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 .................................................. 2
CensMmix .............................................................. 3
concentration .......................................................... 5
fm.smn.cr .............................................................. 6
fmr.smn.cr ............................................................ 8
imm.fm.smn.cr ......................................................... 9
initial.values.fm.smn.cr ............................................ 10
initial.values.fmr.smn.cr ........................................... 11
rmmixcr .............................................................. 12
wage.rates .......................................................... 13

Index 16
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


CensMmix

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 param-
eters.

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

<table>
<thead>
<tr>
<th>cc</th>
<th>Vector of censoring indicators. For each observation: 0 if non-censored, 1 if censored.</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>Vector of responses in case of right censoring.</td>
</tr>
<tr>
<td>nu</td>
<td>Initial value for the EM algorithm, nu it’s degrees of freedom. Value of one size 1 (If Student’s t)</td>
</tr>
<tr>
<td>mu</td>
<td>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)</td>
</tr>
<tr>
<td>Sigma</td>
<td>a list of g arguments of matrices of initial values (dimension p xp) for the scale parameters</td>
</tr>
<tr>
<td>pii</td>
<td>Initial value for the EM algorithm. The vector of initial values (dimension g) for the weights for each cluster. Must sum one!</td>
</tr>
<tr>
<td>g</td>
<td>Numbers of components</td>
</tr>
<tr>
<td>get.init</td>
<td>TRUE or FALSE. It indicates if the program (TRUE) is get the initial values or if the user (FALSE) entered manually the initial values.</td>
</tr>
<tr>
<td>criteria</td>
<td>It indicates if are calculated the criterion selection methods (AIC, BIC, EDC and ICL)</td>
</tr>
<tr>
<td>group</td>
<td>TRUE or FALSE.</td>
</tr>
</tbody>
</table>
family  "t": fits a t-student regression mixture for censored 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


See Also

CensMmix, concentration

Examples

#See examples for the CensMmix function linked above.
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

Examples
```r
## 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)
```
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

c c 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 censored 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


See Also

fm.smn.cr,wage.rates

Examples

#See examples for the CensMixReg function linked above.
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

```r
fmr.smn.cr(cc, y, x, Abetas, 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 censored 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


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**

```r
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**

```r
## 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 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**

```r
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**

```r
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 p*p) 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**

```r
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"
```
wage.rates

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


Examples

```r
# 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)
c <- c(rep(0.428), rep(1, 325))

# Example for regression modelling of censored data based on Mixtures of Scale Mixtures of Normal (SMN) distributions
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)
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 Student-t 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
initial <- initial.values.fm.smr.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)
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 Student-t 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

*Topic **Finite Mixture**
  CensMix, 3

*Topic **Mixture Regression**
  fm.smnn.cr, 6
  fmr.smnn.cr, 8

*Topic **datasets**
  concentration, 5
  initial.values.fm.smnn.cr, 10
  initial.values.fmr.smnn.cr, 11
  wage.rates, 13

*Topic **methods**
  rmmixcr, 12

*Topic **models**
  imm.fm.smnn.cr, 9

*Topic **package**
  CenMixReg-package, 2
  fm.smnn.cr, 6
  fmr.smnn.cr, 8

CenMixReg (CenMixReg-package), 2
CenMixReg-package, 2
CensMix, 3, 4
concentration, 4, 5

fm.smnn.cr, 3, 6, 7, 9
fmr.smnn.cr, 8, 9

imm.fm.smnn.cr, 9
initial.values.fm.smnn.cr, 10
initial.values.fmr.smnn.cr, 11

rmmixcr, 12

smsn.mmnix, /2

wage.rates, 7, 9, 13