  group.index = NULL,
  group.diff = 100,
  theta.source,
  theta.source.cov,
  tauvec = seq(3, 30, 3),
  Lvec = c(1:6),
  admm.rho = 3,
  ebic.delta = 0,
  ebic.gamma = 1,
  transfer.coef = 1,
  susie.iter = 200,
  pip.thres = 0.5,
  pip.min = 0.1,
  cred.pip.thres = 0.95,
  max.iter = 50,
  coverage.causal = 0.95,
  max.eps = 1e-04,
  reliability.thres = 0.8,
  ridge.diff = 100,
  sampling.time = 100,
  sampling.iter = 10,
  ldsc = NULL,
  gcov = NULL
)
```

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 first one should be the transferred linear predictor and last one should be the outcome.
LD	The linkage disequilibrium (LD) matrix. Default is the identity matrix, assuming independent instrumental variables (IVs).
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.
group.penalize	An indicator of whether using SuSiE to penalize highly correlated exposures. Defaults to F.
group.index	A vector of the group index of exposure. Defaults to NULL.
group.diff	The tuning penalizing difference of highly correlated exposure prediction. Defaults to 10.
theta.source	A vector (p x 1) of the causal effect estimate learning from the source data.
theta.source.cov	A matrix (p x p) of the covariance matrix of the causal effect estimate learning from the source data.

tauvec	The candidate vector of tuning parameters for the MCP penalty function. Default is seq(3, 30, by=3).
Lvec	A vector of the number of single effects used in SuSiE. Default is c(1:6).
admm.rho	When choosing "IPOD", the tuning parameter in the nested ADMM algorithm. Default is 2.
ebic.delta	A scale of tuning parameter of causal effect estimate in extended BIC. Default is 0.
ebic.gamma	A scale of tuning parameter of horizontal pleiotropy in extended BIC. Default is 1.
transfer.coef	A scale of transfer.coef of theta.source to theta.target. Default is 1.
susie.iter	A scale of the maximum number of iterations used in SuSiE. Default is 200.
pip.thres	Posterior inclusion probability (PIP) threshold. Individual PIPs less than this value will be shrunk to zero. Default is 0.5.
pip.min	The minimum empirical PIP used in purifying variables in each credible set. Defaults to 0.1.
cred.pip.thres	The threshold of PIP of each credible set. Defaults to 0.95.
max.iter	Maximum number of iterations for causal effect estimation. Default is 50.
coverage.causal	The coverage of defining a credible set in MRBEEX when use.susie = T. Defaults to 0.95.
max.eps	Tolerance for stopping criteria. Default is 1e-4.
reliability.thres	A scale of 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. Default is 0.8.
ridge.diff	A scale of parameter on the differences of causal effect estimate in one credible set. Defaults to 10.
sampling.time	A scale of number of subsampling in estimating the standard error. Default is 100.
sampling.iter	A scale of iteration in subsampling in estimating the standard error. Default is 10.
ldsc	A vector (n x 1) of the LDSCs of the IVs.
gcov	A matrix (p+1 x p+1) of the per-snp genetic covariance matrix of the p exposures and outcome. The last one should be the outcome.

Value

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

MRBEE_TL_UV *Estimate Non-Transferable Causal Effect with MRBEE and SuSiE in UVMR*

Description

This function estimates the non-transferable causal effect using a bias-correction estimating equation in a univariable MR model, considering potential pleiotropy and measurement errors, and using SuSiE to select the non-transferable causal effect.

Usage

```
MRBEE_TL_UV(
  by,
  bX,
  byse,
  bXse,
  Rxy,
  LD = "identity",
  cluster.index = c(1:length(by)),
  theta.source,
  theta.source.cov,
  tauvec = seq(3, 30, 3),
  admm.rho = 3,
  ebic.delta = 1,
  ebic.gamma = 2,
  transfer.coef = 1,
  susie.iter = 200,
  pip.thres = 0.3,
  max.iter = 50,
  max.eps = 1e-04,
  reliability.thres = 0.8,
  sampling.time = 100,
  sampling.iter = 10,
  ldsc = NULL,
  gcov = NULL
)
```

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 first one should be the transferred linear predictor and last one should be the outcome.
LD	The linkage disequilibrium (LD) matrix. Default is the identity matrix, assuming independent instrumental variables (IVs).

`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.

`theta.source` A vector (p x 1) of the causal effect estimate learning from the source data.

`theta.source.cov`
A matrix (p x p) of the covariance matrix of the causal effect estimate learning from the source data.

`tauvec` The candidate vector of tuning parameters for the MCP penalty function. Default is `seq(3, 30, by=3)`.

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

`ebic.delta` A scale of tuning parameter of causal effect estimate in extended BIC. Default is `1`.

`ebic.gamma` A scale of tuning parameter of horizontal pleiotropy in extended BIC. Default is `2`.

`transfer.coef` A scale of transfer.coef of theta.source to theta.target. Default is `1`.

`susie.iter` A scale of the maximum number of iterations used in SuSiE. Default is `200`.

`pip.thres` A scale of PIP theshold for calibyating causality used in SuSiE. Default is `0.3`.

`max.iter` Maximum number of iterations for causal effect estimation. Default is `50`.

`max.eps` Tolerance for stopping criteria. Default is `1e-4`.

`reliability.thres`
A scale of 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. Default is `0.8`.

`sampling.time` A scale of number of subsampling in estimating the standard error. Default is `100`.

`sampling.iter` A scale of iteration in subsampling in estimating the standard error. Default is `10`.

`ldsc` A vector (n x 1) of the LDSCs of the IVs.

`gcov` A matrix (2 x 2) of the per-snp genetic covariance matrix of the p exposures and outcome. The last one should be the outcome.

Value

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

`Sparse_Prediction` *Sparse Prediction of xQTL Effects using SuSiE and CARMA*

Description

This function performs infomative xQTL selection and sparse prediction using SuSiE and CARMA.

Usage

```
Sparse_Prediction(
  bX,
  bXse,
  LD,
  xQTL.Nvec,
  xQTL.method = "SuSiE",
  xQTL.max.L = 10,
  xQTL.weight = NULL,
  xQTL.cred.thres = 0.95,
  outlier.switch = T,
  Annotation = NULL,
  output.labels = NULL,
  carma.iter = 5,
  carma.inner.iter = 5,
  xQTL.max.num = 10,
  carma.epsilon.threshold = 0.001
)
```

Arguments

bX	A matrix of effect estimates from the exposure GWAS.
bXse	A matrix of standard errors of effect estimates from the exposure GWAS.
LD	The linkage disequilibrium (LD) matrix.
xQTL.Nvec	When choosing "SuSiE", the vector of sample sizes of exposures.
xQTL.method	The method used in purifying the xQTLs. SuSiE or CARMA can be used here, where the latter can be more accurate but much most computationally costly. Defaults is SuSiE.
xQTL.max.L	When choosing "SuSiE", the maximum number of L in estimating the xQTL effects. Defaults to 10.
xQTL.weight	When choosing "SuSiE", the vector of weights used in specifying the prior weights of SuSiE. Defaults to NULL.
xQTL.cred.thres	When choosing "SuSiE", the minimum empirical posterior inclusion probability (PIP) used in getting credible sets of xQTL selection. Defaults to 0.95.
outlier.switch	When choosing "CARMA", an indicator of whether turning on outlier detection. Defaults to F.
Annotation	When choosing "CARMA", the annotation matrix of SNP. Default is NULL.
output.labels	When choosing "CARMA", output directory where output will be written while CARMA is running. Defaults to NULL, meaning that a temporary folder will be created and automatically deleted upon completion of the computation.
carma.iter	When choosing "CARMA", the maximum iterations for EM algorithm to run. Defaults to 5.
carma.inner.iter	When choosing "CARMA", the maximum iterations for Shotgun algorithm to run per iteration within EM algorithm. Defaults to 5.
xQTL.max.num	When choosing "CARMA", the maximum number of causal variants assumed per locus, which is similar to the number of single effects in SuSiE. Defaults to 10.

carma.epsilon.threshold
: When choosing "CARMA", the convergence threshold measured by average of Bayes factors. Defaults to 1e-3.

Value

A list containing the results of the MRBEEX analysis using different methods:

bXest A matrix where each column represents R * beta_PLS for each exposure.

bXestse A matrix where each column contains the standard errors of R * beta_PLS for each exposure.

bXest0 A matrix where each column represents beta_PLS for each exposure.

bXestse0 A matrix where each column contains the standard errors of beta_PLS for each exposure.

xQTLfitList A list containing the xQTL selection results, which can be used in CisMRBEEX by setting xQTLfitList = xQTLfitList.

---

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",
  cluster.index
)
```

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 last one.
Rnn	An ((p+1)x(p+1)) correlation matrix of sample overlap; the outcome is the last one.
Nxy	An ((p+1)x1) vector of GWAS sample sizes; the sample size of the outcome is the last one.
Hxy	An ((p+1)x1) vector of heritabilities; the outcome is the last 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).
cluster.index	The indices of LD block.

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