Package: fastM (via r-universe)

October 13, 2024

Type Package

Title Fast Computation of Multivariate M-Estimators

Version 0.0-4

Date 2018-05-23

Author Lutz Duembgen, Klaus Nordhausen, Heike Schuhmacher

Maintainer Klaus Nordhausen <klaus.nordhausen@tuwien.ac.at>

Imports Rcpp (>= 0.11.0)

LinkingTo Rcpp, RcppArmadillo

Description Implements the new algorithm for fast computation of M-scatter matrices using a partial Newton-Raphson procedure for several estimators. The algorithm is described in Duembgen, Nordhausen and Schuhmacher (2016) <doi:10.1016/j.jmva.2015.11.009>.

License GPL (>= 2)

NeedsCompilation yes

Repository CRAN

Date/Publication 2018-05-24 12:45:41 UTC

Contents

fastM-package *Fast Computation of Multivariate M-Estimators*

Description

Implements the new algorithm for fast computation of M-scatter matrices using a partial Newton-Raphson procedure for several estimators. The algorithm is described in Duembgen, Nordhausen and Schuhmacher (2016) <doi:10.1016/j.jmva.2015.11.009>.

Details

Multivariate M-estimators are usually computed using a fixed-point algorithm. As shown in Duembgen et al. (2016) a partial Newton-Raphson procedure applied to the second order Taylor expansion of the target function can make the computation considerably faster. We implement this new algorithm for the multivariate M-estimator of location and scatter using weights coming from the multivariate t-distribution (Kent et al., 1994), its symmetrized version, Tyler's shape matrix (Tyler, 1987) and Duembgen's shape matrix (Duembgen, 1998). For the symmetrized M-estimators we work with incomplete U-statistics to accelerate our procedures initially.

Author(s)

Lutz Duembgen, Klaus Nordhausen, Heike Schuhmacher

Maintainer: Klaus Nordhausen <klaus.nordhausen@tuwien.ac.at>

References

Duembgen, L. (1998), On Tyler's M-functional of scatter in high dimension, Annals of Institute of Statistical Mathematics, ***50**, 471–491.*

Duembgen, L., Nordhausen, K. and Schuhmacher, H. (2016), New algorithms for M-estimation of multivariate location and scatter, Journal of Multivariate Analysis, ***144**, 200–217. doi: 10.1016/j.jmva.2015.11.009*

Kent, J.T., Tyler, D.E. and Vardi, Y. (1994), A curious likelihood identity for the multivariate t-distribution, Communications in Statistics, Theory and Methods, ***23**, 441–453.*

Tyler, D.E. (1987), A distribution-free M-estimator of scatter, Annals of Statistics, ***15**, 234–251.*

DUEMBGENshape *Duembgen's Shape Matrix*

Description

Iterative algorithm to estimate Duembgen's shape matrix using a partial Newton-Raphson approach.

Usage

```
DUEMBGENshape(X, nmax = 500, eps = 1e-06, maxiter = 100, perm = FALSE)
```

Arguments

X	numeric data matrix or dataframe. Missing values are not allowed.
nmax	integer, if the sample size n (number of rows of X) is smaller than nmax, then all n(n-1)/2 pairwise differences will be computed and used in the algorithm. If n is larger, then the algorithm avoids storing all the pairwise differences and is more memory efficient.
eps	convergence tolerance, which means that the algorithm stops when the Frobenius norm of the gradient is smaller than eps.
maxiter	maximum number of iterations.
perm	logical. If TRUE the rows of X will be randomly permuted before starting the computations. See details.

Details

The estimate is based on the new fast algorithm described in Duembgen et al. (2016). Note that Duembgen's shape matrix is standardized such that it has determinant 1.

The function does not check if there are several identical observations. In that case the function will fail.

To get a good initial value for the algorithm, the estimator is first computed based on the pairwise differences of successive observations. Therefore the order of the rows of X is supposed to be random. If this is not the case, the data should be first permuted using the argument perm.

In case maxiter is reached before convergence, the estimate at that iteration is returned and a warning is given.

Value

A list containing:

Sigma	Estimated shape matrix.
iter	Number of iterations of the algorithm.

Author(s)

Lutz Duembgen and Klaus Nordhausen

References

Duembgen, L. (1998), On Tyler's M-functional of scatter in high dimension, Annals of Institute of Statistical Mathematics, ***50***, *471–491.*

Duembgen, L., Nordhausen, K. and Schuhmacher, H. (2016), New algorithms for M-estimation of multivariate location and scatter, Journal of Multivariate Analysis, ***144***, *200–217. doi: 10.1016/j.jmva.2015.11.009*

See Also

tyler.shape

Examples

```
DUEMBGENshape(longley)
DUEMBGENshape(longley, nmax=10)
# compare to
# library(ICSNP)
# duembgen.shape(longley)
```

MVTMLE *M-estimator of Location and Scatter Using Weights Coming From the Multivariate t-distribution*

Description

The algorithm of this function is based on a partial Newton approach and should be faster than the traditional fixed-point algorithm. If the data follows a multivariate t-distribution with the correctly specified degrees of freedom this function gives the maximum likelihood estimate of location and scatter.

Usage

```
MVTMLE(X, nu = 1, location = TRUE, eps = 1e-06, maxiter = 100)
```

Arguments

X	numeric data matrix or dataframe. Missing values are not allowed.
nu	assumed degrees of freedom of the t-distribution. Default is '1' which corresponds to the Cauchy distribution.
location	logical or numeric. If FALSE, it is assumed that the scatter should be computed wrt to the origin. If TRUE the location will be estimated and if it is a numeric vector it will be computed wrt to this vector.
eps	convergence tolerance, which means that the algorithm stops when the Frobenius norm of the gradient is smaller than eps.
maxiter	maximum number of iterations.

Details

The assumed degree of freedom nu must be at least 1 when the location and scatter should be estimated. If only the scatter is to be estimated, then it needs to be larger than zero only.

In case maxiter is reached before convergence, the estimate at that iteration is returned and a warning is given.

Value

A list containing:

mu	Estimated location if location=TRUE, otherwise the user specified location.
Sigma	Estimated scatter matrix.
iter	Number of iterations of the algorithm.

Author(s)

Lutz Duembgen and Klaus Nordhausen

References

Kent, J.T., Tyler, D.E. and Vardi, Y. (1994), A curious likelihood identity for the multivariate t-distribution, Communications in Statistics, Theory and Methods, ***23***, *441–453.*

Duembgen, L., Nordhausen, K. and Schuhmacher, H. (2016), New algorithms for M-estimation of multivariate location and scatter, Journal of Multivariate Analysis, ***144***, *200–217. doi: 10.1016/j.jmva.2015.11.009*

See Also

`cov.trob`, `tM`, `MVTMLEsymm`

Examples

```
MVTMLE(longley)
# compare to
# library(ICS)
# tM(longley)
# library(MASS)
# cov.trob(longley, nu=1, tol = 1e-06, maxit = 100)
```

`MVTMLE0r` *Different Algorithms for M-estimation of Scatter Using Weights Coming From the Multivariate t-distribution*

Description

The functions below are only for comparison purposes and are all written in R. Each function corresponds to a different algorithm for the scatter only problem for M-estimation using weights coming from the multivariate t-distribution.

Usage

```
MVTMLE0r(X, nu = 0, delta = 1e-06, prewhitened = FALSE, steps = FALSE)
MVTMLE0r_FP(X, nu = 0, delta = 1e-06, steps = FALSE)
MVTMLE0r_FP0(X, nu = 0, delta = 1e-06, steps = FALSE)
MVTMLE0r_G(X, nu = 0, delta = 1e-06, steps = FALSE)
MVTMLE0r_CG(X, nu = 0, delta = 1e-06, steps = FALSE)
```

Arguments

X	numeric data matrix or dataframe. Missing values are not allowed.
nu	assumed degrees of freedom of the t-distribution. Must be 0 or larger. Default is '0' which corresponds to Tyler's shape matrix.
delta	convergence tolerance, which means that the algorithms stop when the Frobenius norm of the gradient is smaller than delta.
prewhitened	logical. Is the data prewhitened or not.
steps	logial. If TRUE intermediate results are printed on the console.

Details

All functions are implemented in R and their purpose is only for demonstration of the differences of the different algorithms. The function MVTMLE0r uses the recommended partial Newton approach as implemented also in (MVTMLE and TYLERshape). MVTMLE0r_FP and MVTMLE0r_FP0 are fixed-point algorithms where MVTMLE0r_FP iterates the fixed point equation with 'iterative whitening' of the data. The function MVTMLE0r_G uses a gradient approach and MVTMLE0r_CG a conjugate gradient approach. Note that MVTMLE0r_CG does not check if the 'next' step is really an improvement and that all functions compute the scatter wrt to the origin.

All functions have a hard coded maximum number of iterations of 1000. If that is reached the functions returns the final estimate, however without a warning.

For general purposes we recommend the functions MVTMLE and TYLERshape.

Value

A list containing at least:

S	Estimated scatter matrix (or shape matrix if nu=0).
iter	Number of iterations of the algorithm.

Author(s)

Lutz Duembgen and Klaus Nordhausen

References

Duembgen, L., Nordhausen, K. and Schuhmacher, H. (2016), New algorithms for M-estimation of multivariate location and scatter, Journal of Multivariate Analysis, ***144***, *200–217. doi: 10.1016/j.jmva.2015.11.009*

See Also

MVTMLE, TYLERshape

Examples

```
MVTMLE0r(longley,nu=1)
MVTMLE0r_FP(longley,nu=1)
MVTMLE0r_FP0(longley,nu=1)
MVTMLE0r_G(longley,nu=1)
MVTMLE0r_CG(longley,nu=1)
```

`MVTMLEsymm` *Symmetrized M-estimator of Scatter Using Weights Coming From the t-distribution*

Description

Based on a partial Newton-Raphson approach offers this function two ways to compute the symmetrized M-estimator of scatter. The user can choose if all pairwise differences are choosen and stored in the memory or if the computation and storage of this large matrix is to be avoided.

Usage

```
MVTMLEsymm(X, nu = 1, nmax = 500, eps = 1e-06, maxiter = 100, perm = FALSE)
```

Arguments

`X`	numeric data matrix or dataframe with more rows than columns. Missing values are not allowed.
`nu`	assumed degrees of freedom of the t-distribution, must be larger than 0. Default is '1'.
`nmax`	integer, if the sample size n (number of rows of `X`) is smaller than `nmax`, then all n(n-1)/2 pairwise differences will be computed and used in the algorithm. If n is larger, then the algorithm avoids storing all the pairwise differences and is more memory efficient.
`eps`	convergence tolerance, which means that the algorithm stops when the Frobenius norm of the gradient is smaller than eps.
`maxiter`	maximum number of iterations.
`perm`	logical. If TRUE the rows of `X` will be randomly permuted before starting the computations. See details.

Details

To get a good initial value for the algorithm, the estimator is first computed based on the pairwise differences of successive observations. Therefore the order of the rows of `X` is supposed to be random. If this is not the case, the data should be first permuted using the argument `perm`.

In case `maxiter` is reached before convergence, the estimate at that iteration is returned and a warning is given.

Value

A list containing:

`Sigma`	Estimated scatter matrix.
`iter`	Number of iterations of the algorithm.

Author(s)

Lutz Duembgen and Klaus Nordhausen

References

Duembgen, L., Nordhausen, K. and Schuhmacher, H. (2016), New algorithms for M-estimation of multivariate location and scatter, Journal of Multivariate Analysis, ***144***, *200–217. doi: 10.1016/j.jmva.2015.11.009*

See Also

MVTMLE

Examples

```
MVTMLEsymm(longley)
MVTMLEsymm(longley, nmax=10)
```

TYLERshape *Tyler's Shape Matrix*

Description

Iterative algorithm to estimate Tyler's shape matrix using a partial Newton-Raphson approach.

Usage

```
TYLERshape(X, location = TRUE, eps = 1e-06, maxiter = 100)
```

Arguments

`X`	numeric data matrix or dataframe. Missing values are not allowed.
`location`	logical or numeric. If FALSE, it is assumed that the scatter should be computed wrt to the origin. If TRUE the location will be estimated as the mean vector and if it is a numeric vector it will be computed wrt to the given vector.
`eps`	convergence tolerance, which means that the algorithm stops when the Frobenius norm of the gradient is smaller than eps.
`maxiter`	maximum number of iterations.

Details

The estimate is based on the new fast algorithm described in Duembgen et al. (2016). Note that Tyler's shape matrix is standardized such that it has determinant 1.

The function does not check if there are observations equal to the mean (if `location=TRUE`), to the provided location vector or to the origin (if `location=FALSE`). In these cases the function will fail.

In case `maxiter` is reached before convergence, the estimate at that iteration is returned and a warning is given.

Value

A list containing:

`mu`	Estimated location if `location=TRUE`, otherwise the user specified location.
`Sigma`	Estimated shape matrix.
`iter`	Number of iterations of the algorithm.

Author(s)

Lutz Duembgen and Klaus Nordhausen

References

Tyler, D.E. (1987), A distribution-free M-estimator of scatter, Annals of Statistics, ***15***, *234–251.*

Duembgen, L., Nordhausen, K. and Schuhmacher, H. (2016), New algorithms for M-estimation of multivariate location and scatter, Journal of Multivariate Analysis, ***144***, *200–217. doi: 10.1016/j.jmva.2015.11.009*

See Also

`tyler.shape`

Examples

```
TYLERshape(longley)
# compare to
# library(ICSNP)
# tyler.shape(longley)

TYLERshape(longley, location=FALSE)
# compare to
# library(ICSNP)
# tyler.shape(longley, location=0)
```

Index