Package 'epistasis'

November 23, 2016

Type Package

Title Detecting Epistatic Selection with Partially Observed Genotype Data

Version 0.0.1-1

Author Pariya Behrouzi, Danny Arends and Ernst C. Wit

Maintainer Pariya Behrouzi <pariya.behrouzi@gmail.com>

Depends R (>= 3.1.0), Matrix, glasso, tmvtnorm, igraph, parallel

Imports methods

Description An efficient multi-core package to reconstruct an underlying network of genomic signatures of high-dimensional epistatic selection from partially observed genotype data. The phenotype that we consider is viability. The network captures the conditional dependent short- and long-range linkage disequilibrium structure of genomes and thus reveals aberrant marker-marker associations that are due to epistatic selection. We target on high-dimensional genotype data where number of variables (markers) is larger than number of sample sizes (p >> n). The computations is memory-optimized using the sparse matrix output.

License GPL-3

Date 2016-11-23

NeedsCompilation no

Repository CRAN

Date/Publication 2016-11-23 09:38:45

R topics documented:

`epistasis-package` *Detecting Epistatic Selection with Partially Observed Genotype Data*

Description

A package for reconstructing an underlying network of genomic signatures of high-dimensional epistatic selection from multi-locus genotype data. The package is implemented the recent improvement in the analysis of high-dimensional partially observed genotype data Behrouzi and Wit (2016). The network captures the conditionally dependent short- and long-range linkage disequilibrium structure of a genomes and reveals aberrant marker-marker associations that are due to epistatic selection rather than gametic linkage.

Author(s)

Pariya Behrouzi, Danny Arends, and Ernst C. Wit
Maintainers: Pariya Behrouzi <pariya.behrouzi@gmail.com>

References

P. Behrouzi and E. C. Wit. Detecting Epistatic Selection with Partially Observed Genotype Data Using Copula Graphical Models. *arXiv*, 2016.

`B73Ki11` *A family from Nested Association Mapping (NAM) in maize*

Description

The genotype data for family B73 x ki11 Recombinant Inbred Line (RIL) from NAM population.

Usage

```
data(B73Ki11)
```

Format

The format is a list containing two matrices. 1. the data matrix with 1106 single-nucleotide polymorphism (SNP) markers for 191 individuals. 2. Information about the SNP markers regarding their location in the genome.

Details

The Nested Association Mapping (NAM) initiative in maize populations is designed to reveal the genetic structure of underlying complex traits in maize. As part of this study, an inbred Ki11 maize line was crossed with the B73 reference line. This genotype data contains 1106 markers genotyped for 193 individuals. The B73 x Ki11 RIL is a diploid population with three possible genotypes, k = 3. This data set can be used to detect epistatic selection, short- and long- range linkage disequilibrium between 1106 SNP markers.

Author(s)

Pariya Behrouzi and Ernst C. Wit
Maintainer: Pariya Behrouzi <pariya.behrouzi@gmail.com>

Source

It is publicly available at http://datacommons.cyverse.org/browse/iplant/home/shared/panzea/genotypes

References

1. P. Behrouzi and E. C. Wit. Detecting Epistatic Selection with Partially Observed Genotype Data Using Copula Graphical Models. *arXiv*, 2016.
2. McMullen, M. D., S. Kresovich, H. S. Villeda, P. Bradbury, H. Li, Q. Sun, S. Flint-Garcia, J. Thornsberry, C. Acharya, C. Bottoms, et al. (2009). Genetic properties of the maize nested association mapping population. Science 325 (5941), 737-740. 3. Rodgers-Melnick, E., P. J. Bradbury, R. J. Elshire, J. C. Glaubitz, C. B. Acharya, S. E. Mitchell, C. Li, Y. Li, and E. S. Buckler (2015). Recombination in diverse maize is stable, predictable, and associated with genetic load. Proceedings of the National Academy of Sciences 112 (12), 3823-3828.

Examples

```
data(B73Ki11)
image(B73Ki11$data, ylab="Markers" , xlab= "Individuals")
B73Ki11$info
```

---

cutoffs *Cut-points*

Description

Calculates cut-points of ordinal variables with respect to the Gaussian copula.

Usage

```
cutoffs(y)
```

Arguments

y	An $(n \times p)$ matrix or a `data.frame` corresponding to the data matrix (n is the sample size and p is the number of variables). It also could be an object of class `"episim"`.

Details

The relationship between jth variable and jth latent variable is expressed through this set of cut-points.

Value

cutoffs	A p by $(k+1)$ matrix representing the cut-point values under the Gaussian copula, where k defines the number of states in the dataset.

Author(s)

Pariya Behrouzi and Ernst C. Wit
Maintainer: Pariya Behrouzi <pariya.behrouzi@gmail.com>

References

P. Behrouzi and E. C. Wit. Detecting Epistatic Selection with Partially Observed Genotype Data Using Copula Graphical Models. *arXiv*, 2016.

See Also

`lower.upper` and `epistasis-package`.

Examples

```
## Not run:
D <- episim(p = 100, n = 50, k = 3)
cutoffs(D$data)

## End(Not run)
```

CviCol *Arabidopsis thaliana genotype data*

Description

The genotype data of the Cvi-0 x Col-0 Recombinant Inbred Line (RIL) population.

Usage

```
data(CviCol)
```

Format

The format is a matrix containing 90 single-nucleotide polymorphism (SNP) markers for 367 individuals.

Details

The *Arabidopsis thaliana* genotype data is derived from a RIL cross between Columbia-0 (Col-0) and the Cape Verde Island (Cvi-0), where 367 individuals were genotyped for 90 genetic markers. This is a diploid population with three possible genotpe states (`k = 3`), where the genotypes coded as 0, 1, 2, where 0 and 2 represent the homozygous genotypes and 1 defines the heterozygous genotype.
This data set can be used to detect epistatic selection, short- and long- range linkage disequilibrium between 90 SNP markers.

Author(s)

Pariya Behrouzi and Ernst C. Wit
Maintainer: Pariya Behrouzi <pariya.behrouzi@gmail.com>

Source

It is publicly available at http://publiclines.versailles.inra.fr/page/8

References

1. P. Behrouzi and E. C. Wit. Detecting Epistatic Selection with Partially Observed Genotype Data Using Copula Graphical Models. *arXiv*, 2016.
2. Simon, M., et al. "QTL mapping in five new large RIL populations of Arabidopsis thaliana genotyped with consensus SNP markers." Genetics 178 (2008): 2253-2264.

Examples

```
## Not run:
data(CviCol)

# Graph path
epi <- epistasis(CviCol, method="approx", rho.ratio=0.2)
epi
plot(epi)

sel <- episelect(epi)
plot(sel)

## End(Not run)
```

---

episelect *Model selection*

Description

Estimate the optimal regularization parameter at EM convergence based on different information criteria .

Usage

```
episelect(epi.object, criteria = NULL, ebic.gamma = 0.5, loglik_Y = FALSE, ncores = NULL)
```

Arguments

epi.object	An object with S3 class "epi"
criteria	Model selection criteria. "ebic" and "aic" are available. BIC model selection can be calculated by fixing ebic.gamma = 0.
ebic.gamma	The tuning parameter for ebic. Theebic.gamma = 0 results in bic model selection. The default value is 0.5.
loglik_Y	Model selection based on either log-likelihood of observed data (loglik_Y = TRUE), or the joint log-likelihood of observed and latent variables (loglik_Y = FALSE).
ncores	The number of cores to use for the calculations. Using ncores = NULL automatically detects number of available cores and runs the computations in parallel.

Details

This function computes extended Bayesian information criteria (ebic), Bayesian information criteria, Akaike information criterion (aic) at EM convergence based on observed or joint log-likelihood. The observed log-likelihood can be obtained through

$$\ell_Y(\widehat{\Theta}_\lambda) = Q(\widehat{\Theta}_\lambda|\widehat{\Theta}^{(m)}) - H(\widehat{\Theta}_\lambda|\widehat{\Theta}^{(m)}),$$

Where Q can be calculated from `epistasis` function and H function is

$$H(\widehat{\Theta}_\lambda|\widehat{\Theta}_\lambda^{(m)}) = E_z[\ell_{Z|Y}(\widehat{\Theta}_\lambda)|Y;\widehat{\Theta}_\lambda] = E_z[\log f(z)|Y;\widehat{\Theta}_\lambda] - \log p(y).$$

The "ebic" and "aic" model selection criteria can be obtained as follow

$$ebic(\lambda) = -2\ell(\widehat{\Theta}_\lambda) + (\log n + 4\gamma \log p)df(\lambda)$$

$$aic(\lambda) = -2\ell(\widehat{\Theta}_\lambda) + 2df(\lambda)$$

where df refers to the number of non-zeros offdiagonal elements of $\hat{\Theta}_\lambda$, and $\gamma \in [0, 1]$. Typical value for for `ebic.gamma` is 1/2, but it can also be tuned by experience. Fixing `ebic.gamma = 0` results in bic model selection.

Value

An object with S3 class "episelect" is returned:

`opt.path`	The optimal graph selected from the graph path
`opt.theta`	The optimal precision matrix from the graph path
`opt.Sigma`	The optimal covariance matrix from the graph path
`ebic.scores`	Extended BIC scores for regularization parameter selection at the EM convergence.
`opt.index`	The index of optimal regularization parameter.
`opt.rho`	The selected regularization parameter.

and anything else that is included in the input `epi.object`.

Author(s)

Pariya Behrouzi and Ernst C.Wit
Maintainer: Pariya Behrouzi <pariya.behrouzi@gmail.com>

References

1. P. Behrouzi and E. C. Wit. Detecting Epistatic Selection with Partially Observed Genotype Data Using Copula Graphical Models. *arXiv*, 2016.
2. Ibrahim, Joseph G., Hongtu Zhu, and Niansheng Tang. "Model selection criteria for missing-data problems using the EM algorithm." Journal of the American Statistical Association (2012). 3. D. Witten and J. Friedman. New insights and faster computations for the graphical lasso. *Journal of Computational and Graphical Statistics*, to appear, 2011.
4. J. Friedman, T. Hastie and R. Tibshirani. Sparse inverse covariance estimation with the lasso, *Biostatistics*, 2007.
5. Foygel, R. and M. Drton (2010). Extended bayesian information criteria for Gaussian graphical models. In Advances in Neural Information Processing Systems, pp. 604-612.

See Also

epistasis

Examples

```
## Not run:
#simulate data
D <- episim(p=50, n=100, k= 3, adjacent = 3, alpha = 0.06 , beta = 0.06)
plot(D)

#detect epistatic selection path
out  <-  epistasis(D$data, method="gibbs", n.rho= 5, ncores= 1)

#different graph selection methods
sel.ebic1 <- episelect(out, criteria="ebic")
plot(sel.ebic1)

sel.ebic2 <- episelect(out, criteria="ebic", loglik_Y=TRUE)
plot(sel.ebic2)

sel.aic <- episelect(out, criteria="aic")
plot(sel.aic)

sel.bic <- episelect(out, criteria="ebic", ebic.gamma = 0)
plot(sel.bic)

## End(Not run)
```

`episim` *Generate discrete ordinal data*

Description

Generating discrete ordinal data based on underlying "genome-like" graph structure. The procedure of simulating data relies on a continues variable, which can be simulated from either multivariate normal distribution, or multivariate t-distribution with d degrees of freedom.

Usage

```
episim ( p = 90, n = 200, k = NULL, g = NULL, adjacent = NULL, alpha =
             NULL , beta = NULL, con.dist = "Mnorm", d = NULL, vis = FALSE)
```

Arguments

p	The number of variables. The default value is 90.
n	The number of sample size (observations). The default value is 200.
k	The number of states (categories). The default value is 3.

g	The number of groups (chromosomes) in the graph. The default value is about $p/20$ if $p >= 40$ and 2 if $p < 40$.
adjacent	The number of adjacent variable(s) to be linked to a variable. For example, if adjacent = 1 indicates a variable is linked via an edge with its adjacent variable on the left hand side, and its adjacent variable on the right hand side. The adjacent = 2 defines a variable is linked via an edge with its 2 adjacent variables on its left hand side, and 2 adjacent variables on its right hand side. The default value is 1.
alpha	A probability that a pair of non-adjacent variables in the same group is given an edge. The default value is 0.01.
beta	A probability that variables in different groups are linked with an edge. The default value is 0.02.
con.dist	The distribution of underlying continuous variable. If con.dist = "Mnorm", a multivariate normal distribution with mean 0 is applied. If con.dist = "Mt", the t-distribution with a degrees of freedom is applied. The default distribution is con.dist = "Mnorm".
d	The degrees of freedom of the continuous variable, only applicable when code-con.dist = "Mt". The default value is 3.
vis	Visualize the graph pattern and the adjacency matrix of the true graph structure. The default value is FALSE.

Details

The graph pattern is generated as below:

"genome-like": The p variables are evenly partitions variables into g disjoint groups; the adjacent variables within each group are linked via an edge. With a probability alpha a pair of non-adjacent variables in the same group is given an edge. Variables in different groups are linked with an edge with a probability of beta.

Value

An object with S3 class "episim" is returned:

data	The generated data as an n by p matrix.
Theta	A p by p matrix corresponding to the inverse of covariance.
adj	A p by p matrix corresponding to the adjacency matrix of the true graph structure.
Sigma	A p by p covariance matrix for the generated data.
n.groups	The number of groups.
groups	A vector that indicates each variable belongs to which group.
sparsity	The sparsity levels of the true graph.

Author(s)

Pariya Behrouzi and Ernst C. Wit
Maintainer: Pariya Behrouzi <pariya.behrouzi@gmail.com>

References

P. Behrouzi and E. C. Wit. Detecting Epistatic Selection with Partially Observed Genotype Data Using Copula Graphical Models. *arXiv*, 2016.

See Also

epistasis, and epistasis-package

Examples

```
## Not run:
#genome-like graph structure
sim1 <- episim(alpha = 0.01, beta = 0.02)
plot(sim1)

#genome-like graph structure with more edges between variables in a same or different groups
sim2 <- episim(adjacent = 3, alpha = 0.02 , beta = 0.03)
plot(sim2)

## End(Not run)
```

`epistasis` *Detecting epistatic selection from multi-loci genotype data*

Description

This is the main function of the **epistasis** package. Two methods are available to detect epistatic selection, including (1) approximation method, and (2) gibbs sampling within the Gaussian copula graphical model. Both methods are able to deal with missing genotypes.

Usage

```
epistasis(y, method = "approx", rho = NULL, n.rho = NULL, rho.ratio = NULL,
          ncores = NULL, em.iter = 10, em.tol = .001, verbose = TRUE)
```

Arguments

`y`	An $(n \times p)$ matrix or a `data.frame` corresponding to the data matrix (n is the sample size and p is the number of variables). It also could be an object of class `"episim"`. Input data can contain missing values.
`method`	Detecting epistatic selection with two methods: "gibbs" and "approx". The default method is "approx".

rho	A decreasing sequence of non-negative numbers that control the sparsity level. Leaving the input as rho = NULL, the program automatically computes a sequence of rho based on n.rho and rho.ratio. Users can also supply a decreasing sequence values to override this.
n.rho	The number of regularization parameters. The default value is 10.
rho.ratio	Determines distance between the elements of rho sequence. A small value of rho.ratio results in a large distance between the elements of rho sequence. And a large value of rho.ratio results into a small distance between elements of rho. The default value is 0.3.
ncores	The number of cores to use for the calculations. Using ncores = NULL automatically detects number of available cores and runs the computations in parallel on (available cores - 1).
em.tol	A criteria to stop the EM iterations. The default value is .001.
em.iter	The number of EM iterations. The default value is 10.
verbose	Providing a detail message for tracing output. The default value is TRUE.

Details

Viability is the phenotype that is considered. This function detects the conditional dependent short- and long-range linkage disequilibrium structure of genomes and thus reveals aberrant marker-marker associations that are due to epistatic selection. This function can be used to estimate conditional independence relationships between partially observed data that not follow Gaussianity assumption (e.g. continuous non-Gaussian, discrete, or mixed dataset).

Value

An object with S3 class "epi" is returned:

Theta	A list of estimated p by p precision matrices corresponding to rho.
path	A list of estimated p by p adjacency matrices. This is the graph path corresponding to rho.
Sigma	A list of estimated p by p covariance matrices corresponding to rho.
ES	A list of estimated p by p conditional expectation corresponding to rho.
Z	A list of n by p transformed data based on Gaussian copula.
rho	A n.rho dimensional vector containing the penalty terms.
loglik	A n.rho dimensional vector containing the maximized log-likelihood values along the graph path.
data	The n by p input data matrix.

Note

This function estimates the graph path . To select an optimal graph please refer to episelect.

Author(s)

Pariya Behrouzi and Ernst C. Wit
Maintainers: Pariya Behrouzi <pariya.behrouzi@gmail.com>

References

1. P. Behrouzi and E. C. Wit. Detecting Epistatic Selection with Partially Observed Genotype Data Using Copula Graphical Models. *arXiv*, 2016.
2. D. Witten and J. Friedman. New insights and faster computations for the graphical lasso. *Journal of Computational and Graphical Statistics*, to appear, 2011.
3. J. Friedman, T. Hastie and R. Tibshirani. Sparse inverse covariance estimation with the lasso, *Biostatistics*, 2007.
4. Guo, Jian, et al. "Graphical models for ordinal data." Journal of Computational and Graphical Statistics 24.1 (2015): 183-204.

See Also

episelect

Examples

```
## Not run:
#simulate data
D <- episim(p=50, n=100, k= 3, adjacent = 3, alpha = 0.06 , beta = 0.06)
plot(D)

#epistasis path estimation using approx
out1 <-  epistasis(D$data, method="approx", n.rho= 5)
plot(out1)

#epistasis path estimation using gibbs
out2  <-  epistasis(D$data, method="gibbs", n.rho= 5, ncores= 1)
plot(out2)

## End(Not run)
```

---

lower.upper *Calculates lower band and upper band*

Description

Calculates lower and upper bands for each data point, using a set of cut-points which is obtained from the Gaussian copula.

Usage

```
lower.upper(y)
```

Arguments

y	An $(n \times p)$ matrix or a `data.frame` corresponding to the data matrix (n is the sample size and p is the number of variables). It also could be an object of class `"episim"`.

Value

lower	A n by p matrix representing the lower band for each data point.
upper	A n by p matrix representing the upper band for each data point.

Author(s)

Pariya Behrouzi and Ernst C. Wit
Maintainer: Pariya Behrouzi <pariya.behrouzi@gmail.com>

References

P. Behrouzi and E. C. Wit. Detecting Epistatic Selection with Partially Observed Genotype Data Using Copula Graphical Models. *arXiv*, 2016.

See Also

`cutoffs` and `epistasis-package`.

Examples

```
## Not run:
D <- episim(p = 100, n = 50, k = 3)
lower.upper(D$data)

## End(Not run)
```

---

plot.epi *plot for S3 class "epi"*

Description

Plot the graph path which is the output of two functions `epistasis`.

Usage

```
## S3 method for class 'epi'
plot( x, n.markers=NULL , ... )
```

Arguments

x	An object from "epi" class.
n.markers	A vector containing number of variables/markers in each group/chromosome. For example, the CviCol dataset that is provided in the package contains 5 chromosomes/ groups which the total number of markers is $p = 90$, where the first 24 markers belong into chromosome 1, the next 14 markers into chromosome 2, ..., and chromosome 5 contains 19 markers. Thus, n.mrkr = c(24,14,17,16,19). If `n.mrkr = NULL`, in the graph visualization all markers are represented same colour.
...	System reserved (No specific usage)

Author(s)

Pariya Behrouzi and Ernst C. Wit
Maintainer: Pariya Behrouzi <pariya.behrouzi@gmail.com>

References

P. Behrouzi and E. C. Wit. Detecting Epistatic Selection with Partially Observed Genotype Data Using Copula Graphical Models. *arXiv*, 2016.

See Also

epistasis.

---

plot.episelect *Plot function for S3 class "episelect"*

Description

Plot the optimal graph by model selection

Usage

```
## S3 method for class 'episelect'
plot(x, n.markers = NULL, vertex.size=NULL, vertex.label = FALSE, ...)
```

Arguments

x	An object with S3 class "episelect"
n.markers	A vector containing number of variables/markers in each group/chromosome. For example, the CviCol dataset that is provided in the package contains 5 chromosomes/ groups which the total number of markers is $p = 90$, where the first 24 markers belong into chromosome 1, the next 14 markers into chromosome 2, ..., and chromosome 5 contains 19 markers. Thus, n.mrkr = c(24,14,17,16,19). If `n.mrkr = NULL`, in the graph visualization all markers are represented same colour.

vertex.size	The size of vertices in the graph visualization. The default value is 7.
vertex.label	Assign names to the vertices. Default is FALSE.
...	System reserved (No specific usage)

Author(s)

Pariya Behrouzi and Ernst C. Wit
Maintainer: Pariya Behrouzi <pariya.behrouzi@gmail.com>

References

P. Behrouzi and E. C. Wit. Detecting Epistatic Selection with Partially Observed Genotype Data Using Copula Graphical Models. *arXiv*, 2016.

See Also

episelect

---

plot.episim *Plot function for* S3 *class* "episim"

Description

Visualizes the pattern of the true graph, the adjacency matrix, precison matrix and the covariance matrix of the simulated data.

Usage

```
## S3 method for class 'episim'
plot(x, layout = layout.fruchterman.reingold, ...)
```

Arguments

x	An object of S3 class "episim", from function episim.
layout	The default is "layout.fruchterman.reingold".
...	System reserved (No specific usage)

Author(s)

Pariya Behrouzi and Ernst C. Wit
Maintainer: Pariya Behrouzi <pariya.behrouzi@gmail.com>

References

P. Behrouzi and E. C. Wit. Detecting Epistatic Selection with Partially Observed Genotype Data Using Copula Graphical Models. *arXiv*, 2016.

See Also

episim

Examples

```
## Not run:
# Generating discrete ordinal data with "genome-like" graph structure
data.sim <- episim(alpha = 0.01, beta = 0.02)
plot( data.sim )

## End(Not run)
```

`print.epi` *Print function for S3 class "epi"*

Description

Print a summary of simulated data from function epistasis.

Usage

```
## S3 method for class 'epi'
print(x, ...)
```

Arguments

x	An object with S3 class "epi"
...	System reserved (No specific usage)

Author(s)

Pariya Behrouzi and Ernst C. Wit
Maintainer: Pariya Behrouzi <pariya.behrouzi@gmail.com>

References

P. Behrouzi and E. C. Wit. Detecting Epistatic Selection with Partially Observed Genotype Data Using Copula Graphical Models. *arXiv*, 2016.

See Also

epistasis

`print.episim` *Print function for S3 class "episim"*

Description

Print a summary of simulated data from function `episim`.

Usage

```
## S3 method for class 'episim'
print(x, ...)
```

Arguments

`x`	An object with S3 class `"episim"`
`...`	System reserved (No specific usage)

Author(s)

Pariya Behrouzi and Ernst C. Wit
Maintainer: Pariya Behrouzi <pariya.behrouzi@gmail.com>

References

P. Behrouzi and E. C. Wit. Detecting Epistatic Selection with Partially Observed Genotype Data Using Copula Graphical Models. *arXiv*, 2016.

See Also

episim

R.approx *The expectation of covariance using approximation method*

Description

This function implements the approximation method within the Gaussian copula graphical model to estimate the conditional expectation for the data that not follow Gaussianity assumption (e.g. ordinal, discrete, continuous non-Gaussian, or mixed dataset).

Usage

```
R.approx(y, Z = NULL, Sigma=NULL, rho = NULL, ncores = NULL )
```

Arguments

y	An $(n \times p)$ matrix or a `data.frame` corresponding to the data matrix (n is the sample size and p is the number of variables). It also could be an object of class `"episim"`.
Z	A $(n \times p)$ matrix which is a transformation of the data via the Gaussian copula. If `Z = NULL` internally calculates an initial value for Z.
Sigma	The covariance matrix of the latent variable given the data. If `Sigma = NULL` the Sigma matrix is calculated internally with a desired penalty term, `rho`.
rho	A (non-negative) regularization parameter to calculate `Sigma`. rho=0 means no regularization.
ncores	If `ncores = NULL`, the algorithm internally detects number of available cores and run the calculations in parallel on (available cores - 1). Typical usage is to fix `ncores = 1` when p is small ($p < 500$), and `ncores = NULL` when p is large.

Value

ES	Expectation of covariance matrix(diagonal scaled to 1) of the Gaussian copula graphical model.
Z	New transformation of the data based on given or default `Sigma`.

Author(s)

Pariya Behrouzi and Ernst C. Wit
Maintainer: Pariya Behrouzi <pariya.behrouzi@gmail.com>

References

P. Behrouzi and E. C. Wit. Detecting Epistatic Selection with Partially Observed Genotype Data Using Copula Graphical Models. *arXiv*, 2016.

See Also

epistasis

Examples

```
## Not run:
D <- episim(p = 100, n = 50, k = 3)
R.approx(D$data)

## End(Not run)
```

R.gibbs *The expectation of covariance matrix using Gibbs sampling*

Description

This function implements the Gibbs sampling method within Gaussian copula graphical model to estimate the conditional expectation for the data that not follow Gaussianity assumption (e.g. ordinal, discrete, continuous non-Gaussian, or mixed dataset).

Usage

```
R.gibbs(y, theta, gibbs.iter = 1000, mc.iter = 500,
                  ncores = NULL, verbose = TRUE)
```

Arguments

y	An $(n \times p)$ matrix or a data.frame corresponding to the data matrix (n is the sample size and p is the number of variables). It also could be an object of class "episim".
theta	A $p \times p$ precision matrix. Default is a diagonal matrix.
gibbs.iter	The number of burn-in for the Gibbs sampling. The default value is 1000.
mc.iter	The number of Monte Carlo samples to calculate the conditional expectation. The default value is 500.
ncores	If ncores = NULL, the algorithm internally detects number of available cores and run the calculations in parallel on (available cores - 1). Typical usage is to fix ncores = 1 when p is small $(p < 500)$, and ncores = NULL when p is very large.
verbose	If verbose = FALSE, printing information is disabled. The default value is TRUE.

Details

This function calculates $\bar{R}$ using Gibbs sampling method within the E-step of EM algorithm, where

$$\bar{R} = n^{-1} \sum_{i=1}^{n} E(Z^{(i)} Z^{(i)t} | y^{(i)}, \hat{\Theta}^{(m)})$$

which n is the number of sample size and Z is the latent variable which is obtained from Gaussian copula graphical model.

Value

ES	Expectation of covariance matrix (diagonal scaled to 1) of the Gaussian copula graphical model

Author(s)

Pariya Behrouzi, Danny Arends and Ernst C. Wit
Maintainers: Pariya Behrouzi <pariya.behrouzi@gmail.com>

References

P. Behrouzi and E. C. Wit. Detecting Epistatic Selection with Partially Observed Genotype Data Using Copula Graphical Models. *arXiv*, 2016.

See Also

epistasis.

Examples

```
## Not run:
D <- episim(p = 100, n = 50, k = 3)
R.gibbs(D$data, ncores=1)

## End(Not run)
```

Index