Package 'shapeNA'

February 12, 2021

Title M-Estimation of Shape for Data With Missing Values

Version 0.0.0.9000

Description M-estimators of location and shape following the recently
suggested power family (see Frahm, Nordhausen, Oja (2020)
<doi:10.1016/j.jmva.2019.104569>) are provided in the case of complete data
and also when observations have missing values together with functions
aiding their visualization.

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

R topics documented:

---

barplot.shapeNA   *Barplot Showcasing Missingness Proportion of the Original Data*

Description

Visualize the proportion of missingness per variable in a barplot.

Usage

```
## S3 method for class 'shapeNA'
barplot(height, sortNA = FALSE, ...)
```

Arguments

height	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 passed to [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)
barplot(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 such that the first block consists of complete observations, and the following blocks are ordered from most frequent missingness pattern to least frequent missingness pattern.

Usage

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

Arguments

x   A matrix with missing 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.

Details

In case of ties, that is two patterns occur with the same frequency, the block whose pattern occurs first will be ordered in front of the other block.

This method may fail if the missingness is too strong or if the number of observations is too low (the number of observations has to exceed the number of variables), as it has been designed as a preprocessing step for shape estimations.

Value

A list of class naBlocks with components:

x   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

Function to visualize the missing patterns for objects of class naBlocks.

Usage

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

Arguments

x	A naBlocks object.
...	Additional parameters passed on to [rect]().

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

Function to visualize the shape matrix from objects of class shapeNA by plotting a heatmap where light colored cells indicate small values and dark colored cells indicate high 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 passed to [image]().

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)
## default plot
plot(res)
## plot result in gray scale - reverse order to get a palette starting
## with the lightest instead of the darkest color
plot(res, col = gray.colors(9, rev = TRUE))
```

---

powerShape   *M-estimators of Shape from the Power Family.*

Description

Power M-estimators of shape and location were recently suggested in Frahm et al. (2020). They have a tuning parameter alpha taking values in [0,1]. The extreme case alpha = 1 corresponds to Tyler's shape matrix and alpha = 0 to the classical covariance matrix. The special cases have own, more efficient functions. If the true location is known, it should be supplied as center, otherwise it is estimated simultaneously with the shape.

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 with the shape estimate.
normalization	A string determining how the shape matrix is standardized. The possible values are • 'det' such that the returned shape estimate has determinant 1. • 'trace' such that the returned shape estimate has trace. • 'one' such that the returned shape estimate's top left entry ($S[1,1]$) is 1.
maxiter	A positive integer, restricting the maximum number of iterations.
eps	A numeric, specifying the tolerance level of when the iteration stops.

Details

These functions assume, that the data were generated from an generalized elliptical distribution.

For multivariate normally distributed data, `classicShape` is the maximum likelihood 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.

The function `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.

The function `powerShape` requires an additional parameter, the so-called tail index `alpha`. 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 `NA`, as Tyler's shape matrix has no natural scale.
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., 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   *M-estimators of Shape from the Power Family when Data is Missing*

Description

Power M-estimators of shape and location were recently suggested in Frahm et al. (2020). They have a tuning parameter alpha taking values in [0,1]. The extreme case alpha = 1 corresponds to Tyler's shape matrix and alpha = 0 to the classical covariance matrix. The special cases have own, more efficient functions. If the true location is known, it should be supplied as center, otherwise it is estimated simultaneously with the shape.

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.
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 with the shape estimate.
normalization	A string determining how the shape matrix is standardized. The possible values are

- 'det' such that the returned shape estimate has determinant 1.
- 'trace' such that the returned shape estimate has trace.
- 'one' such that the returned shape estimate's top left entry (S[1,1]) is 1.

maxiter	A positive integer, restricting the maximum number of iterations.
eps	A numeric, specifying tolerance level of when the iteration stops.

Details

These functions assume, that the data were generated from an generalized elliptical distribution and that the missingness mechanism is MCAR.

For multivariate normally distributed data, `classicShapeNA` is the maximum likelihood 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.

The function `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.

The function `powerShapeNA` requires an additional parameter, the so-called tail index alpha. 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 will be NA, as Tyler's shape matrix has no natural scale.

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](https://doi.org/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](https://doi.org/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)
barplot(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	An naBlocks object.
...	Additional parameters passed to [print]().

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

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

Description

Prints the used value of alpha and the estimated shape and location for objects of class shapeNA.

Usage

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

Arguments

x	A shapeNA object
...	Additional parameters passed to lower level [print]().

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, which will be ignored.

Value

Invisibly return NULL.

Examples

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

---

shape2scatter *Scatter Estimates from* shapeNA *Objects*

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 only NA if alpha == 1.

Examples

```
S <- toeplitz(c(1, 0.3, 0.7))
set.seed(123)
x <- mvtnorm::rmvt(100, S, df = 3)
obj_det <- powerShape(x, alpha = 0.85, normalization = 'det')
shape2scatter(obj_det)
obj_tr <- powerShape(x, alpha = 0.85, normalization = 'trace')
shape2scatter(obj_tr)
obj_one <- powerShape(x, alpha = 0.85, normalization = 'one')
shape2scatter(obj_one)
```

---

summary.shapeNA *Summary Method For Class* shapeNA

Description

Summary methods for objects from 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, which will be ignored.

Examples

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

Index