Package 'bsem'

August 14, 2020

Type Package

Title Bayesian Structural Equation Models

Version 1.0.0

Maintainer Renato Panaro <renatovp@ime.usp.br>

Description Flexible routines to allow structural equation modeling particular cases using 'rstan' integration. 'bsem' includes Bayesian semi Confirmatory Factor Analysis, Confirmatory Factor Analysis, and Structural Equation Model. VD Mayrink (2013) <doi:10.1214/12-AOAS607>.

License GPL-3

Biarch true

Depends R (>= 3.5.0), coda (>= 0.19-3), lattice (>= 0.20-38), magrittr (>= 1.5), viridis (>= 0.5.1)

Imports visNetwork (>= 2.0.9), shiny (>= 1.5.0), methods, Rcpp (>= 0.12.0), rstan (>= 2.18.1), rstantools (>= 1.5.1), DiagrammeR (>= 1.0.5)

LinkingTo BH (>= 1.66.0), Rcpp (>= 0.12.0), RcppEigen (>= 0.3.3.3.0), rstan (>= 2.18.1), StanHeaders (>= 2.18.0)

Suggests knitr, devtools, roxygen2, testthat (>= 2.1.0), covr, rmarkdown, bayesplot, tidybayes, ggplot2, gridExtra (>= 2.3), shinythemes (>= 1.1.2), ggfortify (>= 0.4.1), shinyjs (>= 1.1), shinycssloaders (>= 0.3), plotly (>= 4.9.2.1), ggridges (>= 0.5.2), fmsb (>= 0.5.3), visdat (>= 0.5.3), DT (>= 0.14), tidyr (>= 1.1.0), dplyr (>= 1.0.0), reshape2 (>= 1.4.4)

Encoding UTF-8

LazyData true

SystemRequirements GNU make

BugReports https://github.com/rvpanaro/bsem/issues

VignetteBuilder knitr

RoxygenNote 7.1.1

URL https://github.com/rvpanaro/bsem

NeedsCompilation yes

Author Renato Panaro [aut, cre] (<https://orcid.org/0000-0002-1903-2091>,
https://rvpanaro.github.io/),
Vinícius Mayrink [aut] (<https://orcid.org/0000-0002-5683-8326>,
http://www.est.ufmg.br/~vdinizm/),
Marcelo Costa [aut]
(http://somos.ufmg.br/professor/marcelo-azevedo-costa)

Repository CRAN

Date/Publication 2020-08-14 13:40:03 UTC

R topics documented:

`bsem-package` *The 'bsem' package.*

Description

The bsem package allows Bayesian analysis for particular cases of structural equation models (SEMs) based on rstan integration. Examples include confirmatory factor analysis and confirmatory SEM. The full SEM model (outer and inner models), enables the evaluation of user-defined latent variables along with the analysis of established linear relationships among the latent scores.

References

Mayrink, V. D., & Lucas, J. E. (2013). Sparse latent factor models with interactions: Analysis of gene expression data. The Annals of Applied Statistics, 7(2), 799-822.

Mayrink, V. D., & Lucas, J. E. (2015). Bayesian factor models for the detection of coherent patterns in gene expression data. Brazilian Journal of Probability and Statistics, 29(1), 1-33.

arrayplot *Array Plotting*

Description

Graphical representation for matrix class objects using lattice package

Usage

```
arrayplot(
  y,
  mini = -max(abs(min(y)), max(y)),
  maxi = max(abs(min(y)), max(y)),
  colors = 1,
  ylab = "",
  xlab = "",
  main = ""
)
```

Arguments

y	a numerical data.frame or matrix object
mini,	minimum color range (might cause blank spots if misspecified)
maxi,	maximum color range (might cause blank spots if misspecified)
colors	1 = blue/red, 0 = white/black
ylab	y axis label passed to `lattice::levelplot` function
xlab	x axis label passed to `lattice::levelplot` function
main	plot title passed to `lattice::levelplot` function

Author(s)

VD Mayrink

See Also

simdata, summary.bsem, plot.bsem, sem, runShiny

Examples

```
dt <- simdata()

arrayplot(dt$real$alpha)
arrayplot(dt$real$alpha, colors = 0)
```

`plot.bsem` *'bsem' object plot*

Description

'bsem' object plot

Usage

```
## S3 method for class 'bsem'
plot(x, digits = 2, fontsize = 15, width = 5, size = 10, ...)
```

Arguments

`x`	an object of class bsem
`digits`	number of digits to display
`fontsize`	edge fontsize (estimates character size)
`width`	lines width
`size`	nodes size (for both: ellipses and boxes)
`...`	further arguments passed to `visNetwork::visNetwork` function

Value

a visNetwork graph:

The ellipsoidal nodes represent the latent variables

The boxes represent the manifest variables

The dashed lines represent the linear relations between latent scores or between latent scores and exogenous variables

The solid lines represent the relationship between the manifest an latent variables

The recursive solid lines refers to the error variance estimate of each manifest or exogenous variable

Author(s)

RVPanaro

See Also

`sem`, `simdata`, `arrayplot`, `summary.bsem`, `print.bsem`

Examples

```
dt <- bsem::simdata()
names(dt)


semfit <- bsem::sem(
  data = dt$data,
  blocks = dt$blocks,
  paths = dt$paths,
  exogenous = dt$exogenous,
  signals = dt$signals,
  cores = 1
)
plot(semfit)
```

`print.bsem` *'bsem' object print*

Description

'bsem' object print

Usage

```
## S3 method for class 'bsem'
print(x, digits = 4, ...)
```

Arguments

`x`	an object of class bsem
`digits`	number of digits to display
`...`	further arguments passed to or from print methods

Value

none

See Also

`simdata`, `arrayplot`, `summary.bsem`, `sem`, `runShiny`

runShiny *Run a shiny app*

Description

An introduction to the bsem package @details

signals this option is not allowed, it is only available in `sem`

stanfit S4 object of class stanfit

Usage

```
runShiny()
```

Author(s)

RV Panaro

See Also

`sem`, `simdata`, `arrayplot`, `summary.bsem`, `print.bsem`

Examples

```
## Not run:
library("bsem")
runShiny()

## End(Not run)
```

sem *sem: The SEM Function*

Description

Structural Equation Models (SEM) and particular cases using rstan interface

Usage

```
sem(
  data,
  blocks,
  paths,
  exogenous,
  signals,
```

```
  row_names = rownames(data),
 prior_specs = list(beta = c("normal(0,1)"), sigma2 = c("inv_gamma(2.1, 1.1)"), gamma0
  = c("normal(0,1)"), gamma = c("normal(0,1)"), tau2 = c("inv_gamma(2.1, 1.1)")),
  cores = parallel::detectCores(),
  pars = c("alpha", "lambda", "sigma2"),
  iter = 2000,
  chains = 4,
  scaled = FALSE,
  verbose = FALSE,
  refresh = 100,
  ...
)
```

Arguments

data	a mandatory 'matrix' object where the columns are variables and the rows are observations
blocks	a mandatory named list of colnames (or integers in 1:ncol(data)) indicating the manisfest variables corresponding to each block; generic names are assumed for latent variables internally if not defined
paths	list referring to the inner model paths; a list of characters or integers referring to the scores relationship; the jth first latent variable are explained if names(paths) is NULL
exogenous	list referring to the inner model exogenous; a list of characters or integers referring to relationship between exogenous and latent variables; the lth first columns are explained if names(exogenous) is NULL
signals	list referring to the signals of the factor loadings initial values; must be true: (length(signals) == length(blocks)) && (lengths(signals) == lengths(blocks)); (not allowed in runShiny)
row_names	optional identifier for the observations (observation = row)
prior_specs	prior settings for the Bayesian approach; only 'normal' and 'cauchy' for gamma0, gamma and beta; 'gamma', 'lognormal' and 'inv_gamma' for sigma2 and tau2 are available, those prior specifications are ignored if not needed (FA or SEM)
cores	number of core threads to be used
pars	allows parameters to omitted in the outcome; options are any subset of default c("alpha", "lambda", "sigma2")
iter	number of iterations
chains	number of chains
scaled	logical; indicates whether to center and scale the data; default FALSE
verbose	logical; see sampling; default FALSE
refresh	defaults to 100; see sampling;
...	further arguments passed to Stan such as warmup, adapt_delta and others, see sampling.

Details

Fits the SEM to specific data

Consider:

- the outer model as: – outer blocks:

$$X_{pxn} = \alpha_{pxk}\lambda_{kxn} + \epsilon_{pxn}$$

where X is the data matrix with variables in the rows and sample elements in the columns, α_{pxj} is the column vector of loadings for the jth latent variable and λ_{jxn} is the row vector of scores for the jth unobserved variable, $j = 1, \ldots, k$. Normality is assumed for the errors as ϵ_{ij} $N(0, \sigma_i^2)$ for $i = 1, \ldots, p$.

- the inner model as:

– inner paths:

$$\lambda_{jxn} = \beta\lambda^(-j) + \nu$$

where β is a column vector of constant coefficients and $\lambda^(-j)_{(k-1)xn}$ represents a subset of the matrix of scores, i.e. at least excluding the jth row scores. The error assumes ν_j $N(0, 1)$.

– inner exogenous:

$$Y_{lxn} = \gamma_0 + \gamma\lambda + \xi$$

where γ is a column vector of constant coefficients and γ_0 is the intercept. λ_{kxn} is the matrix of scores and the error assumes ξ_l $N(0, \tau_l^2)$.

Value

An object of class `bsem`; a list of 14 to 19:

stanfit S4 object of class stanfit

posterior the list of posterior draws separated by chains

model character; pointer to pre-defined stan model

mean_alpha matrix of factor loadings posterior means

mean_lambda matrix of factor scores posterior means

mean_sigma2 vector of error variances posterior means

mean_beta vector of regression coefficients posterior means

mean_tau2 vector of inner paths error variances posterior means

mean_gamma vector of inner paths regression coefficients posterior means

mean_gamma0 vector of inner paths intercept posterior means

stats posterior descriptives statistics

blocks list of blocks

paths list of paths

credint Highest posterior density intervals (HPD)

h vector of posterior communalities

PTVE vector of total variance proportions

R2 adjusted coefficient of determination

SQE explained sums of squares

SQT total sums of squares

Author(s)

Renato Panaro

See Also

plot.bsem, simdata, arrayplot, summary.bsem, print.bsem

Examples

```
dt <- bsem::simdata()
names(dt)

semfit <- bsem::sem(
  data = dt$data,
  blocks = dt$blocks,
  paths = dt$paths,
  exogenous = dt$exogenous,
  signals = dt$signals,
  cores = 1
)
summary(semfit)
```

---

simdata *Simulated data*

Description

Function to generate artificial data from a structural equation model

Usage

```
simdata(
  paths = list(3:4, 4:5),
  blocks = list(1:3, 4:7, 8:10, 11:16, 17:21),
  sigma2 = runif(Nv, 0.1, 0.9),
  exogenous = list(1:2),
  beta = list(c(1, -0.5), c(-1, 0.5)),
  gamma0 = list(c(1.5)),
  gamma = list(c(0.5, -1)),
  tau2 = list(c(0.49)),
  Nv = 21,
  Ne = 30,
  Nna = 0,
  save = FALSE,
  name = "dt"
)
```

Arguments

paths	list referring to the inner model paths; a list of integers referring to the scores relationship; the jth first latent variable are explained
blocks	list of integers in 1:ncol(data) indicating the manisfest variables corresponding to each block
sigma2	outer model error variances
exogenous,	minimum color range (might cause blank spots if misspecified)
beta	inner paths regression coefficients
gamma0	inner exogenous intercept
gamma	inner exogenous regression coefficients
tau2	inner exogenous error variance
Nv	number of endogenous variables in the database (default = 21)
Ne	number of sample elements in the database (default = 30)
Nna	number of missing observations in each database (default = 0)
save	logical (default = FALSE); whether to save in the local filepath
name	rdata name used if save = TRUE

See Also

plot.bsem, sem, arrayplot, summary.bsem, print.bsem

Examples

```
dt <- bsem::simdata()

arrayplot(dt$real$alpha)
arrayplot(dt$real$alpha, colors = 0)
arrayplot(dt$real$alpha, colors = 1)
arrayplot(dt$real$alpha, -4, 4)
```

---

summary.bsem *'bsem' object summary*

Description

'bsem' object summary

Usage

```
## S3 method for class 'bsem'
summary(object, digits = 4, ...)
```

Arguments

`object`	an object of class spbp
`digits`	number of digits to display
`...`	further arguments passed to or from summary methods

Value

none

Index