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. Gal-laugher[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 .................................................. 2
mpe .......................................................... 2
mspe .......................................................... 4
print.pemix .................................................... 6
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(verbsoe = FALSE, dat = NULL, seedno = 1, G = 1:4, start = "kmeans", kmeansinit = 10, eps = 0.005, maxit = 5000, label = NULL, modelnames = c("EEIE", "VIEE", "EEIE", "VVEI", "EEVV", "EVVE", "EEVV", "EEV"))

Arguments

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

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", "VVVE", "VVEE", "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 (\( \lambda \)) for each group, equal orthogonal matrix of eigenvectors (\( \Gamma \)), variable diagonal
matrices with values proportional to the eigenvalues of each component scale matrix (\( \Delta_g \)),
and equal shape parameter (\( \beta \)).

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

```r
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**

```r
mspe(verbosel = 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", "VVEE", "EEEE", "EEVE", "VVVE", "EEIV", "VIVV", "EEIV",
        "VIVV", "EEVV", "EEVV", "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.
Threshold for convergence for the GEM algorithm used in the Aitken’s stopping criterion.

Maximum number of GEM iterations allowed

For deterministic annealing based initialization. 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.

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

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.

A total of 16 models are provided: "EIIE", "VIIE", "EEIE", "VVIE", "EEEE", "EEVE", "VVVE", "EIV", "VIIV", "EEIV", "VVIV", "EEEV", "EENV", "VVVE", "VVEV", "VVVV".

The component scale matrix is decomposed using an eigen-decomposition:
\[
\Sigma_g = \lambda_g \Gamma_g \Delta_g \Gamma_g^T
\]

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

Function call.

Time taken.

Models fitted.

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

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

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

Model selected by the BIC including estimates.

Model selected by the ICL including estimates.

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

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

See Also mpe.

Examples

```r
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

```r
## 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

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 $\beta$ that determines the kurtosis of the distribution.
mean A $p$-dimensional vector. $\mu$.
scale A $p$-dimensional square scale matrix $\Sigma$.

Value
A matrix with rows representing the $p$-dimensional observations.
Author(s)
Utkarsh J. Dang, Ryan P. Browne, and Paul D. McNicholas

References

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

```r
rspe(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
```r
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. $\mu$.
- **scale**: A $p$-dimensional square scale matrix $\Sigma$.
- **beta**: A positive shape parameter $\beta$ that determines the kurtosis of the distribution.
- **psi**: A $p$-dimensional vector determining skewness. $\mu$.

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

*Topic package
  mixSPE-package, 2

mixSPE (mixSPE-package), 2
mixSPE-package, 2
mpe, 2, 6
mspe, 4, 4

print.pemix, 6
print.spmix, 7

rpe, 7
rspe, 8