Package 'shapeNA'

November 3, 2020

Title M-Estimation of Shape for Data With Missing Values

Version 0.0.0.9000

Description This package allows to compute power M-estimators for shape and location. Power M-estimators generalize both the empirical covariance estimates and Tyler's M-estimate of scatter. Methods for both complete data, as well as data with missing values are supplied.
Additionally, the package contains some basic plotting functions, which aid in visualization of the results.

License GPL-3

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

Depends R (>= 3.6)

Imports graphics, stats, qpdf

Suggests knitr, rmarkdown, mvtnorm, mice

VignetteBuilder knitr

Author Katharina Riemer [cre, aut],
Gabriel Frahm [aut] (<https://orcid.org/0000-0001-7507-730X>),
Klaus Nordhausen [aut] (<https://orcid.org/0000-0002-3758-8501>),
Una Radojicic [aut] (<https://orcid.org/0000-0003-0329-0595>)

Maintainer Katharina Riemer <katharinemer.maintainer@gmail.com>

R topics documented:

`barplotMissProp` *Barplot Showcasing Missingness Proportion of the Original Data*

Description

Visualize the proportion of missingness per variable in a barplot.

Usage

```
barplotMissProp(obj, sortNA = FALSE, ...)
```

Arguments

`obj`	A `shapeNA` object
`sortNA`	A logical. If `FALSE` the original variable order is kept, else the variables are ordered from least to most missingness
`...`	Additional graphical arguments for `barplot`

See Also

`barplot`

Examples

```
S <- toeplitz(seq(1, 0.1, length.out = 3))
x <- mvtnorm::rmvt(100, S, df = 5)
y <- mice::ampute(x, mech='MCAR')$amp
res <- classicShapeNA(y)
barplotMissProp(res)
```

naBlocks *Reorder Data With Missing Values*

Description

Reorder a data set with NA entries to form blocks of missing values. The resulting data will have increasing missingness along the rows and along the columns. The rows are ordered s.t. the first block consists of complete observations, and the following blocks are ordered from most frequent missingness pattern to least frequent missingness pattern. This method may fail, as it has been designed as a preprocessing step for shape estimations.

Usage

```
naBlocks(data, cleanup = TRUE, plot = FALSE)
```

Arguments

data	A matrix with NA values
cleanup	A logical flag. If TRUE, observations with less than 2 responses are discarded
plot	A logical flag. If true, a plot of the missingness pattern is produced.

Value

a naBlocks object, which is a list with

- data the reordered data matrix
- permutation the permutation of the columns, which was applied to reorder the columns according to the number of NAs
- rowPermutation the permutation of the rows, which generates the blocks
- N a vector of all row indices. Each row number points to the beginning of a new missingness pattern
- D a vector specifying the missingness pattern for each block.
- P a vector specifying the number of observed variables per block.
- kn a vector specifying the percentage of observed responses per variable.

plot.naBlocks *Plot Missingness Pattern of Data*

Description

Plot Missingness Pattern of Data

Usage

```
## S3 method for class 'naBlocks'
plot(x, ...)
```

Arguments

x	A naBlocks object
...	additional parameters

Examples

```
x <- mvtnorm::rmvt(100, toeplitz(seq(1, 0.1, length.out = 3)), df = 5)
y <- mice::ampute(x, mech='MCAR')$amp
res <- classicShapeNA(y)
plot(res$naBlocks)
```

plot.shapeNA *Visualization of Shape Estimate*

Description

Plot each matrix entry as a cell, with dark cells indicating high values and light values indicate small values.

Usage

```
## S3 method for class 'shapeNA'
plot(x, message = TRUE, ...)
```

Arguments

x	A shapeNA oopbject
message	A logical, If TRUE, the percentage of observed values per variable is printed in the console.
...	Additional parameters.

Examples

```
x <- mvtnorm::rmvt(100, toeplitz(seq(1, 0.1, length.out = 3)), df = 5)
y <- mice::ampute(x, mech='MCAR')$amp
res <- tylerShapeNA(y)
plot(res)
```

---

powerShape *Compute M-Estimators of Shape for Data Without Missing Values*

Description

powerShape, tylerShape and classicShape compute power M-estimators of shape. Using powerShape and classicShape it is even possible to compute M-estimators of covariance matrices. These functions also compute estimates of location if no center is supplied.

Usage

```
powerShape(x, alpha, center = NULL,
    normalization = c("det", "trace", "one"), maxiter = 1e4, eps = 1e-6)

tylerShape(x, center = NULL,
    normalization = c("det", "trace", "one"), maxiter = 1e4, eps = 1e-6)

classicShape(x, center = NULL,
    normalization = c("det", "trace", "one"), maxiter = 1e4, eps = 1e-6)
```

Arguments

x	A numeric data matrix or data.frame without missing data.
alpha	Tail index, a numeric value from the interval [0, 1]. Determines the power function. For more information see 'Details'.
center	An optional vector of the data's center, if NULL the center will be estimated simultaneously to the shape estimate.
normalization	A string, determines scale of returned shape estimate. The possible values are • 'det' s.t. the returned shape estimate has determinant 1. • 'trace' s.t. the returned shape estimate has trace p. • 'one' s.t. the returned shape estimate's first entry is 1.
maxiter	A positive integer, restricting the maximum number of iterations.
eps	A numeric, specifying tolerance level of when the iteration stops.

Details

For multivariate normally distributed data, `classicShape` is an ML-estimator. This is a special case of the power M-estimator with tail index `alpha` = 0 and returns the empirical covariance matrix and the empirical mean vector.

`tylerShape` maximizes the likelihood function after projecting the observed data of each individual onto the unit hypersphere, in which case we obtain an angular central Gaussian distribution. This is a special case of the power M-estimator with tail index `alpha` = 1 and returns Tyler's M-estimator of scatter and an affine equivariant multivariate median.

`powerShape` requires an additional parameter, the so-called tail index `alpha`. For asymptotic normality, this index should be chosen taking into consideration the data. For heavy tailed data, the index should be closer to 1, for light tailed data the index should be chosen closer to 0.

Value

A list with class 'shapeNA' containing the following components:

S the estimated shape matrix.

scale the scale with which the shape matrix may be scaled to obtain a scatter estimate. If `alpha` == 1, then this value is meaningless.

mu the location parameter, either provided by the user or estimated.

alpha the tail index with which the Power M-estimator has been called.

naBlocks `NULL`, since `powerShape` operates only on complete data.

iterations number of computed iterations before convergence.

call the matched call.

References

Tyler, D.E. (1987). A Distribution-Free M-Estimator of Multivariate Scatter. The Annals of Statistics, 15, 234.251. doi:10.1214/aos/1176350263.

Frahm, G., & Jaekel, U. (2010). A generalization of Tyler's M-estimators to the case of incomplete data. Computational Statistics & Data Analysis, 54, 374-393. doi:10.1016/j.csda.2009.08.019.

Frahm, G., Nordhausen, K., & Oja, H. (2020). M-estimation with incomplete and dependent multivariate data. Journal of Multivariate Analysis, 176, 104569. doi:10.1016/j.jmva.2019.104569.

See Also

powerShapeNA, tylerShapeNA and classicShapeNA for the corresponding functions for data with missing values.

Examples

```
## Generate example data
S <- toeplitz(c(1, 0.1))
x <- mvtnorm::rmvt(100, S)
## Compute some M-estimators
```

```
res0 <- classicShape(x, center = c(0, 0))
res1 <- powerShape(x, alpha = 0.67, normalization = 'one')
res2 <- tylerShape(x, normalization = 'trace')
## Get location estimates
res1$mu
res2$mu
## Get shape estimates
res0$S
res1$S
res2$S
## Print summary
summary(res0)
```

---

powerShapeNA *Compute M-Estimators of Shape for Data With Missing Values*

Description

powerShapeNA, tylerShapeNA and classicShapeNA compute power M-estimators of shape for data with missing values. The underlying missingness mechanism should be MCAR. These functions also compute estimates of location if no center is supplied.

Usage

```
powerShapeNA(x, alpha, center = NULL, normalization = c("det", "trace", "one"),
         maxiter = 1e4, eps = 1e-6)

tylerShapeNA(x, center = NULL, normalization = c("det", "trace", "one"),
          maxiter = 1e4, eps = 1e-6)

classicShapeNA(x, center = NULL, normalization = c("det", "trace", "one"),
         maxiter = 1e4, eps = 1e-6)
```

Arguments

x	A data matrix or data.frame with missing data and p > 2 columns. Representing sample from generalized elliptical distribution and MCAR missingness
alpha	Tail index, a numeric value from the interval [0, 1]. Determines the power function. For more information see 'Details'.
center	An optional vector of the data's center, if NULL the center will be estimated simultaneously to the shape estimate.
normalization	A string, determines scale of returned shape estimate. The possible values are • 'det' s.t. the returned shape estimate has determinant 1. • 'trace' s.t. the returned shape estimate has trace p. • 'one' s.t. the returned shape estimate's first entry is 1.
maxiter	A positive integer, restricting the maximum number of iterations.
eps	A numeric, specifying tolerance level of when the iteration stops.

Details

For multivariate normally distributed data, `classicShapeNA` is an ML-estimator. This is a special case of the power M-estimator with tail index `alpha` = 0 and returns the empirical covariance matrix and the empirical mean vector.

`tylerShapeNA` maximizes the likelihood function after projecting the observed data of each individual onto the unit hypersphere, in which case we obtain an angular central Gaussian distribution. This is a special case of the power M-estimator with tail index `alpha` = 1 and returns Tyler's M-estimator of scatter and an affine equivariant multivariate median.

`powerShapeNA` requires an additional parameter, the so-called tail index `alpha`. For asymptotic normality, this index should be chosen taking into consideration the data. For heavy tailed data, the index should be closer to 1, for light tailed data the index should be chosen closer to 0.

Value

A list with class 'shapeNA' containing the following components:

S the estimated shape matrix.

scale the scale with which the shape matrix may be scaled to obtain a scatter estimate. If `alpha` == 1, then this value is meaningless.

mu the location parameter, either provided by the user or estimated.

alpha the tail index with which the Power M-estimator has been called.

naBlocks an `naBlocks` object, with information about the missingness of the data.

iterations number of computed iterations before convergence.

call the matched call.

References

Frahm, G., & Jaekel, U. (2010). A generalization of Tyler's M-estimators to the case of incomplete data. Computational Statistics & Data Analysis, 54, 374-393. doi:10.1016/j.csda.2009.08.019.

Frahm, G., Nordhausen, K., & Oja, H. (2020). M-estimation with incomplete and dependent multivariate data. Journal of Multivariate Analysis, 176, 104569. doi:10.1016/j.jmva.2019.104569.

See Also

powerShape, tylerShape and classicShape for the corresponding functions for data without missing values.

Examples

```
## Generate a data set with missing values
x <- mvtnorm::rmvt(100, toeplitz(seq(1, 0.1, length.out = 3)), df = 5)
y <- mice::ampute(x, mech='MCAR')$amp

## Compute some M-estimators
res0 <- classicShapeNA(y, center = c(0, 0, 0))
res1 <- powerShapeNA(y, alpha = 0.67, normalization = 'one')
```

```
res2 <- tylerShapeNA(y, normalization = 'trace')

## Get location estimates
res1$mu
res2$mu
## Get shape estimates
res0$S
res1$S
res2$S

## Print summary
summary(res0)
## Inspect missingness pattern
plot(res0$naBlocks)
barplotMissProp(res0)
```

`print.naBlocks` *Print Missingness Pattern*

Description

Print the pattern of missingness in the supplied data, as a matrix for 1s, representing a column vector of responses and 0s, representing a column vector of NAs.

Usage

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

Arguments

x	A naBlocks object
...	Additional parameters.

Details

The first row shows the column names. The leftmost column, without column names, shows the number of rows per block and the rightmost column, titled with # shows the number of observed variables in the block.

Examples

```
x <- mvtnorm::rmvnorm(200, mean = c(0, 0))
classicShape(x)
```

print.shapeNA *Print Method for Elements of Class* shapeNA

Description

Print alpha level, shape estimate and center.

Usage

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

Arguments

x	A shapeNA object
...	Additional parameters.

Examples

```
x <- mvtnorm::rmvt(100, toeplitz(seq(1, 0.1, length.out = 3)), df = 5)
res <- tylerShape(x)
res ## equivalent to call print(res)
```

print.summary.shapeNA *Print Method for Class* summary.shapeNA

Description

Print Method for Class summary.shapeNA

Usage

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

Arguments

x	object returned from summary.shapeNA
...	further arguments

Value

invisibly return NULL

Examples

```
obj <- tylerShape(mvtnorm::rmvt(100, diag(3)))
print(summary(obj))
```

shape2scatter *Scatter Estimates from shapeNA object*

Description

For Power M-estimates with tail index alpha < 1, the resulting estimate has a scale. For these cases, a scatter estimate can be computed. Results from tylerShape and tylerShapeNA give no scatter estimates.

Usage

```
shape2scatter(obj)
```

Arguments

obj	shapeNA object, resulting from a call to powerShape() and other functions from the same family.

Value

scatter matrix estimate, or NA if alpha == 1

Examples

```
S <- toeplitz(c(1, 0.3, 0.7))
set.seed(123)
x <- mvtnorm::rmvt(100, S, df = 3)
obj <- powerShape(x, alpha = 0.85)
shape2scatter(obj)
```

summary.shapeNA *Summary Method for Class* shapeNA

Description

Summary Method for Class shapeNA

Usage

```
## S3 method for class 'shapeNA'
summary(object, ...)
```

Arguments

object	an object of class shapeNA, usually from a call to powerShape or similar functions
...	further arguments

Examples

```
obj <- tylerShape(mvtnorm::rmvt(100, diag(3)))
summary(obj)
```

Index