

Package ‘mixSPE’

January 25, 2021

Type Package

Title Mixtures of Power Exponential and Skew Power Exponential
Distributions for Use in Model-Based Clustering and
Classification

Version 0.9.1

Date 2021-01-19

Author Ryan P. Browne[aut, cre], Utkarsh J. Dang[aut, cre], Michael P. B. Gal-
lauger[ctb], and Paul D. McNicholas[aut]

Maintainer Utkarsh J. Dang <udang@binghamton.edu>

Description Mixtures of skewed and elliptical distributions are implemented using mixtures of multi-
variate skew
power exponential and power exponential distributions, respectively. A generalized expectation-
maximization
framework is used for parameter estimation. Methodology for mixtures of power exponential dis-
tributions is
from Dang et al. (2015) <doi: 10.1111/biom.12351>.

License GPL (>= 2)

Encoding UTF-8

LazyData true

Imports mvtnorm

Suggests testthat

NeedsCompilation no

Repository CRAN

Date/Publication 2021-01-25 05:50:06 UTC

R topics documented:

mixSPE-package	2
EMGr	2
print.spemix	4
rpe	5
rspe	6

Index

7

mixSPE-package	<i>Mixtures of skew power exponential or power exponential distributions.</i>
----------------	---

Description

An implementation of skewed and elliptical mixture distributions for use in model-based clustering.

Details

Package: mixSPE
 Type: Package
 Version: 0.9.1
 Date: 2021-01-19
 License: GPL (>= 2)

EMGr	<i>Function for model-based clustering with the multivariate power exponential (MPE) or the skew power exponential (MSPE) distribution.</i>
------	---

Description

For fitting of a family of 16 mixture models based on mixtures of multivariate skew power exponential distributions with eigen-decomposed covariance structures.

Usage

```
EMGr(data = NULL, initialization = NULL, iModel = "EIIE", G = 2, max.iter = 500,
      epsilon = 0.01, label = NULL, modelSet = "all", skewness = FALSE,
      keepResults = FALSE, seedno = 1, scale = TRUE)
```

Arguments

data	A matrix such that rows correspond to observations and columns correspond to variables.
initialization	0 means a k-means start. A single number indicates number of random starts in addition to a k-means and heirarchical clustering start. A z matrix can be provided directly here as well. Finally, a list can be provided with the same format as <code>modelfit\$bestmod\$gpar</code> .
iModel	Initialization model used to generate initial parameter estimates.

G	A sequence of integers corresponding to the number of components to be fitted.
max.iter	Maximum number of GEM iterations allowed
epsilon	Threshold for convergence for the GEM algorithm used in the Aitken's stopping criterion.
label	Used for model-based classification aka semi-supervised classification. This is a vector of group labels with 0 for unlabelled observations.
modelSet	A total of 16 models are provided: "EIE", "VIE", "EEIE", "VVIE", "EEEE", "EEVE", "VVEE", "VVVE", "EIV", "VIV", "EEIV", "VVIV", "EEEV", "EEVV", "VVEV", "VVVV". modelSet="all" fits all models automatically. Otherwise, a character vector of a subset of these models can be provided.
skewness	If FALSE (default), fits mixtures of multivariate power exponential distributions that cannot model skewness. If TRUE, fits mixtures of multivariate skewed power exponential distributions that can model skewness.
keepResults	Keep results from all models
seedno	Seed number for initialization of k-means or random starts.
scale	If TRUE, scales the data before model fitting.

Details

The component scale matrix is decomposed using an eigen-decomposition:

$$\Sigma_g = \lambda_g \Gamma_g \Delta_g \Gamma_g'$$

The nomenclature is as follows: a EEVE model denotes a model with equal constants associated with the eigenvalues (λ) for each group, equal orthogonal matrix of eigenvectors (Γ), variable diagonal matrices with values proportional to the eigenvalues of each component scale matrix (Δ_g), and equal shape parameter (β).

Value

allModels	Output for each model run.
bestmod	Output for the best model chosen by the BIC.
loglik	Maximum log likelihood for each model
num.iter	Number of iterations required for convergence for each model
num.par	Number of parameters fit for each model
BIC	BIC for each model
maxBIC	Which model was selected by the BIC in the BIC matrix?

Author(s)

Ryan P. Browne, Utkarsh J. Dang, Michael P. B. Gallagher, and Paul D. McNicholas

Examples

```

set.seed(1)
Nobs1 <- 200
Nobs2 <- 250
X1 <- rpe(n = Nobs1, mean = c(0,0), scale = diag(2), beta = 1)
X2 <- rpe(n = Nobs2, mean = c(3,0), scale = diag(2), beta = 2)
x <- as.matrix(rbind(X1, X2))
membership <- c(rep(1, Nobs1), rep(2, Nobs2))
mperun <- EMGr(data=x, initialization=0, iModel="EIIIV", G=2:3,
max.iter=500, epsilon=5e-3, label=NULL, modelSet=c("EIIIV"),
skewness=FALSE, keepResults=TRUE, seedno=1, scale=FALSE)
print(mperun)
print(table(membership,mperun$bestmod$map))
msperun <- EMGr(data=x, initialization=0, iModel="EIIIV", G=2:3,
max.iter=500, epsilon=5e-3, label=NULL, modelSet=c("EIIIV"),
skewness=TRUE, keepResults=TRUE, seedno=1, scale=FALSE)
#print(msperun)
#print(table(membership,msperun$bestmod$map))

set.seed(1)
data(iris)
membership <- as.numeric(factor(iris[, "Species"]))
label <- membership
label[sample(x = 1:length(membership),size = ceiling(0.75*length(membership)),replace = FALSE)] <- 0
dat <- data.matrix(iris[, 1:4])
semisup_class_skewed = EMGr(data=dat, initialization=0, iModel="EIIIV",
G=3, max.iter=500, epsilon=5e-3, label=label, modelSet=c("VVVE"),
skewness=TRUE, keepResults=TRUE, seedno=1, scale=TRUE)
#table(membership,semisup_class_skewed$bestmod$map)

```

```
print.spemix
```

```
Print a summary of the model fit.
```

Description

Print a summary of the model fit including the number of components and the scale structure selected by the BIC and the ICL.

Usage

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

Arguments

```
x           An object of class "spemix".
...         Ignore this
```

Author(s)

Utkarsh J. Dang, Michael P. B. Gallagher, Ryan P. Browne, and Paul D. McNicholas

rpe

Simulate data from the multivariate power exponential distribution.

Description

Simulate data from the multivariate power exponential distribution given the mean, scale matrix, and the shape parameter.

Usage

```
rpe(n = NULL, beta = NULL, mean = NULL, scale = NULL)
```

Arguments

n	Number of observations to simulate.
beta	A positive shape parameter β that determines the kurtosis of the distribution.
mean	A p -dimensional vector. μ .
scale	A p -dimensional square scale matrix Σ .

Value

A matrix with rows representing the p -dimensional observations.

Author(s)

Utkarsh J. Dang, Ryan P. Browne, and Paul D. McNicholas

References

For simulating from the MPE distribution, a modified version of the function `rmvpowerexp` from package `MNM` (Nordhausen and Oja, 2011) is used. The function was modified due to a typo in the `rmvpowerexp` code, as mentioned in the publication (Dang et al., 2015). This program utilizes the stochastic representation of the MPE distribution (Gómez et al., 1998) to generate data. Dang, Utkarsh J., Ryan P. Browne, and Paul D. McNicholas. "Mixtures of multivariate power exponential distributions." *Biometrics* 71, no. 4 (2015): 1081-1089. Gómez, E., M. A. Gomez-Viilegas, and J. M. Marin. "A multivariate generalization of the power exponential family of distributions." *Communications in Statistics-Theory and Methods* 27, no. 3 (1998): 589-600. Nordhausen, Klaus, and Hannu Oja. "Multivariate L1 methods: the package `MNM`." *Journal of Statistical Software* 43, no. 5 (2011): 1-28.

Examples

```
dat <- rpe(n = 1000, beta = 2, mean = rep(0,5), scale = diag(5))
dat <- rpe(n = 1000, beta = 0.8, mean = rep(0,5), scale = diag(5))
```

rspe	<i>Simulate data from the multivariate skew power exponential distribution.</i>
------	---

Description

Simulate data from the multivariate power exponential distribution given the location, scale matrix, shape, and skewness parameter.

Usage

```
rspe(n, location = rep(0, nrow(scale)), scale = diag(length(location)),  
beta = 1, psi = c(0, 0))
```

Arguments

n	Number of observations to simulate.
location	A p -dimensional vector. μ .
scale	A p -dimensional square scale matrix Σ .
beta	A positive shape parameter β that determines the kurtosis of the distribution.
psi	A p -dimensional vector determining skewness. μ .

Details

Based on a Metropolis-Hastings rule.

Value

A matrix with rows representing the p -dimensional observations.

Author(s)

Utkarsh J. Dang, Ryan P. Browne, and Paul D. McNicholas

Examples

```
dat <- rspe(n = 1000, beta = 0.75, location = c(0,0), scale =  
matrix(c(1,0.7,0.7,1),2,2), psi = c(5,5))
```

Index

* **package**

 mixSPE-package, [2](#)

EMGr, [2](#)

mixSPE (mixSPE-package), [2](#)

mixSPE-package, [2](#)

print.spemix, [4](#)

rpe, [5](#)

rspe, [6](#)