Package 'CensMixReg'

April 14, 2018

Type Package

Title Censored Linear Mixture Regression Models

Version 3.1

Date 2018-04-13

Author Luis Benites Sanchez[cre, aut], Victor Hugo Lachos[cre, aut], Edgar J. Lopez Moreno[ctb]

Maintainer Luis Benites Sanchez <lbenitesanchez@gmail.com>

Imports mixsmsn, mnormt, mvtnorm, ClusterR, trimcluster

Description Fit censored linear regression models where the random errors follow a finite mixture of Scale Mixture Normal distributions.
Fit censored linear models of finite mixture multivariate Student-t and Normal distributions.
Fit censored mixture regression models based on scale mixture of normal distributions.

License GPL (>= 2)

LazyData TRUE

NeedsCompilation no

Repository CRAN

Date/Publication 2018-04-14 06:12:09 UTC

R topics documented:

CenMixReg-package *Censored Linear Mixture Regression Models*

Description

This package contains a principal function that performs to estimate the parameters of a regression model considering an error that follows a finite mixture of Scale mixture of normal distributions, using an analytically simple and efficient EM-type algorithm for iteratively computing maximum likelihood estimates of the parameters. Also contains a function for estimate the parameters of a censored linear models of finite mixture multivariate Student-t and Normal distributions.

Details

Package:	CensMixReg
Type:	Package
Version:	1.5
Date:	2017-07-26
License:	GPL (>=2)

Author(s)

Luis Benites Sanchez <lbenitesanchez@gmail.com>, Victor Hugo Lachos <hlachos@ime.unicamp.br> and Edgar J. Lopez Moreno <edgar.javier.lopez.moreno@gmail.com>

Maintainer: Luis Benites Sanchez <lbenitesanchez@gmail.com>

References

Benites, L., Lachos, V.H., Cabral, C.R.B. (2015). Robust Regression Modeling for Censored Data Based on Mixtures of Student-t Distributions. Technical Report 5, Universidade Estadual de Campinas. http://www.ime.unicamp.br/sites/default/files/rp05-15.pdf

Benites, L., Lachos, V.H., H. Bolfarine (2017). Robust Regression Modeling of Censored Databased on Mixtures of Scale Mixtures of Normal Distributions. Technical Report 1, University of Connecticut.

Karlsson, M. & Laitila, T. (2014). Finite mixture modeling of censored regression models. Statistical papers, 55(3), 627-642.

Massuia, M. B., Cabral, C. R. B., Matos, L. A. & Lachos, V. H. (2014). Influence diagnostics for student-t censored linear regression models. Statistics, (ahead-of-print), 1-21.

Arellano-Valle, R., Castro, L., Gonzalez-Farias, G. & Munoz-Gajardo, K. (2012). Student-t censored regression model: properties and inference. Statistical Methods & Applications, 21, 453-473.

Garay, A. M., Lachos, V. H., Bolfarine, H. & Cabral, C. R. (2015). Linear censored regression models with scale mixtures of normal distributions. Statistical Papers, pages 1-32.

See Also

fm.smn.cr

Examples

```
#See examples for the CensMixReg function linked above.
```

CensMmix *Censored multivariate finite mixture model*

Description

Performs a Finite Mixture Censored multivariate (FM-MC) Student-t and Normal distribution using using EM-type algorithm for iteratively computing maximum likelihood estimates of the parameters.

Usage

```
CensMmix(cc, y, nu=3, mu=NULL, Sigma = NULL, pii = NULL, g = NULL, get.init = TRUE,
         criteria = TRUE, group = FALSE, family = "Normal", error = 0.0001,
        iter.max = 300, uni.Sigma = FALSE, obs.prob= FALSE, kmeans.param = NULL)
```

Arguments

cc	Vector of censoring indicators. For each observation: 0 if non-censored, 1 if censored.
y	Vector of responses in case of right censoring.
nu	Initial value for the EM algorithm, nu it's degrees of freedom. Value of one size 1 (If Student's t)
mu	Initial value for the EM algorithm. Each of them must be a vector of length g.(the algorithm considers the number of components to be fitted based on the size of these vectors)
Sigma	a list of g arguments of matrices of initial values (dimension pxp) for the scale parameters
pii	Initial value for the EM algorithm. The vector of initial values (dimension g) for the weights for each cluster. Must sum one!
g	Numbers of components
get.init	TRUE or FALSE. It indicates if the program (TRUE) is get the initial values or if the user (FALSE) entered manually the initial values.
criteria	It indicates if are calculated the criterion selection methods (AIC, BIC, EDC and ICL)
group	TRUE or FALSE.

family	"t": fits a t-student regression mixture for censured data or "Normal": fits a Normal regression mixture censored data
error	define the stopping criterion of the algorithm
iter.max	the maximum number of iterations of the EM algorithm
uni.Sigma	TRUE: if the covariance matrix are equals or FALSE if are not equal
obs.prob	TRUE or FALSE.
kmeans.param	Parameters for the k-means clustering algorithm

Author(s)

Luis Benites Sanchez <lbenitesanchez@gmail.com>, Victor Hugo Lachos <hlachos@ime.unicamp.br>, Edgar J. Lopez Moreno <edgar.javier.lopez.moreno@gmail.com>

References

Arellano-Valle, R. B., Castro, L., Gonzalez-Farias, G. & Munos Gajardo, K. (2012). Student-t censored regression model: properties and inference. Statistical Methods and Applications, 21(4), 453-473.

Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B,39, 1-38.

Peel, D. & McLachlan, G. J. (2000). Robust mixture modelling using the t distribution. Statistics and Computing,10(4), 339-348.

Karlsson, M. & Laitila, T. (2014). Finite mixture modeling of censored regression models. Statistical Papers,55(3), 627-642.

Basso,R.M.,Lachos,V.H.,Cabral,C.R.B. & Ghosh,P. (2010). Robust mixture modeling based on scale mixtures of skew-normal distributions. Computational Statistics & Data Analysis, 54(12), 2926-2941.

Basford, K., Greenway, D.,McLachlan,G. & Peel,D. (1997). Standard errors of fitted component means of normal mixtures. Computational Statistics,12, 1-18.

See Also

CensMmix,concentration

Examples

```
#See examples for the CensMmix function linked above.
```

concentration *Concentration levels*

Description

The dataset corresponds to concentration levels of certain dissolved trace metals in freshwater streams across the Commonwealth of Virginia. This dataset consist of the concentration levels of the dissolved trace metals copper (Cu), lead (Pb), zinc (Zn), calcium (Ca) and magnesium (Mg) from 184 independent randomly selected sites in freshwater streams across Virginia. The Cu, Pb, and Zn concentrations are reported in ug/L of water, whereas Ca and Mg concentration are suitably reported in mg/L of water. Since the measurements are taken at different times, the presence of multiple limit of detection values are possible for each trace metal (VDEQ (2003)). The limit of detection but Cu and Pb is the 0.1ug/L, 1.0mg/L for Zn while Ca and Mg has limit of 0.5mg/L and 1.0mg/L. The percentage of left-censored values of 2.7% for (Ca), 4.9% for (Cu), 9.8% for (Mg) are small in comparison to 78.3% for (Pb) and 38.6% for (Zn), also note that 17.9% of the streams had 0 nondetected trace metals, 39.1% had 1, 37.0% had 2, 3.8% had 3, 1.1% had 4 and 1.1% had 5.

Usage

```
data(concentration)
```

Format

`concentration` is a data frame with 184 cases (rows) with 5 variables (columns).

Details

For a complete description of data concentration levels see VDEQ (2003).

Source

Hoffman, H. & Johnson, R. (2014). Pseudo-likelihood estimation of multivariate normal parameters in the presence of left-censored data. Journal of Agricultural, Biological, and Environmental Statistics, pages 156-171.

VDEQ (2003). The quality of virginia non-tidal streams: First year report. VDEQ Technical Bulletin WQA/2002-2001, Office of Water Quality and Assessments, Virginia Department of Environmental Quality, pages 13-16.

Examples

```
## Not run:
data(concentration)
p  <- 5
y  <- as.matrix(concentration[,1:p])
cc <- as.matrix(concentration[,(p+1):(2*p)])

fitN <- CensMmix(cc, y, nu=3, mu=NULL, Sigma = NULL,  pii = NULL, g = 2,
```

```
                  get.init = TRUE, criteria = TRUE, group = TRUE,
                  family = "Normal", error = 0.00001, iter.max = 350,
                  uni.Sigma = TRUE, obs.prob= FALSE, kmeans.param = NULL)

#Standard error (SE)
fitN$res$SE

## End(Not run)
```

`fm.smn.cr` *Censored linear mixture regression models*

Description

Performs a Finite Mixture Censored (FM-CR) using using EM-type algorithm for iteratively computing maximum likelihood estimates of the parameters.

Usage

```
fm.smn.cr(cc, y, x1, Abetas = NULL, medj = NULL, sigma2 = NULL, pii = NULL,
nu=NULL, g = NULL, family = "Normal", error = 0.00001, iter.max = 100, aitken = TRUE)
```

Arguments

`cc`	Vector of censoring indicators. For each observation: 0 if non-censored, 1 if censored.
`y`	Vector of responses in case of right censoring.
`x1`	Matrix or vector of covariates.
`Abetas`	Parameters of vector regression dimension $(p + 1)$ include intercept
`medj`	Initial value for the EM algorithm. Each of them must be a vector of length g.(the algorithm considers the number of components to be fitted based on the size of these vectors)
`sigma2`	Initial value for the EM algorithm. Each of them must be a vector of length g.(the algorithm considers the number of components to be adjusted based on the size of these vectors)
`pii`	Initial value for the EM algorithm. Each of them must be a vector of length g.(the algorithm considers the number of components to be adjusted based on the size of these vectors)
`nu`	Initial value for the EM algorithm, nu it's degrees of freedom. Value of one size 1 (If Student's t)
`g`	Numbers of components
`family`	"T": fits a t-student regression mixture for censured data or "Normal": fits a Normal regression mixture censored data or "Slash": fits a Slash regression mixture censored data
`error`	define the stopping criterion of the algorithm
`iter.max`	the maximum number of iterations of the EM algorithm
`aitken`	Aitken acceleration: TRUE or FALSE.

Author(s)

Luis Benites Sanchez <lbenitesanchez@gmail.com>, Victor Hugo Lachos <hlachos@ime.unicamp.br>, Edgar J. Lopez Moreno <edgar.javier.lopez.moreno@gmail.com>

References

Benites, L., Lachos, V.H., Cabral, C.R.B. (2015). Robust Regression Modeling for Censored Data Based on Mixtures of Student-t Distributions. Technical Report 5, Universidade Estadual de Campinas. http://www.ime.unicamp.br/sites/default/files/rp05-15.pdf

Karlsson, M. & Laitila, T. (2014). Finite mixture modeling of censored regression models. Statistical papers, 55(3), 627-642.

Massuia, M. B., Cabral, C. R. B., Matos, L. A. & Lachos, V. H. (2014). Influence diagnostics for student-t censored linear regression models. Statistics, (ahead-of-print), 1-21.

Arellano-Valle, R., Castro, L., Gonzalez-Farias, G. & Munoz-Gajardo, K. (2012). Student-t censored regression model: properties and inference. Statistical Methods & Applications, 21, 453-473.

Garay, A. M., Lachos, V. H., Bolfarine, H. & Cabral, C. R. (2015). Linear censored regression models with scale mixtures of normal distributions. Statistical Papers, pages 1-32.

Arellano-Valle, R. B., Castro, L., Gonzalez-Farias, G. & Munos Gajardo, K. (2012). Student-t censored regression model: properties and inference. Statistical Methods and Applications, 21(4), 453-473.

Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B,39, 1-38.

Peel, D. & McLachlan, G. J. (2000). Robust mixture modelling using the t distribution. Statistics and Computing,10(4), 339-348.

Karlsson, M. & Laitila, T. (2014). Finite mixture modeling of censored regression models. Statistical Papers,55(3), 627-642.

Basso,R.M.,Lachos,V.H.,Cabral,C.R.B. & Ghosh,P. (2010). Robust mixture modeling based on scale mixtures of skew-normal distributions. Computational Statistics & Data Analysis, 54(12), 2926-2941.

Basford, K., Greenway, D.,McLachlan,G. & Peel,D. (1997). Standard errors of fitted component means of normal mixtures. Computational Statistics,12, 1-18.

See Also

fm.smn.cr,wage.rates

Examples

```
#See examples for the CensMixReg function linked above.
```

`fmr.smn.cr` *Censored mixture regression models based in the Scale Mixture of Normal (SMN) distribution*

Description

Performs a Finite Mixture Regression (FMR) with censored based in the SMN using EM-type algorithm for iteratively computing maximum likelihood estimates of the parameters.

Usage

```
fmr.smn.cr(cc, y, x, Abetas = NULL, sigma2 = NULL, pii = NULL, nu=NULL, g = NULL,
              family = "Normal", error = 0.00001, iter.max = 100)
```

Arguments

`cc`	Vector of censoring indicators. For each observation: 0 if non-censored, 1 if censored.
`y`	Vector of responses in case of right censoring.
`x`	Matrix or vector of covariates for each component
`Abetas`	Parameters of vector regression dimension $(p_j + 1)$ include or not intercept, j=1,...,G
`sigma2`	Initial value for the EM algorithm. Each of them must be a vector of length g.(the algorithm considers the number of components to be adjusted based on the size of these vectors)
`pii`	Initial value for the EM algorithm. Each of them must be a vector of length g.(the algorithm considers the number of components to be adjusted based on the size of these vectors)
`nu`	Initial value for the EM algorithm, nu it's degrees of freedom. Value of one size 1 (If Student's t or Slash) or size 2 (if Contaminated Normal)
`g`	Numbers of components
`family`	"T": fits a t-student regression mixture for censured data or "Normal": fits a Normal regression mixture censored data or "Slash": fits a Slash regression mixture censored data or "NormalC": fits a Contaminated Normal regression mixture censored data
`error`	define the stopping criterion of the algorithm
`iter.max`	the maximum number of iterations of the EM algorithm

Author(s)

Luis Benites Sanchez <lbenitesanchez@gmail.com>, Victor Hugo Lachos <hlachos@ime.unicamp.br>

References

Zeller, C. B., Cabral, C. R. B. and Lachos, V. H. (2016). Robust mixture regression modeling based on scale mixtures of skew-normal distributions. Test, 25, 375-396.

See Also

fmr.smn.cr,wage.rates

Examples

```
#See examples for the fmr.smn.cr function linked above.
```

---

imm.fm.smn.cr *Information matrix*

Description

Calculate the information matrix of returned analysis based on the model `family` choice (univariate case, p=1).

Usage

```
imm.fm.smn.cr(cc, y,x1,model)
```

Arguments

y	Vector of responses in case of right censoring.
x1	Matrix or vector of covariates.
cc	Vector of censoring indicators. For each observation: 0 if non-censored, 1 if censored.
model	a variable returned by fm.smn.cr

Value

Estimate the Information Matrix of the parameters.

Author(s)

Luis Benites Sanchez <lbenitesanchez@gmail.com>, Victor Hugo Lachos <hlachos@ime.unicamp.br>, Edgar J. Lopez Moreno <edgar.javier.lopez.moreno@gmail.com>

See Also

fm.smn.cr

Examples

```
## see \code{\link{wage.rates}}
```

`initial.values.fm.smn.cr` *Initial values for the FMR-SMN-CR*

Description

Obtained the initial values for the parameter beta_j,sigma_j, p_j and nu the SMN distribution.

Usage

```
initial.values.fm.smn.cr(cc, y, x, g=2, algorithm="k-medoids", family="T",
lower=1, upper=20, space=0.1, plotLog = TRUE, searchNU=TRUE, printNU=TRUE,
  saveFigure = FALSE)
```

Arguments

`cc`	Vector of censoring indicators. For each observation: 0 if non-censored, 1 if censored.
`y`	Vector of responses in case of right censoring.
`x`	Matrix or vector of covariates for each component
`g`	Numbers of components
`algorithm`	It indicates the algorithm: "trim-kmeans","MinMax_kmeans", "k-means" and "k-medoids"
`family`	"T", "Normal", "Slash" or "NormalC"
`lower`	lower value of the search of the nu or gamma parameter
`upper`	upper value of the search of the nu or gamma parameter
`space`	number: increment of the sequence
`plotLog`	TRUE or FALSE the figure of profile
`searchNU`	TRUE or FALSE the search the nu or gamma parameter
`printNU`	TRUE or FALSE the print the nu or gamma parameter
`saveFigure`	TRUE or FALSE save the figure of profile

Examples

```
#See examples for the wage.rates function linked above.
```

initial.values.fmr.smn.cr

Initial values for the FMR-SMN-CR

Description

Obtained the initial values for the parameter beta_j,sigma_j, p_j and nu the SMN distribution.

Usage

```
initial.values.fmr.smn.cr(cc, y, x, g=2, algorithm="k-medoids", family="T",
lower=1, upper=20, space=0.1, plotLog = TRUE, searchNU=TRUE, printNU=TRUE,
saveFigure = FALSE)
```

Arguments

cc	Vector of censoring indicators. For each observation: 0 if non-censored, 1 if censored.
y	Vector of responses in case of right censoring.
x	Matrix or vector of covariates for each component
g	Numbers of components
algorithm	It indicates the algorithm: "trim-kmeans","MinMax_kmeans", "k-means" and "k-medoids"
family	"T", "Normal", "Slash" or "NormalC"
lower	lower value of the search of the nu or gamma parameter
upper	upper value of the search of the nu or gamma parameter
space	number: increment of the sequence
plotLog	TRUE or FALSE the figure of profile
searchNU	TRUE or FALSE the search the nu or gamma parameter
printNU	TRUE or FALSE the print the nu or gamma parameter
saveFigure	TRUE or FALSE save the figure of profile

Examples

```
#See examples for the wage.rates function linked above.
```

rmmixcr *Random multivariate Finite Mixture Censored generator*

Description

Random generator of multivariate FM-SMSN distributions.

Usage

```
rmmixcr(n, pii, mu, Sigma, shape, nu, percCensu, family)
```

Arguments

n	number of observations
pii	a vector of weights for the mixture (dimension of the number g of clusters). Must sum to one!
mu	a list of g lists with each list containing the necessary parameters of the selected family
Sigma	a list of g arguments of matrices of initial values (dimension pxp) for the scale parameters.
shape	a list of g arguments of vectors of initial values (dimension p)for the skewness parameters.
nu	nu it's degrees of freedom.
percCensu	Matrix of censoring indicators. For each observation: 0 if non-censored, 1 if censored.
family	distribution family to be used in fitting ("t", "Skew.t", "Skew.slash", "Skew.normal", "Normal")

Author(s)

Luis Benites Sanchez <lbenitesanchez@gmail.com>, Victor Hugo Lachos <hlachos@ime.unicamp.br>, Edgar J. Lopez Moreno <edgar.javier.lopez.moreno@gmail.com>

See Also

smsn.mmix

Examples

```
mu          <- Sigma <- shape <- family <- list()
mu[[1]]     <- c(-3,-1)
mu[[2]]     <- c(2,3)
Sigma[[1]]  <- matrix(c(8,1,1,5.5), 2,2)
Sigma[[2]]  <- matrix(c(2,1,1,2.5), 2,2)
family[[1]] <- "Skew.t"
family[[2]] <- "Skew.normal"
```

```
shape[[1]]  <- c(4,5)
shape[[2]]  <- c(1,1)
nu          <- c(3,3)
pii         <- c(0.65,0.35)
percCensu   <- c(0.15,0.15)

data        <- rmmixcr(100, pii , mu , Sigma, shape, nu, percCensu, family)
```

`wage.rates` *Wage Rates of 753 Women*

Description

Wage rates of 753 married white women with left censoring.

Usage

```
data(wage.rates)
```

Format

A data frame with 753 observed wage rates of married white women in 1975. Some wage rates are set equal to zero, this means that those wives did not work in 1975, therefore, these observations are considered left censored at zero.

inlf =1 if in labor force, 1975

hours hours worked, 1975

kidslt6 # kids < 6 years

kidsge6 # kids 6-18

age woman's age in yrs

educ years of schooling

wage estimated wage from earns., hours

huseduc husband's years of schooling

faminc family income, 1975

motheduc mother's years of schooling

fatheduc father's years of schooling

unem unem. rate in county of resid.

city =1 if live in SMSA

exper actual labor mkt exper

nwifeinc (faminc - wage*hours)/1000

expersq exper^2

Source

Mroz, T.A. 1987. "The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions".

Examples

```
#Load the data
 data(wage.rates)

#Set the response y and covariate x
 y    <- wage.rates$wage
 x1   <- cbind(1,wage.rates$age,wage.rates$educ,wage.rates$hours/1000)
 cc   <- c(rep(0,428),rep(1,325))

#########################################################################
#Example for regression modelling of censored data based on
#Mixtures of Scale Mixtures of Normal (SMN) distributions
#########################################################################
#Obtain the initial values
 initial    <- initial.values.fm.smn.cr(cc, y,x1,g=2,algorithm="k-means"
 ,family="T",lower=1,upper=20,space=1,plotLog = TRUE,searchNU=TRUE,
 printNU=FALSE, saveFigure = FALSE)

##Fits a left mixture censored Student-t model to the data
 fitT <- fm.smn.cr(cc, y, x1, Abetas = initial$Abetas, medj = initial$medj
 , sigma2 = initial$sigma2, pii = initial$pii, nu=initial$nu, g = 2,
 family = "T", error = 0.0001, iter.max = 500)

##Fits a left mixture censored Normal model to the data
 fitN <- fm.smn.cr(cc, y, x1, Abetas = initial$Abetas, medj = initial$medj
 , sigma2 = initial$sigma2, pii = initial$pii, nu=initial$nu, g = 2,
 family = "Normal", error = 0.0001, iter.max = 500)

#########################################################################
#Example for finite mixture of regression models for censored data
#based on scale mixtures of modelling of censored data based on
#Mixtures of SMN distributions
#########################################################################
#Obtain the initial values
 initial    <- initial.values.fmr.smn.cr(cc, y,x1,g=2,algorithm="k-means"
 ,family="T",lower=1,upper=20,space=1,plotLog = TRUE,searchNU=TRUE,
 printNU=FALSE, saveFigure = FALSE)

##Fits a left mixture censored Student-t model to the data
 fitT <- fmr.smn.cr( cc, y, x1, Abetas = initial$Abetas,
 sigma2 = initial$sigma2, pii = initial$pii, nu=initial$nu,
 g = 2, family = "T", error = 10^-4, iter.max = 500)

##Fits a left mixture censored Normal model to the data
 fitN <- fmr.smn.cr(cc, y, x1, Abetas = initial$Abetas,
 sigma2 = initial$sigma2, pii = initial$pii, nu=initial$nu,
```

```
g = 2, family = "Normal", error = 10^-4, iter.max = 500)
```

Index