Package ‘FBFsearch’

November 23, 2016

Type Package

Title Algorithm for Searching the Space of Gaussian Directed Acyclic Graph Models Through Moment Fractional Bayes Factors

Version 1.1

Date 2016-11-21

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 graph (DAG) models. The algorithm proposed makes use of moment fractional Bayes factors (MFBF) and thus it is suitable for learning sparse graph. The algorithm is implemented by using Armadillo: an open-source C++ linear algebra library.

License GPL (>= 2)

Imports Rcpp (>= 0.12.7)

LinkingTo Rcpp, RcppArmadillo

NeedsCompilation yes

Repository CRAN

Date/Publication 2016-11-23 01:29:18

R topics documented:

dataHuman .................................................. 2
dataPub ..................................................... 2
dataSim100 .................................................. 3
dataSim200 .................................................. 4
dataSim50 .................................................... 4
dataSim6 ..................................................... 5
dataSimHuman .............................................. 6
FBF_GS ..................................................... 6
FBF_LS ...................................................... 8
FBF_RS ...................................................... 9

Index 12
**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**

```r
data(HumanPw)
```

**Format**

- `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**


**References**

D. Altomare, G. Consonni and L. La Rocca (2012). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Article submitted to Biometric Methodology*.


---

**dataPub**  
*Publishing productivity data*

**Description**

Data on publishing productivity among academics.

**Usage**

```r
data(PubProd)
```
**dataSim100**

**Format**

- **dataPub**: contains the following objects:
  - **Corr**: Matrix (7x7) with the correlation matrix of the variables.
  - **nobs**: Scalar with the number of observations.

**Source**


**References**


D. Altomare, G. Consonni and L. La Rocca (2012). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Article submitted to Biometric Methodology*.

**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(dataSim100)`

**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 (2012). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Article submitted to Biometric Methodology*.

**References**

dataSim200

**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 (2012). Objective Bayesian search of gaussian directed acyclic graphical models for ordered variables with non-local priors. *Article submitted to Biometric Methodology*.

**References**


dataSim50

**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)
**dataSim6**

**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 (2012). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Article submitted to Biometric Methodology*.

**References**


---

**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(dataSim6)

**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 (2012). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Article submitted to Biometric Methodology*. 
**DataSimHuman**

*Simulated cell signalling pathway data*

**Description**

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

**Usage**

```r
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 (2012). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Article submitted to Biometric Methodology*.

**References**


---

**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**

```r
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`: Constant 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 (2012). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Article submitted to Biometric Methodology*.

Examples

```r
## Not run:
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
```
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, nob, G_base, h, C, n_tot_mod)

Arguments
Corr  
qxq correlation matrix.
nob  
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 (2012). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Article submitted to Biometric Methodology.*

Examples

```r
## Not run:
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

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

## End(Not run)
```

---

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

**Description**

Estimate the edge inclusion probabilities for a regression model (Y(q) on Y(q-1),...,Y(1)) with q variables from observational data, using the moment fractional Bayes factor approach.

**Usage**

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

**Arguments**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>Corr</code></td>
<td>qxq correlation matrix.</td>
</tr>
<tr>
<td><code>nobs</code></td>
<td>Number of observations.</td>
</tr>
<tr>
<td><code>G_base</code></td>
<td>Base model.</td>
</tr>
<tr>
<td><code>h</code></td>
<td>Parameter prior.</td>
</tr>
</tbody>
</table>
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 (q*q) with the estimated edge inclusion probabilities.
M_G Matrix (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. LaRocca (2012). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. Article submitted to Biometric Methodology.

Examples

```r
## Not run:

data(SimDag6)

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

Res_search=FBF_RS(Corr, nobs, matrix(0,1,(q-1)), 1, 0, 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
```
#Structural Hamming Distance between the true DAG and the median probability DAG

sum(sum(abs(M_med-Mt)))

## End(Not run)
Index

*Topic **dag**
  FBF_GS, 6
  FBF_LS, 8
  FBF_RS, 9

*Topic **datasets**
  dataHuman, 2
  dataPub, 2
  dataSim100, 3
  dataSim200, 4
  dataSim50, 4
  dataSim6, 5
  dataSimHuman, 6

*Topic **models**
  FBF_GS, 6
  FBF_LS, 8
  FBF_RS, 9

*Topic **multivariate**
  FBF_GS, 6
  FBF_LS, 8
  FBF_RS, 9

*Topic **stochastic search**
  FBF_GS, 6
  FBF_LS, 8
  FBF_RS, 9

  dataHuman, 2
  dataPub, 2
  dataSim100, 3
  dataSim200, 4
  dataSim50, 4
  dataSim6, 5
  dataSimHuman, 6

  FBF_GS, 6
  FBF_LS, 8
  FBF_RS, 9