Package ‘CensRegSMSN’

November 19, 2025

Title Censored Linear Regression Models under Heavy-tailed Distributions

Version 0.0.1

Maintainer Yessenia Alvarez Gil <`yessenia.alvarez@ufpe.br`>

Description Functions for fitting univariate linear regression models under Scale Mixtures of Skew-Normal (SMSN) distributions, considering left, right or interval censoring and missing responses. Estimation is performed via an EM-type algorithm. Includes selection criteria, sample generation and envelope. For details, see Gil, Y.A., Garay, A.M., and Lachos, V.H. (2025) <doi:10.1007/s10260-025-00797-x>.

License GPL-3

Encoding UTF-8

RoxygenNote 7.3.2

Depends R (>= 3.5.0)

Imports mvtnorm, mnormt, cubature, ggplot2

NeedsCompilation no

Author Yessenia Alvarez Gil [aut, cre],
Aldo M. Garay [aut],
Victor H. Lachos [aut]

Repository CRAN

Date/Publication 2025-11-19 19:40:12 UTC

Contents

CensRegSMSN 2
gen_SMSNCens_sample 4

Index **7**

`CensRegSMSN` *Fit Censored Linear Regression Model under Scale Mixtures of Skew-Normal Distributions*

Description

Fits a univariate linear regression model with censoring and/or missing values in the response variable, assuming it follows a distribution from the Scale Mixtures of Skew-Normal (SMSN) family. Computes standard errors using the empirical information matrix and provides model selection criteria (AIC, BIC, CAIC, HQ). Optionally generates envelope plots based on martingale residuals.

Usage

```
CensRegSMSN(
  cc,
  x,
  y,
  beta = NULL,
  sigma2 = NULL,
  lambda = NULL,
  nu = NULL,
  cens = "Int",
  UL = NULL,
  get.init = TRUE,
  show.envelope = FALSE,
  error = 1e-04,
  iter.max = 300,
  family = "ST",
  verbose = TRUE
)
```

Arguments

`cc`	Indicator vector for incomplete observations of length `n`. Each element should be `0` if the observation is fully observed, or `1` if it is incomplete (either censored or missing).
`x`	Design matrix (of dimension `n x p`) corresponding to the covariates in the linear predictor.
`y`	Response vector of length `n`. For fully observed data, it contains the observed values. In the case of right or left censoring, it represents the censoring limit. For interval censoring, it corresponds to the lower bound of the censoring interval. Missing values (`NA`) are allowed.
`beta`	Optional initial values for the regression coefficients. Default is `NULL`.
`sigma2`	Optional initial value for the scale parameter. Default is `NULL`.
`lambda`	Optional initial value for the shape parameter (for skewed distributions). Default is `NULL`.

`nu`	Optional initial value for the distribution-specific parameter. Required for `T`, `ST`, `CN`, and `SCN` families. Must be a two-dimensional vector for `CN` and `SCN`. Should not be provided for `N` or `SN`. Default is `NULL`.
`cens`	Character indicating the type of censoring. Should be one of `"Left"`, `"Right"` or `"Int"`. Default is `"Int"`.
`UL`	Vector of upper limits of length `n` for interval-censored observations. Must be provided when `cens = "Int"`.
`get.init`	Logical; if `TRUE`, initial values are automatically computed. If `FALSE`, initial values must be provided. Default is `TRUE`.
`show.envelope`	Logical; if `TRUE`, an envelope plot based on transformed martingale residuals is produced. Default is `FALSE`.
`error`	Convergence threshold for the algorithm. Default is `0.0001`.
`iter.max`	Maximum number of iterations allowed in the algorithm. Default is `300`.
`family`	Character string indicating the distribution family. Possible values include: `"SN"` (Skew-Normal), `"ST"` (Skew-t), `"SCN"` (Skew Contaminated Normal), `"N"` (Normal), `"T"` (Student-t), `"CN"` (Contaminated Normal). Default is `"ST"`.
`verbose`	Logical indicating whether results should be printed to the console. Default is `TRUE`.

Details

The model assumes that the response variable follows a distribution from the Scale Mixtures of Skew-Normal (SMSN) family, which allows for heavy tails and/or asymmetry.

Interval censoring is a general framework that includes left and right censoring and missing responses, providing a unified treatment for all cases.

For the Skew Contaminated Normal (`"SCN"`) and the Contaminated Normal (`"CN"`) distributions, the `nu` parameter must be a two-dimensional vector with values in the interval $(0, 1)$.

Value

A list with the following components:

`beta`	Estimated regression coefficients.
`sigma2`	Estimated scale parameter.
`lambda`	Estimated shape parameter. For symmetric distributions ("N", "T", "CN"), this is zero.
`nu`	Estimated parameters of the scale mixture distribution. `NULL` for "SN" and "N" families. A scalar for "ST" and "T", and a vector for "SCN" and "CN".
`SE`	Standard errors of the estimated parameters.
`iter`	Number of iterations until convergence.
`logver`	Value of the log-likelihood function at convergence, computed under the fitted model.
`AIC`, `BIC`, `CAIC`, `HQ`	Information criteria for model selection.
`residual`	Transformed martingale residuals used for envelope plots. Returned only if `show.envelope = TRUE`; otherwise `NULL`.

References

Gil, Y. A., Garay, A. M. & Lachos, V. H. Likelihood-based inference for interval censored regression models under heavy-tailed distributions. *Stat Methods Appl* 34, 519–544 (2025). doi:10.1007/s1026002500797x.

Examples

```
# See examples in ?gen_SMSNCens_sample  for a complete workflow
# illustrating data generation and model fitting.
```

`gen_SMSNCens_sample` *Generate simulated censored data under heavy-tailed Distributions*

Description

Simulates a univariate linear regression dataset with censoring and/or missing values in the response variable, considering that the error follows a SMSN distribution.

Usage

```
gen_SMSNCens_sample(
  n,
  x,
  beta,
  sigma2,
  lambda,
  nu,
  cens = "Int",
  pcens = 0,
  pna = 0,
  family = "ST"
)
```

Arguments

`n`	Integer. Sample size to be generated.
`x`	Numeric matrix of covariates (dimension `n` x `p`). Not contain missing values.
`beta`	Numeric vector of regression coefficients of length `p`.
`sigma2`	Positive numeric scalar. Scale parameter of SMSN class.
`lambda`	Numeric scalar. Shape parameter that controls the skewness in the SMSN class. Ignored when `family` = `"N"`, `"T"` or `"CN"`.
`nu`	Distribution-specific parameter: for `"ST"` or `"T"`, `nu` is a scalar > 2 (degrees of freedom); for `"SCN"` or `"CN"`, a vector `(nu1, nu2)` with values in (0,1). Ignored for `"SN"` and `"N"`.

`cens`	Character string indicating the type of censoring: `"Left"`, `"Right"` or `"Int"`. Default is `"Int"`.
`pcens`	Proportion of censored observations. Must be between 0 and 1. Default is `0`.
`pna`	Proportion of missing values (treated as extreme interval censoring). Must be between 0 and 1. Only allowed when `cens = "Int"`. Default is `0`.
`family`	Character string indicating the error distribution family. Possible values: `"SN"` (Skew-Normal), `"ST"` (Skew-t), `"SCN"` (Skew Contaminated Normal), `"N"` (Normal), `"T"` (Student-t) and `"CN"` (Contaminated Normal). Default is `"ST"`.

Details

The following procedures are applied to the generated response variable with incomplete observation:

- **Left censoring**: values below a cutoff point (defined based on the `pcens`) are replaced by that cutoff, indicating that the true value is less than or equal to it.
- **Right censoring**: values above a cutoff point (also based on the `pcens`) are replaced by that value, indicating that the true value is greater than or equal to it.
- **Interval censoring**: a subset of observations is randomly selected (based on the `pcens`), and each value is replaced by an interval centered at the true value.
- **Missing data**: an additional subset of observations (defined based on the `pna`) is replaced by unbounded intervals of the form `(-Inf, Inf)`, representing complete uncertainty about the true value.

Value

A list with the following components:

`y`	Fully observed response values (uncensored).
`yc`	Incomplete response values.
`cc`	Censoring indicator. `0` for observed data and `1` for censored or missing case.
`UL`	Vector of upper limits of the censoring interval. Equal to `NULL` for left or right censoring. For missing data, equal to `Inf`.

Examples

```
set.seed(1997)

# Generate covariates and true parameter values
n      <- 500
x      <- cbind(1, rnorm(n))
beta   <- c(2, -1)
sigma2 <- 1
lambda <- 3
nu     <- 3

# Generate a simulated dataset under SMSN-ICR model, with interval censoring and/or missing values
sample <- gen_SMSNCens_sample(n = n, x = x, beta = beta, sigma2 = sigma2,
```

```
                         lambda = lambda, nu = nu, cens = "Int",
                         pcens = 0.1, pna = 0.05, family = "ST")

# Fit the SMSN-ICR model using the generated data
fit <- CensRegSMSN(sample$cc, x, sample$yc, cens = "Int", UL = sample$UL, get.init = TRUE,
                   show.envelope = TRUE, family = "ST")
```

Index

CensRegSMSN, 2

gen_SMSNCens_sample, 4