Package ‘FBFsearch’

February 19, 2015

Type Package

Title Algorithm for searching the space of Gaussian directed acyclic graphical models through moment fractional Bayes factors

Version 1.0

Date 2013-04-12

Author Davide Altomare, Guido Consonni and Luca La Rocca

Maintainer Davide Altomare <davide.altomare@gmail.com>

Description We propose an objective Bayesian algorithm for searching the space of Gaussian directed acyclic graphical models when the variables are assumed to satisfy a given ordering. The approach used is based on non-local parameter priors and thus it is suitable for learning sparse graphs. The algorithm is implemented in C++ using the open-source library Armadillo.

License GPL (>= 2)

Depends Rcpp (>= 0.9.13), RcppArmadillo (>= 0.3.2.4)

LinkingTo Rcpp, RcppArmadillo

NeedsCompilation yes

Repository CRAN

Date/Publication 2013-04-13 06:39:46

R topics documented:

`dataHuman` *Cell signalling pathway data*

Description

Data on a set of flow cytometry experiments on signaling networks of human immune system cells. The dataset includes p=11 proteins and n=7466 samples.

Usage

```
data(HumanPw)
```

Format

`dataHuman` contains the following objects:

`Obs` Matrix (7466x11) with the observations.

`Perms` List of 5 matrices (1x11) each of which with a permutation of the nodes.

`TDag` Matrix (11x11) with the adjacency matrix of the known regulatory network.

Source

Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D., and Nolan, G. (2003). Casual protein- signaling networks derived from multiparameter single-cell data. *Science* 308, 504-6.

References

D. Altomare, G. Consonni and L. La Rocca (2013). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Biometrics*.

Shojaie, A. and Michailidis, G. (2010). Penalized likelihood methods for estimation of sparse high-dimensional directed acyclic graphs. *Biometrika* 97, 519-538.

`dataPub` *Publishing productivity data*

Description

Data on publishing productivity among academics.

Usage

```
data(PubProd)
```

Format

`dataPub` contains the following objects:

`Corr` Matrix (7x7) with the correlation matrix of the variables.

`nobs` Scalar with the number of observations.

Source

Spirtes, P., Glymour, C., and Scheines, R. (2000). Causation, prediction and search (2nd edition). *Cambridge, MA: The MIT Press*. pages 1-16.

References

Drton, M. and Perlman, M. D. (2008). A SINful approach to Gaussian graphical model selection. *J. Statist. Plann. Inference* 138, 1179-1200.

D. Altomare, G. Consonni and L. La Rocca (2013). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Biometrics*.

`dataSim100` *DAG model with 100 nodes and 100 edges*

Description

`dataSim100` is a `list` with the adjacency matrix of a randomly generated DAG with 100 nodes and 100 edges, 10 samples generated from the DAG and 5 permutations of the nodes.

Usage

```
data(SimDag100)
```

Format

`dataSim100` contains the following objects:

`Obs` List of 10 matrices (100x100) each of which with 100 observations generated from the DAG.

`Perms` List of 5 matrices (1x100) each of which with a permutation of the nodes.

`TDag` Matrix (100x100) with the adjacency matrix of the DAG.

Source

D. Altomare, G. Consonni and L. La Rocca (2013). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Biometrics*.

References

Shojaie, A. and Michailidis, G. (2010). Penalized likelihood methods for estimation of sparse high-dimensional directed acyclic graphs. *Biometrika* 97, 519-538.

`dataSim200` *DAG model with 200 nodes and 100 edges*

Description

`dataSim200` is a `list` with the adjacency matrix of a randomly generated DAG with 200 nodes and 100 edges, 10 samples generated from the DAG and 5 permutations of the nodes.

Usage

```
data(SimDag200)
```

Format

`dataSim200` contains the following objects:

`Obs` List of 10 matrices (100x200) each of which with 100 observations simulated from the DAG.

`Perms` List of 5 matrices (1x200) each of which with a permutation of the nodes.

`TDag` Matrix (200x200) with the adjacency matrix of the DAG.

Source

D. Altomare, G. Consonni and L. La Rocca (2013). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Biometrics*.

References

Shojaie, A. and Michailidis, G. (2010). Penalized likelihood methods for estimation of sparse high-dimensional directed acyclic graphs. *Biometrika* 97, 519-538.

`dataSim50` *DAG model with 50 nodes and 100 edges*

Description

`dataSim50` is a `list` with the adjacency matrix of a randomly generated DAG with 50 nodes and 100 edges, 10 samples generated from the DAG and 5 permutations of the nodes.

Usage

```
data(SimDag50)
```

Format

`dataSim50` contains the following objects:

`Obs` List of 10 matrices (100x50) each of which with 100 observations simulated from the DAG.

`Perms` List of 5 matrices (1x50) each of which with a permutation of the nodes.

`TDag` Matrix (50x50) with the adjacency matrix of the DAG.

Source

D. Altomare, G. Consonni and L. La Rocca (2013). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Biometrics*.

References

Shojaie, A. and Michailidis, G. (2010). Penalized likelihood methods for estimation of sparse high-dimensional directed acyclic graphs. *Biometrika* 97, 519-538.

`dataSim6` *DAG model with 6 nodes and 5 edges*

Description

`dataSim6` is a `list` with the adjacency matrix of a randomly generated DAG with 6 nodes and 5 edges and 100 correlation matrices generated from the DAG.

Usage

```
data(SimDag6)
```

Format

`dataSim6` contains the following objects:

`Corr` List of 100 matrices (6x6) each of which with a correlation matrix generated from the DAG.

`TDag` Matrix (6x6) with the adjacency matrix of the DAG.

References

D. Altomare, G. Consonni and L. La Rocca (2013). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Biometrics*.

`dataSimHuman` *Simulated cell signalling pathway data*

Description

Data generated from the known regulatory network of human cell signalling data.

Usage

```
data(SimHumanPw)
```

Format

`dataSimHuman` contains the following objects:

`Obs` List of 100 matrices (100x11) each of which with 100 observations simulated from the known regulatory network.

`Perms` List of 5 matrices (1x11) each of which with a permutation of the nodes.

`TDag` Matrix (11x11) with the adjacency matrix of the known regulatory network.

Source

D. Altomare, G. Consonni and L. La Rocca (2013). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Biometrics*.

References

Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D., and Nolan, G. (2003). Casual protein- signaling networks derived from multiparameter single-cell data. *Science* 308, 504-6.

Shojaie, A. and Michailidis, G. (2010). Penalized likelihood methods for estimation of sparse high-dimensional directed acyclic graphs. *Biometrika* 97, 519-538.

`FBF_GS` *Moment Fractional Bayes Factor Stochastic Search with Global Prior for Gaussian DAG Models*

Description

Estimate the edge inclusion probabilities for a Gaussian DAG with q nodes from observational data, using the moment fractional Bayes factor approach with global prior.

Usage

```
FBF_GS(Corr, nobs, G_base, h, C, n_tot_mod, n_hpp)
```

Arguments

Corr	qxq correlation matrix.
nobs	Number of observations.
G_base	Base DAG.
h	Parameter prior.
C	Costant who keeps the probability of all local moves bounded away from 0 and 1.
n_tot_mod	Maximum number of different models which will be visited by the algorithm, for each equation.
n_hpp	Number of the highest posterior probability models which will be returned by the procedure.

Value

An object of `class` `list` with:

- `M_q` Matrix (qxq) with the estimated edge inclusion probabilities.
- `M_G` Matrix (n*n_hpp)xq with the n_hpp highest posterior probability models returned by the procedure.
- `M_P` Vector (n_hpp) with the n_hpp posterior probabilities of the models in M_G.

Author(s)

Davide Altomare (`<davide.altomare@gmail.com>`).

References

D. Altomare, G. Consonni and L. La Rocca (2013). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Biometrics*.

Examples

```
data(SimDag6)

Corr=dataSim6$SimCorr[[1]]
nobs=50
q=ncol(Corr)
Gt=dataSim6$TDag

Res_search=FBF_GS(Corr, nobs, matrix(0,q,q), 1, 0.01, 1000, 10)
M_q=Res_search$M_q
M_G=Res_search$M_G
M_P=Res_search$M_P

G_med=M_q
G_med[M_q>=0.5]=1
G_med[M_q<0.5]=0 #median probability DAG
```

```
G_high=M_G[1:q,1:q] #Highest Posterior Probability DAG (HPP)
pp_high=M_P[1] #Posterior Probability of the HPP

sum(sum(abs(G_med-Gt))) #Structural Hamming Distance between the true DAG and the median probability DAG
sum(sum(abs(G_high-Gt))) #Structural Hamming Distance between the true DAG and the highest probability DAG
```

FBF_LS *Moment Fractional Bayes Factor Stochastic Search with Local Prior for DAG Models*

Description

Estimate the edge inclusion probabilities for a directed acyclic graph (DAG) from observational data, using the moment fractional Bayes factor approach with local prior.

Usage

```
FBF_LS(Corr, nobs, G_base, h, C, n_tot_mod)
```

Arguments

Corr	qxq correlation matrix.
nobs	Number of observations.
G_base	Base DAG.
h	Parameter prior.
C	Costant who keeps the probability of all local moves bounded away from 0 and 1.
n_tot_mod	Maximum number of different models which will be visited by the algorithm, for each equation.

Value

An object of `class` `matrix` with the estimated edge inclusion probabilities.

Author(s)

Davide Altomare (<davide.altomare@gmail.com>).

References

D. Altomare, G. Consonni and L. La Rocca (2013). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Biometrics*.

Examples

```
data(SimDag6)

Corr=dataSim6$SimCorr[[1]]
nobs=50
q=ncol(Corr)
Gt=dataSim6$TDag

M_q=FBF_LS(Corr, nobs, matrix(0,q,q), 0, 0.01, 1000)

G_med=M_q
G_med[M_q>=0.5]=1
G_med[M_q<0.5]=0 #median probability DAG

sum(sum(abs(G_med-Gt))) #Structural Hamming Distance between the true DAG and the median probability DAG
```

FBF_RS *Moment Fractional Bayes Factor Stochastic Search for Regression Models*

Description

Estimate the edge inclusion probabilities for a regression model, using the moment fractional Bayes factor approach.

Usage

```
FBF_RS(Corr, nobs, G_base, h, C, n_tot_mod, n_hpp)
```

Arguments

Corr	qxq correlation matrix.
nobs	Number of observations.
G_base	Base model.
h	Parameter prior.
C	Costant who keeps the probability of all local moves bounded away from 0 and 1.
n_tot_mod	Maximum number of different models which will be visited by the algorithm, for each equation.
n_hpp	Number of the highest posterior probability models which will be returned by the procedure.

Value

An object of class list with:

- M_q Matrix (qxq) with the estimated edge inclusion probabilities.
- M_G Matrix (n*n_hpp)xq with the n_hpp highest posterior probability models returned by the procedure.
- M_P Vector (n_hpp) with the n_hpp posterior probabilities of the models in M_G.

Author(s)

Davide Altomare (<davide.altomare@gmail.com>).

References

D. Altomare, G. Consonni and L. La Rocca (2013). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Biometrics*.

Examples

```
data(SimDag6)

Corr=dataSim6$SimCorr[[1]]
nobs=50
q=ncol(Corr)
Gt=dataSim6$TDag

# Regression of Y(q) on Y(q-1),...,Y(1)

Res_search=FBF_RS(Corr, nobs, matrix(0,1,(q-1)), 1, 0.01, 1000, 10)
M_q=Res_search$M_q
M_G=Res_search$M_G
M_P=Res_search$M_P


Mt=rev(matrix(Gt[1:(q-1),q],1,(q-1))) #True Model

M_med=M_q
M_med[M_q>=0.5]=1
M_med[M_q<0.5]=0 #median probability model

sum(sum(abs(M_med-Mt))) #Structural Hamming Distance between the true DAG and the median probability DAG
```

Index