Package 'CensMFM'

May 2, 2020

Type Package

Title Finite Mixture of Multivariate Censored/Missing Data

Version 2.11

Date 2020-04-13

Imports MomTrunc (>= 5.87), mvtnorm (>= 1.0.11), gridExtra, ggplot2, tlrmvnmvt (>= 1.1.0)

Suggests mixsmsn

Description It fits finite mixture models for censored or/and missing data using several multivariate distributions. Point estimation and asymptotic inference (via empirical information matrix) are offered as well as censored data generation. Pairwise scatter and contour plots can be generated. Possible multivariate distributions are the well-known normal, Student-t and skew-normal distributions. This package is an complement of Lachos, V. H., Moreno, E. J. L., Chen, K. & Cabral, C. R. B. (2017) <doi:10.1016/j.jmva.2017.05.005> for the multivariate skew-normal case.

License GPL (>= 2)

Encoding UTF-8

LazyData true

NeedsCompilation no

Author Francisco H. C. de Alencar [aut, cre],
Christian E. Galarza [aut],
Larissa A. Matos [ctb],
Victor H. Lachos [ctb]

Maintainer Francisco H. C. de Alencar <hildemardealencar@gmail.com>

Repository CRAN

Date/Publication 2020-05-02 14:20:06 UTC

R topics documented:

1

CensMFM-package   *Finite Mixture of Multivariate Censored/Missing Data*

Description

It fits finite mixture models for censored or/and missing data using several multivariate distributions. Point estimation and asymptotic inference (via empirical information matrix) are offered as well as censored data generation. Pairwise scatter and contour plots can be generated. Possible multivariate distributions are the well-known normal, Student-t and skew-normal distributions. This package is an complement of Lachos, V. H., Moreno, E. J. L., Chen, K. & Cabral, C. R. B. (2017) <doi:10.1016/j.jmva.2017.05.005> for the multivariate skew-normal case.

Details

The DESCRIPTION file:

Package:	CensMFM
Type:	Package
Title:	Finite Mixture of Multivariate Censored/Missing Data
Version:	2.11
Date:	2020-04-13
Authors@R:	c(person("Francisco H. C.", "de Alencar", email = "hildemardealencar@gmail.com", role = c("aut", "cre")), p
Imports:	MomTrunc (>= 5.87), mvtnorm (>= 1.0.11), gridExtra, ggplot2, tlrmvnmvt (>= 1.1.0)
Suggests:	mixsmsn
Description:	It fits finite mixture models for censored or/and missing data using several multivariate distributions. Point esti
License:	GPL (>=2)
Encoding:	UTF-8
LazyData:	true
Author:	Francisco H. C. de Alencar [aut, cre], Christian E. Galarza [aut], Larissa A. Matos [ctb], Victor H. Lachos [ctb
Maintainer:	Francisco H. C. de Alencar <hildemardealencar@gmail.com>

Index of help topics:

```
CensMFM-package         Finite Mixture of Multivariate Censored/Missing
                        Data
fit.FMMSNC              Fitting Finite Mixture of Multivariate
                        Distributions.
rMMSN                   Random Generator of Finite Mixture of
                        Multivariate Distributions.
rMMSN.contour           Pairwise Scatter Plots and Histograms for
                        Finite Mixture of Multivariate Distributions.
rMSN                    Generating from Multivariate Skew-normal and
                        Normal Random Distributions.
```

~~ An overview of how to use the package, including the most important functions ~~

Author(s)

NA

Maintainer: NA

References

Cabral, C. R. B., Lachos, V. H., & Prates, M. O. (2012). Multivariate mixture modeling using skew-normal independent distributions. Computational Statistics & Data Analysis, 56(1), 126-142.

Prates, M. O., Lachos, V. H., & Cabral, C. (2013). mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. Journal of Statistical Software, 54(12), 1-20.

C.E. Galarza, L.A. Matos, D.K. Dey & V.H. Lachos. (2019) On Moments of Folded and Truncated Multivariate Extended Skew-Normal Distributions. Technical report. ID 19-14. University of Connecticut.

F.H.C. de Alencar, C.E. Galarza, L.A. Matos & V.H. Lachos. (2019) Finite Mixture Modeling of Censored and Missing Data Using the Multivariate Skew-Normal Distribution. echnical report. ID 19-31. University of Connecticut.

See Also

fit.FMMSNC, rMSN, rMMSN and rMMSN.contour

---

fit.FMMSNC   *Fitting Finite Mixture of Multivariate Distributions.*

Description

It adjusts a finite mixture of censored and/or missing multivariate distributions (FM-MC). These are the Skew-normal, normal and Student-t multivariate distributions. It uses a EM-type algorithm for iteratively computing maximum likelihood estimates of the parameters.

Usage

```
fit.FMMSNC(cc, LI, LS, y, mu = NULL, Sigma = NULL, shape = NULL, pii = NULL,
nu = NULL, g = NULL, get.init = TRUE, criteria = TRUE, family = "SN", error = 1e-05,
iter.max = 350, uni.Gama = FALSE, kmeans.param = NULL, cal.im = FALSE)
```

Arguments

cc	vector of censoring indicators. For each observation it takes 0 if non-censored, 1 if censored.
LI	the matrix of lower limits of dimension nxp. See details section.
LS	the matrix of upper limits of dimension nxp. See details section.
y	the response matrix with dimension nxp.
mu	a list with g entries, where each entry represents location parameter per group, being a vector of dimension. p.

Sigma	a list with g entries, where each entry represents a scale parameter per group, a matrix with dimension. pxp.
shape	a list with g entries, where each entry represents a skewness parameter, being a vector of dimension p.
pii	a vector of weights for the mixture (dimension of the number g of clusters). Must sum to one!
nu	the degrees of freedom for the Student-t distribution case, being a vector with dimension g.
g	number of mixture components.
get.init	Logical, TRUE or FALSE. If (get.init==TRUE) the function computes the initial values, otherwise (get.init==FALSE) the user should enter the initial values manually.
criteria	Logical, TRUE or FALSE. It indicates if likelihood-based criteria selection methods (AIC, BIC and EDC) are computed for comparison purposes.
family	distribution family to be used. Available distributions are the Skew-normal ("SN"), normal ("Normal") or Student-t ("t") distribution.
error	relative error for stopping criterion of the algorithm. See details.
iter.max	the maximum number of iterations of the EM algorithm.
uni.Gama	Logical, TRUE or FALSE. If uni.Gama==TRUE, the scale matrices per group are considered to be equals.
kmeans.param	a list with alternative parameters for the kmeans function when generating initial values. List by default is list(iter.max = 10,n.start = 1,algorithm = "Hartigan-Wong").
cal.im	Logical, TRUE or FALSE. If cal.im==TRUE, the information matrix is calculated and the standard errors are reported.

Details

The information matrix is calculated with respect to the entries of the square root matrix of Sigma, this using the Empirical information matrix. Disclaimer: User must be careful since the inference is asymptotic, so it must be used for decent sample sizes. Stopping criterion is abs((loglik/loglik-1))<epsilon.

Value

It returns a list that depending of the case, it returns one or more of the following objects:

mu	a list with g components, where each component is a vector with dimension p containing the estimated values of the location parameter.
Sigma	a list with g components, where each component is a matrix with dimension pxp containing the estimated values of the scale matrix.
Gamma	a list with g components, where each component is a matrix with dimension pxp containing the estimated values of the $Gamma$ scale matrix.
shape	a list with g components, where each component is a vector with dimension p containing the estimated values of the skewness parameter.

nu	a vector with one element containing the value of the degreees of freedom nu parameter.
pii	a vector with g elements containing the estimated values of the weights pii.
Zij	a n x p matrix containing the estimated weights values of the subjects for each group.
yest	a n x p matrix containing the estimated values of y.
MI	a list with the standard errors for all parameters.
logLik	the log-likelihood value for the estimated parameters.
aic	the AIC criterion value for the estimated parameters.
bic	the BIC criterion value for the estimated parameters.
edc	the EDC criterion value for the estimated parameters.
iter	number of iterations until the EM algorithm converges.
group	a n x p matrix containing the classification for the subjects to each group.
time	time in minutes until the EM algorithm converges.

Note

The uni.Gama parameter refers to the Γ matrix for the Skew-normal distribution, while for the normal and student-t distribution, this parameter refers to the Σ matrix.

Author(s)

Francisco H. C. de Alencar <hildemardealencar@gmail.com>, Christian E. Galarza <cgalarza88@gmail.com>, Victor Hugo Lachos <hlachos@uconn.edu> and Larissa A. Matos <larissam@ime.unicamp.br>

Maintainer: Francisco H. C. de Alencar <hildemardealencar@gmail.com>

References

Cabral, C. R. B., Lachos, V. H., & Prates, M. O. (2012). Multivariate mixture modeling using skew-normal independent distributions. Computational Statistics & Data Analysis, 56(1), 126-142.

Prates, M. O., Lachos, V. H., & Cabral, C. (2013). mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. Journal of Statistical Software, 54(12), 1-20.

C.E. Galarza, L.A. Matos, D.K. Dey & V.H. Lachos. (2019) On Moments of Folded and Truncated Multivariate Extended Skew-Normal Distributions. Technical report. ID 19-14. University of Connecticut.

F.H.C. de Alencar, C.E. Galarza, L.A. Matos & V.H. Lachos. (2019) Finite Mixture Modeling of Censored and Missing Data Using the Multivariate Skew-Normal Distribution. echnical report. ID 19-31. University of Connecticut.

See Also

rMSN, rMMSN and rMMSN.contour

Examples

```
mu         <- Sigma <- shape <- list()
mu[[1]]    <- c(-3,-4)
mu[[2]]    <- c(2,2)
Sigma[[1]] <- matrix(c(3,1,1,4.5), 2,2)
Sigma[[2]] <- matrix(c(2,1,1,3.5), 2,2)
shape[[1]] <- c(-2,2)
shape[[2]] <- c(-3,4)
nu         <- c(0,0)
pii        <- c(0.6,0.4)
percen <- c(0.1,0.2)
n <- 200
g <- 2
seed <- 654678

set.seed(seed)
test = rMMSN(n = n, pii = pii,mu = mu,Sigma = Sigma,shape = shape,
percen = percen, each = TRUE, family = "SN")

Zij <- test$G
cc <- test$cc
y <- test$y

## left censoring ##
LI <-cc
LS <-cc
LI[cc==1]<- -Inf
LS[cc==1]<- y[cc==1]


#full analysis may take a few seconds more...

test_fit.cc0 = fit.FMMSNC(cc, LI, LS, y, mu=mu,
Sigma = Sigma, shape=shape, pii = pii, g = 2, get.init = FALSE,
criteria = TRUE, family = "Normal", error = 0.0001,
iter.max = 200, uni.Gama = FALSE, cal.im = FALSE)


test_fit.cc = fit.FMMSNC(cc, LI, LS, y, mu=mu,
Sigma = Sigma, shape=shape, pii = pii, g = 2, get.init = FALSE,
criteria = TRUE, family = "SN", error = 0.00001,
iter.max = 350, uni.Gama = FALSE, cal.im = TRUE)

## missing data ##
pctmiss <- 0.2 # 20% of missing data in the whole data
missing <- matrix(runif(n*g), nrow = n) < pctmiss
y[missing] <- NA

cc <- matrix(nrow = n,ncol = g)
cc[missing] <- 1
cc[!missing] <- 0
```

```
LI <- cc
LS <-cc
LI[cc==1]<- -Inf
LS[cc==1]<- +Inf

test_fit.mis = fit.FMMSNC(cc, LI, LS, y, mu=mu,
Sigma = Sigma, shape=shape, pii = pii, g = 2, get.init = FALSE,
criteria = TRUE, family = "SN", error = 0.00001,
iter.max = 350, uni.Gama = FALSE, cal.im = TRUE)
```

---

rMMSN	*Random Generator of Finite Mixture of Multivariate Distributions.*

Description

It generates random realizations following a multivariate finite mixture of Skew-normal (`family ==`
`"SN"`) and normal (`family == "Normal"`) distributions under censoring. Censoring level can be set
as a percentage and it can be adjusted per group if desired.

Usage

```
rMMSN(n = NULL, mu = NULL, Sigma = NULL, shape = NULL, percent = NULL,
each = FALSE, pii = NULL, family = "SN")
```

Arguments

n	number of observations
mu	a list with g entries, where each entry represents location parameter per group, being a vector of dimension p.
Sigma	a list with g entries, where each entry represents a scale parameter per group, a matrix with dimension pxp.
shape	a list with g entries, where each entry represents a skewness parameter, being a vector of dimension p.
percent	Percentage of censored data in each group or data as a whole (see next item).
each	If each == TRUE, the data will be censored in each group, where `percent` must be a vector of dimension p. Besides, if each == FALSE (by default), the data will be censored in the whole set, then `percent` must be a vector of dimension 1.
pii	a vector of weights for the mixture of dimension g, the number of clusters. It must sum to one!
family	distribution family to be used for fitting. Options are "SN" for the Skew-normal and "Normal" for the Normal distribution respectively.

Value

It returns a list that depending of the case, it returns one or more of the following objects:

y   a n x p matrix containing the generated random realizations.

G   a vector of length n containing the group classification per subject.

cutoff   a vector containing the censoring cutoffs per group.

Author(s)

Francisco H. C. de Alencar <hildemardealencar@gmail.com>, Christian E. Galarza <cgalarza88@gmail.com>, Victor Hugo Lachos <hlachos@uconn.edu> and Larissa A. Matos <larissam@ime.unicamp.br>

Maintainer: Francisco H. C. de Alencar <hildemardealencar@gmail.com>

References

Cabral, C. R. B., Lachos, V. H., & Prates, M. O. (2012). Multivariate mixture modeling using skew-normal independent distributions. Computational Statistics & Data Analysis, 56(1), 126-142.

Prates, M. O., Lachos, V. H., & Cabral, C. (2013). mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. Journal of Statistical Software, 54(12), 1-20.

C.E. Galarza, L.A. Matos, D.K. Dey & V.H. Lachos. (2019) On Moments of Folded and Truncated Multivariate Extended Skew-Normal Distributions. Technical report. ID 19-14. University of Connecticut.

F.H.C. de Alencar, C.E. Galarza, L.A. Matos & V.H. Lachos. (2019) Finite Mixture Modeling of Censored and Missing Data Using the Multivariate Skew-Normal Distribution. echnical report. ID 19-31. University of Connecticut.

See Also

fit.FMMSNC, rMSN and rMMSN.contour

Examples

```
mu <- Sigma <- shape <- list()
mu[[1]]     <- c(-3,-4)
mu[[2]]     <- c(2,2)
shape[[1]]  <- c(-2,2)
shape[[2]]  <- c(-3,4)
Sigma[[1]]  <- matrix(c(3,1,1,4.5), 2,2)
Sigma[[2]]  <- matrix(c(2,1,1,3.5), 2,2)
pii         <- c(0.6,0.4)
percent     <- c(0.1,0.1)
family <- "SN"
n <-100

set.seed(20)
rMMSN(n = n,pii = pii, mu = mu, Sigma = Sigma, shape = shape,
percent = percent, each = TRUE, family = family)
```

| rMMSN.contour | *Pairwise Scatter Plots and Histograms for Finite Mixture of Multivariate Distributions.* |

Description

It plots the scatter plots with density contours for different multivariate distributions. Possible options are the Skew-normal (`family == "SN"`), Normal (`family == "Normal"`) and Student-t (`family == "t"`) distribution. Different colors are used by groups. Histograms are shown in the diagonal.

Usage

```
rMMSN.contour(model = NULL, y = NULL, mu = NULL, Sigma = NULL,
shape = NULL, nu = NULL, pii = NULL, Zij = NULL,
contour = FALSE, hist.Bin = 30, contour.Bin = 10,
slice = 100, col.names = NULL, length.x = c(0.5, 0.5),
length.y = c(0.5, 0.5), family = "SN")
```

Arguments

model	is an object resultant from the codefit.FMMSNC function.
y	the response matrix with dimension nxp.
mu	a list with g entries, where each entry represents location parameter per group, being a vector of dimension p.
Sigma	a list with g entries, where each entry represents a scale parameter per group, a matrix with dimension pxp.
shape	a list with g entries, where each entry represents a skewness parameter, being a vector of dimension p.
nu	the degrees of freedom for the Student-t distribution case, being a vector with dimension g.
pii	a vector of weights for the mixture of dimension g, the number of clusters. It must sum to one!
Zij	a matrix of dimension nxp indicating the group for each observation.
contour	If `contour == TRUE` the density contour will be shown, if `contour == FALSE` the density contour must be not returned.
hist.Bin	number of bins in the histograms. Default is 30.
contour.Bin	creates evenly spaced contours in the range of the data. Default is 10.
slice	desired length of the sequence for the variables grid. This grid is build for the contours.
col.names	names passed to the data matrix y of dimension p.
length.x	a vector of dimension 2 with the value to be subtracted and added from the minimum and maximum observation in the x-axis respectively. Default is `c(0.5,0.5)`.

length.y   a vector of dimension 2 with the value to be subtracted and added from the mini-
                  mum and maximum observation in the y-axis respectively. Default is `c(0.5,0.5)`.

family   distribution family to be used. Available distributions are the Skew-normal
                  ("SN"), normal ("Normal") or Student-t ("t") distribution.

Details

If the `model` object is used, the user still has the option to choose the `family`. If the `model` object
is not used, the user must input all other parameters. User may use the rMMSN function to generate
data.

Note

This functions works well for any length of g and p, but contour densities are only shown for $p = 2$.

Author(s)

Francisco H. C. de Alencar <hildemardealencar@gmail.com>, Christian E. Galarza <cgalarza88@gmail.com>,
Victor Hugo Lachos <hlachos@uconn.edu> and Larissa A. Matos <larissam@ime.unicamp.br>

Maintainer: Francisco H. C. de Alencar <hildemardealencar@gmail.com>

References

Cabral, C. R. B., Lachos, V. H., & Prates, M. O. (2012). Multivariate mixture modeling using
skew-normal independent distributions. Computational Statistics & Data Analysis, 56(1), 126-142.

Prates, M. O., Lachos, V. H., & Cabral, C. (2013). mixsmsn: Fitting finite mixture of scale mixture
of skew-normal distributions. Journal of Statistical Software, 54(12), 1-20.

C.E. Galarza, L.A. Matos, D.K. Dey & V.H. Lachos. (2019) On Moments of Folded and Truncated
Multivariate Extended Skew-Normal Distributions. Technical report. ID 19-14. University of
Connecticut.

F.H.C. de Alencar, C.E. Galarza, L.A. Matos & V.H. Lachos. (2019) Finite Mixture Modeling of
Censored and Missing Data Using the Multivariate Skew-Normal Distribution. echnical report. ID
19-31. University of Connecticut.

See Also

fit.FMMSNC, rMMSN and fit.FMMSNC

Examples

```
mu          <- Sigma <- shape <- list()
mu[[1]]     <- c(-3,-4)
mu[[2]]     <- c(2,2)
Sigma[[1]]  <- matrix(c(3,1,1,4.5), 2,2)
Sigma[[2]]  <- matrix(c(2,1,1,3.5), 2,2)
shape[[1]]  <- c(-2,2)
shape[[2]]  <- c(-3,4)
nu          <- 0
pii         <- c(0.6,0.4)
```

```
percent       <- c(0.1,0.2)
n <- 100
seed <- 654678

set.seed(seed)
test = rMMSN(n = n, pii = pii,mu = mu,Sigma = Sigma,shape = shape,
percent = percent, each = TRUE, family = "SN")


## SN ##
SN.contour = rMMSN.contour(model = NULL, y = test$y, Zij = test$G
,mu = mu, Sigma = Sigma, shape = shape, pii = pii, family = "SN")

#Plotting contours may take some time...

## SN ##
SN.contour = rMMSN.contour(model = NULL, y = test$y, Zij = test$G
,mu = mu, Sigma = Sigma, shape = shape, pii = pii, contour = TRUE,
family = "SN")

## Normal ##
N.contour = rMMSN.contour(model = NULL,y = test$y, Zij = test$G
,mu = mu, Sigma = Sigma, shape = shape, pii = pii, contour = TRUE,
family = "Normal")

## t ##
t.contour = rMMSN.contour(model = NULL,y = test$y, Zij = test$G
,mu = mu, Sigma = Sigma, shape = shape, pii = pii, nu = c(4,3),
contour = TRUE, family = "t")
```

rMSN	*Generating from Multivariate Skew-normal and Normal Random Distributions.*

Description

It generates random realizations from a multivariate Skew-normal and Normal distribution.

Usage

```
rMSN(n, mu, Sigma, shape)
```

Arguments

n	number of observations.
mu	a numeric vector of length p representing the location parameter.
Sigma	a numeric positive definite matrix with dimension pxp representing the scale parameter.

shape   a numeric vector of length p representing the skewness parameter for Skew-
                         normal(SN) case. If shape == 0, the SN case reduces to a normal (symmetric)
                         distribution.

Value

It returns a n x p matrix containing the generated random realizations.

Author(s)

Francisco H. C. de Alencar <hildemardealencar@gmail.com>, Christian E. Galarza <cgalarza88@gmail.com>,
Victor Hugo Lachos <hlachos@uconn.edu> and Larissa A. Matos <larissam@ime.unicamp.br>

Maintainer: Francisco H. C. de Alencar <hildemardealencar@gmail.com>

References

Cabral, C. R. B., Lachos, V. H., & Prates, M. O. (2012). Multivariate mixture modeling using
skew-normal independent distributions. Computational Statistics & Data Analysis, 56(1), 126-142.

Prates, M. O., Lachos, V. H., & Cabral, C. (2013). mixsmsn: Fitting finite mixture of scale mixture
of skew-normal distributions. Journal of Statistical Software, 54(12), 1-20.

C.E. Galarza, L.A. Matos, D.K. Dey & V.H. Lachos. (2019) On Moments of Folded and Truncated
Multivariate Extended Skew-Normal Distributions. Technical report. ID 19-14. University of
Connecticut.

F.H.C. de Alencar, C.E. Galarza, L.A. Matos & V.H. Lachos. (2019) Finite Mixture Modeling of
Censored and Missing Data Using the Multivariate Skew-Normal Distribution. echnical report. ID
19-31. University of Connecticut.

See Also

fit.FMMSNC, rMMSN and rMMSN.contour

Examples

```
mu     <- c(-3,-4)
Sigma  <- matrix(c(3,1,1,4.5), 2,2)
shape <- c(-3,2)
rMSN(10,mu = mu,Sigma = Sigma,shape = shape)
```

Index