

Package ‘fitHeavyTail’

July 22, 2025

Title Mean and Covariance Matrix Estimation under Heavy Tails

Version 0.2.0

Date 2023-5-1

Description Robust estimation methods for the mean vector, scatter matrix, and covariance matrix (if it exists) from data (possibly containing NAs) under multivariate heavy-tailed distributions such as angular Gaussian (via Tyler's method), Cauchy, and Student's t distributions. Additionally, a factor model structure can be specified for the covariance matrix. The latest revision also includes the multivariate skewed t distribution. The package is based on the papers: Sun, Babu, and Palomar (2014); Sun, Babu, and Palomar (2015); Liu and Rubin (1995); Zhou, Liu, Kumar, and Palomar (2019); Pascal, Ollila, and Palomar (2021).

Maintainer Daniel P. Palomar <daniel.p.palomar@gmail.com>

URL <https://CRAN.R-project.org/package=fitHeavyTail>,
<https://github.com/convexfi/fitHeavyTail>,
<https://www.danielppalomar.com>,
<https://doi.org/10.1109/TSP.2014.2348944>,
<https://doi.org/10.1109/TSP.2015.2417513>,
<https://doi.org/10.23919/EUSIPCO54536.2021.9616162>

BugReports <https://github.com/convexfi/fitHeavyTail/issues>

License GPL-3

Encoding UTF-8

RoxygenNote 7.2.1

Imports ICSNP, mvtnorm, ghyp, numDeriv, stats

Suggests ggplot2, reshape2, knitr, rmarkdown, R.rsp, testthat

VignetteBuilder knitr, rmarkdown, R.rsp

NeedsCompilation no

Author Daniel P. Palomar [cre, aut],
Rui Zhou [aut],
Xiwen Wang [aut],
Frédéric Pascal [ctb],
Esa Ollila [ctb]

Repository CRAN

Date/Publication 2023-05-01 14:20:05 UTC

Contents

fitHeavyTail-package	2
fit_Cauchy	3
fit_mvst	5
fit_mvt	8
fit_Tyler	12
nu_OPP_estimator	14
nu_POP_estimator	15
Index	18

fitHeavyTail-package	<i>fitHeavyTail: Mean and Covariance Matrix Estimation under Heavy Tails</i>
----------------------	--

Description

Robust estimation methods for the mean vector, scatter matrix, and covariance matrix (if it exists) from data (possibly containing NAs) under multivariate heavy-tailed distributions such as angular Gaussian (via Tyler’s method), Cauchy, and Student’s t distributions. Additionally, a factor model structure can be specified for the covariance matrix. The latest revision also includes the multivariate skewed t distribution. The package is based on the papers: Sun, Babu, and Palomar (2014); Sun, Babu, and Palomar (2015); Liu and Rubin (1995); Zhou, Liu, Kumar, and Palomar (2019); Pascal, Ollila, and Palomar (2021).

Functions

[fit_Tyler](#), [fit_Cauchy](#), [fit_mvt](#), and [fit_mvst](#).

Help

For a quick help see the README file: [GitHub-README](#).
For more details see the vignette: [CRAN-vignette](#).

Author(s)

Daniel P. Palomar and Rui Zhou

References

- Ying Sun, Prabhu Babu, and Daniel P. Palomar, "Regularized Tyler's Scatter Estimator: Existence, Uniqueness, and Algorithms," IEEE Trans. on Signal Processing, vol. 62, no. 19, pp. 5143-5156, Oct. 2014. <<https://doi.org/10.1109/TSP.2014.2348944>>
- Ying Sun, Prabhu Babu, and Daniel P. Palomar, "Regularized Robust Estimation of Mean and Covariance Matrix Under Heavy-Tailed Distributions," IEEE Trans. on Signal Processing, vol. 63, no. 12, pp. 3096-3109, June 2015. <<https://doi.org/10.1109/TSP.2015.2417513>>
- Chuanhai Liu and Donald B. Rubin, "ML estimation of the t-distribution using EM and its extensions, ECM and ECME," Statistica Sinica (5), pp. 19-39, 1995.
- Chuanhai Liu, Donald B. Rubin, and Ying Nian Wu, "Parameter Expansion to Accelerate EM: The PX-EM Algorithm," Biometrika, Vol. 85, No. 4, pp. 755-770, Dec., 1998
- Rui Zhou, Junyan Liu, Sandeep Kumar, and Daniel P. Palomar, "Robust factor analysis parameter estimation," Lecture Notes in Computer Science (LNCS), 2019. <<https://arxiv.org/abs/1909.12530>>
- Esa Ollila, Daniel P. Palomar, and Frédéric Pascal, "Shrinking the Eigenvalues of M-estimators of Covariance Matrix," IEEE Trans. on Signal Processing, vol. 69, pp. 256-269, Jan. 2021. <<https://doi.org/10.1109/TSP.2020.3043952>>
- Frédéric Pascal, Esa Ollila, and Daniel P. Palomar, "Improved estimation of the degree of freedom parameter of multivariate t-distribution," in Proc. European Signal Processing Conference (EUSIPCO), Dublin, Ireland, Aug. 23-27, 2021. <<https://doi.org/10.23919/EUSIPCO54536.2021.9616162>>

fit_Cauchy

Estimate parameters of a multivariate elliptical distribution to fit data under a Cauchy distribution

Description

Estimate parameters of a multivariate elliptical distribution, namely, the mean vector and the covariance matrix, to fit data. Any data sample with NAs will be simply dropped. The estimation is based on the maximum likelihood estimation (MLE) under a Cauchy distribution and the algorithm is obtained from the majorization-minimization (MM) optimization framework. The Cauchy distribution does not have second-order moments and the algorithm actually estimates the scatter matrix. Nevertheless, assuming that the observed data has second-order moments, the covariance matrix is returned by computing the missing scaling factor with a very effective method.

Usage

```
fit_Cauchy(
  X,
  initial = NULL,
  max_iter = 200,
  ptol = 0.001,
  ftol = Inf,
  return_iterates = FALSE,
  verbose = FALSE
)
```

Arguments

<code>X</code>	Data matrix containing the multivariate time series (each column is one time series).
<code>initial</code>	List of initial values of the parameters for the iterative estimation method. Possible elements include: <ul style="list-style-type: none"> • <code>mu</code>: default is the data sample mean, • <code>cov</code>: default is the data sample covariance matrix, • <code>scatter</code>: default follows from the scaled sample covariance matrix.
<code>max_iter</code>	Integer indicating the maximum number of iterations for the iterative estimation method (default is 100).
<code>ptol</code>	Positive number indicating the relative tolerance for the change of the variables to determine convergence of the iterative method (default is $1e-3$).
<code>ftol</code>	Positive number indicating the relative tolerance for the change of the log-likelihood value to determine convergence of the iterative method (default is Inf, so it is not active). Note that using this argument might have a computational cost as a convergence criterion due to the computation of the log-likelihood (especially when <code>X</code> is high-dimensional).
<code>return_iterates</code>	Logical value indicating whether to record the values of the parameters (and possibly the log-likelihood if <code>ftol</code> < Inf) at each iteration (default is FALSE).
<code>verbose</code>	Logical value indicating whether to allow the function to print messages (default is FALSE).

Value

A list containing possibly the following elements:

<code>mu</code>	Mean vector estimate.
<code>cov</code>	Covariance matrix estimate.
<code>scatter</code>	Scatter matrix estimate.
<code>converged</code>	Boolean denoting whether the algorithm has converged (TRUE) or the maximum number of iterations <code>max_iter</code> has reached (FALSE).
<code>num_iterations</code>	Number of iterations executed.
<code>cpu_time</code>	Elapsed CPU time.
<code>log_likelihood</code>	Value of log-likelihood after converge of the estimation algorithm (if <code>ftol</code> < Inf).
<code>iterates_record</code>	Iterates of the parameters (<code>mu</code> , <code>scatter</code> , and possibly <code>log_likelihood</code> (if <code>ftol</code> < Inf)) along the iterations (if <code>return_iterates</code> = TRUE).

Author(s)

Daniel P. Palomar

References

Ying Sun, Prabhu Babu, and Daniel P. Palomar, "Regularized Robust Estimation of Mean and Covariance Matrix Under Heavy-Tailed Distributions," IEEE Trans. on Signal Processing, vol. 63, no. 12, pp. 3096-3109, June 2015.

See Also

[fit_Tyler](#) and [fit_mvt](#)

Examples

```
library(mvtnorm)      # to generate heavy-tailed data
library(fitHeavyTail)

X <- rmvt(n = 1000, df = 6) # generate Student's t data
fit_Cauchy(X)
```

fit_mvst	<i>Estimate parameters of a multivariate (generalized hyperbolic) skewed t distribution to fit data</i>
----------	---

Description

Estimate parameters of a multivariate (generalized hyperbolic) skewed Student's t distribution to fit data, namely, the location vector, the scatter matrix, the skewness vector, and the degrees of freedom. The estimation is based on the maximum likelihood estimation (MLE) and the algorithm is obtained from the expectation-maximization (EM) method.

Usage

```
fit_mvst(
  X,
  nu = NULL,
  gamma = NULL,
  initial = NULL,
  max_iter = 500,
  ptol = 0.001,
  ftol = Inf,
  PXEM = TRUE,
  return_iterates = FALSE,
  verbose = FALSE
)
```

Arguments

<code>X</code>	Data matrix containing the multivariate time series (each column is one time series).
<code>nu</code>	Degrees of freedom of the skewed t distribution (otherwise it will be iteratively estimated).
<code>gamma</code>	Skewness vector of the skewed t distribution (otherwise it will be iteratively estimated).
<code>initial</code>	List of initial values of the parameters for the iterative estimation method. Possible elements include: <ul style="list-style-type: none"> • <code>nu</code>: default is 4, • <code>mu</code>: default is the data sample mean, • <code>gamma</code>: default is the sample skewness vector, • <code>scatter</code>: default follows from the scaled sample covariance matrix,
<code>max_iter</code>	Integer indicating the maximum number of iterations for the iterative estimation method (default is 500).
<code>ptol</code>	Positive number indicating the relative tolerance for the change of the variables to determine convergence of the iterative method (default is $1e-3$).
<code>ftol</code>	Positive number indicating the relative tolerance for the change of the log-likelihood value to determine convergence of the iterative method (default is Inf , so it is not active). Note that using this argument might have a computational cost as a convergence criterion due to the computation of the log-likelihood (especially when X is high-dimensional).
<code>PXEM</code>	Logical value indicating whether to use the parameter expansion (PX) EM method to accelerating the convergence.
<code>return_iterates</code>	Logical value indicating whether to record the values of the parameters (and possibly the log-likelihood if <code>ftol</code> < Inf) at each iteration (default is <code>FALSE</code>).
<code>verbose</code>	Logical value indicating whether to allow the function to print messages (default is <code>FALSE</code>).

Details

This function estimates the parameters of a (generalized hyperbolic) multivariate Student's t distribution (`mu`, `scatter`, `gamma` and `nu`) to fit the data via the expectation-maximization (EM) algorithm.

Value

A list containing (possibly) the following elements:

<code>mu</code>	Location vector estimate (not the mean).
<code>gamma</code>	Skewness vector estimate.
<code>scatter</code>	Scatter matrix estimate.
<code>nu</code>	Degrees of freedom estimate.
<code>mean</code>	Mean vector estimate:

```

              mean = mu + nu/(nu-2) * gamma
cov           Covariance matrix estimate:
              cov = nu/(nu-2) * scatter + 2*nu^2 / (nu-2)^2 / (nu-4) * gamma*gamma'
converged     Boolean denoting whether the algorithm has converged (TRUE) or the maximum
              number of iterations max_iter has been reached (FALSE).
num_iterations Number of iterations executed.
cpu_time      Elapsed overall CPU time.
log_likelihood_vs_iterations
              Value of log-likelihood over the iterations (if ftol < Inf).
iterates_record
              Iterates of the parameters (mu, scatter, nu, and possibly log_likelihood (if
              ftol < Inf)) along the iterations (if return_iterates = TRUE).
cpu_time_at_iter
              Elapsed CPU time at each iteration (if return_iterates = TRUE).

```

Author(s)

Rui Zhou, Xiwen Wang, and Daniel P. Palomar

References

Aas Kjersti and Ingrid Hobæk Haff. "The generalized hyperbolic skew Student's t-distribution," Journal of financial econometrics, pp. 275-309, 2006.

See Also

[fit_mvt](#)

Examples

```

library(mvtnorm)      # to generate heavy-tailed data
library(fitHeavyTail)

# parameter setting
N <- 5
T <- 200
nu <- 6
mu <- rnorm(N)
scatter <- diag(N)
gamma <- rnorm(N)    # skewness vector

# generate GH Skew t data
taus <- rgamma(n = T, shape = nu/2, rate = nu/2)
X <- matrix(data = mu, nrow = T, ncol = N, byrow = TRUE) +
      matrix(data = gamma, nrow = T, ncol = N, byrow = TRUE) / taus +
      rmvnorm(n = T, mean = rep(0, N), sigma = scatter) / sqrt(taus)

# fit skew t model

```

```

fit_mvst(X)

# setting lower limit for nu (e.g., to guarantee existence of co-skewness and co-kurtosis matrices)
options(nu_min = 8.01)
fit_mvst(X)

```

fit_mvt	<i>Estimate parameters of a multivariate Student's t distribution to fit data</i>
---------	---

Description

Estimate parameters of a multivariate Student's t distribution to fit data, namely, the mean vector, the covariance matrix, the scatter matrix, and the degrees of freedom. The data can contain missing values denoted by NAs. It can also consider a factor model structure on the covariance matrix. The estimation is based on the maximum likelihood estimation (MLE) and the algorithm is obtained from the expectation-maximization (EM) method.

Usage

```

fit_mvt(
  X,
  na_rm = TRUE,
  nu = c("iterative", "kurtosis", "MLE-diag", "MLE-diag-resampled", "cross-cumulants",
        "all-cumulants", "Hill"),
  nu_iterative_method = c("POP", "OPP", "OPP-harmonic", "ECME", "ECM", "POP-approx-1",
                          "POP-approx-2", "POP-approx-3", "POP-approx-4", "POP-exact", "POP-sigma-corrected",
                          "POP-sigma-corrected-true"),
  initial = NULL,
  optimize_mu = TRUE,
  weights = NULL,
  scale_covmat = FALSE,
  PX_EM_acceleration = TRUE,
  nu_update_start_at_iter = 1,
  nu_update_every_num_iter = 1,
  factors = ncol(X),
  max_iter = 100,
  ptol = 0.001,
  ftol = Inf,
  return_iterates = FALSE,
  verbose = FALSE
)

```

Arguments

X	Data matrix containing the multivariate time series (each column is one time series).
---	---

na_rm	Logical value indicating whether to remove observations with some NAs (default is TRUE). Otherwise, the NAs will be imputed at a higher computational cost.
nu	Degrees of freedom of the t distribution. Either a number (>2) or a string indicating the method to compute it: <ul style="list-style-type: none"> "iterative": iterative estimation (with method to be specified in argument nu_iterative_method) with the rest of the parameters (default method); "kurtosis": one-shot estimation based on the kurtosis obtained from the sampled moments; "MLE-diag": one-shot estimation based on the MLE assuming a diagonal sample covariance; "MLE-diag-resampled": like method "MLE-diag" but resampled for better stability.
nu_iterative_method	String indicating the method for iteratively estimating nu (in case nu = "iterative"): <ul style="list-style-type: none"> "ECM": maximization of the Q function [Liu-Rubin, 1995]; "ECME": maximization of the log-likelihood function [Liu-Rubin, 1995]; "OPP": estimator from paper [Ollila-Palomar-Pascal, TSP2021, Alg. 1]; "OPP-harmonic": variation of "OPP"; "POP": improved estimator as in paper [Pascal-Ollila-Palomar, EUSIPCO2021, Alg. 1] (default method).
initial	List of initial values of the parameters for the iterative estimation method (in case nu = "iterative"). Possible elements include: <ul style="list-style-type: none"> mu: default is the data sample mean, cov: default is the data sample covariance matrix, scatter: default follows from the scaled sample covariance matrix, nu: can take the same values as argument nu, default is 4, B: default is the top eigenvectors of initial\$cov multiplied by the sqrt of the eigenvalues, psi: default is $\text{diag}(\text{initial}\\$cov - \text{initial}\\$B \% \% t(\text{initial}\\$B))$.
optimize_mu	Boolean indicating whether to optimize mu (default is TRUE).
weights	Optional weights for each of the observations (the length should be equal to the number of rows of X).
scale_covmat	Logical value indicating whether to scale the scatter and covariance matrices to minimize the MSE estimation error by introducing bias (default is FALSE). This is particularly advantageous when the number of observations is small compared to the number of variables.
PX_EM_acceleration	Logical value indicating whether to accelerate the iterative method via the PX-EM acceleration technique (default is TRUE) [Liu-Rubin-Wu, 1998].
nu_update_start_at_iter	Starting iteration (default is 1) for iteratively estimating nu (in case nu = "iterative").
nu_update_every_num_iter	Frequency (default is 1) for iteratively estimating nu (in case nu = "iterative").

factors	Integer indicating number of factors (default is <code>ncol(X)</code> , so no factor model assumption).
max_iter	Integer indicating the maximum number of iterations for the iterative estimation method (default is 100).
ptol	Positive number indicating the relative tolerance for the change of the variables to determine convergence of the iterative method (default is $1e-3$).
ftol	Positive number indicating the relative tolerance for the change of the log-likelihood value to determine convergence of the iterative method (default is <code>Inf</code> , so it is not active). Note that using this argument might have a computational cost as a convergence criterion due to the computation of the log-likelihood (especially when X is high-dimensional).
return_iterates	Logical value indicating whether to record the values of the parameters (and possibly the log-likelihood if <code>ftol</code> < <code>Inf</code>) at each iteration (default is <code>FALSE</code>).
verbose	Logical value indicating whether to allow the function to print messages (default is <code>FALSE</code>).

Details

This function estimates the parameters of a multivariate Student's t distribution (`mu`, `cov`, `scatter`, and `nu`) to fit the data via the expectation-maximization (EM) algorithm. The data matrix X can contain missing values denoted by NAs. The estimation of `nu` is very flexible: it can be directly passed as an argument (without being estimated), it can be estimated with several one-shot methods (namely, "kurtosis", "MLE-diag", "MLE-diag-resampled"), and it can also be iteratively estimated with the other parameters via the EM algorithm.

Value

A list containing (possibly) the following elements:

<code>mu</code>	Mu vector estimate.
<code>scatter</code>	Scatter matrix estimate.
<code>nu</code>	Degrees of freedom estimate.
<code>mean</code>	Mean vector estimate: <code>mean = mu</code>
<code>cov</code>	Covariance matrix estimate: <code>cov = nu/(nu-2) * scatter</code>
<code>converged</code>	Boolean denoting whether the algorithm has converged (<code>TRUE</code>) or the maximum number of iterations <code>max_iter</code> has been reached (<code>FALSE</code>).
<code>num_iterations</code>	Number of iterations executed.
<code>cpu_time</code>	Elapsed CPU time.
<code>B</code>	Factor model loading matrix estimate according to <code>cov = (B %*% t(B) + diag(psi))</code> (only if factor model requested).

psi Factor model idiosyncratic variances estimates according to $\text{cov} = (B \% \% t(B) + \text{diag}(\text{psi}))$ (only if factor model requested).

log_likelihood_vs_iterations Value of log-likelihood over the iterations (if $\text{ftol} < \text{Inf}$).

iterates_record Iterates of the parameters (μ , scatter , ν , and possibly log_likelihood (if $\text{ftol} < \text{Inf}$)) along the iterations (if $\text{return_iterates} = \text{TRUE}$).

Author(s)

Daniel P. Palomar and Rui Zhou

References

Chuanhai Liu and Donald B. Rubin, "ML estimation of the t-distribution using EM and its extensions, ECM and ECME," *Statistica Sinica* (5), pp. 19-39, 1995.

Chuanhai Liu, Donald B. Rubin, and Ying Nian Wu, "Parameter Expansion to Accelerate EM: The PX-EM Algorithm," *Biometrika*, Vol. 85, No. 4, pp. 755-770, Dec., 1998

Rui Zhou, Junyan Liu, Sandeep Kumar, and Daniel P. Palomar, "Robust factor analysis parameter estimation," *Lecture Notes in Computer Science (LNCS)*, 2019. <<https://arxiv.org/abs/1909.12530>>

Esa Ollila, Daniel P. Palomar, and Frédéric Pascal, "Shrinking the Eigenvalues of M-estimators of Covariance Matrix," *IEEE Trans. on Signal Processing*, vol. 69, pp. 256-269, Jan. 2021. <<https://doi.org/10.1109/TSP.2020.3043952>>

Frédéric Pascal, Esa Ollila, and Daniel P. Palomar, "Improved estimation of the degree of freedom parameter of multivariate t-distribution," in *Proc. European Signal Processing Conference (EUSIPCO)*, Dublin, Ireland, Aug. 23-27, 2021. <<https://doi.org/10.23919/EUSIPCO54536.2021.9616162>>

See Also

[fit_Tyler](#), [fit_Cauchy](#), [fit_mvst](#), [nu OPP_estimator](#), and [nu POP_estimator](#)

Examples

```
library(mvtnorm)      # to generate heavy-tailed data
library(fitHeavyTail)

X <- rmvt(n = 1000, df = 6) # generate Student's t data
fit_mvt(X)

# setting lower limit for nu
options(nu_min = 4.01)
fit_mvt(X, nu = "iterative")
```

fit_Tyler	<i>Estimate parameters of a multivariate elliptical distribution to fit data via Tyler's method</i>
-----------	---

Description

Estimate parameters of a multivariate elliptical distribution, namely, the mean vector and the covariance matrix, to fit data. Any data sample with NAs will be simply dropped. The algorithm is based on Tyler's method, which normalizes the centered samples to get rid of the shape of the distribution tail. The data is first demeaned (with the geometric mean by default) and normalized. Then the estimation is based on the maximum likelihood estimation (MLE) and the algorithm is obtained from the majorization-minimization (MM) optimization framework. Since Tyler's method can only estimate the covariance matrix up to a scaling factor, a very effective method is employed to recover the scaling factor.

Usage

```
fit_Tyler(
  X,
  initial = NULL,
  estimate_mu = TRUE,
  max_iter = 200,
  ptol = 0.001,
  ftol = Inf,
  return_iterates = FALSE,
  verbose = FALSE
)
```

Arguments

X	Data matrix containing the multivariate time series (each column is one time series).
initial	List of initial values of the parameters for the iterative estimation method. Possible elements include: <ul style="list-style-type: none"> • mu: default is the data sample mean, • cov: default is the data sample covariance matrix.
estimate_mu	Boolean indicating whether to estimate mu (default is TRUE).
max_iter	Integer indicating the maximum number of iterations for the iterative estimation method (default is 100).
ptol	Positive number indicating the relative tolerance for the change of the variables to determine convergence of the iterative method (default is 1e-3).
ftol	Positive number indicating the relative tolerance for the change of the log-likelihood value to determine convergence of the iterative method (default is Inf, so it is not active). Note that using this argument might have a computational cost as a convergence criterion due to the computation of the log-likelihood (especially when X is high-dimensional).

return_iterates	Logical value indicating whether to record the values of the parameters (and possibly the log-likelihood if <code>ftol < Inf</code>) at each iteration (default is FALSE).
verbose	Logical value indicating whether to allow the function to print messages (default is FALSE).

Value

A list containing possibly the following elements:

mu	Mean vector estimate.
scatter	Scatter matrix estimate.
nu	Degrees of freedom estimate (assuming an underlying Student's t distribution).
cov	Covariance matrix estimate.
converged	Boolean denoting whether the algorithm has converged (TRUE) or the maximum number of iterations <code>max_iter</code> has reached (FALSE).
num_iterations	Number of iterations executed.
cpu_time	Elapsed CPU time.
log_likelihood	Value of log-likelihood after converge of the estimation algorithm (if <code>ftol < Inf</code>).
iterates_record	Iterates of the parameters (mu, scatter, and possibly log_likelihood (if <code>ftol < Inf</code>)) along the iterations (if <code>return_iterates = TRUE</code>).

Author(s)

Daniel P. Palomar

References

Ying Sun, Prabhu Babu, and Daniel P. Palomar, "Regularized Tyler's Scatter Estimator: Existence, Uniqueness, and Algorithms," IEEE Trans. on Signal Processing, vol. 62, no. 19, pp. 5143-5156, Oct. 2014.

See Also

[fit_Cauchy](#) and [fit_mvt](#)

Examples

```
library(mvtnorm)      # to generate heavy-tailed data
library(fitHeavyTail)

X <- rmvt(n = 1000, df = 6) # generate Student's t data
fit_Tyler(X)
```

nu_opp_estimator	<i>Estimate the degrees of freedom of a heavy-tailed t distribution based on the OPP estimator</i>
------------------	---

Description

This function estimates the degrees of freedom of a heavy-tailed t distribution based on the OPP estimator from paper [Ollila-Palomar-Pascal, TSP2021, Alg. 1]. Traditional nonparametric methods or likelihood methods provide erratic estimations of the degrees of freedom unless the number of observations is very large. The POP estimator provides a stable estimator based on random matrix theory. A number of different versions are provided, but the default POP method will most likely be the desired choice.

Usage

```
nu_opp_estimator(var_X, trace_scatter, r2, method = c("OPP", "OPP-harmonic"))
```

Arguments

var_X	Vector with the sample variance of the columns of the data matrix.
trace_scatter	Trace of the scatter matrix.
r2	Vector containing the values of $\text{diag}(X_c \%*\% \text{inv}(\text{scatter}) \%*\% t(X_c))$, where X_c is the centered data matrix.
method	String indicating the version of the OPP estimator (default is just "OPP"). Other option is the variation: "OPP-harmonic".

Value

Estimated value of the degrees of freedom ν of a heavy-tailed t distribution.

Author(s)

Esa Ollila, Frédéric Pascal, and Daniel P. Palomar

References

Esa Ollila, Daniel P. Palomar, and Frédéric Pascal, "Shrinking the Eigenvalues of M-estimators of Covariance Matrix," IEEE Trans. on Signal Processing, vol. 69, pp. 256-269, Jan. 2021. <<https://doi.org/10.1109/TSP.2020.3043952>>

Examples

```
library(mvtnorm)      # to generate heavy-tailed data
library(fitHeavyTail)

# parameters
N <- 5
```

```

T <- 100
nu_true <- 4          # degrees of freedom
mu_true <- rep(0, N)  # mean vector
Sigma_true <- diag(N) # scatter matrix

# generate data
X <- rmvt(n = T, sigma = Sigma_true, delta = mu_true, df = nu_true) # generate Student's t data
mu <- colMeans(X)
Xc <- X - matrix(mu, T, N, byrow = TRUE) # center data

# usage #1
nu OPP_estimator(var_X = 1/(T-1)*colSums(Xc^2), trace_scatter = sum(diag(Sigma_true)))

# usage #2
r2 <- rowSums(Xc * (Xc %*% solve(Sigma_true)))
nu OPP_estimator(var_X = 1/(T-1)*colSums(Xc^2), trace_scatter = sum(diag(Sigma_true)),
                 method = "OPP-harmonic", r2 = r2)

```

nu_POP_estimator	<i>Estimate the degrees of freedom of a heavy-tailed t distribution based on the POP estimator</i>
------------------	---

Description

This function estimates the degrees of freedom of a heavy-tailed t distribution based on the POP estimator from paper [Pascal-Ollila-Palomar, EUSIPCO2021, Alg. 1]. Traditional nonparametric methods or likelihood methods provide erratic estimations of the degrees of freedom unless the number of observations is very large. The POP estimator provides a stable estimator based on random matrix theory. A number of different versions are provided, but the default POP method will most likely be the desired choice.

Usage

```

nu_POP_estimator(
  Xc = NULL,
  N = NULL,
  T = NULL,
  Sigma = NULL,
  nu = NULL,
  r2 = NULL,
  method = c("POP", "POP-approx-1", "POP-approx-2", "POP-approx-3", "POP-approx-4",
             "POP-exact", "POP-sigma-corrected", "POP-sigma-corrected-true"),
  alpha = 1
)

```

Arguments

Xc	Centered data matrix (with zero mean) containing the multivariate time series (each column is one time series).
N	Number of variables (columns of data matrix) in the multivariate time series.
T	Number of observations (rows of data matrix) in the multivariate time series.
Sigma	Current estimate of the scatter matrix.
nu	Current estimate of the degrees of freedom of the t distribution.
r2	Vector containing the values of <code>diag(Xc %*% inv(scatter) %*% t(Xc))</code> .
method	String indicating the version of the POP estimator (default is just "POP" and should work well in all cases). Other versions include: "POP-approx-1", "POP-approx-2", "POP-approx-3", "POP-approx-4", "POP-exact", "POP-sigma-corrected", "POP-sigma-corrected-true".
alpha	Value for the acceleration technique (cf. <code>fit_mvt()</code>).

Value

Estimated value of the degrees of freedom ν of a heavy-tailed t distribution.

Author(s)

Frédéric Pascal, Esa Ollila, and Daniel P. Palomar

References

Frédéric Pascal, Esa Ollila, and Daniel P. Palomar, "Improved estimation of the degree of freedom parameter of multivariate t-distribution," in Proc. European Signal Processing Conference (EU-SIPCO), Dublin, Ireland, Aug. 23-27, 2021. <<https://doi.org/10.23919/EUSIPCO54536.2021.9616162>>

Examples

```
library(mvtnorm)      # to generate heavy-tailed data
library(fitHeavyTail)

# parameters
N <- 5
T <- 100
nu_true <- 4          # degrees of freedom
mu_true <- rep(0, N)  # mean vector
Sigma_true <- diag(N) # scatter matrix

# generate data
X <- rmvt(n = T, sigma = Sigma_true, delta = mu_true, df = nu_true) # generate Student's t data
mu <- colMeans(X)
Xc <- X - matrix(mu, T, N, byrow = TRUE) # center data

# usage #1
nu_POP_estimator(Xc = Xc, nu = 10, Sigma = Sigma_true)
```



```
# usage #2
r2 <- rowSums(Xc * (Xc %*% solve(Sigma_true)))
nu_POP_estimator(r2 = r2, nu = 10, N = N)

# usage #3
nu_POP_estimator(r2 = r2, nu = 10, N = N, method = "POP-approx-1")
```

Index

`fit_Cauchy`, [2](#), [3](#), [11](#), [13](#)
`fit_mvst`, [2](#), [5](#), [11](#)
`fit_mvt`, [2](#), [5](#), [7](#), [8](#), [13](#), [16](#)
`fit_Tyler`, [2](#), [5](#), [11](#), [12](#)
`fitHeavyTail-package`, [2](#)

`nu OPP_estimator`, [11](#), [14](#)
`nu POP_estimator`, [11](#), [15](#)