Package 'MvBinary'

January 20, 2025

Type Package

Title Modelling Multivariate Binary Data with Blocks of Specific
One-Factor Distribution

Version 1.1

Date 2016-12-15

Author Matthieu Marbac and Mohammed Sedki

Maintainer Mohammed Sedki <mohammed.sedki@u-psud.fr>

Description Modelling Multivariate Binary Data with Blocks of Specific One-Factor Distribu-
tion. Variables are grouped into independent blocks. Each variable is described by two continu-
ous parameters (its marginal probability and its dependency strength with the other block vari-
ables), and one binary parameter (positive or negative dependency). Model selection con-
sists in the estimation of the repartition of the variables into blocks. It is carried out by the maxi-
mization of the BIC criterion by a deterministic (faster) algorithm or by a stochas-
tic (more time consuming but optimal) algorithm. Tool functions facilitate the model interpreta-
tion.

License GPL (>= 2)

Imports methods, mgcv, parallel

Depends R (>= 3.0.2)

Repository CRAN

Repository/R-Forge/Project mvbinary

Repository/R-Forge/Revision 31

Repository/R-Forge/DateTimeStamp 2016-12-14 14:16:05

Date/Publication 2016-12-15 16:46:38

NeedsCompilation no

Contents

Index   **8**

MvBinary-package   *MvBinary a package for Multivariate Binary data*

Description

MvBinary is a tool for fitting the distribution of correlated multivariate binary data.

Details

Package:	MvBinary
Type:	Package
Version:	1.0.0
Date:	2015-11-03
License:	GPL-2
LazyLoad:	yes

Author(s)

Author: Marbac M., and Sedki S.

References

Matthieu Marbac, Mohammed Sedki (2015). A Family of Blockwise One-Factor Distributions for Modelling High-Dimensional Binary Data. arXiv:1511.01343

Examples

```
# Package loading
rm(list=ls())
require(MvBinary)

# Data loading
data(MvBinaryExample)

# Parameter estimation by the HAC-based algorithm on 2 cores
# where the EM algorithms are initialized 10 times
```

```
res.CAH <- MvBinaryEstim(MvBinaryExample, 2, nbinit.EM = 10)

# Summary of the estimated model
summary(res.CAH)

# Print the parameters of the estimated model
print(res.CAH)
```

ComputeEmpiricCramer   *Computation of the Empiric Cramer'v.*

Description

This function computes the Empiric Cramer's V for a binary data set.

Usage

```
ComputeEmpiricCramer(x)
```

Arguments

x   a binary matrix.

Value

Return the matrix of the Empiric Cramer's V.

ComputeMvBinaryCramer   *Computation of the model Cramer'v.*

Description

This function computes the model Cramer's V for a binary data set.

Usage

```
ComputeMvBinaryCramer(results)
```

Arguments

results   an instance of S4 class MvBinaryResult (provided by the function MvBinaryEstim)

Value

Return the matrix of the Empiric Cramer's V.

MvBinaryEstim   *Create an instance of the [MvBinaryResult] class*

Description

This function performs the model selection and the parameter inference.

Usage

```
MvBinaryEstim(x, nbcores = 1, algorithm = "HAC", modelslist = NULL,
  tol.EM = 0.01, nbinit.EM = 40, nbiter.MH = 50, nbchains.MH = 10)
```

Arguments

x	matrix of the binary observation.
nbcores	number of cores used for the model selection (only for Linux). Default is 1.
algorithm	algorithm used for the model selection ("HAC": deterministic algorithm based on the HAC of the variables, "MH": stochastic algorithm for optimizing the BIC criterion, "List": comparison of the models provided by the users). Default is "HAC".
modelslist	list of models provided by the user (only used when algorithm="List"). Default is NULL
tol.EM	stopping criterion for the EM algorithm. Default is 0.01
nbinit.EM	number of random initializations for the EM algorithm. Default is 40.
nbiter.MH	number of successive iterations without finding a model having a better BIC criterion which involves the stopping of the Metropolis-Hastings algorithm (only used when algorithm="MH"). Default is 50.
nbchains.MH	number of radom initializations for the stochastic algorithm (only used when algorithm="MH"). Default is 10.

Value

Returns an instance of the [MvBinaryResult] class.

Examples

```
# Data loading
data(MvBinaryExample)

# Parameter estimation by the HAC-based algorithm on 2 cores
# where the EM algorithms are initialized 10 times
res.CAH <- MvBinaryEstim(MvBinaryExample, 2, nbinit.EM = 10)

# Parameter estimation for two competing models
res.CAH <- MvBinaryEstim(MvBinaryExample, algorithm="List",
 modelslist=list(c(1,1,2,2,3,4), c(1,1,1,2,2,2)), nbinit.EM = 10)
```

```
# Summary of the estimated model
summary(res.CAH)

# Print the parameters of the estimated model
print(res.CAH)
```

MvBinaryExample   *Simulated binary data: MvBinaryExample*

Description

The file MvBinaryExample.rda describes 400 individuals by 6 binary variables.

Format

A matrix with 400 observations on the 6 variables.

Details

This data set has been simulated from the MvBinary model. The first three variables are dependent. The last three variables are dependent.

Examples

```
data(MvBinaryExample)
```

MvBinaryProbaPost   *Computation of the model Cramer'v.*

Description

This function computes the model Cramer's V for a binary data set.

Usage

```
MvBinaryProbaPost(x, param)
```

Arguments

x	a binary matrix.
param	an instance of S4 class MvBinaryResult (provided by the function MvBinaryEstim)

Value

Return the logprobability for each row of matrix x conditionally on the model defined by param.

MvBinaryResult-class   *Constructor of [MvBinaryResult] class*

Description

This S4 class contains the results from the function MvBinaryEstim.

Details

alpha numeric. It indicates the marginal probability of that each variables are equal to 1.

epsilon numeric. It indicates the dependency strength of each variables (between 0 and 1) with the other block variables.

delta binary. It indicates the kind of dependency: two variables affiliated into the same block are positively correlated if their delta are equal and they are negatively correlated otherwise.

blocks numeric. It indicates the block of each variable.

nbparam numeric. It indicates the number of continuous parameters.

loglike numeric. The model likelihood.

bic numeric. The model BIC.

Examples

```
getSlots("MvBinaryResult")
```

---

plants   *Real binary data: Plants*

Description

The file plants.rda describes 35583 plants by indicating if they occur (1) or not (2) in 69 states of the Norht America.

Format

A matrix with 35583 observations on the 69 variables.

Details

This data set been extracted from the USA plants database, July 29, 2015.

Examples

```
data(plants)
```

print	*Summary function.*

Description

This function prints the parameters resulting from MvBinaryEstim.

Usage

```
## S4 method for signature 'MvBinaryResult'
print(x)
```

Arguments

x   output object from [MvBinaryEstim](#).

summary	*Summary function.*

Description

This function gives the summary of output from MvBinaryEstim.

Usage

```
## S4 method for signature 'MvBinaryResult'
summary(object)
```

Arguments

object   output object from [MvBinaryEstim](#).

Index