Package 'bnmr'

December 7, 2023

Type Package

Title A package to conduct Bayesian Network-based Mendelian randomization using individual-level GWAS data

Version 0.2.2

Author Jianle Sun

Maintainer The package maintainer <sjl-2017@sjtu.edu.cn>

Description A package to conduct causal estimation between exposure and outcome across GWAS data using BNMR model, including the variants filtering via random graph forest in the learning stage, as well as the effect size estimation through Bayesian Mendelian randomization with shrinkage prior in the inference stage.

License GPL

Encoding UTF-8

LazyData False

Imports bnlearn,
rstan,
MendelianRandomization,
plyr,
dplyr,
AER,
ivmodel,
parallel,
purrr

RoxygenNote 7.2.3

R topics documented:

bn *Getting suitable genetic IVs through random graph forest, which is based on Bayesian network structure learning*

Description

Getting suitable genetic IVs through random graph forest, which is based on Bayesian network structure learning

Usage

```
bn(
  df,
  snp,
  exposureName,
  bn_method = "hc",
  repeats = 1000,
  selectNum = NA,
  alpha = 0.5,
  nsam = 1000,
  psam = 100,
  sample_replace = TRUE
)
```

Arguments

df	a data frame which contains data of SNPs and specified exposure. The values of snps in the data frame should be either numeric or factors (not integers) for BN learning.
snp	a vector of string belonging to column names of df, which is the name of SNPs included in BN structure learning.
exposureName	a string which is a colname of df corresponding to the exposure studied.
bn_method	method for BN structure learning. Possible values are the function name of structure learning algorithm implemented in bnlearn. Default is "hc".
repeats	an integer standing for the number of subsamples or bootstraps. Default is 1000.
selectNum	the number of instrument to select. Default is NA.
alpha	a number between 0 and 1 to specify the threshold for IV selection. We will use a threshold for variant selection as alpha*psam/length(snp). If selectNum is specified, the parameter will not be used. Default is 0.5.
nsam	the size of individuals in each subsample of random graph forest. Default is 1000.
psam	the size of variants in each subsample of random graph forest. Default is 100.
sample_replace	is a boolean value to determine the sampling methods for individuals. TRUE with replacement and FALSE without replacement. Default is TRUE.

Value

a list containing:

selectsnp	a vector of string containing the colnames of df corresponding to selected SNPs.
dfscore	a data frame containing the score calculated for each SNP.

Examples

```
n <- 2000
p <- 200
snps <- replicate(p,sample(0:2,n,replace = TRUE))
snps <- apply(snps,2,as.numeric)
snpname <- paste0("g",1:p)
df <- as.data.frame(snps)
colnames(df) <- snpname
truesnp <- paste0("g",sample(1:p,50))
df$x <- as.matrix(df[,truesnp])%*%rnorm(50,0.05,0.05)+rnorm(n,0,1)
df$y <- 0.5*df$x+rnorm(n,0,1)

model <- bn(df,snpname,"x")
```

---

bnmr *Causal inference between traits using the Bayesian Network-based Mendelian randomization*

Description

Causal inference between traits using the Bayesian Network-based Mendelian randomization

Usage

```
bnmr(
  df,
  snp,
  exposureName,
  outcomeName,
  bn_method = "hc",
  repeats = 1000,
  selectNum = NA,
  alpha = 0.5,
  nsam = 1000,
  psam = 100,
  sample_replace = TRUE,
  mr_model = "linear",
  prior = "horseshoe",
  init = "median",
  n.iter = 5000,
  n.chain = 4
)
```

Arguments

df	a data frame which contains data of SNPs and specified exposure. The values of snps in the data frame should be either numeric or factors (not integers) for BN learning.
snp	a vector of string belonging to column names of df, which is the name of SNPs included in BN structure learning.

exposureName	a string which is a colname of df corresponding to the exposure studied.
outcomeName	a string which is a column name of df corresponding to the outcome studied.
bn_method	method for BN structure learning. Possible values are the function name of structure learning algorithm implemented in bnlearn. Default is "hc".
repeats	an integer standing for the number of subsamples or bootstraps. Default is 1000.
selectNum	the number of instrument to select. Default is NA.
alpha	a number between 0 and 1 to specify the threshold for IV selection. We will use a threshold for variant selection as alpha*psam/length(snp). If selectNum is specified, the parameter will not be used. Default is 0.5.
nsam	the size of individuals in each subsample of random graph forest. Default is 1000.
psam	the size of variants in each subsample of random graph forest. Default is 100.
sample_replace	is a boolean value to determine the sampling methods for individuals. TRUE with replacement and FALSE without replacement. Default is TRUE.
mr_model	model for MR. Possible values are "linear" or "logit". Default is "linear".
prior	a string represented shrinkage prior used in estimation. It can be "horseshoe", "lasso", "hyperlasso", "spikeslabBernoulli", "spikeslabUniform". Default is "horseshoe".
init	the init value of theta for MCMC estimation. It can be a specific numeric or a string of "TSLS", "median", "egger" and "ivw", which means the initial value of the iteration will be calculated automatically by the above method. Default is "median".
n.iter	an integer standing for the number of iterations. Default is 5000.
n.chain	the number of chains in MCMC sampling. Default is 4.

Value

a list containing:

selectsnp	a vector of string containing the colnames of df corresponding to selected SNPs.
dfscore	a data frame containing the score calculated for each SNP.
betaList	a vector cantaining the posterior of the causal parameter of interest using MCMC sampling.
mean	the mean estimate of the causal parameter.
se	the standard error of the estimation.
lower	the lower boundary of the 95% CI of the causal estimation.
upper	the upper boundary of the 95% CI of the causal estimation.
Rhat	an indicator to measure the convergence (at convergence, Rhat <= 1.1).
fit_detail	an S4 class stanfit object containing the details of Bayesian MR estimation

Examples

```
n <- 2000
p <- 200
snps <- replicate(p,sample(0:2,n,replace = TRUE))
snps <- apply(snps,2,as.numeric)
snpname <- paste0("g",1:p)
```

```
df <- as.data.frame(snps)
colnames(df) <- snpname
truesnp <- paste0("g",sample(1:p,50))
df$x <- as.matrix(df[,truesnp])%*%rnorm(50,0.05,0.05)+rnorm(n,0,1)
df$y <- 0.5*df$x+rnorm(n,0,1)
model <- bnmr(df,snpname,"x","y")
```

`mr` *Causal estimation by Bayesian Mendelian randomization with shrinkage prior to cope with pleiotropy*

Description

Causal estimation by Bayesian Mendelian randomization with shrinkage prior to cope with pleiotropy

Usage

```
mr(
  df,
  selectsnp,
  exposureName,
  outcomeName,
  mr_model = "linear",
  prior = "horseshoe",
  init = "median",
  n.iter = 5000,
  n.chain = 4
)
```

Arguments

`df`	a data frame which contains data of IVs, specified exposure and outcome.
`selectsnp`	a vector of string containing the column names of df corresponding to the IV used in MR.
`exposureName`	a string which is a column name of df corresponding to the exposure studied.
`outcomeName`	a string which is a column name of df corresponding to the outcome studied.
`mr_model`	model for MR. Possible values are "linear" or "logit". Default is "linear".
`prior`	a string represented shrinkage prior used in estimation. It can be "horseshoe", "lasso", "hyperlasso", "spikeslabBernoulli", "spikeslabUniform". Default is "horseshoe".
`init`	the init value of theta for MCMC estimation. It can be a specific numeric or a string of "TSLS", "median", "egger" and "ivw", which means the initial value of the iteration will be calculated automatically by the above method. Default is "median".
`n.iter`	an integer standing for the number of iterations. Default is 5000.
`n.chain`	the number of chains in MCMC sampling. Default is 4.

Value

a list containing:

betaList	a vector cantaining the posterior of the causal parameter of interest using MCMC sampling.
mean	the mean estimate of the causal parameter.
se	the standard error of the estimation.
lower	the lower boundary of the 95% CI of the causal estimation.
upper	the upper boundary of the 95% CI of the causal estimation.
Rhat	an indicator to measure the convergence (at convergence, Rhat <= 1.1).
fit_detail	an S4 class stanfit object containing the details of Bayesian MR estimation

Examples

```
n <- 2000
p <- 200
snps <- replicate(p,sample(0:2,n,replace = TRUE))
snps <- apply(snps,2,as.numeric)
snpname <- paste0("g",1:p)
df <- as.data.frame(snps)
colnames(df) <- snpname
truesnp <- paste0("g",sample(1:p,50))
df$x <- as.matrix(df[,truesnp])%*%rnorm(50,0.05,0.05)+rnorm(n,0,1)
df$y <- 0.5*df$x+rnorm(n,0,1)
model <- mr(df,truesnp,"x","y")
```

---

mr_split *Split the whole dataset into samll subsets and conduct MCMC sampling parallelly and combine the posterior.*

Description

Split the whole dataset into samll subsets and conduct MCMC sampling parallelly and combine the posterior.

Usage

```
mr_split(
  df,
  selectsnp,
  exposureName,
  outcomeName,
  mr_model = "linear",
  prior = "horseshoe",
  init = "median",
  n.iter = 4000,
  n.chain = 4,
  n.split = 4
)
```

Arguments

df	a data frame which contains data of IVs, specified exposure and outcome.
selectsnp	a vector of string containing the column names of df corresponding to the IV used in MR.
exposureName	a string which is a column name of df corresponding to the exposure studied.
outcomeName	a string which is a column name of df corresponding to the outcome studied.
mr_model	model for MR. Possible values are "linear" or "logit". Default is "linear".
prior	a string represented shrinkage prior used in estimation. It can be "horseshoe", "lasso", "hyperlasso", "spikeslabBernoulli", "spikeslabUniform". Default is "horseshoe".
init	the init value of theta for MCMC estimation. It can be a specific numeric or a string of "TSLS", "median", "egger" and "ivw", which means the initial value of the iteration will be calculated automatically by the above method. Default is "median".
n.iter	an integer standing for the number of iterations. Default is 5000.
n.chain	the number of chains in MCMC sampling. Default is 4.
n.split	the number of subsets to be split. Default is 4.

Value

a list containing:

betaList	a vector cantaining the combined posterior of the causal parameter of interest using MCMC sampling.
mean	the mean estimate of the causal parameter.
se	the standard error of the estimation.
lower	the lower boundary of the 95% CI of the causal estimation.
upper	the upper boundary of the 95% CI of the causal estimation.
Rhat	an indicator to measure the convergence (at convergence, Rhat <= 1.1).

Examples

```
n <- 2000
p <- 200
snps <- replicate(p,sample(0:2,n,replace = TRUE))
snps <- apply(snps,2,as.numeric)
snpname <- paste0("g",1:p)
df <- as.data.frame(snps)
colnames(df) <- snpname
truesnp <- paste0("g",sample(1:p,50))
df$x <- as.matrix(df[,truesnp])%*%rnorm(50,0.05,0.05)+rnorm(n,0,1)
df$y <- 0.5*df$x+rnorm(n,0,1)
model <- mr_split(df,truesnp,"x","y",n.split=4)
```

Index