Package ‘mixSPE’

June 19, 2019

Type Package

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

Version 0.1.1

Date 2019-06-18

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

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

Description Mixtures of skewed and elliptical distributions are implemented using mixtures of multivariate skew
power exponential and power exponential distributions, respectively. A generalized expectation-maximization
framework is used for parameter estimation. Methodology for mixtures of power exponential distributions 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 2019-06-19 09:10:02 UTC

R topics documented:

`mixSPE-package` *Mixtures of skew power exponential or power exponential distributions.*

Description

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

Details

Package:	mixSPE
Type:	Package
Version:	0.1.1
Date:	2019-06-18
License:	GPL (>= 2)

`mpe` *Function for model-based clustering with the multivariate power exponential (PE) distribution.*

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

```
mpe(verbose = FALSE, dat = NULL, seedno = 1, G = 1:4, start = "kmeans", kmeansinit = 10,
eps = 0.005, maxit = 5000, label = NULL, modelnames = c("EIIE", "VIIE", "EEIE", "VVIE",
"EEEE", "EEVE", "VVEE", "VVVE", "EIIV", "VIIV", "EEIV", "VVIV", "EEEV", "EEVV", "VVEV",
"VVVV"))
```

Arguments

`verbose`	A short progress indicator.
`dat`	A matrix such that rows correspond to observations and columns correspond to variables.

seedno	Seed number for initialization of k-means or random starts.
G	A sequence of integers corresponding to the number of components to be fitted.
start	Inputting "kmeans" initializes the component labels for each observation from a k-means classification. Option "random" results in a random hard initialization for the component label for each observation.
kmeansinit	Number of random starts to the k-means initialization function.
eps	Threshold for convergence for the GEM algorithm used in the Aitken's stopping criterion.
maxit	Maximum number of GEM iterations allowed.
label	Used for model-based classification aka semi-supervised classification.
modelnames	A total of 16 models are provided: "EIIE", "VIIE", "EEIE", "VVIE", "EEEE", "EEVE", "VVEE", "VVVE", "EIIV", "VIIV", "EEIV", "VVIV", "EEEV", "EEVV", "VVEV", "VVVV".

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

call	Function call.
time	Time taken.
modelnames	Models fitted.
msc	Matrix of results with BIC, ICL, and log-likelihood values achieved for each model.
bicclassification	Maximum a posteriori component label indicators of each observation from the model selected by the BIC.
iclclassification	Maximum a posteriori component label indicators of each observation from the model selected by the ICL.
bicselection	Model selected by the BIC including estimates.
iclselection	Model selected by the ICL including estimates.
zlist	List of initial labels for each observation from the initialization function for each number of components.

Author(s)

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

See Also

See Also mspe.

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 <- mpe(verbose = TRUE, dat = x, seedno = 1, G = 1:2, start="kmeans",
               modelnames = c("EIIV", "EEEV", "VVVV"))
print(mperun)
print(table(membership,mperun$bicclassification))
```

---

mspe — *Function for model-based clustering with the multivariate skew power exponential (SPE) distribution.*

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

```
mspe(verbose = FALSE, dat = NULL, seedno = 1, G = 1:4, start = "kmeans", kmeansinit = 10,
eps = 0.005, maxit = 2000, anneal = NULL, label = NULL, psistart = "zero", modelnames =
c("EIIE", "VIIE", "EEIE", "VVIE", "EEEE", "EEVE", "VVEE", "VVVE", "EIIV", "VIIV", "EEIV",
"VVIV", "EEEV", "EEVV", "VVEV", "VVVV"))
```

Arguments

verbose	A short progress indicator.
dat	A matrix such that rows correspond to observations and columns correspond to variables.
seedno	Seed number for initialization of k-means or random starts.
G	A sequence of integers corresponding to the number of components to be fitted.
start	Inputting "kmeans" initializes the component labels for each observation from a k-means classification. Option "random" results in a random hard initialization for the component label for each observation.
kmeansinit	Number of random starts to the k-means initialization function.

`eps` Threshold for convergence for the GEM algorithm used in the Aitken's stopping criterion.

`maxit` Maximum number of GEM iterations allowed

`anneal` For deterministic annealing based initilization. Provide a non-decreasing vector of numbers rising from a small number to 1. Example: rep(seq(.05, 1, length.out=6),each=2). Takes experimentation.

`label` Used for model-based classification aka semi-supervised classification.

`psistart` Default is a vector of zeros for each group. If "est" is used, a non-parameteric estimate using the mean and median of the inferred cluster based on initialized labels is used.

`modelnames` A total of 16 models are provided: "EIIE", "VIIE", "EEIE", "VVIE", "EEEE", "EEVE", "VVEE", "VVVE", "EIIV", "VIIV", "EEIV", "VVIV", "EEEV", "EEVV", "VVEV", "VVVV".

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

`call` Function call.

`time` Time taken.

`modelnames` Models fitted.

`msc` Matrix of results with BIC, ICL, and log-likelihood values achieved for each model.

`bicclassification`
Maximum a posteriori component label indicators of each observation from the model selected by the BIC.

`iclclassification`
Maximum a posteriori component label indicators of each observation from the model selected by the ICL.

`bicselection` Model selected by the BIC including estimates.

`iclselection` Model selected by the ICL including estimates.

`zlist` List of initial labels for each observation from the initialization function for each number of components.

Author(s)

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

See Also

See Also mpe.

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))
msperun <- mspe(verbose = TRUE, dat = x, seedno = 1, G = 1:2, start="kmeans",
               modelnames = c("EIIV"))
print(msperun)
print(table(membership,msperun$bicclassification))
```

---

print.pemix *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 'pemix'
print(x, ...)
```

Arguments

x	An object of class "pemix".
...	Ignore this

Author(s)

Utkarsh J. Dang, Ryan P. Browne, and Paul D. McNicholas `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. Gallaugher, 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 — *Simulate data from the multivariate skew power exponential distribution.*

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