Package: MRMix (via r-universe)

May 28, 2026

Type Package

Title Mendelian Randomization Analysis Using Mixture Models (MRMix)

Version 0.1.0

Author Guanghao Qi and Nilanjan Chatterjee

Maintainer Guanghao Qi <gqi1@jhu.edu>

Description This package gives robust estimation of causal effects by conducting Mendelian randomization analysis using a mixture model approach.

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 6.1.0

Repository https://remlapmot.r-universe.dev

Date/Publication 2020-01-27 23:41:06 UTC

RemoteUrl https://github.com/gqi/MRMix

RemoteRef HEAD

RemoteSha 56afdb2bc96760842405396f5d3f02e60e305039

Contents

MRMix *Two-sample Mendelian randomization analysis using mixture models*

Description

This function conducts Mendelian randomization analysis using an underlying mixture model incorporating a fraction of the genetic instruments to have direct effect on the outcome (horizontal pleiotropy). MRMix takes GWAS summary statistics as inputs to estimate causal effects of one trait on another. For stability of the method, we recommend using summary statistics in the standardized scale: 1) For both binary and continuous traits, summary-statistics should be standardized by genotypic variance; 2) In addition, for continuous phenotype, summary-statistics should be standardized by phenotypic variance. If the data are not in the standardized scale, users may use the standardize function to standardize their data. See Details and Examples for more information.

Usage

```
MRMix(betahat_x, betahat_y, sx, sy, theta_temp_vec = seq(-1, 1, by =
  0.01), pi_init = 0.6, sigma_init = 1e-05, profile = FALSE)
```

Arguments

`betahat_x`	GWAS effect estimates of the exposure, recommended to be in standardized scale. Vector of length `K`, where `K` is the number of instruments (SNPs).
`betahat_y`	GWAS effect estimates of the outcome, recommended to be in standardized scale. Vector of length `K`.
`sx`	Standard error of `betahat_x`, recommended to be in standardized scale. Vector of length `K`.
`sy`	Standard error of `betahat_y`, recommended to be in standardized scale. Vector of length `K`.
`theta_temp_vec`	A vector of the grid search values for the causal effect `theta`. Default to be `seq(-1,1,by=0.01)`. Users may adjust the grid if larger effects are possible.
`pi_init`	Initial value of the probability mass at the null component of the mixture model corresponding to underlying valid instruments. Default to be 0.6. See Details.
`sigma_init`	Initial value of the variance of the non-null component of the mixture model which corresponds to underlying invalid instruments with pleiotropic effect. Default to be 1e-5. See Details.
`profile`	Whether to include the profile matrix. Default to be `FALSE`. If `TRUE`, include the profile matrix in the output. See Value `profile_mat` for details.

Details

The algorithm searches over a grid of possible values of the causal effect `theta`. For each fixed `theta`, it fits mixture model `pi0*N(0,sy^2+theta^2*sx^2)+(1-pi0)*N(0,sigma2)` on the residual `betahat_y-theta*betahat_x`. It then chooses the value of `theta` that leads to the maximum `pi0` as the estimate of causal effect. Summary statistics can be standardized using the `standardize()` function if they are estimates from linear or logistic regression. Do not use `Standardize()` for other models.

Value

A list that contains

theta	Estimate of causal effect. Assuming the summary statistics are standardized, theta represents increase in mean value of Y in s.d. unit of Y (for continuous outcomes) or log-OR of Y (for binary outcomes) associated with per s.d. unit increase in values of X (for continuous exposures) or values of X changing from 0 to 1 (for binary exposures).
pi0	The probability mass of the null component corresponding to the estimated theta.
sigma2	The variance of the non-null component corresponding to the estimated theta.
SE_theta	Standard error of causal effect estimate.
zstat_theta	Z-statistic for test of the causal effect estimate.
pvalue_theta	P-value from the z test for the causal effect.
profile_mat	A matrix of 3 columns containing details of the grid search. The first column is theta_temp_vec. The second and third columns are the corresponding pi0 and sigma2 values. Only returned if profile=TRUE.

References

1. Qi, Guanghao, and Nilanjan Chatterjee. "Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects." Nature Communications 10.1 (2019): 1941.

2. Qi, Guanghao, and Nilanjan Chatterjee. "A Comprehensive Evaluation of Methods for Mendelian Randomization Using Realistic Simulations of Genome-wide Association Studies." bioRxiv (2019): 702787.

Examples

```
data("sumstats", package = "MRMix")
# Convert summary statistics to standardized scale
data_std = standardize(sumstats$betahat_x, sumstats$betahat_y, sumstats$sx, sumstats$sy, xtype = "continuous", yt
# MRMix analysis
est = MRMix(data_std$betahat_x_std, data_std$betahat_y_std, data_std$sx_std, data_std$sy_std)
str(est) # True causal effect is 0.2.
# Include profile matrix
est = MRMix(data_std$betahat_x_std, data_std$betahat_y_std, data_std$sx_std, data_std$sy_std, profile = TRUE)
str(est)
```

MRMix_se *Standard error of the MRMix estimator*

Description

This function calculates the standard error of the MRMix estimator using asymptotic theory.

Usage

```
MRMix_se(betahat_x, betahat_y, sx, sy, theta, pi0, sigma2)
```

Arguments

betahat_x	GWAS effect estimates of the exposure, recommended to be in standardized scale. Vector of length K, where K is the number of instruments.
betahat_y	GWAS effect estimates of the outcome, recommended to be in standardized scale. Vector of length K.
sx	Standard error of betahat_x, recommended to be in standardized scale. Vector of length K.
sy	Standard error of betahat_y, recommended to be in standardized scale. Vector of length K.
theta	Estimate of causal effect.
pi0	The probability mass of the null component corresponding to the estimated theta.
sigma2	The variance of the non-null component corresponding to the estimated theta.

Value

The standard error of MRMix estimator.

References

Qi, Guanghao, and Nilanjan Chatterjee. "Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects." Nature Communications 10.1 (2019): 1941.

---

standardize *Standardize summary statistics for MRMix analysis*

Description

1) For both binary and continuous traits, this function standardizes GWAS summary statistics by genotypic variance; 2) In addition, for continuous phenotype, this function standardizes summary statistics by phenotypic variance. This function is designed for GWAS estimates from linear or logistic regression. Do not use for other models.

Usage

```
standardize(betahat_x, betahat_y, sx, sy, xtype, ytype, nx, ny, MAF)
```

Arguments

betahat_x	GWAS effect estimates of the exposure. Vector of length K, where K is the number of instruments (SNPs).
betahat_y	GWAS effect estimates of the outcome. Vector of length K.
sx	Standard error of betahat_x. Vector of length K.
sy	Standard error of betahat_y. Vector of length K.
xtype	Is the exposure a continuous or binary trait? Set to xtype="continuous" or xtype="binary". Or set to xtype="n" if exposure summary statistics do not need to be standardized.
ytype	Is the outcome a continuous or binary trait? Set to ytype="continuous" or ytype="binary". Or set to ytype="n" if outcome summary statistics do not need to be standardized.
nx	SNP-specific sample size (recommended) or total sample size of the study associated with the exposure. Vector of length K or a single number. Set to NULL if trait is binary. Summary statistics for binary traits are standardized by the genotypic variance which can be calculated using the minor allele frequency (MAF) under Hardy-Weinberg equilibrium. Hence sample size is not needed for binary traits.
ny	SNP-specific sample size (recommended) or total sample size of the study associated with the outcome. Vector of length K or a single number. Set to NULL if trait is binary for the same reason as for nx.
MAF	Minor allele frequency. Vector of length K. Set to NULL if both traits are continuous. Summary statistics for continuous traits are standardized as z statistics rescaled by sample size, hence MAF is not needed.

Details

Using the exposure X as an example: 1) For continuous phenotypes analyzed with linear regression, data are standardized by betahat_x_std=betahat_x/(sx*sqrt(nx)); sx_std=1/sqrt(nx). Note that this standardization assumes that GWAS was conducted without covariate adjustment or the covariates do not have strong effects on Y. If the covariates have strong effects on Y, set nx equal to the effective sample size, which can be approximated by N/(1-R2), where N is the sample size associated with the study for X and R2 is the R-squared for the covariates. 2) For binary phenotypes analyzed with logistic regression, data are standardized by betahat_x_std=betahat_x*sqrt(2*MAF*(1-MAF)); sx_std=sx*sqrt(2*MAF*(1-MAF)). Same formulas apply to the outcome Y.

Value

A list that contains

betahat_x_std	Standardized betahat for the exposure.
betahat_y_std	Standardized betahat for the outcome.
sx_std	Standard error of betahat_x_std.
sy_std	Standard error of betahat_y_std.

References

1. Qi, Guanghao, and Nilanjan Chatterjee. "Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects." Nature Communications 10.1 (2019): 1941.

2. Qi, Guanghao, and Nilanjan Chatterjee. "A Comprehensive Evaluation of Methods for Mendelian Randomization Using Realistic Simulations of Genome-wide Association Studies." bioRxiv (2019): 702787.

Index