Package: leptokurticMixture (via r-universe)

September 20, 2024

Title Implements Parsimonious Finite Mixtures of Multivariate Elliptical Leptokurtic-Normals

Version 1.1

Description A way to fit Parsimonious Finite Mixtures of Multivariate Elliptical Leptokurtic-Normals. Two methods of estimation are implemented.

Date 2023-09-08

Encoding UTF-8

License GPL (>= 2)

Imports stats

NeedsCompilation yes

RoxygenNote 7.2.3

Author Ryan Browne [aut, cre] (0000-0003-4543-0218), Luca Bagnato [ctb], Antonio Punzo [ctb]

Maintainer Ryan Browne <rpbrowne@uwaterloo.ca>

Repository CRAN

Date/Publication 2023-09-09 12:00:02 UTC

Contents

`compareEstimation` *Compare the two methods of estimation*

Description

Compare the two methods of estimation for fitting a finite mixture of multivariate elliptical leptokurtic-normal distributions; fixed point iterations and MM algorithm.

Usage

```
compareEstimation(
  mod = NULL,
  data = NULL,
  G = NULL,
  n = 10^4,
  tol = 1e-06,
  wt = NULL,
  n0 = 25,
  lab = NULL
)
```

Arguments

`mod`	A character of length 4 such as "VVVV", indicating the model; the covariance and beta parameters.
`data`	A n x p matrix of observations.
`G`	The number of components to fit.
`n`	The maximum number of EM iterations.
`tol`	The tolerance for the stopping rule; lack of progress. The default is 1e-6 but it depends on the dataset.
`wt`	a (n x d) matrix of weights for initialization if NULL, then a random weight matrix is generated.
`n0`	Given wt, the number of iterations used to obtain the initial parameters
`lab`	Using given labels (lab) as starting values.

Value

A vector of times, number of iterations and log-likelihood values.

EM *EM for the finite mixtures of MLN*

Description

Performs a number of iterations of the EM for the multivariate elliptical leptokurtic-normal (MLN) distribution until the tolerance for the lack progress or the maximum number of iterations is reached. An implementation of parsimonious clustering models via the eigen-decomposition of the scatter matrix and allowing the concentration parameter to be varying, equal or fixed across components.

Usage

```
EM(
  data = NULL,
  G = 2,
  model = NULL,
  kml = c(1, 0, 1),
  n = 10,
  epsilon = 0.01,
  gpar0 = NULL,
  estimation = 1,
  label = NULL
)
```

Arguments

data	A n x p matrix of observations.
G	A integer determine the number of components of the mixture model.
model	a character of length 4 such as "VVVV", indicating the model; the covariance and beta parameters. The 1st position controls, lambda, the volume; "V" varying across components or "E" equal across components. The 2nd position controls the eigenvalues; V" varying across components, "E" equal across components or "I" the identity matrix. The 3rd position controls the orientation; "V" varying across components, "E" equal across components or "I" the identity matrix. The 4th position controls the concentration, beta; "V" varying across components, "E" equal across components, "F" fixed at the maximum value.
kml	a vector of length 3 indicating, the number of k-means starts, number of random starts and the number of EM iterations used for each start
n	The maximum number of EM iterations.
epsilon	The tolerance for the stopping rule; lack of progress. The default is 1e-6 but it depends on the dataset.
gpar0	A list of model parameters .
estimation	If 1 (default) use the fixed point iterations and if 2 the MM algorithm.
label	If NULL then the data has no known groups. If is.integer then some of the observations have known groups. If label[i]=k then observation belongs to group k. If label[i]=0 then observation has no known group.

Value

A list with following items

- loglik - A vector of the loglikelihood values
- gpar - A list containing the parameters values
- z - A n x G matrix of the posterior probabilities
- map - A vector the maximum a posteriori derived from z
- label - The input provided.
- numpar - The number of free parameters in the fitted model.
- maxLoglik - The largest value from loglik.

Examples

```
x1 = rmln(n=100, d=4, mu=rep(5,4), diag(4), beta=2)
x2 = rmln(n=100, d=4, mu=rep(-5,4), diag(4), beta=2)
x = rbind( x1,x2)
mlnFit = EM(data=x, G=2, model="VVVF")
```

---

pmln *Parsimonious model-based clustering with the multivariate elliptical leptokurtic-normal*

Description

Performs parsimonious clustering with the multivariate elliptical leptokurtic-normal (MLN). There are 14 possible scale matrix structure and 2 for the kurtosis parameter for a total of 28 models.

Usage

```
pmln(
  data = NULL,
  G = 1:3,
  covModels = NULL,
  betaModels = "B",
  kml = c(1, 0, 1),
  label = NULL,
  scale.data = TRUE,
  veo = FALSE,
  iterMax = 1000,
  tol = 1e-08,
  pprogress = FALSE,
  method = "FP"
)
```

Arguments

data	A n x p matrix of observations.
G	A integer determine the number of components of the mixture model.
covModels	if NULL fit 14 possible scale matrix structures. Otherwise a character vector where each element has length 3. e.g. c("VVV", "EEE") A character of length 4 such as "VVVV", indicating the model; the covariance and beta parameters. The 1st position controls, lambda, the volume; "V" varying across components or "E" equal across components. The 2nd position controls the eigenvalues; V" varying across components, "E" equal across components or "I" the identity matrix. The 3rd position controls the orientation; "V" varying across components, "E" equal across components or "I" the identity matrix.
betaModels	set to "V", "E", "B", "F". "V" varying across components, "E" equal across components, "B" consider both "V" & "E", "F" fixed at the maximum value.
kml	a vector of length 3 indicating, the number of k-means starts, number of random starts and the number of EM iterations used for each start
label	If NULL then the data has no known groups. If is.integer then some of the observations have known groups. If label[i]=k then observation belongs to group k. If label[i]=0 then observation has no known group.
scale.data	Should the data be scaled before clustering. The default is TRUE.
veo	"Variables exceed observations". If TRUE, fit the model even though the number variables in the model exceeds the number of observations.
iterMax	The maximum number of EM iterations for each model fitted.
tol	The tol for the stopping rule; lack of progress. The default is 1e-6 but it depends on the data set.
pprogress	If TRUE print the progress of the function.
method	If FP use the fixed point iteration method otherwise if MM use the MM method.

Value

A list of

- startobject - A statement on how the models were initialized
- gpar - A list of parameter values for the model choosen by the BIC
- loglik - A vector of the log-likelihoods values
- z - A n x G matrix of the posterior probabilities from the model choosen by the BIC
- map - A vector the maximum a posteriori derived from z
- BIC - An array with dimensions (G, number of fitted models, 3). The last dimension indices the loglik, number of free parameters and BIC for each fitted model.
- bicModel - Information as list on the model choosen by the BIC.

Examples

```
x1 = rmln(n=100, d=4, mu=rep(5,4), diag(4), beta=2)
x2 = rmln(n=100, d=4, mu=rep(-5,4), diag(4), beta=2)
x = rbind( x1,x2)
mlnFit = pmln(data=x, G=2, covModels=c("VVV", "EEE"), betaModels="B")
```

`rmln` *Generate realizations from the multivariate elliptical leptokurtic-normal distribution*

Description

This function calculates the log cumulative density function for the multivariate-t with scale matrix equal to the identity matrix. It finds the mode and then uses Gaussian quadrature to estimate the integral.

Usage

```
rmln(n = NULL, d = NULL, mu = NULL, Sigma = NULL, beta = NULL)
```

Arguments

`n`	number of observations
`d`	the dimension of the observations
`mu`	location parameter of length d
`Sigma`	(d x d) scatter matrix
`beta`	the concentration parameter

Value

A (n x d) matrix of realizations

Examples

```
x = rmln(n=10, d=4, mu=rep(0,4), diag(4), beta=2)
```

Index