Package: MR2 (via r-universe)

Title Bayesian multi-response Mendelian randomisation

Version 0.1.0

Date 2023-02-05

Maintainer Leonardo Bottolo <lb664@cam.ac.uk>

Description This package performs two-sample summary-level data
Bayesian multi-response Mendelian randomisation with the
selection of important exposures shared or distinct across
responses. Post-processing and selection of exposures based on
an in-house FDR implementation are also provided.

BugReports https://github.com/lb664/MR2/issues

License GPL-2 | file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3.9000

RdMacros Rdpack

Imports BDgraph, graphics, grDevices, MASS, MCMCpack, mvnfast,
mvtnorm, Rdpack, truncnorm, utils

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

RemoteUrl https://github.com/lb664/MR2

RemoteRef HEAD

RemoteSha 568dd7c9504306a455a74a006d27933f42076811

Contents

Index   **[25](#)**

BMM_EM   *Beta density finite mixture model*

Description

Post-processing of MR2 Markov chain Monte Carlo output. Beta density finite mixture model of the marginal posterior probability of inclusion (mPPI) or edge marginal posterior of inclusion (ePPI)

Usage

```
BMM_EM(
  PPI,
  K = 2,
  iter_max = 1000,
  tol = 1e-06,
  plotting = 0,
  printing = FALSE,
  digits = 3,
  cex_list = list(cex = 1, cex.axis = 1.375, cex.lab = 1.375)
)
```

Arguments

PPI	Marginal posterior probability of inclusion (mPPI) or edge marginal posterior of inclusion (ePPI) return by MR2
K	Number of components. Default values set at 2
iter_max	Max number of iterations of the EM algorithm. Default values set at 1000
tol	Stopping criteria tolerance of the EM algorithm. Default values set at 1e-6
plotting	• 0 (default) No plots • 1 Plot after convergence • 2 Plot at each iteration of the EM algorithm
printing	Printing parameters value at each iteration of the EM algorithm. Default values set at FALSE
digits	Integer indicating the number of decimal places
cex_list	List of printing options used list(cex, cex.axis, cex.lab)

Details

For details regarding the EM algorithm to classify posterior probabilities of inclusion, see in Zuber et al. (2023)

Value

The value returned is a list object `list(PPI, w, a, b, p, Llik, aStore, bStore, pStore, LlikStore, stopcode, opt)`

- `PPI` Ordered vector of the posterior probabilities of inclusion
- `w` Matrix of allocation probabilities
- `a` Vector of estimated first parameter of the beta components
- `b` Vector of estimated second parameter of the beta components
- `Llik` Final log-likelihood estimated by the EM algorithm
- `aStore` Stored vectors of the estimated first parameter of the beta components for each iteration of the EM algorithm
- `bStore` Stored vectors of the estimated second parameter of the beta components for each iteration of the EM algorithm
- `pStore` Stored vectors of estimated probability of each beta component for each iteration of the EM algorithm
- `LlikStoreStore` Stored value of the log-likelihood for each iteration of the EM algorithm
- `stopcode` Message that describes how the EM algorithm has terminated
- `opt` List of options used `list(iter_max, tol, plotting, printing)`

References

Zuber V, Lewin A, Levin M, Haglund A, Ben-Aicha Gonzalez S, Emanueli C, Damrauer S, Burgess S, Gill D, Bottolo L (2023). "Multi-response Mendelian randomization: Identification of shared and distinct exposures for multimorbidity and multiple related disease outcomes." *Submitted*. https://www.biorxiv.org/content/10.1101/2023.02.01.526689v1.

Examples

```
# Example 1: Analysis of one replication of simulated Scenario II-Counfounding with q = 5
# responses, p = 15 exposures and nIV = 100 genetic variants used as IVs. The number of
# expected exposures directly associated with each response is set at 2 and its variance at 2,
# with a priori range of direct causal association ranging between 0 and 8

Sim_Counfounding <- Scenario_Counfounding(nRep = 1, q = 5, p = 15, nIV = 100, seed = 31122021)
beta_Y <- Sim_Counfounding$Y
beta_X <- Sim_Counfounding$X

MR2_output <- MR2(beta_Y, beta_X, EVgamma = c(2, 2),
                  niter = 7500, burnin = 2500, thin = 5, monitor = 1000, seed = 31122021)

PostProc_output <- PostProc(MR2_output, beta_Y, beta_X)
```

```
BMM_EM_output <- BMM_EM(PostProc_output$gammaPost)


# Example 2: Analysis of one replication of simulated Scenario IV-Directed pleiotropy with q =
# 5 responses, p = 15 exposures and nIV = 100 genetic variants used as IVs and the effect of
# the shared pleiotropic pathway on the responses set as \code{2}. The number of expected
# exposures directly associated with each response is set at 1 and its variance at 2,
# with a priori range of direct causal associations ranging between 0 and 7. A non-
# decomposable graph for the inverse of the residual correlation between responses is selected

Sim_Pleiotropy <- Scenario_Pleiotropy(nRep = 1, q = 5, p = 15, nIV = 100, undirectedA = FALSE,
                                      seed = 280610971)
beta_Y <- Sim_Pleiotropy$Y
beta_X <- Sim_Pleiotropy$X

MR2_output <- MR2(beta_Y, beta_X, EVgamma = c(1, 3),
                  niter = 15000, burnin = 5000, thin = 10, monitor = 1000,
                  nonDecomp = TRUE, seed = 280610971)

PostProc_output <- PostProc(MR2_output, beta_Y, beta_X)

BMM_EM_output <- BMM_EM(PostProc_output$gammaPost)
```

BVS *Bayesian Variable Selection*

Description

Internal function that performs one Markov chain Monte Carlo iteration to sample from the posterior distribution of all unknowns involved in the Bayesian Variable Selection step. For details, see Alexopoulos and Bottolo (2021) and Zuber et al. (2023)

Usage

```
BVS(
  Y,
  NA_Y,
  X,
  XB,
  gammaCurr,
  BCurr,
  ZCurr,
  RCurr,
  invRCurr,
  d,
  responseType,
  pFixed,
  nCat,
  cutPoint,
```

```
  negBinomPar,
  hyperPar,
  samplerPar
)
```

References

Alexopoulos A, Bottolo L (2021). "Bayesian Variable Selection for Gaussian copula regression models." *J. Comput. Graph. Stat.*, **30**(3), 578–593. doi:10.1080/10618600.2020.1840997.

Zuber V, Lewin A, Levin M, Haglund A, Ben-Aicha Gonzalez S, Emanueli C, Damrauer S, Burgess S, Gill D, Bottolo L (2023). "Multi-response Mendelian randomization: Identification of shared and distinct exposures for multimorbidity and multiple related disease outcomes." *Submitted.* https://www.biorxiv.org/content/10.1101/2023.02.01.526689v1.

FDR_BMM   *False Discovery Rate based on Beta densities finite mixture model*

Description

Post-processing of MR2 Markov chain Monte Carlo output. Estimation of the posterior probability of inclusion cutoff for a given nominal False Discovery Rate level

Usage

```
FDR_BMM(PPI, w, FDR_nominal = 0.05, plotting = FALSE)
```

Arguments

PPI	Ordered marginal posterior probability of inclusion (mPPI) or edge marginal posterior of inclusion (ePPI) return by MBB_EM()
w	Matrix of allocation probabilities
FDR_nominal	Nominal level of False Discovery Rate to be controlled. Default values set at 0.05
plotting	Logical with no plots as default FALSE

Details

For details regarding the False Discovery Rate control based on the Beta densities finite mixture model, see in Zuber et al. (2023)

Value

The value returned is a list object `list(cutoff, cutoff_idx, FDR_est, FNR_est)`

- `cutoff` Posterior probability of inclusion cutoff for a given nominal FDR level
- `cutoff_idx` Index of the ordered posterior probability of inclusion that corresponds to the estimated cutoff
- `FDR_est` Estimated False Discovery Rate
- `FNR_est` Estimated False Negative Rate

References

Zuber V, Lewin A, Levin M, Haglund A, Ben-Aicha Gonzalez S, Emanueli C, Damrauer S, Burgess S, Gill D, Bottolo L (2023). "Multi-response Mendelian randomization: Identification of shared and distinct exposures for multimorbidity and multiple related disease outcomes." *Submitted*. https://www.biorxiv.org/content/10.1101/2023.02.01.526689v1.

Examples

```
# Example 1: Analysis of one replication of simulated Scenario II-Counfounding with q = 5
# responses, p = 15 exposures and nIV = 100 genetic variants used as IVs. The number of
# expected exposures directly associated with each response is set at 2 and its variance at 2,
# with a priori range of direct causal association ranging between 0 and 8

Sim_Counfounding <- Scenario_Counfounding(nRep = 1, q = 5, p = 15, nIV = 100, seed = 31122021)
beta_Y <- Sim_Counfounding$Y
beta_X <- Sim_Counfounding$X

MR2_output <- MR2(beta_Y, beta_X, EVgamma = c(2, 2),
                  niter = 7500, burnin = 2500, thin = 5, monitor = 1000, seed = 31122021)

PostProc_output <- PostProc(MR2_output, beta_Y, beta_X)

BMM_EM_output <- BMM_EM(PostProc_output$gammaPost)
FDR_BMM_output <- FDR_BMM(BMM_EM_output$PPI, BMM_EM_output$w)
PostProc_output$thetaPost* (PostProc_output$gammaPost>FDR_BMM_output$cutoff) * 1


# Example 2: Analysis of one replication of simulated Scenario IV-Directed pleiotropy with q =
# 5 responses, p = 15 exposures and nIV = 100 genetic variants used as IVs and the effect of
# the shared pleiotropic pathway on the responses set as \code{2}. The number of expected
# exposures directly associated with each response is set at 1 and its variance at 2,
# with a priori range of direct causal associations ranging between 0 and 7. A non-decomposable
# graph for the inverse of the residual correlation between responses is selected

Sim_Pleiotropy <- Scenario_Pleiotropy(nRep = 1, q = 5, p = 15, nIV = 100, undirectedA = FALSE,
                                       seed = 280610971)
beta_Y <- Sim_Pleiotropy$Y
beta_X <- Sim_Pleiotropy$X

MR2_output <- MR2(beta_Y, beta_X, EVgamma = c(1, 3),
                  niter = 15000, burnin = 5000, thin = 10, monitor = 1000,
```

```
                    nonDecomp = TRUE, seed = 280610971)

PostProc_output <- PostProc(MR2_output, beta_Y, beta_X)

BMM_EM_output <- BMM_EM(PostProc_output$gammaPost)
FDR_BMM_output <- FDR_BMM(BMM_EM_output$PPI, BMM_EM_output$w)
PostProc_output$thetaPost* (PostProc_output$gammaPost>FDR_BMM_output$cutoff) * 1
```

---

mod_BDgraph   *Birth and Death sampler of non-decomposable graphs*

Description

Internal function that performs one Markov chain Monte Carlo iteration to sample from the posterior distribution of a non-decomposable graph. The R code of this function is a modified version of the BDgraph R-package (Mohammadi and Wit (2019)). For details, see also Mohammadi and Wit (2015) and Alexopoulos and Bottolo (2021)

Usage

```
mod_BDgraph(
  data,
  g.start,
  K.start,
  samplerPar,
  n = NULL,
  not.cont = NULL,
  df.prior = 2,
  g.prior = 0.5,
  threshold = 1e-08
)
```

References

Alexopoulos A, Bottolo L (2021). "Bayesian Variable Selection for Gaussian copula regression models." *J. Comput. Graph. Stat.*, **30**(3), 578–593. doi:10.1080/10618600.2020.1840997.

Mohammadi A, Wit EC (2015). "Bayesian structure learning in sparse Gaussian graphical models." *Bayesian Anal.*, **10**(1), 109–138. doi:10.1214/14BA889.

Mohammadi R, Wit EC (2019). "BDgraph: An R package for Bayesian structure learning in graphical models." *J. Stat. Softw.*, **89**(3), 1–30. doi:10.18637/jss.v089.i03.

---

mod_BDgraphInit   *Birth and Death sampler of non-decomposable graphs*

Description

Internal function that performs one Markov chain Monte Carlo iteration to sample from the posterior distribution of a non-decomposable graph and initialises the Markov chain Monte Carlo. The R code of this function is a modified version of the BDgraph R-package (Mohammadi and Wit (2019)). For details, see also Mohammadi and Wit (2015) and Alexopoulos and Bottolo (2021)

Usage

```
mod_BDgraphInit(
  data,
  samplerPar,
  n = NULL,
  not.cont = NULL,
  df.prior = 2,
  g.prior = 0.5,
  threshold = 1e-08
)
```

References

Alexopoulos A, Bottolo L (2021). "Bayesian Variable Selection for Gaussian copula regression models." *J. Comput. Graph. Stat.*, **30**(3), 578–593. doi:10.1080/10618600.2020.1840997.

Mohammadi A, Wit EC (2015). "Bayesian structure learning in sparse Gaussian graphical models." *Bayesian Anal.*, **10**(1), 109–138. doi:10.1214/14BA889.

Mohammadi R, Wit EC (2019). "BDgraph: An R package for Bayesian structure learning in graphical models." *J. Stat. Softw.*, **89**(3), 1–30. doi:10.18637/jss.v089.i03.

MR2   *MR2: Multi-response Mendelian randomization*

Description

Markov chain Monte Carlo implementation of Bayesian Multi-response Mendelian randomization model for two-sample summary-level data

Usage

```
MR2(
  Y,
  X,
  EVgamma = c(2, 2),
  tau = 1,
  niter,
  burnin,
  thin,
  monitor,
  fullCov = FALSE,
  nonDecomp = FALSE,
  alpha = 0.05,
  probAddDel = 0.9,
  probAdd = 0.5,
  probMix = 0.25,
  geomMean = EVgamma[1],
  graphParNiter = max(16, dim(Y)[2] * dim(Y)[2]),
  std = TRUE,
  seed = 31122021
)
```

Arguments

Y	(number of IVs) times (number of responses) matrix of summary-level responses
X	(number of IVs) times (number of exposures) matrix of summary-level exposures
EVgamma	A priori number of expected exposures directly associated with each response and its variance. For details, see Kohn et al. (2001), Alexopoulos and Bottolo (2021) and Zuber et al. (2023). Default value is EVgamma = c(2, 2) which implies a priori range of important exposures associated with each response between 0 and 6
tau	Prior variance of the (direct causal effect) theta prior with the default value set at 1. If std = TRUE, then tau = 1
niter	Number of Markov chain Monte Carlo iterations (including burn-in)
burnin	Number of Markov chain Monte Carlo iterations to be discarded as burn-in
thin	Store the Markov chain Monte Carlo output at every thin-th iteration
monitor	Display the monitored-th iteration
fullCov	Logical parameter to select full (TRUE) or sparse (FALSE default) inverse correlation matrix Markov chain Monte Carlo sampler
nonDecomp	Logical, TRUE if a non-decomposable graphical model for the inverse correlation matrix is selected
alpha	Level of significance (0.05 default) to select important exposures for each response by using one at-a-time response, univariable MR. Used in the construction of the proposal distribution of gamma

probAddDel	Probability (`0.9` default) of adding/deleting one exposure to be directly associated in a model with a response during Markov chain Monte Carlo exploration
probAdd	Probability (`0.9` default) of adding one exposure to be directly associated in a model with a response during Markov chain Monte Monte Carlo exploration
probMix	Probability (`0.25` default) of selecting a geometric proposal distribution that samples the index of the exposure being added/deleted. For details, see Alexopoulos and Bottolo (2021)
geomMean	Mean of the geometric proposal distribution that samples the index of the exposure being added/deleted. Default value is the number of expected exposures directly associated with each response and specified in `EVgamma`. For details, see Alexopoulos and Bottolo (2021)
graphParNiter	Number of Markov chain Monte Carlo internal iterations to sample the graphical model between responses. Default value is the max between 16 and the square of the number of responses
std	Logical parameter to standardise the exposures before the analysis. Default value is set at `TRUE`
seed	Seed used to initialise the Markov chain Monte Carlo algorithm

Details

For details regarding the model and the algorithm, see Zuber et al. (2023) and Alexopoulos and Bottolo (2021), respectively

Value

The value returned is a list object `list(theta, G, R, D, postMean, hyperPar, samplerPar, opt)`

- `theta` Matrix of the (thinned) samples drawn from the posterior distribution of the direct causal effects

- `G` 3D array of the (thinned) samples drawn from the posterior distribution of the graphs

- `R` 3D array of the (thinned) samples drawn from the posterior distribution of the correlation matrix between the responses

- `D` Matrix of the (thinned) samples drawn from the posterior distribution of the standard deviations of each responses

- `postMean` List of the posterior means of `list(theta, G, R, D)`

- `hyperPar` List of the hyper-parameters `list(tau)` and the parameters of the beta-binomial distribution derived from `EVgamma`

- `samplerPar` List of parameters used in the Markov chain Monte Carlo algorithm `list(alpha, probAddDel, probAdd, probMix, geomMean, graphParNiter)`

- `opt` List of options used `list(std, seed)`

References

Alexopoulos A, Bottolo L (2021). "Bayesian Variable Selection for Gaussian copula regression models." *J. Comput. Graph. Stat.*, **30**(3), 578–593. doi:10.1080/10618600.2020.1840997.

Kohn R, Smith M, Chan D (2001). "Nonparametric regression using linear combinations of basis functions." *Stat. Comput.*, **11**(4), 313–322. doi:10.1023/A:1011916902934.

Zuber V, Lewin A, Levin M, Haglund A, Ben-Aicha Gonzalez S, Emanueli C, Damrauer S, Burgess S, Gill D, Bottolo L (2023). "Multi-response Mendelian randomization: Identification of shared and distinct exposures for multimorbidity and multiple related disease outcomes." *Submitted*. https://www.biorxiv.org/content/10.1101/2023.02.01.526689v1.

Examples

```
# Example 1: Analysis of one replication of simulated Scenario II-Counfounding with q = 5
# responses, p = 15 exposures and nIV = 100 genetic variants used as IVs. The number of expected
# exposures directly associated with each response is set at 2 and its variance at 2, with a
# priori range of direct causal associations ranging between 0 and 8

Sim_Counfounding <- Scenario_Counfounding(nRep = 1, q = 5, p = 15, nIV = 100, seed = 280610971)
beta_Y <- Sim_Counfounding$Y
beta_X <- Sim_Counfounding$X

MR2_output <- MR2(beta_Y, beta_X, EVgamma = c(2, 2),
                  niter = 7500, burnin = 2500, thin = 5, monitor = 1000, seed = 280610971)
head(MR2_output$postMean$theta)


# Example 2: Analysis of one replication of simulated Scenario IV-Directed pleiotropy with q =
# 5 responses, p = 15 exposures and nIV = 100 genetic variants used as IVs and the effect of
# the shared pleiotropic pathway on the responses set as \code{2}. The number of expected
# exposures directly associated with each response is set at 1 and its variance at 2,
# with a priori range of direct causal associations ranging between 0 and 7. A non-decomposable
# graph for the inverse of the residual correlation between responses is selected

Sim_Pleiotropy <- Scenario_Pleiotropy(nRep = 1, q = 5, p = 15, nIV = 100, undirectedA = FALSE,
                                       seed = 280610971)
beta_Y <- Sim_Pleiotropy$Y
beta_X <- Sim_Pleiotropy$X

MR2_output <- MR2(beta_Y, beta_X, EVgamma = c(1, 3),
                  niter = 15000, burnin = 5000, thin = 10, monitor = 1000,
                  nonDecomp = TRUE, seed = 280610971)
```

PostProc   *Post-processing of MR2 output*

Description

Post-processing of MR2 Markov chain Monte Carlo output

Usage

```
PostProc(output, Y, X, alpha = 0.05, digits = 3)
```

Arguments

output	Object returned by MR2
Y	(number of IVs) times (number of responses) matrix of summary-level responses
X	(number of IVs) times (number of exposures) matrix of summary-level exposures
alpha	Area of the tails of a symmetric credible interval (0.05 default) centred around the estimated direct causal effects and the estimated covariance and (partial) residual correlation matrices
digits	Integer indicating the number of decimal places

Details

For details regarding the post-processing of the Markov chain Monte Carlo output, see details in Zuber et al. (2023)

Value

The value returned is a list object list(gammaPost, thetaPost, thetaPost_CI, DPost, GPost, SigmaPost, SigmaPost_CI,RPost, RPost_CI, Rm1Post, Rm1Post_CI, decomposability_freq, CPO, scaleCPO, freq_ouliers, PLML, DIC, opt)

- gammaPost Marginal posterior probability of inclusion (mPPI)
- thetaPost Posterior mean of the direct causal effects
- thetaPost_CI of the direct causal effects
- DPost Posterior mean of the standard deviations of each response
- GPost Edge posterior probability of inclusion (ePPI)
- SigmaPost Posterior mean of the variance-covariance matrix between the responses after conditioning on the direct causal effects
- SigmaPost_CI $(1 - \alpha)$ credible interval of the variance-covariance matrix between the responses after conditioning on the direct causal effects
- RPost Posterior mean of the correlation matrix between the responses after conditioning on the direct causal effects
- RPost_CI $(1 - \alpha)$ credible interval of the correlation matrix between the responses after conditioning on the direct causal effects
- Rm1Post Posterior mean of the partial correlation matrix between the responses after conditioning on the direct causal effects

- Rm1Post_CI $(1 - \alpha)$ credible interval of the partial correlation matrix between the responses after conditioning on the direct causal effects
- decomposability_freq Frequency of visited decomposability models during the Markov chain Monte Carlo. If nonDecomp = TRUE, frequency can be less than 1
- CPO Conditional predictive ordinate for each instrumental variable, see Ntzoufras (2008)
- scaleCPO Scaled conditional predictive ordinate for each instrumental variable, see Congdon (2005)
- PLML Pseudo log-marginal likelihood
- scaleCPO Scaled conditional predictive ordinate for each instrumental variable, see Congdon (2005)
- freq_ouliers Frequency outliers or high-leverage and influential instrumental variables defined as the frequency of scale CPO < 0.01, see Congdon (2005)
- PLML Pseudo log-marginal likelihood
- stability_PLML Variance Pseudo log-marginal likelihood
- DIC Deviation Information Criteria, see Spiegelhalter et al. (2002)
- opt List of options used list(α, digits)

References

Congdon P (2005). *Bayesian Models for Categorical Data*. John Wiley \& Sons, West Sussex, England. doi:10.1002/0470092394.

Ntzoufras I (2008). *Bayesian Modeling Using WinBUGS*. John Wiley \& Sons, West Sussex, England. doi:10.1002/9780470434567.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002). "Bayesian measures of model complexity and fit (with discussion)." *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, **64**(4), 583–639. doi:10.1111/14679868.00353.

Zuber V, Lewin A, Levin M, Haglund A, Ben-Aicha Gonzalez S, Emanueli C, Damrauer S, Burgess S, Gill D, Bottolo L (2023). "Multi-response Mendelian randomization: Identification of shared and distinct exposures for multimorbidity and multiple related disease outcomes." *Submitted*. https://www.biorxiv.org/content/10.1101/2023.02.01.526689v1.

Examples

```
# Example 1: Analysis of one replication of simulated Scenario II-Counfounding with q = 5
# responses, p = 15 exposures and nIV = 100 genetic variants used as IVs. The number of
# expected exposures directly associated with each response is set at 2 and its variance at 2,
# with a priori range of direct causal association ranging between 0 and 8

Sim_Counfounding <- Scenario_Counfounding(nRep = 1, q = 5, p = 15, nIV = 100, seed = 31122021)
beta_Y <- Sim_Counfounding$Y
beta_X <- Sim_Counfounding$X

MR2_output <- MR2(beta_Y, beta_X, EVgamma = c(2, 2),
                  niter = 7500, burnin = 2500, thin = 5, monitor = 1000, seed = 31122021)
```

```
PostProc_output <- PostProc(MR2_output, beta_Y, beta_X)


# Example 2: Analysis of one replication of simulated Scenario IV-Directed pleiotropy with q =
# 5 responses, p = 15 exposures and nIV = 100 genetic variants used as IVs and the effect of
# the shared pleiotropic pathway on the responses set as \code{2}. The number of expected
# exposures directly associated with each response is set at 1 and its variance at 3,
# with a priori range of direct causal associations ranging between 0 and 7. A non-decomposable
# graph for the inverse of the residual correlation between responses is selected

Sim_Pleiotropy <- Scenario_Pleiotropy(nRep = 1, q = 5, p = 15, nIV = 100, undirectedA = FALSE,
                                      seed = 280610971)
beta_Y <- Sim_Pleiotropy$Y
beta_X <- Sim_Pleiotropy$X

MR2_output <- MR2(beta_Y, beta_X, EVgamma = c(1, 3),
                  niter = 15000, burnin = 5000, thin = 10, monitor = 1000,
                  nonDecomp = TRUE, seed = 280610971)

PostProc_output <- PostProc(MR2_output, beta_Y, beta_X)
```

Sample_G   *Sampler of decomposable graphs*

Description

Internal functions that perform one Markov chain Monte Carlo iteration to sample from the posterior distribution of a decomposable graph. The R code of this function is a modified version of the Matlab code in Talhouk et al. (2012). For details, see Alexopoulos and Bottolo (2021)

Usage

```
Sample_G(W, GCurr, samplerPar, delta = 2)
```

References

Alexopoulos A, Bottolo L (2021). "Bayesian Variable Selection for Gaussian copula regression models." *J. Comput. Graph. Stat.*, **30**(3), 578–593. doi:10.1080/10618600.2020.1840997.

Talhouk A, Doucet A, Murphy K (2012). "Efficient Bayesian inference for multivariate Probit models with sparse inverse correlation matrices." *J. Comput. Graph. Stat.*, **21**(3), 739–757. doi:10.1080/10618600.2012.679239.

Sample_gamma *Sampler of the binary latent variable*

Description

Internal function for the proposal distribution of the binary latent vector for each response based on Guan and Stephens (2011). For details, see Alexopoulos and Bottolo (2021)

Usage

```
Sample_gamma(gammaCurr, p, mlogpval, samplerPar)
```

References

Alexopoulos A, Bottolo L (2021). "Bayesian Variable Selection for Gaussian copula regression models." *J. Comput. Graph. Stat.*, **30**(3), 578–593. doi:10.1080/10618600.2020.1840997.

Guan Y, Stephens M (2011). "Bayesian variable selection regression for genome-wide association studies and other large-scale problems." *Ann. Appl. Stat.*, 1780–1815. doi:10.1214/11AOAS455.

Scenario_Counfounding *Simulated scenario: Confounding*

Description

Simulator to generate multiple responses and multiple exposures two-sample summary-level data with confounding effects

Usage

```
Scenario_Counfounding(
  nRep = 2,
  q = 5,
  p = 10,
  nIV = 100,
  nSubject = 50000,
  MAF = 0.05,
  thetaRange = c(-2, 2),
  thetaSparsness = 0.3,
  heritY = 0.25,
  betaXRange = c(-2, 2),
  heritX = 0.1,
  rhoX = 0.6,
  thetaUY = 1,
  thetaUX = 2,
```

```
      std = TRUE,
      seed = 31122021
)
```

Arguments

nRep	Number of simulated replicates
q	Number of simulated summary-level responses
p	Number of simulated summary-level exposures
nIV	Number of genetic variants used as instrumental variables
nSubject	Number of individuals in each sample of individual-level data
MAF	Major allele frequency used in the simulation of the genotypes in each sample individual-level data. Default value set at `0.05`
thetaRange	Sought range of the simulated direct causal effects. Default value set at `(-2,2)`
thetaSparsness	Proportion of simulated causal associations between exposures and responses. Default value set at `0.30`
heritY	Heriditability of the responses. Default value set at `0.25`
betaXRange	Sought range of the simulated genetic effects on the exposures. Default value set at `(-2,2)`
heritX	Heriditability of the exposures. Default value set at `0.10`
rhoX	Correlation between the exposures. Default value set at `0.60`
thetaUY	Impact of the confounder on the responses. Default value set at `1`
thetaUX	Impact of the confounder on the exposures. Default value set at `2`
std	Logical parameter to standardise the summary-level exposures before the analysis. Default value is set at (`TRUE`)
seed	Seed used to initialise the simulation

Details

For details regarding the simulated scenario, see details in Zuber et al. (2023)

Value

The value returned is a list object `list(Y, X, theta, par)`

- `Y` 3D array (IVs times responses times replicates) of the simulated (summary-level) regression coefficients of the genetic effects on each response
- `X` 3D array (IVs times exposures times replicates) of the simulated (summary-level) regression coefficients of the genetic effects on each exposure
- `theta` Matrix ((exposures times replicates) times replicates) of the simulated direct causal effects
- `par` List of all parameters used in the simulation `list(nRep, q, p, nIV, nSubject, MAF, thetaRange, thetaSparsness, heritY, betaXRange, heritX, rhoX, thetaUY, thetaUX, std, seed)`

References

Zuber V, Lewin A, Levin M, Haglund A, Ben-Aicha Gonzalez S, Emanueli C, Damrauer S, Burgess S, Gill D, Bottolo L (2023). "Multi-response Mendelian randomization: Identification of shared and distinct exposures for multimorbidity and multiple related disease outcomes." *Submitted*. https://www.biorxiv.org/content/10.1101/2023.02.01.526689v1.

Examples

```
# Example: Simulation of one replication of simulated Scenario II-Counfounding with q = 5
# responses, p = 15 exposures and nIV = 100 genetic variants used as IVs

Sim_Counfounding <- Scenario_Counfounding(nRep = 1, q = 5, p = 15, nIV = 100, seed = 280610971)
head(Sim_Counfounding$theta)
Sim_Counfounding$par
```

Scenario_Dependence   *Simulated scenario: Dependence*

Description

Simulator to generate multiple responses and multiple exposures two-sample summary-level data with confounding and dependence between the responses not due to any unmeasured pleiotropic pathway

Usage

```
Scenario_Dependence(
  nRep = 2,
  q = 5,
  p = 10,
  nIV = 100,
  nSubject = 50000,
  MAF = 0.05,
  thetaRange = c(-2, 2),
  thetaSparsness = 0.3,
  heritY = 0.25,
  rhoY = 0.6,
  betaXRange = c(-2, 2),
  heritX = 0.1,
  rhoX = 0.6,
  thetaUY = 1,
  thetaUX = 2,
  std = TRUE,
  seed = 31122021
)
```

Arguments

nRep	Number of simulated replicates
q	Number of simulated summary-level responses
p	Number of simulated summary-level exposures
nIV	Number of genetic variants used as instrumental variables
nSubject	Number of individuals in each sample of individual-level data
MAF	Major allele frequency used in the simulation of the genotypes in each sample individual-level data. Default value set at `0.05`
thetaRange	Sought range of the simulated direct causal effects. Default value set at `(-2,2)`
thetaSparsness	Proportion of simulated causal associations between exposures and responses. Default value set at `0.30`
heritY	Heriditability of the responses. Default value set at `0.25`
rhoY	Correlation between the responses not due to any unmeasured pleiotropic pathway. Default value set at `0.60`
betaXRange	Sought range of the simulated genetic effects on the exposures. Default value set at `(-2,2)`
heritX	Heriditability of the exposures. Default value set at `0.10`
rhoX	Correlation between the exposures. Default value set at `0.60`
thetaUY	Impact of the confounder on the responses. Default value set at 1
thetaUX	Impact of the confounder on the exposures. Default value set at 2
std	Logical parameter to standardise the summary-level exposures before the analysis. Default value is set at (`TRUE`)
seed	Seed used to initialise the simulation

Details

For details regarding the simulated scenario, see details in Zuber et al. (2023)

Value

The value returned is a list object `list(Y, X, theta, par)`

- `Y` 3D array (IVs times responses times replicates) of the simulated (summary-level) regression coefficients of the genetic effects on each response

- `X` 3D array (IVs times exposures times replicates) of the simulated (summary-level) regression coefficients of the genetic effects on each exposure

- `theta` Matrix ((exposures times replicates) times replicates) of the simulated direct causal effects

- `par` List of all parameters used in the simulation `list(nRep, q, p, nIV, nSubject, MAF, thetaRange, thetaSparsness, heritY, rhoX, betaXRange, heritX, rhoX, thetaUY, thetaUX, std, seed)`

References

Zuber V, Lewin A, Levin M, Haglund A, Ben-Aicha Gonzalez S, Emanueli C, Damrauer S, Burgess S, Gill D, Bottolo L (2023). "Multi-response Mendelian randomization: Identification of shared and distinct exposures for multimorbidity and multiple related disease outcomes." *Submitted*. `https://www.biorxiv.org/content/10.1101/2023.02.01.526689v1`.

Examples

```
# Example: Simulation of one replication of simulated Scenario V-Dependence with q = 5
# responses, p = 15 exposures and nIV = 100 genetic variants used as IVs

Sim_Dependence <- Scenario_Dependence(nRep = 1, q = 5, p = 15, nIV = 100, seed = 280610971)
head(Sim_Dependence$theta)
Sim_Dependence$par
```

Scenario_Mediation   *Simulated scenario: Mediation*

Description

Simulator to generate multiple responses and multiple exposures two-sample summary-level data with confounding and pleiotropic effects as well as the effect of one response on another one

Usage

```
Scenario_Mediation(
  nRep = 2,
  q = 5,
  p = 10,
  nIV = 100,
  nSubject = 50000,
  MAF = 0.05,
  thetaRange = c(-2, 2),
  thetaSparsness = 0.3,
  heritY = 0.25,
  rhoY = 0.6,
  betaXRange = c(-2, 2),
  heritX = 0.1,
  rhoX = 0.6,
  undirectedA = TRUE,
  thetaA = 1,
  thetaM = 1,
  thetaUY = 1,
  thetaUX = 2,
  std = TRUE,
  seed = 31122021
)
```

Arguments

nRep	Number of simulated replicates
q	Number of simulated summary-level responses
p	Number of simulated summary-level exposures
nIV	Number of genetic variants used as instrumental variables
nSubject	Number of individuals in each sample of individual-level data
MAF	Major allele frequency used in the simulation of the genotypes in each sample individual-level data. Default value set at `0.05`
thetaRange	Sought range of the simulated direct causal effects. Default value set at `(-2,2)`
thetaSparsness	Proportion of simulated causal associations between exposures and responses. Default value set at `0.30`
heritY	Heriditability of the responses. Default value set at `0.25`
rhoY	Correlation between the responses not due to any unmeasured pleiotropic pathway. Default value set at `0.60`
betaXRange	Sought range of the simulated genetic effects on the exposures. Default value set at `(-2,2)`
heritX	Heriditability of the exposures. Default value set at `0.10`
rhoX	Correlation between the exposures. Default value set at `0.60`
undirectedA	Logical parameter for simulating an undirected (`TRUE`, default) or directed pleiotropic pathway
thetaA	Effect of the shared pleiotropic pathway on the responses. Default value set at `1`
thetaM	Effect of one response on another one, both selected at random. Default value set at `1`
thetaUY	Impact of the confounder on the responses. Default value set at `1`
thetaUX	Impact of the confounder on the exposures. Default value set at `2`
std	Logical parameter to standardise the summary-level exposures before the analysis. Default value is set at (`TRUE`)
seed	Seed used to initialise the simulation

Details

For details regarding the simulated scenario, see details in Zuber et al. (2023)

Value

The value returned is a list object `list(Y, X, theta, par)`

- `Y` 3D array (IVs times responses times replicates) of the simulated (summary-level) regression coefficients of the genetic effects on each response
- `X` 3D array (IVs times exposures times replicates) of the simulated (summary-level) regression coefficients of the genetic effects on each exposure
- `theta` Matrix ((exposures times replicates) times replicates) of the simulated direct causal effects

- `mediation` Matrix (two times replicates) times replicates) of the selected responses to be the mediated response and the mediator
- `par` List of all parameters used in the simulation `list(nRep, q, p, nIV, nSubject, MAF, thetaRange, thetaSparsness, heritY, rhoY, betaXRange, heritX, rhoX, undirectedA, thetaA, thetaM, thetaUY, thetaUX, std, seed)`

References

Zuber V, Lewin A, Levin M, Haglund A, Ben-Aicha Gonzalez S, Emanueli C, Damrauer S, Burgess S, Gill D, Bottolo L (2023). "Multi-response Mendelian randomization: Identification of shared and distinct exposures for multimorbidity and multiple related disease outcomes." *Submitted*. https://www.biorxiv.org/content/10.1101/2023.02.01.526689v1.

Examples

```
# Example 1: Simulation of one replication of simulated Scenario Undirected pleiotropy with
# Mediation and with q = 5 responses, p = 15 exposures and nIV = 100 genetic variants used as
# IVs

Sim_Mediation <- Scenario_Mediation(nRep = 1, q = 5, p = 15, nIV = 100, seed = 280610971)
Sim_Mediation$mediation
head(Sim_Mediation$theta)
Sim_Mediation$par


# Example 2: Simulation of one replication of simulated Scenario Directed pleiotropy with
# Mediation and with q = 5 responses, p = 15 exposures and nIV = 100 genetic variants used as
# IVs and the effect of the shared pleiotropic pathway on the responses set as \code{2}

Sim_Mediation <- Scenario_Mediation(nRep = 1, q = 5, p = 15, nIV = 100, undirectedA = FALSE,
                                    thetaA = 2, seed = 280610971)
Sim_Mediation$mediation
head(Sim_Mediation$theta)
Sim_Mediation$par
```

Scenario_Pleiotropy   *Simulated scenario: Pleiotropy*

Description

Simulator to generate multiple responses and multiple exposures two-sample summary-level data with confounding and pleiotropic effects

Usage

```
Scenario_Pleiotropy(
  nRep = 2,
  q = 5,
  p = 10,
```

```
    nIV = 100,
    nSubject = 50000,
    MAF = 0.05,
    thetaRange = c(-2, 2),
    thetaSparsness = 0.3,
    heritY = 0.25,
    betaXRange = c(-2, 2),
    heritX = 0.1,
    rhoX = 0.6,
    undirectedA = TRUE,
    thetaA = 1,
    thetaUY = 1,
    thetaUX = 2,
    std = TRUE,
    seed = 31122021
)
```

Arguments

nRep	Number of simulated replicates
q	Number of simulated summary-level responses
p	Number of simulated summary-level exposures
nIV	Number of genetic variants used as instrumental variables
nSubject	Number of individuals in each sample of individual-level data
MAF	Major allele frequency used in the simulation of the genotypes in each sample individual-level data. Default value set at `0.05`
thetaRange	Sought range of the simulated direct causal effects. Default value set at `(-2,2)`
thetaSparsness	Proportion of simulated causal associations between exposures and responses. Default value set at `0.30`
heritY	Heriditability of the responses. Default value set at `0.25`
betaXRange	Sought range of the simulated genetic effects on the exposures. Default value set at `(-2,2)`
heritX	Heriditability of the exposures. Default value set at `0.10`
rhoX	Correlation between the exposures. Default value set at `0.60`
undirectedA	Logical parameter for simulating an undirected (`TRUE`, default) or directed pleiotropic pathway
thetaA	Effect of the shared pleiotropic pathway on the responses. Default value set at 1
thetaUY	Impact of the confounder on the responses. Default value set at 1
thetaUX	Impact of the confounder on the exposures. Default value set at 2
std	Logical parameter to standardise the summary-level exposures before the analysis. Default value is set at (`TRUE`)
seed	Seed used to initialise the simulation

Details

For details regarding the simulated scenario, see details in Zuber et al. (2023)

Value

The value returned is a list object `list(Y, X, theta, par)`

- `Y` 3D array (IVs times responses times replicates) of the simulated (summary-level) regression coefficients of the genetic effects on each response

- `X` 3D array (IVs times exposures times replicates) of the simulated (summary-level) regression coefficients of the genetic effects on each exposure

- `theta` Matrix ((exposures times replicates) times replicates) of the simulated direct causal effects

- `par` List of all parameters used in the simulation `list(nRep, q, p, nIV, nSubject, MAF, thetaRange, thetaSparsness, heritY, betaXRange, heritX, rhoX, undirectedA, thetaA, thetaUY, thetaUX, std, seed)`

References

Zuber V, Lewin A, Levin M, Haglund A, Ben-Aicha Gonzalez S, Emanueli C, Damrauer S, Burgess S, Gill D, Bottolo L (2023). "Multi-response Mendelian randomization: Identification of shared and distinct exposures for multimorbidity and multiple related disease outcomes." *Submitted*. https://www.biorxiv.org/content/10.1101/2023.02.01.526689v1.

Examples

```
# Example 1: Simulation of one replication of simulated Scenario III-Undirected pleiotropy with
# q = 5 responses, p = 15 exposures and nIV = 100 genetic variants used as IVs

Sim_Pleiotropy <- Scenario_Pleiotropy(nRep = 1, q = 5, p = 15, nIV = 100, seed = 280610971)
head(Sim_Pleiotropy$theta)
Sim_Pleiotropy$par


# Example 2: Simulation of one replication of simulated Scenario IV-Directed pleiotropy with
# q = 5 responses, p = 15 exposures and nIV = 100 genetic variants used as IVs and the effect
# of the shared pleiotropic pathway on the responses set as \code{2}

Sim_Pleiotropy <- Scenario_Pleiotropy(nRep = 1, q = 5, p = 15, nIV = 100, undirectedA = FALSE,
                                      thetaA = 2, seed = 280610971)
head(Sim_Pleiotropy$theta)
Sim_Pleiotropy$par
```

Sim_HIW   *Sampler of Hyper-Inverse Wishart*

Description

Internal functions to simulate from the Hyper Inverse-Wishart distribution given a decomposable graph. The R code of this function is a modified version of the Matlab code of Carvalho et al. (2007) and Talhouk et al. (2012). For details, see details, see Alexopoulos and Bottolo (2021)

Usage

```
Sim_HIW(G, S, n)
```

References

Alexopoulos A, Bottolo L (2021). "Bayesian Variable Selection for Gaussian copula regression models." *J. Comput. Graph. Stat.*, **30**(3), 578–593. doi:10.1080/10618600.2020.1840997.

Carvalho CM, Massam H, West M (2007). "Simulation of hyper-inverse Wishart distributions in graphical models." *Biometrika*, **94**(3), 647–659. doi:10.1093/biomet/asm056.

Talhouk A, Doucet A, Murphy K (2012). "Efficient Bayesian inference for multivariate Probit models with sparse inverse correlation matrices." *J. Comput. Graph. Stat.*, **21**(3), 739–757. doi:10.1080/10618600.2012.679239.

Index