Package ‘MixSAL’

May 18, 2018

Type  Package
Title  Mixtures of Multivariate Shifted Asymmetric Laplace (SAL) Distributions
Version 1.0
Date  2018-05-010
Author  Brian C. Franczak [aut, cre], Ryan P. Browne [aut, cph], Paul D. McNicholas [aut, cph], Katherine L. Burak [ctb]
Depends MASS (>= 3.1.3)
Maintainer Brian C. Franczak <franczakb@macewan.ca>
Description The current version of the 'MixSAL' package allows users to generate data from a multivariate SAL distribution or a mixture of multivariate SAL distributions, evaluate the probability density function of a multivariate SAL distribution or a mixture of multivariate SAL distributions, and fit a mixture of multivariate SAL distributions using the Expectation-Maximization (EM) algorithm (see Franczak et. al, 2014, <doi:10.1109/TPAMI.2013.216>, for details).
License GPL (>= 2)
NeedsCompilation no
Repository CRAN
Date/Publication 2018-05-18 12:10:24 UTC

R topics documented:

- MixSAL-package ................................................................. 2
- dmsal ................................................................. 3
- dsal ................................................................. 4
- msal ................................................................. 5
- rmsal ................................................................. 7
- rsal ................................................................. 8
- yeast ................................................................. 9

Index 10
Description

The current version of the MixSAL package allows users to generate data from a multivariate SAL distribution or a mixture of multivariate SAL distributions, evaluate the probability density function of a multivariate SAL distribution or a mixture of multivariate SAL distributions, and fit a mixture of multivariate SAL distributions using the Expectation-Maximization (EM) algorithm (see Franczak et. al, 2014 for details).

Details

Package:             MixSAL
Type:               Package
Version:            1.0
Date:               2018-05-09
License:            GPL (>=3.1.3)

This package contains the function msal for carrying about model based clustering using mixtures of SAL distributions; the functions rsal and rmsal for generating data from a multivariate SAL or mixture of multivariate SAL distributions, and hte functions dsal and dmsal for evaluating the model based clustering and classification using the mixture of generalized hyperbolic factor analyzers; the function MCGHD for model based clustering using the mixture of coalesced generalized hyperbolic distributions, and some real data sets.

Author(s)

Brian C. Franczak [aut, cre], Ryan P. Brown [aut, ctb], and Paul D. McNicholas [aut, ctb]
Maintainer: Brian C. Franczak <franczakb@macewan.ca>

References


Examples

```r
## Clustering Simulated Data
alpha <- matrix(c(2,2,1,2),2,2)
sig <- array(NA,dim=c(2,2,2))
sig[,1] <- diag(2)
sig[,2] <- matrix(c(1,0.5,0.5,1),2,2)
mu <- matrix(c(0,0,-2,5),2,2)
pi.g <- rep(1/2,2)
x <- rmsal(n=500,p=2,alpha=alpha,sig=sig,mu=mu,pi.g=pi.g)
```
dmsal

msal.ex1 <- msal(x=x[,,-1],G=2)
table(x[,1],msal.ex1$cluster)

## Evaluate the probability density function of the specified mixture of SAL distributions
df.sal <- dmsal(x=x[,,-1],alpha=alpha,sig=sig,mu=mu,pi.g=pi.g)
df.sal[1:10]

dmsal  

Probability Density Function for a Mixture of SAL Distributions

Description

Evaluates the probability density function of a mixture of multivariate SAL distribution.

Usage

dmsal(x, alpha, sig, mu, pi.g)

Arguments

x  A n by p matrix where each row corresponds a p-dimensional observation.
alpha  A matrix where each row specifies the direction of skewness in each variable for each mixture component.
sig  An array where each matrix specifies the covariance matrix for each mixture component.
mu  A matrix where each row gives the mean vector for each mixture component.
pi.g  A vector specifying the mixing components.

Value

A vector of length n that gives the value of the probability density function for each observation in the matrix x and the specified parameter values.

Author(s)

Brian C. Franczak [aut, cre], Ryan P. Browne [aut, ctb], Paul D. McNicholas [aut, ctb]
Maintainer: Brian C. Franczak <franczakb@macewan.ca>

References

Examples

```r
## For this illustration, consider the following dataset generated from a mixture of bivariate SAL distributions with the specified parameter set:
alpha <- matrix(c(2,2,1,2),2,2)
sig <- array(NA,dim=c(2,2,2))
sig[,][,1] <- diag(2)
sig[,][,2] <- matrix(c(1,0.5,0.5,1),2,2)
mu <- matrix(c(0,0,-2,5),2,2)
pi.g <- rep(1/2,2)
x <- rmsal(n=10,p=2,alpha=alpha,sig=sig,mu=mu,pi.g=pi.g)
## The value of the probability density function for each of the simulated values are given by:
dmsal(x=x[,1],alpha=alpha,sig=sig,mu=mu,pi.g=pi.g)
```

---

dsal

Probability Density Function for a Multivariate SAL Distribution

Description

Evaluates the probability density function of a multivariate SAL distribution.

Usage

```
dsal(x, alpha, sig, mu)
```

Arguments

- `x`: A n by p matrix where each row corresponds a p-dimensional observation.
- `alpha`: A vector specifying the direction of skewness in each variable.
- `sig`: A matrix specifying the covariance matrix of the variables.
- `mu`: A vector specifying the mean vector.

Value

A vector of length n that gives the value of the probability density function for each observation in the matrix x and the specified parameter values.

Author(s)

Brian C. Franczak [aut, cre], Ryan P. Browne [aut, ctb], Paul D. McNicholas [aut, ctb]
Maintainer: Brian C. Franczak <franczakb@macewan.ca>

References

Examples

```r
# For this illustration, consider bivariate SAL data from the specified distribution:
x <- rsal(n=10, p=2, alpha=c(2,2), sig=diag(2), mu=c(0,0))
# The value of the probability density function for each of the simulated values are given by:
dsal(x=x, alpha=c(2,2), sig=diag(2), mu=c(0,0))
```

---

**msal**

*Model-Based Clustering using a Mixture of SAL Distributions*

---

**Description**

Performs model-based clustering using a mixture of SAL distributions. The expectation-maximization (EM) algorithm is used for parameter estimation, the Aitken’s acceleration criterion is used to determine convergence, both the BIC and ICL values are given for the considered mixtures.

**Usage**

```r
msal(x, G, start = 1, max.it = 10000, eps = 0.01, print.it = F, print.warn = F, print.prmtrs = F)
```

**Arguments**

- `x`: A n by p matrix where each row corresponds a p-dimensional observation.
- `G`: The desired number of mixture components.
- `start`: Specifies how to initialize the zig matrix. If start equals 1, k-means clustering is used. If start equals 2, a random start is used. If start is a vector of length n, then the zig matrix is constructed based from this vector.
- `max.it`: The desired number of iterations for the EM algorithm.
- `eps`: The desired difference between the asymptotic estimate of the log-likelihood and the current log-likelihood value.
- `print.it`: If True, the iteration number of the EM algorithm is printed.
- `print.warn`: If True, the observation number that the mean vector is closest too is given.
- `print.prmtrs`: If True, the parameter set is printed on each iteration of the EM algorithm.

**Details**

The mixture of SAL distributions are fitted using an EM algorithm with a “Set-Back” procedure to deal with the issue of Infinite Log-Likelihood Values that arise when updating the mean vector (see Section 3.4.2 of Franczak et.al (2014) for details).
Value

The `msal` function outputs a list with the following components:

- **loglik**: A vector giving the log-likelihood values from each iteration of the considered EM algorithm.
- **alpha**: A matrix where each row specifies the direction of skewness in each variable for each mixture component.
- **sig**: An array where each matrix specifies the covariance matrix for each mixture component.
- **mu**: A matrix where each row gives the mean vector for each mixture component.
- **piNg**: A vector specifying the mixing components.
- **bic**: An integer giving the Bayesian Information Criterion (BIC) for the fitted model.
- **icl**: An integer giving the Integrated Completed Likelihood (ICL) for the fitted model.
- **cluster**: A vector of length n giving the group label for each observation in the considered data set.

Author(s)

Brian C. Franczak [aut, cre], Ryan P. Browne [aut, ctb], Paul D. McNicholas [aut, ctb]

Maintainer: Brian C. Franczak <franczakb@macewan.ca>

References


Examples

```r
## Clustering Simulated Data
alpha <- matrix(c(2,2,1,2),2,2)
sig <- array(NA,dim=c(2,2,2))
sig[,]1 <- diag(2)
sig[,]2 <- matrix(c(1,0.5,0.5,1),2,2)
mu <- matrix(c(0,0,-2,5),2,2)
pi.g <- rep(1/2,2)
x <- rmsal(n=500,p=2,alpha=alpha,sig=sig,mu=mu,pi=g=pi.g)
msal.ex1 <- msal(x=x[,]1,G=2)
table(x[,]1,msal.ex1$cluster)

## Clustering the Old Faithful Geyser Data
data(faithful)
msal.ex2 <- msal(x=faithful,G=2)
plot(x=faithful,col=msal.ex2$cluster)

## Clustering the Yeast Data
data(yeast)
msal.ex3 <- msal(x=yeast[,]1,G=2)
table(yeast[,]1,msal.ex3$cluster)
```
Simulate from a Mixture of Multivariate SAL Distributions

Description

Generates data from a mixture of multivariate shifted asymmetric Laplace (SAL) distributions.

Usage

rmsal(n, p, alpha, sig, mu, pi.g)

Arguments

n  The number of observations required.
p  The dimension of the data.
alpha  A matrix where each row specifies the direction of skewness in each variable for each mixture component.
sig  An array where each matrix specifies the covariance matrix for each mixture component.
mu  A matrix where each row gives the mean vector for each mixture component.
pi.g  A vector specifying the mixing components.

Value

An n by p + 1 matrix where each row corresponds to one observation from the specified mixture of SAL distributions. The first column gives the component (or group) label for each observation and columns 2 to p + 1 give the values of the p-dimensional observation.

Author(s)

Brian C. Franczak [aut, cre], Ryan P. Browne [aut, ctb], Paul D. McNicholas [aut, ctb]
Maintainer: Brian C. Franczak <franczakb@macewan.ca>

References


Examples

alpha <- matrix(c(2,2,1,2),2,2)
sig <- array(NA, dim=c(2,2,2))
sig[,]1 <- diag(2)
sig[,]2 <- matrix(c(1,0.5,0.5,1),2,2)
mu <- matrix(c(0,0,-2.5,2,2)
pi.g <- rep(1/2,2)
x <- rmsal(n=500,p=2,alpha=alpha,sig=sig,mu=mu,pi.g=pi.g)
plot(x[,1],col=x[,1],pch=x[,1])
Simulate from a Multivariate SAL Distribution

Description

Generates data from a multivariate shifted asymmetric Laplace (SAL) distributions.

Usage

rsal(n, p, alpha, sig, mu)

Arguments

n
The number of observations required.

p
The dimension of the data.

alpha
A vector specifying the direction of skewness in each variable.

sig
A matrix specifying the covariance matrix of the variables.

mu
A vector specifying the mean vector.

Value

An n by p matrix where each row corresponds to one observation from the specified multivariate SAL distribution.

Author(s)

Brian C. Franczak [aut, cre], Ryan P. Browne [aut, ctb], Paul D. McNicholas [aut, ctb]

Maintainer: Brian C. Franczak <franczakb@macewan.ca>

References


Examples

x <- rsal(n=500, p=2, alpha=c(2, 2), sig=diag(2), mu=c(0, 0))
plot(x)
Yeast Data

Description

Subset of the yeast dataset from Nakai and Kanehisa (1991, 1992). This subset contains three variables: McGeoch’s method for signal sequence recognition (mcg), the score of the ALOM membrane spanning region prediction program (alm), and the score of discriminant analysis of the amino acid content of vacuolar and extracellular proteins (vac).

Usage

data(yeast)

Format

A vector containing 141 observations.

Source

UCI machine learning repository.

References


Examples

data(yeast) # Loads the subset of the yeast data set
head(yeast) # Displays the first six rows of this subset of the yeast data set
Index

*Topic EM Algorithm
  msal, 5

*Topic Finite Mixture Models
  MixSAL-package, 2

*Topic Mixtures of SAL Distributions
  dmsal, 3
  msal, 5
  rmsal, 7

*Topic Model-Based Clustering
  MixSAL-package, 2
  msal, 5

*Topic Multivariate SAL Distribution
  dsal, 4
  rsal, 8

*Topic Probability Density Function
  dmsal, 3
  dsal, 4

*Topic Shifted Asymmetric Laplace Distributions
  MixSAL-package, 2

*Topic Simulation
  rmsal, 7
  rsal, 8

*Topic datasets
  yeast, 9

*Topic yeast
  yeast, 9

dmsal, 3
dsal, 4

MixSAL (MixSAL-package), 2
MixSAL-package, 2
msal, 5

rmsal, 7
rsal, 8

yeast, 9