Package: mr.simss (via r-universe)

September 25, 2024

Type Package

Title MR-SimSS: Mendelian Randomisation (MR) method that combats
Winner's Curse using a simulated sample splitting approach

Version 0.1.0

Description Designed to provide users with a method, namely MR-SimSS,
which uses simulated sample splitting in order to alleviate
Winner's Curse bias in MR causal effect estimates. This
approach also takes into account sample overlap between the
exposure and outcome genome-wide association studies. It uses
summary statistics from genome-wide association studies and
works in combination with existing MR methods, such as IVW and
MR-RAPS.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Roxygen list(markdown=TRUE)

RoxygenNote 7.2.3

URL <https://github.com/amandaforde/mr.simss>

BugReports <https://github.com/amandaforde/mr.simss/issues>

Imports dplyr, doParallel, foreach, magrittr, mr.raps, stats,
parallel, pracma, tibble, TwoSampleMR, abind

Remotes MRCIEU/TwoSampleMR

Suggests knitr, rmarkdown

VignetteBuilder knitr

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

RemoteUrl https://github.com/amandaforde/mr.simss

RemoteRef HEAD

RemoteSha 27d81e0378372cd62ddef45272c4ccdb82c88072

Contents

---

mr.simss-package	*mr.simss: Mendelian Randomisation (MR) method that combats Winner's Curse using a simulated sample splitting approach*

Description

A package designed to provide users with a method, namely **MR-SimSS**, which uses simulated sample splitting in order to alleviate *Winner's Curse* bias in MR causal effect estimates. This approach also takes into account sample overlap between the exposure and outcome genome-wide association studies. It uses summary statistics from genome-wide association studies and works in combination with existing MR methods, such as IVW and MR-RAPS.

Details

Full documentation available here: https://amandaforde.github.io/mr.simss/

---

est_lambda	*Estimating correlation between SNP-exposure and SNP-outcome effect sizes*

Description

est_lambda is a function which allows users to obtain an *unbiased* estimate for *lambda*, a term used to describe the correlation between the SNP-outcome and SNP-exposure effect sizes, using a conditional log-likelihood approach. This correlation is affected by the number of overlapping samples between the two GWASs and the correlation between the exposure and the outcome. Thus, when using the function mr_simss, if the fraction of overlap and the correlation between exposure and outcome are unknown, it is recommended to employ est_lambda and use the value returned from est_lambda in mr_simss. **Note:** For greater accuracy in the estimation of *lambda*, it is advisable to use summary statistics of the entire set of *unpruned* SNPs from the exposure and outcome GWASs.

Usage

```
est_lambda(data, z.threshold = 0.5)
```

Arguments

`data`	A data frame to be inputted by the user containing summary statistics from the exposure and outcome GWASs. It must have at least five columns with column names SNP, `beta.exposure`, `beta.outcome`, `se.exposure`, and `se.outcome`. Each row must correspond to a unique SNP, identified by SNP.
`z.threshold`	A value which is used to obtain a subset of SNPs which have absolute *z*-statistics for both exposure and outcome GWASs less than this value. The method then assumes that both of the true SNP-outcome and SNP-exposure effect sizes of each SNP in this subset are approximately 0. The default setting is `z.threshold=0.5`.

Value

A value which is an estimate of *lambda*, the correlation between the SNP-outcome and SNP-exposure effect sizes, using a conditional log-likelihood approach. Note that this estimate is unbiased but potentially has a high degree of variance.

See Also

<https://amandaforde.github.io/mr.simss/articles/perform-MR-SimSS.html> for illustration of the use of est_lambda with a toy data set and https://amandaforde.github.io/mr.simss/articles/derive-MR-SimSS.html for the theoretical derivation of this method based on a conditional log-likelihood approach for estimating *lambda*.

`mr_simss`	*MR-SimSS main function*

Description

`mr_simss` is the main function for the method, **MR-SimSS**, which is a method based on simulated sample splitting in order to alleviate *Winner's Curse* bias in MR causal effect estimates. It also takes into account sample overlap between the exposure and outcome GWASs. It uses GWAS summary statistics and works in combination with existing MR methods, such as IVW and MR-RAPS.

Usage

```
mr_simss(
  data,
  subset = FALSE,
  sub.cut = 0.05,
  est.lambda = TRUE,
  n.exposure = 1,
  n.outcome = 1,
  n.overlap = 1,
  cor.xy = 0,
  n.iter = 1000,
  splits = 2,
```

```
  pi = 0.5,
  pi2 = 0.5,
  threshold = 5e-08,
  mr_method = "mr_ivw",
  parallel = TRUE,
  n.cores = NULL,
  lambda.thresh = 0.5
)
```

Arguments

data	A data frame to be inputted by the user containing summary statistics from the exposure and outcome GWASs. It must have at least five columns with column names SNP, beta.exposure, beta.outcome, se.exposure and se.outcome. Each row must correspond to a unique SNP, identified by SNP.
subset	A logical which permits the user to perform this method with either the original complete set of SNPs or a subset of SNPs in order to reduce computational time. The default setting is subset=FALSE.
sub.cut	A numerical value required if subset=TRUE, which ensures that for a single iteration of our method, the number of instruments selected if the full set of SNPs is used and the number of instruments if merely the subset is used will be equal with probability at least 1-sub.cut.
est.lambda	A logical which allows the user to specify if they want to use the function, est_lambda, to obtain an estimate for *lambda*, a term used to describe the correlation between the SNP-outcome and SNP-exposure effect sizes. This correlation is affected by the number of overlapping samples between the two GWASs and the correlation between the exposure and the outcome. Thus, it is recommended to use est_lambda if the fraction of overlap and the correlation between exposure and outcome are unknown. The default setting is est.lambda=TRUE.
n.exposure	A numerical value to be specified by the user which is equal to the number of individuals that were in the exposure GWAS. It should be specified by the user if est.lambda=FALSE. The default setting is n.exposure=1.
n.outcome	A numerical value to be specified by the user which is equal to the number of individuals that were in the outcome GWAS. It should be specified by the user if est.lambda=FALSE. The default setting is n.outcome=1.
n.overlap	A numerical value to be specified by the user which is equal to the number of individuals that were in both the exposure and outcome GWAS. It should be specified by the user if est.lambda=FALSE. The default setting is n.overlap=1. The function requires that this value is less than or equal to the minimum of n.exposure and n.outcome.
cor.xy	A numerical value to be specified by the user which is equal to the observed correlation between the exposure and the outcome. This value must be between -1 and 1. It should be specified by the user if est.lambda=FALSE. The default setting is cor.xy=0. If this value is unknown, the user is encouraged to use the function est_lambda.

n.iter	A numerical value which specifies the number of iterations of the method, i.e. the number of times sample splits are randomly simulated. The default setting is n.iter=1000.
splits	A numerical value that must be equal to 2 or 3, indicating whether splits of 2 or 3 should be simulated. It is recommended that in the case of no overlap between the two GWASs that splits of 2 should be used while in the presence of overlap, especially full overlap, splits of 3 should be used. The default setting is splits=2.
pi	A numerical value which determines the fraction of the first split in both the 2 and 3 split approaches. This is the fraction that will be used for SNP selection. The default setting is pi=0.5. This value must be between 0 and 1.
pi2	A numerical value which determines the fraction of the second split in the 3 split approach. The default setting is pi2=0.5. This value must be between 0 and 1.
threshold	A numerical value which specifies the threshold used to select instrument SNPs for MR at each iteration. The default setting is threshold=5e-8. This value must be between 0 and 1.
mr_method	A string which specifies the MR method that MR-SimSS works in combination with. It is possible to use any method outputted in the list TwoSampleMR::mr_method_list()$obj. However, it is currently advised that the user chooses "mr_ivw" or "mr_raps". The default setting is mr_method="mr_ivw".
parallel	A logical value which allows the user to specify if they wish to use this function in parallel or in series. The default setting is parallel=TRUE. It is advisable to use this default, especially when n.iter is large.
n.cores	A numerical value which determines how many cores will be used if parallel=TRUE. This value should be supplied by the user if they wish to use less cores than the output of parallel::detectCores()-1. The default setting is n.cores=NULL.
lambda.thresh	A value which is used when estimating *lambda* to obtain a subset of SNPs which have absolute *z*-statistics for both exposure and outcome GWASs less than this value. The method then assumes that both of the true SNP-outcome and SNP-exposure effect sizes of each SNP in this subset are approximately 0. The default setting is lambda.thresh=0.5.

Value

A list containing two elements, summary and results. summary is a data frame with one row which outputs b, the estimated causal effect of exposure on outcome obtained using the **MR-SimSS** method, as well as se, the associated standard error of this estimate and pval, corresponding *p*-value. It also contains the MR method used, the average number of instrument SNPs used in each iteration and the number of iterations used. results is a data frame which contains the output from each iteration. It is in a similar style as the output from using the function mr from the TwoSampleMR R package.

See Also

<https://amandaforde.github.io/mr.simss/articles/perform-MR-SimSS.html> for illustration of the use of mr_simss with a toy data set and further information regarding this MR method.

---

split2	*Simulating 2 splits with MR-SimSS*

Description

split2 is a function which is used by the main function, mr_simss in order to perform the 2 split approach of the method, **MR-SimSS**.

Usage

```
split2(data, lambda.val = 0, pi = 0.5, mr_method = "mr_ivw", threshold = 5e-08)
```

Arguments

data
: A data frame to be inputted by the user containing summary statistics from the exposure and outcome GWASs. It must have at least five columns with column names SNP, beta.exposure, beta.outcome, se.exposure and se.outcome. Each row must correspond to a unique SNP, identified by SNP.

lambda.val
: A numerical value which is computed within the main function, mr_simss. It is an estimate of *lambda*, a term used to describe the correlation between the SNP-outcome and SNP-exposure effect sizes. The default setting is lambda.val=0.

pi
: A numerical value which determines the fraction of the first split. This is the fraction that will be used for SNP selection. The default setting is pi=0.5.

mr_method
: A string which specifies the MR method that MR-SimSS works in combination with. It is possible to use any method outputted in the list TwoSampleMR::mr_method_list()$obj. However, it is currently advised that the user chooses "mr_ivw" or "mr_raps". The default setting is mr_method="mr_ivw".

threshold
: A numerical value which specifies the threshold used to select instrument SNPs for MR at each iteration. The default setting is threshold=5e-8.

Value

A data frame which contains the output from one iteration. It is in a similar style as the output from using the function mr from the TwoSampleMR R package. The MR method used, the number of instrument SNPs, the causal effect estimate, it associated standard error and *p*-value are all outputted.

---

split3 *Simulating 3 splits with MR-SimSS*

Description

split3 is a function which is used by the main function, mr_simss in order to perform the 3 split approach of the method, **MR-SimSS**.

Usage

```
split3(
  data,
  lambda.val = 0,
  pi = 0.5,
  pi2 = 0.5,
  mr_method = "mr_ivw",
  threshold = 5e-08
)
```

Arguments

data	A data frame to be inputted by the user containing summary statistics from the exposure and outcome GWASs. It must have at least five columns with column names SNP, beta.exposure, beta.outcome, se.exposure and se.outcome. Each row must correspond to a unique SNP, identified by SNP.
lambda.val	A numerical value which is computed within the main function, mr_simss. It is an estimate of *lambda*, a term used to describe the correlation between the SNP-outcome and SNP-exposure effect sizes. The default setting is lambda.val=0.
pi	A numerical value which determines the fraction of the first split. This is the fraction that will be used for SNP selection. The default setting is pi=0.5.
pi2	A numerical value which determines the fraction of the second split. The default setting is pi2=0.5.
mr_method	A string which specifies the MR method that MR-SimSS works in combination with. It is possible to use any method outputted in the list TwoSampleMR::mr_method_list()$obj. However, it is currently advised that the user chooses "mr_ivw" or "mr_raps". The default setting is mr_method="mr_ivw".
threshold	A numerical value which specifies the threshold used to select instrument SNPs for MR at each iteration. The default setting is threshold=5e-8.

Value

A data frame which contains the output from one iteration. It is in a similar style as the output from using the function mr from the TwoSampleMR R package. The MR method used, the number of instrument SNPs, the causal effect estimate, it associated standard error and *p*-value are all outputted.

Index