

Package ‘mgm’

June 20, 2017

Type Package

Title Estimating Time-Varying k-Order Mixed Graphical Models

Version 1.2-1

Date 2017-06-20

Author Jonas Haslbeck

Maintainer Jonas Haslbeck <jonashaslbeck@gmail.com>

Description Estimation of k-Order time-varying Mixed Graphical Models and mixed VAR(p) models via elastic-net regularized neighborhood regression.

URL <http://arxiv.org/abs/1510.06871v2>

BugReports <https://github.com/jmbh/mgm/issues>

License GPL (>= 2)

Imports matrixcalc, glmnet, stringr, Hmisc

LazyData true

NeedsCompilation no

Repository CRAN

Date/Publication 2017-06-20 08:49:52 UTC

R topics documented:

mgm-package	2
bwSelect	3
datasets	6
mgm	7
mgmsampler	14
mvar	19
mvarsampler	25
predict.mgm	27
print.mgm	30
tvmgm	31

tvmgmsampler	35
tvmvar	39
tvmvarsampler	43

Index	47
--------------	-----------

mgm-package	<i>Estimating Time-Varying k-order Mixed Graphical Models</i>
-------------	---------------------------------------------------------------

Description

Estimation of time-varying Mixed Graphical models and mixed VAR models via elastic-net regularized neighborhood regression.

Details

Package: mgm
 Type: Package
 Version: 1.2-1
 Date: 2017-06-20
 License: GPL-2

Author(s)

Jonas Haslbeck

Maintainer: <jonashaslbeck@gmail.com>

References

Haslbeck, J., & Waldorp, L. J. (2016). mgm: Structure estimation for time-varying mixed graphical models in high-dimensional data. arXiv preprint <http://arxiv.org/abs/1510.06871v2>.

Haslbeck, J., & Waldorp, L. J. (2015). Structure estimation for mixed graphical models in high-dimensional data. arXiv preprint arXiv:1510.05677.

Loh, P. L., & Wainwright, M. J. (2013). Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. *The Annals of Statistics*, 41(6), 3022-3049.

Yang, E., Baker, Y., Ravikumar, P., Allen, G., & Liu, Z. (2014). Mixed graphical models via exponential families. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics* (pp. 1042-1050).

aggregation in the output list (see below). Note that continuous variables are scaled (centered and divided by their standard deviation), hence the absolute error and 0/1-loss are roughly on the scale scale.

Note that selecting the bandwidth with the EBIC is no alternative. This is because the EBIC always selects the intercept model with the lowest bandwidth. The reason is that the unregularized intercept closely models the noise in the data and hence the penalty sets all other parameters to zero. This problem is solved by using out of sample prediction error in the cross validation scheme.

Value

The function returns a list with the following entries:

call	Contains all provided input arguments. If <code>saveData = TRUE</code> , it also contains the data.
bwModels	Contains the models estimated at the time points in the tests set. For details see <code>tvmvar</code> or <code>tvmgm</code> .
fullErrorFolds	List with number of entries equal to the length of <code>bwSeq</code> entries. Each entry contains a list with <code>bwFolds</code> entries. Each of those entries contains a <code>bwFoldsize</code> times <code>p</code> matrix of out of sample prediction errors.
fullError	The same as <code>fullErrorFolds</code> but pooled over folds.
meanError	List with number of entries equal to the length of <code>bwSeq</code> entries. Each entry contains the average prediction error over variables and time points in the test set.
testsets	List with <code>bwFolds</code> entries, which contain the rows of the test sample for each fold.
zeroweights	List with <code>bwFolds</code> entries, which contains the observation weights used to fit the model at the <code>bwFoldsize</code> time points.

Author(s)

Jonas Haslbeck <jonashaslbeck@gmail.com>

References

- Foygel, R., & Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. In *Advances in neural information processing systems* (pp. 604-612).
- Barber, R. F., & Drton, M. (2015). High-dimensional Ising model selection with Bayesian information criteria. *Electronic Journal of Statistics*, 9(1), 567-607.
- Haslbeck, J., & Waldorp, L. J. (2016). `mgm`: Structure Estimation for time-varying Mixed Graphical Models in high-dimensional Data. arXiv preprint arXiv:1510.06871.

Examples

```
## Not run:
```

```
## A) bwSelect for tvmgm()

# A.1) Generate noise data set
p <- 5
n <- 100
data_n <- matrix(rnorm(p*n), nrow=100)
head(data_n)

type <- c('c', 'c', rep('g', 3))
level <- c(2, 2, 1, 1, 1)
x1 <- data_n[,1]
x2 <- data_n[,2]
data_n[x1>0,1] <- 1
data_n[x1<0,1] <- 0
data_n[x2>0,2] <- 1
data_n[x2<0,2] <- 0

head(data_n)

# A.2) Estimate optimal bandwidth parameter

bwobj_mgm <- bwSelect(data = data_n,
                      type = type,
                      level = level,
                      bwSeq = seq(0.05, 1, length=3),
                      bwFolds = 1,
                      bwFoldsize = 3,
                      modeltype = 'mgm',
                      k = 3,
                      pbar = TRUE,
                      overparameterize = TRUE)

print.mgm(bwobj_mgm)

## B) bwSelect for tvnVar()

# B.1) Generate noise data set

p <- 5
n <- 100
data_n <- matrix(rnorm(p*n), nrow=100)
head(data_n)

type <- c('c', 'c', rep('g', 3))
level <- c(2, 2, 1, 1, 1)
x1 <- data_n[,1]
x2 <- data_n[,2]
data_n[x1>0,1] <- 1
data_n[x1<0,1] <- 0
data_n[x2>0,2] <- 1
```

```
data_n[x2<0,2] <- 0

head(data_n)

# B.2) Estimate optimal bandwidth parameter

bwobj_mvar <- bwSelect(data = data_n,
                      type = type,
                      level = level,
                      bwSeq = seq(0.05, 1, length=3),
                      bwFolds = 1,
                      bwFoldsize = 3,
                      modeltype = 'mvar',
                      lags = 1:3,
                      pbar = TRUE,
                      overparameterize = TRUE)

print.mgm(bwobj_mvar)

## End(Not run)
```

datasets

Example Datasets in the mgm Package

Description

The autism dataset (and its short version) are taken from Deserno et al. (2016).

The restingstate fMRI data are taken from Schmittmann et al. (2015).

The gene expression data across the life span of the fruit fly are taken from Gibberd & Nelson (2017), who took a subset of the data first presented by Arbeitman et al. (2002).

The symptom data of the single individual diagnosed with major depression is described in Kosakowski et al. (2017).

The PTSD data is taken from McNally et al. (2015).

The dataset `mgm_data` is generated by example code shown in `?mgmsampler`, and `mvar_data` is generated by example code shown in `?mvarsampler`.

All datasets are loaded automatically. All real data sets come as a list including the data and additional information (names of variables, types of variables, time stamps for time series data, etc.)

References

- Deserno, M. K., Borsboom, D., Begeer, S., & Geurts, H. M. (2016). Multicausal systems ask for multicausal approaches: A network perspective on subjective well-being in individuals with autism spectrum disorder. *Autism*.
- Schmittmann, V. D., Jahfari, S., Borsboom, D., Savi, A. O., & Waldorp, L. J. (2015). Making large-scale networks from fMRI data. *PloS one*, 10(9), e0129074.
- Gibberd, A. J., & Nelson, J. D. (2017). Regularized Estimation of Piecewise Constant Gaussian Graphical Models: The Group-Fused Graphical Lasso. *Journal of Computational and Graphical Statistics*, (just-accepted).
- Arbeitman, M. N., Furlong, E. E., Imam, F., Johnson, E., Null, B. H., Baker, B. S., ... & White, K. P. (2002). Gene expression during the life cycle of *Drosophila melanogaster*. *Science*, 297(5590), 2270-2275.
- Kossakowski, J., Groot, P., Haslbeck, J., Borsboom, D., & Whichers, M. (2017). Data from "Critical Slowing Down as a Personalized Early Warning Signal for Depression". *Journal of Open Psychology Data*, 5(1).
- McNally, R. J., Robinaugh, D. J., Wu, G. W., Wang, L., Deserno, M. K., & Borsboom, D. (2015). Mental disorders as causal systems a network approach to posttraumatic stress disorder. *Clinical Psychological Science*, 3(6), 836-849.

 mgm

Estimating Mixed Graphical Models

Description

Function to estimate k-degree Mixed Graphical Models via neighborhood regression.

Usage

```
mgm(data, type, level, lambdaSeq, lambdaSel, lambdaFolds,
     lambdaGam, alphaSeq, alphaSel, alphaFolds, alphaGam, k,
     ruleReg, weights, threshold, method, binarySign, scale,
     verbatim, pbar, warnings, saveModels, saveData,
     overparameterize, signInfo, ...)
```

Arguments

data	n x p data matrix
type	p vector indicating the type of variable for each column in data. 'g' for Gaussian, 'p' for Poisson, 'c' for categorical.
level	p vector indicating the number of categories of each variable. For continuous variables set to 1.
lambdaSeq	A sequence of lambdas that should be searched (see also lambdaSel). Defaults to NULL, which lets glmnet select a reasonable lambda sequence (recommended).

<code>lambdaSel</code>	Procedure to select the tuning parameter controlling the Lq-penalization. The two options are cross validation 'CV' and the Extended Bayesian Information Criterion (EBIC) 'EBIC'. The EBIC performs well in selecting sparse graphs (see Barber and Drton, 2010 and Foygel and Drton, 2014). Note that when also searching the alpha parameter in the elastic net penalty, cross validation should be preferred, as the parameter vector will not necessarily be sparse anymore. The EBIC tends to be a bit more conservative than CV (see Haslbeck and Waldorp, 2016). CV can sometimes not be performed with categorical variables, because <code>glmnet</code> requires at least 2 events of each category of each categorical variable in each training-fold. Defaults to <code>lambdaSel = 'CV'</code> .
<code>lambdaFolds</code>	Number of folds in cross validation if <code>lambdaSel = 'CV'</code>
<code>lambdaGam</code>	Hyperparameter gamma in the EBIC if <code>lambdaSel = 'EBIC'</code> . Defaults to <code>lambdaGam = .25</code> .
<code>alphaSeq</code>	A sequence of alpha parameters for the elastic net penalty in [0,1] that should be searched (see also <code>alphaSel</code>). Defaults to <code>alphaSeq = 1</code> , which means that the lasso is being used. <code>alphaSeq = 0</code> corresponds to an L2-penalty (Ridge regression). For details see Friedman, Hastie and Tibshirani (2010).
<code>alphaSel</code>	Procedure to select the alpha parameter in the elastic net penalty. The two options are cross validation 'CV' and the Extended Bayesian Information Criterion (EBIC) 'EBIC'. The EBIC performs well in selecting sparse graphs (see Barber and Drton, 2010 and Foygel and Drton, 2014). Note that when also searching the alpha parameter in the elastic net penalty, cross validation should be preferred, as the parameter vector will not necessarily be sparse anymore. The EBIC tends to be a bit more conservative than CV (see Haslbeck and Waldorp, 2016). CV can sometimes not be performed with categorical variables, because <code>glmnet</code> requires at least 2 events of each category of each categorical variable in each training-fold. Defaults to <code>alphaSel = 'CV'</code> .
<code>alphaFolds</code>	Number of folds in cross validation if <code>alphaSel = 'CV'</code>
<code>alphaGam</code>	Hyperparameter gamma in the EBIC if <code>alphaSel = 'EBIC'</code> . Defaults to <code>alphaGam = .25</code> .
<code>k</code>	Order up until including which interactions are included in the model. <code>k = 2</code> means that all pairwise interactions are included, <code>k = 3</code> means that all pairwise and all three-way interactions are included, etc. In previous versions of <code>mgm</code> the order of interactions was specified by the parameter <code>d</code> , the largest size or a neighborhood. Note that <code>k = d + 1</code> .
<code>ruleReg</code>	Rule to combine estimates from neighborhood regression. E.g. for pairwise interactions, two estimates (one from regressing A on B and one from B on A) have to be combined in one edge parameter. <code>ruleReg = 'AND'</code> requires all estimates to be nonzero in order to set the edge to be present. <code>ruleReg = 'OR'</code> requires at least one estimate to be nonzero in order to set the edge to be present. For higher order interactions, <code>k</code> estimates have to be combined with this rule.
<code>weights</code>	A <code>n</code> vector with weights for observations.
<code>threshold</code>	A threshold below which edge-weights are put to zero. This is done in order to guarantee a constant false-positive rate. <code>threshold = 'LW'</code> refers to the threshold in Loh and Wainwright (2013), which was used in all previous versions of <code>mgm</code> . <code>threshold = 'HW'</code> refers to the threshold in Haslbeck and

	Waldorp (2016). If <code>threshold = 'none'</code> no thresholding is applied. Defaults to <code>threshold = 'LW'</code> .
<code>method</code>	Estimation method, currently only <code>method = 'glm'</code> .
<code>binarySign</code>	If <code>binarySign = TRUE</code> a sign for the interaction within binary nodes and between binary and continuous nodes is provided in the output. Note that in this case the two categories of the binary variables have to be coded in 0,1. This is to ensure that the interpretation of the sign is unambiguous: a positive sign of a parameter means that increasing the associated predictor results in a higher probability for category 1.
<code>scale</code>	If <code>scale = TRUE</code> , all Gaussian nodes (specified by 'g' in type) are centered and divided by their standard deviation. Scaling is recommended, because otherwise the penalization of a parameter depends on the variance of the associated predictor.
<code>verbatim</code>	If <code>verbatim = TRUE</code> no warnings and no progress bar is shown. Defaults to <code>verbatim = FALSE</code> .
<code>pbar</code>	If <code>pbar = TRUE</code> a progress bar is shown. Defaults to <code>pbar = TRUE</code> .
<code>warnings</code>	If <code>warnings = TRUE</code> no warnings are returned. Defaults to <code>warnings = FALSE</code> .
<code>saveModels</code>	If <code>saveModels = FALSE</code> , only information about the weighted adjacency matrix, and if $k > 2$ about the factor graph is provided in the output list. If <code>saveModels = TRUE</code> , all fitted parameters are additionally returned.
<code>saveData</code>	If <code>saveData = TRUE</code> , the data is saved in the output list. Defaults to <code>saveData = FALSE</code> .
<code>overparameterize</code>	If <code>overparameterize = TRUE</code> , <code>mgm()</code> uses over-parameterized design-matrices for each neighborhood regression; this means that an interaction between two categorical variables with m and s categories is parameterized by $m*s$ parameters. If <code>overparameterize = FALSE</code> the standard parameterization (in <code>glmnet</code>) with $m*(s-1)$ parameters is used, where the first category of the predicting variable serves as reference category. If all variables are continuous both parameterizations are the same. Note that the default is set to <code>overparameterize = FALSE</code> , to be consistent with the previous <code>mgm</code> versions. However note that in order to recover higher order interactions categorical variables the overparameterized parameterization is often advantageous. See the examples below for an illustration. Note that we can estimate the model despite the colinear columns in the design matrix because we use penalized regression.
<code>signInfo</code>	If <code>signInfo = TRUE</code> , a message is shown in the console, indicating that the sign of estimates is stored separately. Defaults to <code>signInfo = TRUE</code> .
<code>...</code>	Additional arguments.

Details

`mgm()` estimates an exponential mixed graphical model as introduced in Yang and colleagues (2014). Haslbeck and Waldorp (2016) show that the estimation procedure is consistent.

Value

The function returns a list with the following entries:

<code>call</code>	Contains all provided input arguments. If <code>saveData = TRUE</code> , it also contains the data
<code>pairwise</code>	Contains a list with all information about estimated pairwise interactions. <code>wadj</code> contains the $p \times p$ weighted adjacency matrix, if p is the number of variables in the network. <code>signs</code> has the same dimensions as <code>wadj</code> and contains the signs for the entries of <code>wadj</code> : 1 indicates a positive sign, -1 a negative sign and 0 an undefined sign. A sign is undefined if an edge is a function of more than one parameter. This is the case for interactions involving a categorical variable with more than 2 categories. <code>edgecolor</code> also has the same dimensions as <code>wadj</code> contains a color for each edge, depending on <code>signs</code> . It is provided for more convenient plotting. If only pairwise interactions are modeled ($d = 1$), <code>wadj</code> contains all conditional independence relations.
<code>factorgraph</code>	Contains a list with all information about all estimated interactions (factors). <code>graph</code> contains a weighted adjacency matrix of a (bipartite) factor graph. If p is the number of variables and E the number of interactions (factors) in the model, this matrix has dimensions $(p+E) \times (p+E)$. The factor graph is further specified by the following objects: <code>signs</code> is a matrix of the same dimensions as <code>graph</code> that indicates the sign of each interaction, if defined (see <code>pairwise</code> above). <code>edgecolor</code> is a matrix with the same dimension as <code>graph</code> that provides edge colors depending on the sign as above. <code>order</code> is a $(p+E)$ vector indicating the order of interaction. The first p entries are set to zero. Note that <code>factorgraph</code> contains the same information as <code>pairwise</code> if only pairwise interactions are estimated.
<code>rawfactor</code>	A list with three entries that relate each interaction in the model to all its parameters. This is different to the output provided in <code>factorgraph</code> , where one value is assigned to each interaction. <code>indicator</code> contains a list with $k-1$ entries, one for each order of modeled interaction, which contain the estimated (nonzero) interactions. <code>weightsAgg</code> contains a list with $k-1$ entries, which in turn contain R lists, where R is the number of interactions (and rows in the corresponding list entry in <code>indicator</code>) that were estimated (nonzero) in the given entry. Each of these entries contains the mean of the absolute values of all parameters involved in this interaction. <code>weights</code> has the same structure as <code>weightsAgg</code> , but does contain all parameters involved in the interaction instead of the mean of their absolute values. <code>signs</code> has the same structure as <code>weightsAgg/weights</code> and provides the sign of the interaction, if defined.
<code>intercepts</code>	A list with p entries, which contain the intercept/thresholds for each node in the network. In case a given node is categorical with m categories, there are m thresholds for this variable.
<code>nodemodels</code>	A list with p <code>glmnet()</code> models, from which all above output is computed. Also contains the coefficients <code>models</code> for the selected λ and the applied τ threshold τ .

Author(s)

Jonas Haslbeck <jonashaslbeck@gmail.com>

References

- Foygel, R., & Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. In *Advances in neural information processing systems* (pp. 604-612).
- Barber, R. F., & Drton, M. (2015). High-dimensional Ising model selection with Bayesian information criteria. *Electronic Journal of Statistics*, 9(1), 567-607.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1), 1.
- Haslbeck, J., & Waldorp, L. J. (2015). Structure estimation for mixed graphical models in high-dimensional data. arXiv preprint arXiv:1510.05677.
- Haslbeck, J., & Waldorp, L. J. (2016). mgm: Structure Estimation for time-varying Mixed Graphical Models in high-dimensional Data. arXiv preprint arXiv:1510.06871.
- Loh, P. L., & Wainwright, M. J. (2012, December). Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. In *NIPS* (pp. 2096-2104).
- Yang, E., Baker, Y., Ravikumar, P., Allen, G. I., & Liu, Z. (2014, April). Mixed Graphical Models via Exponential Families. In *AISTATS* (Vol. 2012, pp. 1042-1050).

Examples

```
## Not run:

## We fit a pairwise and 3-order MGM to the mixed Autism dataset (?autism_data)

# 1) Fit Pairwise MGM

# Call mgm()
fit_d2 <- mgm(data = autism_data$data,
              type = autism_data$type,
              level = autism_data$lev,
              k = 2) # ad most pairwise interactions

# Weighted adjacency matrix
fit_d2$pairwise$wadj

# Visualize using qgraph()
library(qgraph)
qgraph(fit_d2$pairwise$wadj,
       edge.color = fit_d2$pairwise$edgecolor,
       layout = 'spring',
       labels = autism_data$colnames)

# 2) Fit MGM with pairwise & threeway interactions
fit_d3 <- mgm(data = autism_data$data,
              type = autism_data$type,
              level = autism_data$lev,
              k = 3) # include all interactions up to including order 3
```

```

# List of estimated interactions
fit_d3$rawfactor$indicator

# Compute Label vector
labels <- c(autism_data$colnames, rep('', sum(fit_d3$factorgraph$nodetype)))

# Visualize factor graph using qgraph()
library(RColorBrewer)
nodeColors <- brewer.pal(3, 'Set1')

qgraph(fit_d3$factorgraph$graph,
       color = nodeColors[fit_d3$factorgraph$order+1],
       edge.color = fit_d3$factorgraph$edgecolor,
       layout = 'spring',
       labels = labels,
       shape = c('circle', 'square')[fit_d3$factorgraph$nodetype+1],
       vsize = c(8, 4)[fit_d3$factorgraph$nodetype+1])

# 3) Predict values

pred_obj <- predict(fit_d3, autism_data$data)

head(pred_obj$predicted) # predicted values

pred_obj$errors # Nodewise errors

## Here we illustrate why we need to overparameterize the design matrix to
## recover higher order interactions including categorical variables

# 1) Define Graph (one 3-way interaction between 3 binary variables)

# a) General Graph Info
type = c('c', 'c', 'c')
level = c(2, 2, 2)
# b) Define Interaction
factors <- list()
factors[[1]] <- NULL # no pairwise interactions
factors[[2]] <- matrix(c(1,2,3), ncol=3, byrow = T) # one 3-way interaction
interactions <- list()
interactions[[1]] <- NULL
interactions[[2]] <- vector('list', length = 1)
# threeway interaction no1
interactions[[2]][[1]] <- array(0, dim = c(level[1], level[2], level[3]))
theta <- .7
interactions[[2]][[1]][1, 1, 1] <- theta #weight theta for conf (1,1,1), weight 0 for all others
# c) Define Thresholds
thresholds <- list()
thresholds[[1]] <- c(0, 0)
thresholds[[2]] <- c(0, 0)
thresholds[[3]] <- c(0, 0)

```

```
# 2) Sample from Graph

iter <- 1
set.seed(iter)
N <- 2000
d_iter <- mgmsampler(factors = factors,
                    interactions = interactions,
                    thresholds = thresholds,
                    type = type,
                    level = level,
                    N = N,
                    nIter = 50,
                    pbar = TRUE)

# 3.1) Estimate order 3 MGM usind standard parameterization

d_est_stand <- mgm(data = d_iter$data,
                  type = type,
                  level = level,
                  k = 3,
                  lambdaSel = 'CV',
                  ruleReg = 'AND',
                  pbar = TRUE,
                  overparameterize = FALSE,
                  signInfo = FALSE)

# We look at the nodewise regression for node 1 (same for all)
coefs_stand <- d_est_stand$nodemodels[[1]]$model
coefs_stand
# we see that nonzero-zero pattern of parameter vector does not allow us to infer whether
# interactions are present or not

# 3.2) Estimate order 3 MGM usind overparameterization

d_est_over <- mgm(data = d_iter$data,
                 type = type,
                 level = level,
                 k = 3,
                 lambdaSel = 'CV',
                 ruleReg = 'AND',
                 pbar = TRUE,
                 overparameterize = TRUE,
                 signInfo = FALSE)

# We look at the nodewise regression for node 1 (same for all)
coefs_over <- d_est_over$nodemodels[[1]]$model
coefs_over # recovers exactly the 3-way interaction

# For more examples see https://github.com/jmbh/mgmDocumentation
```

```
## End(Not run)
```

mgmsampler

Sample from k-order Mixed Graphical Model

Description

Generates samples from a k-order Mixed Graphical Model

Usage

```
mgmsampler(factors, interactions, thresholds, sds, type,
           level, N, nIter = 250, pbar = TRUE, divWarning = 10^3)
```

Arguments

- | | |
|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| factors | This object indicates which interactions are present in the model. It is a list in which the first entry corresponds to 2-way interactions, the second entry corresponds to 3-way interactions, etc. and the kth entry to the k+1-way interaction. Each entry contains a matrix with dimensions order x number of interaction of given order. Each row in the matrix indicates an interaction, e.g. (1, 3, 7, 9) in the matrix in list entry three indicates a 4-way interaction between the variables 1, 3, 7 and 9. |
| interactions | This object specifies the parameters associated to the interactions specified in factors. Corresponding to the structure in factors, this object is a list, where the kth entry corresponds to the k+1-way interactions. Each list entry contains another list, with entries equal to the number of rows in the corresponding matrix in factors. Each of these list entries (for a fixed k) contains a k-dimensional array that specifies the parameters of the given k-order interaction. For instance, if we have a 3-way interaction (1, 2, 3) and all variables are binary, we have a 2 x 2 x 2 array specifying the parameters for each of the 2 ³ = 8 possible configurations. If all variables are continuous, we have a 1 x 1 x 1 array, so the interaction is specified by one parameter only. See the examples below for an illustration. |
| thresholds | A list with p entries corresponding to p variables in the model. Each entry contains a vector indicating the threshold for each category (for categorical variables) or a numeric value indicating the threshold/intercept (for continuous variables). |
| sds | A numeric vector with p entries, specifying the variances of Gaussian variables. If variables 6 and 13 are Gaussians, then the corresponding entries of sds have to contain the corresponding variances. Other entries are ignored. |

type	p character vector indicating the type of variable for each column in data. 'g' for Gaussian, 'p' for Poisson, 'c' of each variable.
level	p integer vector indicating the number of categories of each variable. For continuous variables set to 1.
N	Number of samples that should be drawn from the distribution.
nIter	Number of iterations in the Gibbs sampler until a sample is drawn.
pbar	If pbar = TRUE a progress bar is shown. Defaults to pbar = TRUE.
divWarning	mgmsampler() returns a warning message if the absolute value of a continuous variable the chain of the gibbs sampler is larger than divWarning. To our best knowledge there is no theory yet defining a parameter space that ensures a proper probability density and hence a converging chain. Defaults to divWarning = 10^3 .

Details

We use a Gibbs sampler to sample from the joint distribution introduced by Yang and colleagues (2014). Note that the constraints on the parameter space necessary to ensure that the joint distribution is normalizable are to our best knowledge unknown. Yang and colleagues (2014) give these constraints for a number of simple pairwise models. In practice, an 'improper joint density' will lead to a sampling process that approaches infinity, and hence mgmsampler() will return Inf / -Inf values.

Value

A list containing:

call	Contains all provided input arguments.
data	The N x p data matrix of sampled values

Author(s)

Jonas Haslbeck <jonashaslbeck@gmail.com>

References

- Haslbeck, J., & Waldorp, L. J. (2016). mgm: Structure Estimation for time-varying Mixed Graphical Models in high-dimensional Data. arXiv preprint arXiv:1510.06871.
- Yang, E., Baker, Y., Ravikumar, P., Allen, G. I., & Liu, Z. (2014, April). Mixed Graphical Models via Exponential Families. In AISTATS (Vol. 2012, pp. 1042-1050).

Examples

```
## Not run:

# ----- Example 1: p = 10 dimensional Gaussian -----

# ----- 1) Specify Model -----
```

```

# a) General Graph Info
p <- 10 # number of variables
type = rep('g', p) # type of variables
level = rep(1, 10) # number of categories for each variable (1 = convention for continuous)

# b) Define interactions
factors <- list()
factors[[1]] <- matrix(c(1,2,
                        1,3,
                        4,5,
                        7,8), ncol=2, byrow = T) # 4 pairwise interactions

interactions <- list()
interactions[[1]] <- vector('list', length = 4)

# all pairwise interactions have value .5
for(i in 1:4) interactions[[1]][[i]] <- array(.5, dim=c(1, 1))

# c) Define Thresholds
thresholds <- vector('list', length = p)
thresholds <- lapply(thresholds, function(x) 0 ) # all means are zero

# d) Define Variances
sds <- rep(1, p) # All variances equal to 1

# ----- 2) Sample cases -----

data <- mgmsampler(factors = factors,
                  interactions = interactions,
                  thresholds = thresholds,
                  sds = sds,
                  type = type,
                  level = level,
                  N = 500,
                  nIter = 100,
                  pbar = FALSE)

# ----- 3) Recover model from sampled cases -----

set.seed(1)
mgm_obj <- mgm(data = data$data,
              type = type,
              level = level,
              d = 2,
              lambdaSel = 'CV',
              ruleReg = 'AND')

mgm_obj$rawfactor$indicator # worked!

```

```

# ----- Example 2: p = 3 Binary model with one 3-way interaction -----

# ----- 1) Specify Model -----

# a) General Graph Info
type = c('c', 'c', 'c')
level = c(2, 2, 2)

# b) Define Interaction
factors <- list()
factors[[1]] <- NULL # no pairwise interactions
factors[[2]] <- matrix(c(1,2,3), ncol=3, byrow = T) # one 3-way interaction

interactions <- list()
interactions[[1]] <- NULL
interactions[[2]] <- vector('list', length = 2)
# threeway interaction no1
interactions[[2]][[1]] <- array(0, dim = c(level[1], level[2], level[3]))
theta <- 2
interactions[[2]][[1]][1, 1, 1] <- theta # fill in nonzero entries
# thus: high probability for the case that x1 = x2 = x3 = 1

# c) Define Thresholds
thresholds <- list()
thresholds[[1]] <- rep(0, level[1])
thresholds[[2]] <- rep(0, level[2])
thresholds[[3]] <- rep(0, level[3])

# ----- 2) Sample cases -----

set.seed(1)
dlist <- mgmsampler(factors = factors,
                   interactions = interactions,
                   thresholds = thresholds,
                   type = type,
                   level = level,
                   N = 500,
                   nIter = 100,
                   pbar = TRUE)

# ----- 3) Check: Contingency Table -----

dat <- dlist$data
table(dat[,1], dat[,2], dat[,3]) # this is what we expected

# ----- 4) Recover model from sampled cases -----

mgm_obj <- mgm(data = dlist$data,
               type = type,
               level = level,

```

```

        d = 2,
        lambdaSel = 'CV')

mgm_obj$rawfactor$indicator
# Seems difficult to differentiate 2-way interactions for binary 3-way interaction

# ----- Example 3: p = 5 Mixed Graphical Model with two 3-way interaction -----

# ----- 1) Specify Model -----

# a) General Graph Info
type = c('g', 'c', 'c', 'g')
level = c(1, 3, 5, 1)
# b) Define Interaction
factors <- list()
factors[[1]] <- NULL # no pairwise interactions
factors[[2]] <- matrix(c(1,2,3,
                        2,3,4), ncol=3, byrow = T) # no pairwise interactions
interactions <- list()
interactions[[1]] <- NULL
interactions[[2]] <- vector('list', length = 2)
# 3-way interaction no1
interactions[[2]][[1]] <- array(0, dim = c(level[1], level[2], level[3]))
interactions[[2]][[1]][,1:3] <- rep(.8, 3) # fill in nonzero entries
# 3-way interaction no2
interactions[[2]][[2]] <- array(0, dim = c(level[2], level[3], level[4]))
interactions[[2]][[2]][1,1,] <- .3
interactions[[2]][[2]][2,2,] <- .3
interactions[[2]][[2]][3,3,] <- .3
# c) Define Thresholds
thresholds <- list()
thresholds[[1]] <- 0
thresholds[[2]] <- rep(0, level[2])
thresholds[[3]] <- rep(0, level[3])
thresholds[[4]] <- 0
# d) Define Variances
sds <- rep(.1, length(type))

# ----- 2) Sample cases -----

set.seed(1)
data <- mgmsampler(factors = factors,
                  interactions = interactions,
                  thresholds = thresholds,
                  sds = sds,
                  type = type,
                  level = level,
                  N = 500,
                  nIter = 100,
                  pbar = FALSE)

```

```

# ----- 3) Check: Conditional Means -----

# We condition on the categorical variables and check whether
# the conditional means match what we expect from the model:

dat <- data$data

# Check interaction 1
mean(dat[dat[,2] == 1 & dat[,3] == 1, 1]) # (compare with interactions[[2]][[1]])
mean(dat[dat[,2] == 1 & dat[,3] == 5, 1])
# first mean higher, ok!

# Check interaction 2
mean(dat[dat[,2] == 1 & dat[,3] == 1, 4]) # (compare with interactions[[2]][[2]])
mean(dat[dat[,2] == 1 & dat[,3] == 2, 4])
# first mean higher, ok!

## End(Not run)

```

mvar

Estimating mixed Vector Autoregressive Model (mVAR)

Description

Estimates Mixed Vector Autoregressive Model (mVAR) via elastic-net regularized Generalized Linear Models

Usage

```

mvar(data, type, level, lambdaSeq, lambdaSel, lambdaFolds,
      lambdaGam, alphaSeq, alphaSel, alphaFolds, alphaGam, lags,
      consec, weights, threshold, method, binarySign, scale, verbatim,
      pbar, warnings, saveModels, saveData, overparameterize, signInfo, ...)

```

Arguments

data	n x p data matrix
type	p vector indicating the type of variable for each column in data. 'g' for Gaussian, 'p' for Poisson, 'c' for categorical.
level	p vector indicating the number of categories of each variable. For continuous variables set to 1.
lambdaSeq	A sequence of lambdas that should be searched (see also lambdaSel). Defaults to NULL, which lets glmnet select a reasonable lambda sequence (recommended).

lambdaSel	Procedure to select the tuning parameter controlling the Lq-penalization. The two options are cross validation 'CV' and the Extended Bayesian Information Criterion (EBIC) 'EBIC'. The EBIC performs well in selecting sparse graphs (see Barber and Drton, 2010 and Foygel and Drton, 2014). Note that when also searching the alpha parameter in the elastic net penalty, cross validation should be preferred, as the parameter vector will not necessarily be sparse anymore. The EBIC tends to be a bit more conservative than CV (see Haslbeck and Waldorp, 2016). CV can sometimes not be performed with categorical variables, because glmnet requires at least 2 events of each category of each categorical variable in each training-fold. Defaults to lambdaSel = 'CV'.
lambdaFolds	Number of folds in cross validation if lambdaSel = 'CV'
lambdaGam	Hyperparameter gamma in the EBIC if lambdaSel = 'EBIC'. Defaults to lambdaGam = .25.
alphaSeq	A sequence of alpha parameters for the elastic net penalty in [0,1] that should be searched (see also alphaSel). Defaults to alphaSeq = 1, which means that the lasso is being used. alphaSeq = 0 corresponds to an L2-penalty (Ridge regression). For details see Friedman, Hastie and Tibshirani (2010).
alphaSel	Procedure to select the alpha parameter in the elastic net penalty. The two options are cross validation 'CV' and the Extended Bayesian Information Criterion (EBIC) 'EBIC'. The EBIC performs well in selecting sparse graphs (see Barber and Drton, 2010 and Foygel and Drton, 2014). Note that when also searching the alpha parameter in the elastic net penalty, cross validation should be preferred, as the parameter vector will not necessarily be sparse anymore. The EBIC tends to be a bit more conservative than CV (see Haslbeck and Waldorp, 2016). CV can sometimes not be performed with categorical variables, because glmnet requires at least 2 events of each category of each categorical variable in each training-fold. Defaults to alphaSel = 'CV'.
alphaFolds	Number of folds in cross validation if alphaSel = 'CV'
alphaGam	Hyperparameter gamma in the EBIC if alphaSel = 'EBIC'. Defaults to alphaGam = .25.
lags	Vector of positive integers indicating the lags included in the mVAR model (e.g. 1:3 or c(1,3,5))
consec	An integer vector of length n, indicating the measurement points of the rows in data. The time difference of 1 is what we model as a lag with size 1 in the VAR model. This vector is used to delete all rows from the VAR design matrix, for which no max(lags) complete previous measurements are available. This is useful in many applications where some time points are completely missing, or gaps in the time series have structural reasons (respondents only respond during the hours they are awake). The 'trimmed' dataset is returned in call\$data_lagged if saveData = TRUE. Defaults to consec = NULL, which assumes that all measurements are consecutive, i.e. consec = 1:n. In this case only the first max(lags) lags are excluded to obtain the VAR design matrix.
weights	A vector with n - max(lags) entries, indicating the weight for each observation. The mVAR design matrix has with n - max(lags) rows, because the first row must be predictable by the highest lag. The weights have to be on the scale [0, n - max(lags)].

threshold	A threshold below which edge-weights are put to zero. This is done in order to guarantee a constant false-positive rate. <code>threshold = 'LW'</code> refers to the threshold in Loh and Wainwright (2013), which was used in all previous versions of <code>mgm</code> . <code>threshold = 'HW'</code> refers to the threshold in Haslbeck and Waldorp (2016). If <code>threshold = 'none'</code> no thresholding is applied. Defaults to <code>threshold = 'LW'</code> .
method	Estimation method, currently only <code>method = 'glm'</code> .
binarySign	If <code>binarySign = TRUE</code> a sign for the interaction within binary nodes and between binary and continuous nodes is provided in the output. Note that in this case the two categories of the binary variables have to be coded in 0,1. This is to ensure that the interpretation of the sign is unambiguous: a positive sign of a parameter means that increasing the associated predictor results in a higher probability for category 1.
scale	If <code>scale = TRUE</code> , all Gaussian nodes (specified by 'g' in type) are centered and divided by their standard deviation. Scaling is recommended, because otherwise the penalization of a parameter depends on the variance of the associated predictor.
verbatim	If <code>verbatim = TRUE</code> no warnings and no progress bar is shown. Defaults to <code>verbatim = FALSE</code> .
pbar	If <code>pbar = TRUE</code> a progress bar is shown. Defaults to <code>pbar = TRUE</code> .
warnings	If <code>warnings = TRUE</code> no warnings are returned. Defaults to <code>warnings = FALSE</code> .
saveModels	If <code>saveModels = FALSE</code> , only information about the weighted adjacency matrix, and if $d > 1$ about the factor graph is provided in the output list. If <code>saveModels = TRUE</code> , all fitted parameters are additionally returned.
saveData	If <code>saveData = TRUE</code> , the data is saved in the output list. Defaults to <code>saveData = FALSE</code> .
overparameterize	If <code>overparameterize = TRUE</code> , <code>mgm()</code> uses over-parameterized design-matrices for each neighborhood regression; this means that a cross-lagged effect between two categorical variables with m and s categories is parameterized by $m*s$ parameters. If <code>overparameterize = FALSE</code> the standard parameterization (in <code>glmnet</code>) with $m*(s-1)$ parameters is used, where the first category of the predicting variable serves as reference category. If all variables are continuous both parameterizations are the same. The default is set to <code>overparameterize = FALSE</code> .
signInfo	If <code>signInfo = TRUE</code> , a message is shown in the console, indicating that the sign of estimates is stored separately. Defaults to <code>signInfo = TRUE</code> .
...	Additional arguments.

Details

`mgm()` estimates an exponential mixed graphical model as introduced in Yang and colleagues (2014). Haslbeck and Waldorp (2015) show that the estimation procedure is consistent. See Haslbeck and Waldorp (2016) for details about how to rewrite the exponential MGM as a mVAR model.

Value

The function returns a list with the following entries:

call	Contains all provided input arguments. If <code>saveData = TRUE</code> , it also contains the data.
wadj	A $p \times p \times n_lags$ array, in which rows are predicted by columns, i.e. entry <code>wadj[1, 2, 4]</code> corresponds to the parameter(s) of variable 2 at time point t predicting variable 1 at time point $t - z$, where z is the fourth specified lag in <code>lags</code> and <code>n_lags</code> is the number of specified lags in <code>lags</code> . For interactions that involve more than two parameters (e.g. always for categorical variables with more than 2 categories), we take the arithmetic mean of the absolute value of all parameters. The full set of estimated parameters is saved in <code>rawlags</code> (see below).
signs	A $p \times p \times n_lags$ array, specifying the signs corresponding to the entries of <code>wadj</code> (if defined), where <code>n_lags</code> is the number of specified lags in <code>lags</code> . 1/-1 indicate positive and negative relationships, respectively. 0 indicates that no sign is defined, which is the case for interactions that involve a categorical variable where an interaction can have more than one parameter. If <code>binarySign = TRUE</code> , a sign is calculated for interactions between binary variables and binary and continuous variables, where the interaction is still defined by one parameter and hence a sign can be specified. NA indicates that the corresponding parameter in <code>wadj</code> is zero.
edgecolor	A $p \times p \times n_lags$ array of colors indicating the sign of each parameter. This array contains the same information as <code>signs</code> and is included for convenient plotting.
rawlags	List with entries equal to the number of specified lags in <code>lags</code> . Each entry is a nested list, with each <code>p</code> entries: the first level indicates the predicted variable, the second level the predictor variable. In case of categorical variables, interactions have more than one parameter.
intercepts	A list with <code>p</code> entries, which contain the intercept/thresholds for each node in the network. In case a given node is categorical with <code>m</code> categories, there are <code>m</code> thresholds for this variable.
nodemodels	A list with <code>p</code> <code>glmnet()</code> models, from which all above output is computed. Also contains the coefficients <code>models</code> for the selected <code>lambda</code> and the applied tau threshold <code>tau</code> .

Author(s)

Jonas Haslbeck <jonashaslbeck@gmail.com>

References

- Foygel, R., & Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. In *Advances in neural information processing systems* (pp. 604-612).
- Barber, R. F., & Drton, M. (2015). High-dimensional Ising model selection with Bayesian information criteria. *Electronic Journal of Statistics*, 9(1), 567-607.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1), 1.

Haslbeck, J., & Waldorp, L. J. (2015). Structure estimation for mixed graphical models in high-dimensional data. arXiv preprint arXiv:1510.05677.

Haslbeck, J., & Waldorp, L. J. (2016). mgm: Structure Estimation for time-varying Mixed Graphical Models in high-dimensional Data. arXiv preprint arXiv:1510.06871.

Loh, P. L., & Wainwright, M. J. (2012, December). Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. In NIPS (pp. 2096-2104).

Yang, E., Baker, Y., Ravikumar, P., Allen, G. I., & Liu, Z. (2014, April). Mixed Graphical Models via Exponential Families. In AISTATS (Vol. 2012, pp. 1042-1050).

Examples

```
## Not run:

## We generate data from a mixed VAR model and then recover the model using mvar()

# 1) Define mVAR model

p <- 6 # Six variables
type <- c('c', 'c', 'c', 'c', 'g', 'g') # 4 categorical, 2 gaussians
level <- c(2, 2, 4, 4, 1, 1) # 2 categoricals with m=2, 2 categoricals with m=4, two continuous
max_level <- max(level)

lags <- c(1, 3, 9) # include lagged effects of order 1, 3, 9
n_lags <- length(lags)

# Specify thresholds
thresholds[[1]] <- rep(0, level[1])
thresholds[[2]] <- rep(0, level[2])
thresholds[[3]] <- rep(0, level[3])
thresholds[[4]] <- rep(0, level[4])
thresholds[[5]] <- rep(0, level[5])
thresholds[[6]] <- rep(0, level[6])

# Specify standard deviations for the Gaussians
sds <- rep(NULL, p)
sds[5:6] <- 1

# Create coefficient array
coefarray <- array(0, dim=c(p, p, max_level, max_level, n_lags))

# a.1) interaction between continuous 5<-6, lag=3
coefarray[5, 6, 1, 1, 2] <- .4
# a.2) interaction between 1<-3, lag=1
m1 <- matrix(0, nrow=level[2], ncol=level[4])
m1[1,1:2] <- 1
m1[2,3:4] <- 1
coefarray[1, 3, 1:level[2], 1:level[4], 1] <- m1
# a.3) interaction between 1<-5, lag=9
coefarray[1, 5, 1:level[1], 1:level[5], 3] <- c(0, 1)
```

```
# 2) Sample
set.seed(1)
dlist <- mvarsampler(coefarray = coefarray,
                    lags = lags,
                    thresholds = thresholds,
                    sds = sds,
                    type = type,
                    level = level,
                    N = 200,
                    pbar = TRUE)

# 3) Recover

set.seed(1)
mvar_obj <- mvar(data = dlist$data,
                type = type,
                level = level,
                lambdaSel = 'CV',
                lags = c(1, 3, 9),
                signInfo = FALSE,
                overparameterize = F)

# Did we recover the true parameters?
mvar_obj$wadj[5, 6, 2] # cross-lagged effect of 6 on 2 over lag lags[2]
mvar_obj$wadj[1, 3, 1] # cross-lagged effect of 3 on 1 over lag lags[1]
mvar_obj$wadj[1, 5, 3] # cross-lagged effect of 1 on 5 over lag lags[3]

# How to get the exact parameter estimates?
# Example: the full parameters for the crossed-lagged interaction of 2 on 1 over lag lags[1]
mvar_obj$rawlags[[1]][[1]][[2]]

# 4) Predict / Compute nodewise Error

pred_mvar <- predict.mgm(mvar_obj, dlist$data)

head(pred_mvar$predicted) # predicted values

pred_mvar$errors # Nodewise errors

# For more examples see https://github.com/jmbh/mgmDocumentation

## End(Not run)
```

mvarsampler

Sampling from a mixed VAR model

Description

Function to sample from a mixed VAR (mVAR) model

Usage

```
mvarsampler(coefarray, lags, thresholds, sds,
            type, level, N, pbar)
```

Arguments

coefarray	A $p \times p \times \max(\text{level}) \times \max(\text{level}) \times n_lags$ array, where p are the number of variables, level is the input argument <code>level</code> and n_lags is the number of specified lags in <code>lags</code> , so $n_lags = \text{length}(n_lags)$. The first four dimensions specify the parameters involved in the cross-lagged effects of the lag specified in the 5th dimension. I.e. <code>coefarray[5, 6, 1, 1, 3]</code> indicates the cross-lagged effect of variable 6 on variable 5 (if both are continuous), for the third lag specified in <code>lags</code> . If variable 1 and 3 are categorical with $m = 2$ and $= 4$ categories, respectively, then <code>coefarray[1, 3, 1:2, 1:4, 1]</code> indicates the $m*s=8$ parameters specifying this interaction for the first lag specified in <code>lags</code> . See the examples below for an illustration.
lags	A vector indicating the lags in the mVAR model. E.g. <code>lags = c(1, 4, 9)</code> specifies lags of order 1, 3, 9. The number of specified lags has to match the 5th dimension in <code>coefarray</code> .
thresholds	A list with p entries, each consisting of a vector indicating a threshold for each category of the given variable. For continuous variable, the vector has length 1.
sds	A vector of length p indicating the standard deviations of the included Gaussian nodes. If non-Gaussian variables are included in the mVAR model, the corresponding entries are ignored.
type	p vector indicating the type of variable for each column in data. 'g' for Gaussian, 'p' for Poisson, 'c' for categorical.
level	p vector indicating the number of categories of each variable. For continuous variables set to 1.
N	The number of samples to be drawn from the specified mVAR model.
pbar	If <code>pbar = TRUE</code> , a progress bar is shown.

Details

We sample from the mVAR model by separately sampling from its corresponding p conditional distributions.

Value

A list with two entries:

call	The function call
data	The sampled $n \times p$ data matrix

Author(s)

Jonas Haslbeck <jonashaslbeck@gmail.com>

References

Haslbeck, J., & Waldorp, L. J. (2016). mgm: Structure Estimation for time-varying Mixed Graphical Models in high-dimensional Data. arXiv preprint arXiv:1510.06871.

Examples

```
## Not run:

# a) Set up mvar model

p <- 6 # Six variables
type <- c('c', 'c', 'c', 'c', 'g', 'g') # 4 categorical, 2 gaussians
level <- c(2, 2, 4, 4, 1, 1) # 2 categoricals with 2 categories, 2 with 5
max_level <- max(level)

lags <- c(1, 3, 9) # include lagged effects of order 1, 3, 9
n_lags <- length(lags)

# Specify thresholds
thresholds[[1]] <- c(0, 0)
thresholds[[2]] <- c(0, 0)
thresholds[[3]] <- c(0, 0, 0, 0)
thresholds[[4]] <- c(0, 0, 0, 0)
thresholds[[5]] <- c(0)
thresholds[[6]] <- c(0)

# Specify standard deviations for the Gaussians
sds <- rep(NULL, p)
sds[5:6] <- 1

# Create coefficient array
coefarray <- array(0, dim=c(p, p, max_level, max_level, n_lags))

# a.1) interaction between continuous 5<-6, lag=3
coefarray[5, 6, 1, 1, 2] <- .4
# a.2) interaction between 1<-3, lag=1
m1 <- matrix(0, nrow=level[2], ncol=level[4])
m1[1,1:2] <- 1
```

```

m1[2,3:4] <- 1
coefarray[1, 3, 1:level[2], 1:level[4], 1] <- m1
# a.3) interaction between 1<-5, lag=9
coefarray[1, 5, 1:level[1], 1:level[5], 3] <- c(0, 1)

# b) Sample
set.seed(1)
dlist <- mvarsampler(coefarray = coefarray,
                    lags = lags,
                    thresholds = thresholds,
                    sds = sds,
                    type = type,
                    level = level,
                    N = 200,
                    pbar = TRUE)

# c) Recover
set.seed(1)
mvar_obj <- mvar(data = dlist$data,
                type = type,
                level = level,
                lambdaSel = 'CV',
                lags = c(1, 3, 9))

# Did we recover the true parameters?
mvar_obj$wadj[5, 6, 2] # yes
mvar_obj$wadj[1, 3, 1] # yes
mvar_obj$wadj[1, 5, 3] # yes

## End(Not run)

```

predict.mgm

Compute predictions from mgm model objects

Description

Similar to other predict methods, this function computes predicted values and prediction errors from a mgm model-object (mgm, mvar, tvmgm or tvmvar).

Usage

```

## S3 method for class 'mgm'
predict(object, data, errorCon, errorCat, tvMethod, consec, ...)

```

Arguments

object	An mgm model object, i.e. the output of the functions <code>mgm()</code> , <code>mvar()</code> , <code>tvmgm()</code> or <code>tvmmvar()</code>
data	A $n \times p$ data matrix with the same structure (number of variables p and types of variables) as the data used to fit the model.
errorCon	Either a character vector specifying the types of nodewise error that should be computed, where the two provided error functions for continuous variables are <code>errorCon = 'RMSE'</code> , the Root Mean Squared Error and <code>errorCon = 'R2'</code> , the proportion of explained variance. The default is <code>errorCon = c('RMSE' 'R2')</code> . Alternatively, <code>errorCon</code> can be a list, where each list entry is a custom error function of the form <code>foo(true, pred)</code> , where <code>true</code> and <code>pred</code> are the arguments for the vectors of true and predicted values, respectively. If predictions are made for a time-varying model and <code>tvMethod = 'weighted'</code> , the weighted R2 or RMSE are computed. If a custom function is used, an additional argument for the weights has to be provided: <code>foo(true, pred, weights)</code> . Note that custom error functions can also be combined with the built-in functions, i.e. <code>errorCon = list('RMSE', 'CustomError'=foo)</code> .
errorCat	Either a character vector specifying the types of nodewise error that should be computed, where the two provided error functions for categorical variables are <code>errorCat = 'CC'</code> , the proportion of correct classification (accuracy) and <code>errorCat = 'nCC'</code> , the proportion of correct classification normalized by the marginal distribution of the variable at hand. Specifically, <code>nCC = (CC - norm_constant) / (1 - norm_constant)</code> , where <code>norm_constant</code> is the relative frequency of the category with the highest relative frequency. The default is <code>errorCon = c('CC' 'nCC')</code> . Alternatively, <code>errorCat</code> can be a list, where each list entry is a custom error function of the form <code>foo(true, pred)</code> , where <code>true</code> and <code>pred</code> are the arguments for the vectors of true and predicted values, respectively. If predictions are made for a time-varying model and <code>tvMethod = 'weighted'</code> , the weighted R2 or RMSE are computed. If a custom function is used, an additional argument for the weights has to be provided: <code>foo(true, pred, weights)</code> . Note that custom error functions can also be combined with the built-in functions, i.e. <code>errorCon = list('nCC', 'CustomError'=foo)</code> .
tvMethod	The type of error calculated for time-varying models: <code>tvMethod = 'weighted'</code> computes errors by computing a weighted error over all cases in the time series at each estimation point specified in <code>estpoints</code> in <code>tvmmgm()</code> or <code>tvmmvar()</code> . The weighting corresponds to the weighting used for estimation (see <code>?tvmmgm</code> or <code>?tvmmvar</code>). <code>tvMethod = 'closestModel'</code> determines for each time point the closest model and uses that model for prediction. See Details below for a more elaborate explanation.
consec	A integer vector of length <code>nrow(data)</code> , indicating the sequence of measurement points in a time series. For details see <code>?mvar</code> . This is only relevant for mVAR models and time series with unequal time intervals. Defaults to <code>consec = NULL</code> , which assumes equal time intervals. <code>consec</code> is ignored if a <code>mgm</code> or <code>tvmmgm</code> object is provided to <code>predict.mgm()</code>
...	Additional arguments.

Details

In the case of time-varying model nodewise errors can be computed in two different ways.

First, one computes the predicted value for each of the N cases in the time series for all models (estimated at different estimation points, see `?tvmgm` or `?tvmvar`). Then the error of each of the N cases for each of the models is weighted by the weight that has been used to estimate a given model at its estimation point. This means that the error of a case in the end of a time-series gets a high weight for models estimated in the end of the time-series and a low weight for models estimated in the beginning of the time series.

Second, we determine for each case in the time-series the closest estimation point, and use the model estimated at that estimation point to make predictions for that case.

Note that the error function normalized accuracy (nCC) is negative, if the full model performs worse than the intercept model. This can happen if the model overfits the data.

Value

A list with the following entries:

<code>call</code>	Contains all provided input arguments.
<code>predicted</code>	A $n \times p$ matrix with predicted values, matching the dimension of the true values in <code>true</code> .
<code>probabilities</code>	A list with p entries corresponding to p nodes in the data. If a variable is categorical, the corresponding entry contains a $n \times k$ matrix with predicted probabilities, where k is the number of categories of the categorical variable. If a variable is continuous, the corresponding entry is empty.
<code>true</code>	Contains the true values. For <code>mgm</code> and <code>tvmgm</code> objects these are equal to the data provided via <code>data</code> . For <code>mvar</code> and <code>tvmvar</code> objects, these are equal to the rows that can be predicted in a VAR model, depending on the largest specified lag.
<code>errors</code>	A matrix containing the all types of errors specified via <code>errorCon</code> and <code>errorCat</code> , for each variable. If <code>tvMethod = 'weighted'</code> , the matrix becomes an array, with an additional dimension for the estimation point.

Author(s)

Jonas Haslbeck <jonashaslbeck@gmail.com>

References

- Haslbeck, J., & Waldorp, L. J. (2016). `mgm`: Structure Estimation for time-varying Mixed Graphical Models in high-dimensional Data. arXiv preprint arXiv:1510.06871.
- Haslbeck, J., & Waldorp, L. J. (2015). Structure estimation for mixed graphical models in high-dimensional data. arXiv preprint arXiv:1510.05677.

Examples

```
## Not run:  
  
# See examples in ?mgm, ?tvmgm, ?mvar and ?tvmvar.  
  
## End(Not run)
```

print.mgm	<i>Print method for mgm objects</i>
-----------	-------------------------------------

Description

Returns basic information about fit objects, prediction objects and bandwidth-selection objects.

Usage

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

Arguments

x	The output object of mgm(), mvar(), tvmgm(), tvmvar(), predict.mgm() or bwSelect().
...	Additional arguments.

Value

Writes basic information about the object in the console.

Author(s)

Jonas Haslbeck <jonashaslbeck@gmail.com>

Description

Estimates time-varying k-degree Mixed Graphical Models (MGMs) via elastic-net regularized kernel smoothed Generalized Linear Models

Usage

```
tvmgm(data, type, level, timepoints, estpoints, bandwidth, ...)
```

Arguments

<code>data</code>	A $n \times p$ data matrix.
<code>type</code>	p vector indicating the type of variable for each column in data. 'g' for Gaussian, 'p' for Poisson, 'c' for categorical.
<code>level</code>	p vector indicating the number of categories of each variable. For continuous variables set to 1.
<code>timepoints</code>	A strictly increasing numeric vector of length <code>nrow(data)</code> indicating time points for the measurements in data. If <code>timepoints</code> is not specified, it is assumed that the time points are equally spaced. For details, see Haslbeck and Waldorp (2016).
<code>estpoints</code>	Vector indicating estimation points on interval $[0, n]$.
<code>bandwidth</code>	We use a gaussian density on the unit time-interval $[0,1]$ to determine the weights for each observation at each estimated time point. The bandwidth specifies the standard deviation the Gaussian density. To get some intuition, which bandwidth results in the combination of how many data close in time one can plot Gaussians on $[0,1]$ for different bandwidths. The bandwidth can also be selected in a data driven way using the function (see <code>bwSelect</code>).
<code>...</code>	Arguments passed to <code>mgm</code> , specifying how each single model should be estimated. See <code>?mgm</code> .

Details

Estimates a sequence of MGMs at the time points specified at the locations specified via `estpoints`. `tvmgm()` is a wrapper around `mgm()` and estimates a series of MGM with different weightings which are defined by the estimation locations in `estpoints` and the bandwidth parameter specified in `bandwidth`. For details see Haslbeck and Waldorp (2016).

Value

A list with the following entries:

<code>call</code>	Contains all provided input arguments. If <code>saveData = TRUE</code> , it also contains the data.
-------------------	-----------------------------------------------------------------------------------------------------

pairwise	Contains a list with all information about estimated pairwise interactions. <code>wadj</code> contains a $p \times p \times \text{estpoints}$ array containing the weighted adjacency matrix for each estimation point specified in <code>estpoints</code> , if p is the number of variables in the network. <code>signs</code> has the same dimensions as <code>wadj</code> and contains the signs for the entries of <code>wadj</code> : 1 indicates a positive sign, -1 a negative sign and 0 an undefined sign. A sign is undefined if an edge is a function of more than one parameter. This is the case for interactions involving a categorical variable with more than 2 categories. <code>edgecolor</code> also has the same dimensions as <code>wadj</code> contains a color for each edge, depending on <code>signs</code> . It is provided for more convenient plotting. If only pairwise interactions are modeled ($d = 1$), <code>wadj</code> contains all conditional independence relations.
factorgraph	Contains a list with all information about all estimated interactions (factors). <code>graph</code> contains an $(p+E) \times (p+E) \times \text{estpoints}$ array containing weighted adjacency matrix of a (bipartite) factor graph for each estimation point specified in <code>estpoints</code> . If p is the number of variables and E the number of interactions (factors) in the model, each of these matrices (for a fixed estimation point) has dimensions $(p+E) \times (p+E)$. The factor graph is further specified by the following objects: <code>signs</code> is a matrix of the same dimensions as <code>graph</code> that indicates the sign of each interaction, if defined (see <code>pairwise</code> above). <code>edgecolor</code> is a matrix with the same dimension as <code>graph</code> that provides edge colors depending on the sign as above. <code>order</code> contains a list with $\text{length}(\text{estpoints})$ entries; each element is a $(p+E)$ vector indicating the order of interaction. The first p entries are set to zero and correspond to the variables. <code>nodetype</code> contains a list with $\text{length}(\text{estpoints})$ entries; each element is a $(p+E)$ vector indicating whether a row/column corresponds to a variable (0) or a factor (1). The last two vectors are helpful to plot the factor graph (see examples below). Note that <code>factorgraph</code> contains the same information as <code>pairwise</code> if only pairwise interactions are estimated.
rawfactor	Contains a list with entries equal to the number of estimated time points specified in <code>estpoints</code> ; each entry is a list with three entries that relate each interaction in the model to all its parameters. This is different to the output provided in <code>factorgraph</code> , where one value is assigned to each interaction. <code>indicator</code> contains a list with d entries, one for each order of modeled interaction, which contain the estimated (nonzero) interactions. <code>weights</code> contains a list with d entries, which in turn contain R lists, where R is the number of interactions (and rows in the corresponding list entry in <code>indicator</code>) that were estimated (nonzero) in the given entry. <code>signs</code> has the same structure as <code>weights</code> and provides the sign of the interaction, if defined.
intercepts	Contains a list with entries equal to the number of estimated time points specified in <code>estpoints</code> ; each entry is a list with p entries, which contain the intercept/thresholds for each node in the network. In case a given node is categorical with m categories, there are m thresholds for this variable.
tvmodels	Contains the MGM model estimated by <code>mgm()</code> at each time point specified via <code>estpoints</code> . See <code>?mgm</code> for a detailed description of this output.

Author(s)

Jonas Haslbeck <jonashaslbeck@gmail.com>

References

Haslbeck, J., & Waldorp, L. J. (2016). mgm: Structure Estimation for time-varying Mixed Graphical Models in high-dimensional Data. arXiv preprint arXiv:1510.06871.

Haslbeck, J., & Waldorp, L. J. (2015). Structure estimation for mixed graphical models in high-dimensional data. arXiv preprint arXiv:1510.05677.

Examples

```
## Not run:

## We specify a time-varying MGM and recover it using tvmgm()

# 1) Specify Model

# a) Define Graph
p <- 6
type = c('c', 'c', 'g', 'g', 'p', 'p')
level = c(2, 3, 1, 1, 1, 1)
n_timepoints <- 1000

# b) Define Interaction
factors <- list()
factors[[1]] <- matrix(c(1,2,
                        2,3,
                        3,4), ncol=2, byrow = T) # no pairwise interactions
factors[[2]] <- matrix(c(1,2,3,
                        2,3,4), ncol=3, byrow = T) # one 3-way interaction

interactions <- list()
interactions[[1]] <- vector('list', length = 3)
interactions[[2]] <- vector('list', length = 2)
# 3 2-way interactions
interactions[[1]][[1]] <- array(0, dim = c(level[1], level[2], n_timepoints))
interactions[[1]][[2]] <- array(0, dim = c(level[2], level[3], n_timepoints))
interactions[[1]][[3]] <- array(0, dim = c(level[3], level[4], n_timepoints))
# 2 3-way interactions
interactions[[2]][[1]] <- array(0, dim = c(level[1], level[2], level[3], n_timepoints))
interactions[[2]][[2]] <- array(0, dim = c(level[2], level[3], level[4], n_timepoints))
theta <- .3
interactions[[1]][[1]][1, 1, ] <- theta
interactions[[1]][[2]][1, 1, ] <- theta
interactions[[1]][[3]][1, 1, ] <- seq(0, theta, length = n_timepoints)
interactions[[2]][[1]][1, 1, 1, ] <- theta
interactions[[2]][[2]][1, 1, 1, ] <- theta
# c) Define Thresholds
thresholds <- list()
thresholds[[1]] <- matrix(0, nrow = n_timepoints, ncol= level[1])
thresholds[[2]] <- matrix(0, nrow = n_timepoints, ncol= level[2])
thresholds[[3]] <- matrix(0, nrow = n_timepoints, ncol= level[3])
```



```

pred_mgm_cv_w$errors
pred_mgm_cv_cM$errors

# For more examples see https://github.com/jmbh/mgmDocumentation

## End(Not run)

```

tvmgmsampler

Sample from time-varying k-order Mixed Graphical Model

Description

Generates samples from a time-varying k-order Mixed Graphical Model

Usage

```
tvmgmsampler(factors, interactions, thresholds, sds, type,
             level, nIter = 250, pbar = TRUE, ...)
```

Arguments

factors	The same object as factors in mgmsampler(). An interaction is specified in factors if it should be nonzero at least at one time point in the time series. The values of each parameter at each time point is specified via interactions.
interactions	The same object as factors in mgmsampler(), except that each array indicating the parameters of an interaction has an additional (the last) dimension, indicating time. Corresponding to the time vector in factors, the time vector has to be a sequence of integers {1, 2, ..., N}. For an illustration see the examples below.
thresholds	A list with p entries for p variables, each of which contains a N x m matrix. The columns contain the m thresholds for m categories (for continuous variables m = 1 and the entry contains the threshold/intercept). The rows indicate how the thresholds change over time.
sds	N x p matrix indicating the standard deviations of Gaussians specified in type for {1, ..., N} time points. Entries not referring to Gaussians are ignored.
type	p character vector indicating the type of variable for each column in data. 'g' for Gaussian, 'p' for Poisson, 'c' of each variable.
level	p integer vector indicating the number of categories of each variable. For continuous variables set to 1.
nIter	Number of iterations in the Gibbs sampler until a sample is drawn.
pbar	If pbar = TRUE a progress bar is shown. Defaults to pbar = TRUE.
...	Additional arguments.

Details

tvmgmsampler is a wrapper function around mgmsampler. Its input is the same as for mgmsampler, except that each object has an additional dimension for time. The number of time points is specified via entries in the additional time dimension.

Value

A list containing:

call	Contains all provided input arguments.
data	The $N \times p$ data matrix of sampled values

Author(s)

Jonas Haslbeck <jonashaslbeck@gmail.com>

References

Haslbeck, J., & Waldorp, L. J. (2016). mgm: Structure Estimation for time-varying Mixed Graphical Models in high-dimensional Data. arXiv preprint arXiv:1510.06871.

Yang, E., Baker, Y., Ravikumar, P., Allen, G. I., & Liu, Z. (2014, April). Mixed Graphical Models via Exponential Families. In AISTATS (Vol. 2012, pp. 1042-1050).

Examples

```
## Not run:

# ----- Example 1: p = 4 dimensional Gaussian -----

# ----- 1) Specify Model -----

# a) General Graph Info
type = c('g', 'g', 'g', 'g') # Four Gaussians
level = c(1, 1, 1, 1)
n_timepoints = 500 # Number of time points

# b) Define Interaction
factors <- list()
factors[[1]] <- array(NA, dim=c(2, 2)) # two pairwise interactions
factors[[1]][1, 1:2] <- c(3,4)
factors[[1]][2, 1:2] <- c(1,2)

# Two parameters, one linearly increasing from 0 to 0.8, another one lin decreasing from 0.8 to 0
interactions <- list()
interactions[[1]] <- vector('list', length = 2)
interactions[[1]][[1]] <- array(0, dim = c(level[1], level[2], n_timepoints))
interactions[[1]][[1]][1,1, ] <- seq(.8, 0, length = n_timepoints)
interactions[[1]][[2]] <- array(0, dim = c(level[1], level[2], n_timepoints))
```

```

interactions[[1]][[2]][1,1, ] <- seq(0, .8, length = n_timepoints)

# c) Define Thresholds
thresholds <- vector('list', length = 4)
thresholds <- lapply(thresholds, function(x) matrix(0, ncol = level[1], nrow = n_timepoints))

# d) Define Standard deviations
sds <- matrix(1, ncol = length(type), nrow = n_timepoints) # constant across variables and time

# ----- 2) Sample cases -----

set.seed(1)
dlist <- tvmgmsampler(factors = factors,
                     interactions = interactions,
                     thresholds = thresholds,
                     sds = sds,
                     type = type,
                     level = level,
                     nIter = 75,
                     pbar = TRUE)

# ----- 3) Recover model from sampled cases -----

set.seed(1)
tvmgm_obj <- tvmgm(data = dlist$data,
                  type = type,
                  level = level,
                  estpoints = seq(1, n_timepoints, length = 15),
                  bandwidth = .2,
                  d = 2,
                  lambdaSel = 'CV',
                  ruleReg = 'AND')

# How well did we recover those two time-varying parameters?
plot(tvmgm_obj$pairwise$wadj[3,4,], type='l', ylim=c(0,.8))
lines(tvmgm_obj$pairwise$wadj[1,2,], type='l', col='red')
# Looks quite good

# ----- Example 2: p = 5 binary; one 3-way interaction -----

# ----- 1) Specify Model -----

# a) General Graph Info
p <- 5 # number of variables
type = rep('c', p) # all categorical
level = rep(2, p) # all binary
n_timepoints <- 1000

# b) Define Interaction

```

```

factors <- list()
factors[[1]] <- NULL # no pairwise interactions
factors[[2]] <- array(NA, dim = c(1,3)) # one 3-way interaction
factors[[2]][1, 1:3] <- c(1, 2, 3)

interactions <- list()
interactions[[1]] <- NULL # no pairwise interactions
interactions[[2]] <- vector('list', length = 1) # one 3-way interaction
# 3-way interaction no1
interactions[[2]][[1]] <- array(0, dim = c(level[1], level[2], level[3], n_timepoints))
theta <- 2
interactions[[2]][[1]][1, 1, 1, ] <- seq(0, 2, length = n_timepoints) # fill in nonzero entries

# c) Define Thresholds
thresholds <- list()
for(i in 1:p) thresholds[[i]] <- matrix(0, nrow = n_timepoints, ncol = level[i])

# ----- 2) Sample cases -----

set.seed(1)
dlist <- tvmgmsampler(factors = factors,
                    interactions = interactions,
                    thresholds = thresholds,
                    type = type,
                    level = level,
                    nIter = 150,
                    pbar = TRUE)

# ----- 3) Check Marginals -----

dat <- dlist$data[1:round(n_timepoints/2),]
table(dat[,1], dat[,2], dat[,3])

dat <- dlist$data[round(n_timepoints/2):n_timepoints,]
table(dat[,1], dat[,2], dat[,3])

# Observation: much stronger effect in second half of the time-series,
# which is what we expect

# ----- 4) Recover model from sampled cases -----

set.seed(1)
tvmgm_obj <- tvmgm(data = dlist$data,
                  type = type,
                  level = level,
                  estpoints = seq(1, n_timepoints, length = 15),
                  bandwidth = .2,
                  d = 2,
                  lambdaSel = 'CV',
                  ruleReg = 'AND')

```

```
tvmgm_obj$rawfactor$indicator
# Seems very difficult to recover this time-varying 3-way binary interaction
# See also the corresponding stationary problems in the examples of ?mgmsampler
```

```
## End(Not run)
```

tvmvar

Estimating time-varying Mixed Vector Autoregressive Model (mVAR)

Description

Estimates time-varying Mixed Vector Autoregressive Model (mVAR) via elastic-net regularized kernel smoothed Generalized Linear Models

Usage

```
tvmvar(data, type, level, timepoints, estpoints, bandwidth, ...)
```

Arguments

data	A $n \times p$ data matrix.
type	p vector indicating the type of variable for each column in data. 'g' for Gaussian, 'p' for Poisson, 'c' for categorical.
level	p vector indicating the number of categories of each variable. For continuous variables set to 1.
timepoints	A strictly increasing numeric vector of length $nrow(data)$ indicating time points for the measurements in data. If timepoints is not specified, it is assumed that the time points are equally spaced. For details, see Haslbeck and Waldorp (2016).
estpoints	Vector indicating estimation points on interval $[0, n-\max(lags)]$. Lags are passed to via the <code>...</code> arguments. See <code>?mvar</code> for how to specify lags. $n-\max(lags)$ is the number of rows in the design matrix in a VAR model: the first $\max(lags)$ rows of the data matrix have to be omitted, otherwise we have no data to predict the first row.
bandwidth	We use a gaussian density on the unit time-interval $[0, 1]$ to determine the weights for each observation at each estimated time point. The bandwidth specifies the standard deviation the Gaussian density. To get some intuition, which bandwidth results in the combination of how many data close in time one can plot

Gaussians on [0,1] for different bandwidths. The bandwidth can also be selected in a data driven way using the function (see `bwSelect`).

... Arguments passed to `mvar`, specifying how each single model should be estimated. See `?mvar`.

Details

Estimates a sequence of mVAR models at the time points specified at the locations specified via `estpoints`. `tvmvar()` is a wrapper around `mvar()` and estimates a series of MGM with different weightings which are defined by the estimation locations in `estpoints` and the bandwidth parameter specified in `bandwidth`. For details see Haslbeck and Waldorp (2016)

Value

A list with the following entries:

<code>call</code>	Contains all provided input arguments. If <code>saveData = TRUE</code> , it also contains the data.
<code>wadj</code>	A $p \times p \times n_lags \times N$ array, where <code>n_lags</code> is the number of specified lags in <code>lags</code> (see <code>?mvar</code>) and <code>N</code> is the number of estimation points specified in <code>estpoints</code> . For instance, <code>wadj[1, 2, 1, 10]</code> is the cross-lagged predicting variable 1 at time point <code>t</code> by variable 2 at time point <code>t - z</code> , where <code>z</code> is specified by the first lag specified in <code>lags</code> (see <code>?mvar</code>), in the model estimated at estimation point 10.
<code>signs</code>	Has the same structure as <code>wadj</code> and specifies the signs corresponding to the parameters in <code>wadj</code> , if defined. 1/-1 indicate positive and negative relationships, respectively. 0 indicates that no sign is defined, which is the case for interactions that involve a categorical variable where an interaction can have more than one parameter. If <code>binarySign = TRUE</code> , a sign is calculated for interactions between binary variables and binary and continuous variables, where the interaction is still defined by one parameter and hence a sign can be specified. NA indicates that the corresponding parameter in <code>wadj</code> is zero. See also <code>?mvar</code> .
<code>intercepts</code>	A list with <code>S</code> entries, where <code>S</code> is the number of estimated time points. Each entry of that list contains a list <code>p</code> entries with the intercept/thresholds for each node in the network. In case a given node is categorical with <code>m</code> categories, there are <code>m</code> thresholds for this variable.
<code>tvmmodels</code>	Contains the mVAR model estimated by <code>mvar()</code> at each time point specified via <code>estpoints</code> . See <code>?mvar</code> for a detailed description of this output.

Author(s)

Jonas Haslbeck <jonashaslbeck@gmail.com>

References

- Haslbeck, J., & Waldorp, L. J. (2016). *mgm: Structure Estimation for time-varying Mixed Graphical Models in high-dimensional Data*. arXiv preprint arXiv:1510.06871.
- Haslbeck, J., & Waldorp, L. J. (2015). *Structure estimation for mixed graphical models in high-dimensional data*. arXiv preprint arXiv:1510.05677.

Examples

```
## Not run:

## We set up the same model as in the example of mvar(), but
## specify one time-varying parameter, and try to recover it with
## tvmvar()

# a) Specify time-varying VAR model

p <- 6 # Six variables
type <- c('c', 'c', 'c', 'c', 'g', 'g') # 4 categorical, 2 gaussians
level <- c(2, 2, 4, 4, 1, 1) # 2 categoricals with 2 categories, 2 with 5
max_level <- max(level)
n_timepoints <- 4000

lags <- c(1, 3, 9) # include lagged effects of order 1, 3, 9
n_lags <- length(lags)

# Specify thresholds
thresholds[[1]] <- matrix(0, ncol=level[1], nrow=n_timepoints)
thresholds[[2]] <- matrix(0, ncol=level[2], nrow=n_timepoints)
thresholds[[3]] <- matrix(0, ncol=level[3], nrow=n_timepoints)
thresholds[[4]] <- matrix(0, ncol=level[4], nrow=n_timepoints)
thresholds[[5]] <- matrix(0, ncol=level[5], nrow=n_timepoints)
thresholds[[6]] <- matrix(0, ncol=level[6], nrow=n_timepoints)

# Specify standard deviations for the Gaussians
sds <- matrix(NA, ncol=p, nrow=n_timepoints)
sds[, 5:6] <- 1

# Create coefficient array
coefarray <- array(0, dim=c(p, p, max_level, max_level, n_lags, n_timepoints))

# a.1) interaction between continuous 5<-6, lag=3
coefarray[5, 6, 1, 1, 2, ] <- seq(0, .4, length = n_timepoints) # only time-varying parameter
# a.2) interaction between 1<-3, lag=1
m1 <- matrix(0, nrow=level[2], ncol=level[4])
m1[1,1:2] <- 1
m1[2,3:4] <- 1
coefarray[1, 3, 1:level[2], 1:level[4], 1, ] <- m1 # constant across time
# a.3) interaction between 1<-5, lag=9
coefarray[1, 5, 1:level[1], 1:level[5], 3, ] <- c(0, 1) # constant across time

# b) Sample
set.seed(1)
dlist <- tvmvarsampler(coefarray = coefarray,
                      lags = lags,
                      thresholds = thresholds,
```

```

        sds = sds,
        type = type,
        level = level,
        pbar = TRUE)

# c.1) Recover: stationary
set.seed(1)
mvar_obj <- mvar(data = dlist$data,
                type = type,
                level = level,
                lambdaSel = 'CV',
                lags = c(1, 3, 9),
                signInfo = FALSE)

# Did we recover the true parameters?
mvar_obj$wadj[5, 6, 2] # cross-lagged effect of 6 on 2 over lag lags[2]
mvar_obj$wadj[1, 3, 1] # cross-lagged effect of 3 on 1 over lag lags[1]
mvar_obj$wadj[1, 5, 3] # cross-lagged effect of 1 on 5 over lag lags[3]

# c.2) Recover: time-varying
set.seed(1)
mvar_obj <- tvmmvar(data = dlist$data,
                   type = type,
                   level = level,
                   estpoints = seq(0, n_timepoints-9, length=10),
                   bandwidth = .15,
                   lambdaSel = 'CV',
                   lags = c(1, 3, 9),
                   signInfo = FALSE)

# Did we recover the true parameters?
mvar_obj$wadj[5, 6, 2, ] # true increase nicely recovered
mvar_obj$wadj[1, 3, 1, ] # yes
mvar_obj$wadj[1, 5, 3, ] # yes

# Plotting parameter estimates over time
plot(mvar_obj$wadj[5, 6, 2, ],
     type='l', ylim=c(-.2,.7),
     lwd=2, ylab='Parameter value', xlab='Estimation points')
lines(mvar_obj$wadj[1, 3, 1, ], col='red', lwd=2)
lines(mvar_obj$wadj[1, 5, 3, ], col='blue', lwd=2)
legend('bottomright', c('5 <-- 6', '1 <-- 3', '1 <-- 5'),
      lwd = c(2,2,2), col=c('black', 'red', 'blue'))

# d) Predict values / compute nodewise error

mvar_pred_w <- predict.mgm(object=mvar_obj,
                          data=dlist$data,
                          tvMethod = 'weighted')

```

```

mvar_pred_cM <- predict.mgm(object=mvar_obj,
                           data=dlist$data,
                           tvMethod = 'closestModel')

mvar_pred_w$errors
mvar_pred_cM$errors

# For more examples see https://github.com/jmbh/mgmDocumentation

## End(Not run)

```

tvmvarsampler

Sampling from a time-varying mixed VAR model

Description

Function to sample from a time-varying mixed VAR (mVAR) model

Usage

```
tvmvarsampler(coefarray, lags, thresholds,
              sds, type, level, pbar)
```

Arguments

- | | |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| coefarray | A $p \times p \times \max(\text{level}) \times \max(\text{level}) \times n_lags \times N$ array, where p are the number of variables, level is the input argument <code>level</code> and n_lags is the number of specified lags in <code>lags</code> , so $n_lags = \text{length}(n_lags)$, and N is the number of time points in the time series. The first four dimensions specify the parameters involved in the cross-lagged effects of the lag specified in the 5th dimension. I.e. <code>coefarray[5, 6, 1, 1, 3, 100]</code> indicates the cross-lagged effect of variable 6 on variable 5 (if both are continuous), for the third lag specified in <code>lags</code> at time point 100. If variable 1 and 3 are categorical with $m = 2$ and $= 4$ categories, respectively, then <code>coefarray[1, 3, 1:2, 1:4, 1, 250]</code> indicates the $m*s=8$ parameters specifying this interaction for the first lag specified in <code>lags</code> at time point 250. See the examples below for an illustration. |
| lags | A vector indicating the lags in the mVAR model. E.g. <code>lags = c(1, 4, 9)</code> specifies lags of order 1, 3, 9. The number of specified lags has to match the 5th dimension in <code>coefarray</code> . |
| thresholds | A list with p entries, each consisting of a matrix indicating a threshold for each category of the given variable (column) and time point (row). For continuous variable, the matrix has 1 column. |

sds	A $N \times p$ matrix specifying the standard deviation of Gaussian variables (columns) at each time point (rows) If non-Gaussian variables are included in the mVAR model, the corresponding columns are ignored.
type	p vector indicating the type of variable for each column in data. 'g' for Gaussian, 'p' for Poisson, 'c' for categorical.
level	p vector indicating the number of categories of each variable. For continuous variables set to 1.
pbar	If pbar = TRUE, a progress bar is shown.

Details

We sample from the mVAR model by separately sampling from its corresponding p conditional distributions.

Value

A list with two entries:

call	The function call
data	The sampled $n \times p$ data matrix

Author(s)

Jonas Haslbeck <jonashaslbeck@gmail.com>

References

Haslbeck, J., & Waldorp, L. J. (2016). mgm: Structure Estimation for time-varying Mixed Graphical Models in high-dimensional Data. arXiv preprint arXiv:1510.06871.

Examples

```
## Not run:

# a) Set up time-varying mvar model

p <- 6 # Six variables
type <- c('c', 'c', 'c', 'c', 'g', 'g') # 4 categorical, 2 gaussians
level <- c(2, 2, 4, 4, 1, 1) # 2 categoricals with 2 categories, 2 with 5
max_level <- max(level)

lags <- c(1, 3, 9) # include lagged effects of order 1, 3, 9
n_lags <- length(lags)

N <- 5000
```

```

# Specify thresholds
thresholds[[1]] <- matrix(0, ncol=2, nrow=N)
thresholds[[2]] <- matrix(0, ncol=2, nrow=N)
thresholds[[3]] <- matrix(0, ncol=4, nrow=N)
thresholds[[4]] <- matrix(0, ncol=4, nrow=N)
thresholds[[5]] <- matrix(0, ncol=1, nrow=N)
thresholds[[6]] <- matrix(0, ncol=1, nrow=N)

# Specify standard deviations for the Gaussians
sds <- matrix(NA, ncol=6, nrow=N)
sds[,5:6] <- 1

# Create coefficient array
coefarray <- array(0, dim=c(p, p, max_level, max_level, n_lags, N))

# a.1) interaction between continuous 5<-6, lag=3
coefarray[5, 6, 1, 1, 2, ] <- c(rep(.5, N/2), rep(0, N/2))
# a.2) interaction between 1<-3, lag=1
m1 <- matrix(0, nrow=level[2], ncol=level[4])
m1[1, 1:2] <- 1
m1[2, 3:4] <- 1
coefarray[1, 3, 1:level[2], 1:level[4], 1, ] <- m1
# a.3) interaction between 1<-5, lag=9
coefarray[1, 5, 1:level[1], 1:level[5], 3, ] <- c(0, 1)

dim(coefarray)

# b) Sample
set.seed(1)
dlist <- tvmvvarsampler(coefarray = coefarray,
                        lags = lags,
                        thresholds = thresholds,
                        sds = sds,
                        type = type,
                        level = level,
                        pbar = TRUE)

# c) Recover: time-varying mVAR model
set.seed(1)
tvmvar_obj <- tvmvvar(data = dlist$data,
                     type = type,
                     level = level,
                     lambdaSel = 'CV',
                     lags = c(1, 3, 9),
                     estpoints = seq(0, N-9, length=10),
                     bandwidth = .05)

tvmvar_obj$wadj[5, 6, 2, ] # parameter goes down, as specified
tvmvar_obj$wadj[1, 3, 1, ]
tvmvar_obj$wadj[1, 5, 3, ]

```

```
## End(Not run)
```

Index

*Topic **\textasciitildekwd1**

 tvmvarsampler, [43](#)

*Topic **\textasciitildekwd2**

 tvmvarsampler, [43](#)

*Topic **datasets**

 datasets, [6](#)

*Topic **package**

 mgm-package, [2](#)

autism_data (datasets), [6](#)

autism_data_large (datasets), [6](#)

bwSelect, [3](#)

datasets, [6](#)

fruitfly_data (datasets), [6](#)

mgm, [7](#)

mgm-package, [2](#)

mgm_data (datasets), [6](#)

mgmfit (mgm), [7](#)

mgmsampler, [14](#)

mvar, [19](#)

mvar_data (datasets), [6](#)

mvarsampler, [25](#)

predict.mgm, [27](#)

print.mgm, [30](#)

PTSD_data (datasets), [6](#)

restingstate_data (datasets), [6](#)

symptom_data (datasets), [6](#)

tv.mgmfit (tvmgm), [31](#)

tv_var.mgm (tvmvar), [39](#)

tvmgm, [31](#)

tvmgmsampler, [35](#)

tvmvar, [39](#)

tvmvarsampler, [43](#)

var.mgm (mvar), [19](#)