Package ‘GUD’

May 27, 2024

Title Bayesian Modal Regression Based on the GUD Family

Version 0.0.5

Description Provides probability density functions and sampling algorithms for three key distributions from the General Unimodal Distribution (GUD) family: the Flexible Gumbel (FG) distribution, the Double Two-Piece (DTP) Student-t distribution, and the Two-Piece Scale (TPSC) Student-t distribution. Additionally, this package includes a function for Bayesian linear modal regression, leveraging these three distributions for model fitting. The details of the Bayesian modal regression model based on the GUD family can be found at Liu, Huang, and Bai (2022) <doi:10.48550/arXiv.2211.10776>.

URL https://arxiv.org/pdf/2211.10776

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.1

Biarch true

Depends R (>= 3.4.0)

Imports MASS, methods, Rcpp (>= 0.12.0), RcppParallel (>= 5.0.1), rstan (>= 2.18.1), rstantools (>= 2.4.0), posterior (>= 1.5.0)

LinkingTo BH (>= 1.66.0), Rcpp (>= 0.12.0), RcppEigen (>= 0.3.3.3.0), RcppParallel (>= 5.0.1), rstan (>= 2.18.1), StanHeaders (>= 2.18.0)

SystemRequirements GNU make

NeedsCompilation yes

Author Qingyang Liu [aut, cre] (<https://orcid.org/0000-0003-3265-6330>), Xianzheng Huang [aut] (<https://orcid.org/0000-0001-7077-0869>), Ray Bai [aut] (<https://orcid.org/0000-0002-7190-7844>)

Maintainer Qingyang Liu <qingyang@email.sc.edu>

Repository CRAN

Date/Publication 2024-05-27 18:10:13 UTC

R topics documented:

`dDTP` *The DTP-Student-t Distribution*

Description

The DTP-Student-t Distribution

Usage

```
dDTP(x, theta, sigma1, sigma2, delta1, delta2)

rDTP(n, theta, sigma1, sigma2, delta1, delta2)
```

Arguments

`x`	vector of quantiles.
`theta`	vector of the location parameters.
`sigma1`	vector of the scale parameters of the left skewed part.
`sigma2`	vector of the scale parameters of the right skewed part.
`delta1`	the degree of freedom of the left skewed part.
`delta2`	the degree of freedom of the right skewed part.
`n`	number of observations.

Details

The DTP-Student-t distribution has density

$$f_{\mathrm{DTP}}\left(y \mid \theta, \sigma_1, \sigma_2, \delta_1, \delta_2\right) = w f_{\mathrm{LT}}\left(y \mid \theta, \sigma_1, \delta_1\right) + (1-w) f_{\mathrm{RT}}\left(y \mid \theta, \sigma_2, \delta_2\right),$$

where

$$w = \frac{\sigma_1 f\left(0 \mid \delta_2\right)}{\sigma_1 f\left(0 \mid \delta_2\right) + \sigma_2 f\left(0 \mid \delta_1\right)},$$

$f(0 \mid \delta)$ represents

$$f((y-\theta)/\sigma \mid \delta) \text{ evaluated at } y = \theta,$$

$$f_{\mathrm{LT}}(y \mid \theta, \sigma, \delta) = \frac{2}{\sigma} f\left(\left.\frac{y-\theta}{\sigma}\right| \delta\right) \mathbb{I}(y < \theta),$$

and

$$f_{\mathrm{RT}}(y \mid \theta, \sigma, \delta) = \frac{2}{\sigma} f\left(\left.\frac{y-\theta}{\sigma}\right| \delta\right) \mathbb{I}(y \geq \theta).$$

Value

dDTP gives the density. rDTP generates random deviates.

References

See also https://arxiv.org/pdf/2211.10776.

Examples

```
set.seed(100)
require(graphics)

# Random Number Generation
X <- rDTP(n = 1e5,theta = 5,sigma1 = 7,sigma2 = 3,delta1 = 5,delta2 = 6)

# Plot the histogram
hist(X, breaks = 100, freq = FALSE)

# The red dashed line should match the underlining histogram
points(x = seq(-100,40,length.out = 1000),
       y = dDTP(x = seq(-100,40,length.out = 1000),
                theta = 5,sigma1 = 7,sigma2 = 3,delta1 = 5,delta2 = 6),
       type = "l",
       col = "red",
       lwd = 3,
       lty = 2)
```

---

dFG *The Flexible Gumbel Distribution*

Description

The Flexible Gumbel Distribution

Usage

```
dFG(x, w, loc, sigma1, sigma2)

rFG(n, w, loc, sigma1, sigma2)
```

Arguments

x	vector of quantiles.
w	vector of weight parameters.
loc	vector of the location parameters.
sigma1	vector of the scale parameters of the left skewed part.
sigma2	vector of the scale parameters of the right skewed part.
n	number of observations.

Details

The Gumbel distribution has density

$$f_{\text{Gumbel}}\left(y \mid \theta, \sigma\right) = \frac{1}{\sigma} \exp\left\{-\frac{y-\theta}{\sigma} - \exp\left(-\frac{y-\theta}{\sigma}\right)\right\},$$

where $\theta \in \mathbb{R}$ is the mode as the location parameter, $\sigma > 0$ is the scale parameter.

The flexible Gumbel distribution has density

$$f_{\text{FG}}\left(y \mid w, \theta, \sigma_1, \sigma_2\right) = w f_{\text{Gumbel}}\left(-y \mid -\theta, \sigma_1\right) + (1-w) f_{\text{Gumbel}}\left(y \mid \theta, \sigma_2\right).$$

where $w \in [0, 1]$ is the weight parameter, $\sigma_1 > 0$ is the scale parameter of the left skewed part and $\sigma_2 > 0$ is the scale parameter of the right skewed part.

Value

`dFG` gives the density. `rFG` generates random deviates.

References

See also https://arxiv.org/pdf/2211.10776.

Examples

```
set.seed(100)
require(graphics)

# Random Number Generation
X <- rFG(n = 1e5, w = 0.3, loc = 0, sigma1 = 1, sigma2 = 2)

# Plot the histogram
hist(X, breaks = 100, freq = FALSE)

# The red dashed line should match the underlining histogram
points(x = seq(-10,20,length.out = 1000),
       y = dFG(x = seq(-10,20,length.out = 1000),
               w = 0.3, loc = 0, sigma1 = 1, sigma2 = 2),
       type = "l",
       col = "red",
       lwd = 3,
       lty = 2)
```

---

dTPSC *The TPSC-Student-t Distribution*

Description

The TPSC-Student-t Distribution

Usage

```
dTPSC(x, w, theta, sigma, delta)

rTPSC(n, w, theta, sigma, delta)
```

Arguments

`x`	vector of quantiles.
`w`	vector of weight parameters.
`theta`	vector of the location parameters.
`sigma`	vector of the scale parameters.
`delta`	the degree of freedom.
`n`	number of observations.

Details

The TPSC-Student-t distribution has density

$$f_{\mathrm{TPSC}}(y \mid w, \theta, \sigma, \delta) = w f_{\mathrm{LT}}\left(y \mid \theta, \sigma\sqrt{\frac{w}{1-w}}, \delta\right) + (1-w) f_{\mathrm{RT}}\left(y \mid \theta, \sigma\sqrt{\frac{1-w}{w}}, \delta\right),$$

where

$$f_{\mathrm{LT}}(y \mid \theta, \sigma, \delta) = \frac{2}{\sigma} f\left(\frac{y-\theta}{\sigma}\middle|\, \delta\right) \mathbb{I}(y < \theta),$$

and

$$f_{\mathrm{RT}}(y \mid \theta, \sigma, \delta) = \frac{2}{\sigma} f\left(\frac{y-\theta}{\sigma}\middle|\, \delta\right) \mathbb{I}(y \geq \theta).$$

Value

`dTPSC` gives the density. `rTPSC` generates random deviates.

References

See also https://arxiv.org/pdf/2211.10776.

Examples

```
set.seed(100)
require(graphics)

# Random Number Generation
X <- rTPSC(n = 1e5,w = 0.7,theta = -1,sigma = 3,delta = 5)

# Plot the histogram
hist(X, breaks = 100, freq = FALSE)

# The red dashed line should match the underlining histogram
```

```
points(x = seq(-70,50,length.out = 1000),
       y = dTPSC(x = seq(-70,50,length.out = 1000),
                 w = 0.7,theta = -1,sigma = 3,delta = 5),
       type = "l",
       col = "red",
       lwd = 3,
       lty = 2)
```

GUD *The 'GUD' package.*

Description

This R package encompasses the probability density functions of three key distributions: the flexible Gumbel distribution, the double two-piece Student-t distribution, and the two-piece scale Student-t distribution, all belonging to the general unimodal distribution family, along with their corresponding sampling algorithms. Additionally, the package offers a function for Bayesian linear modal regression, leveraging these three distributions for model fitting.

Author(s)

Maintainer: Qingyang Liu <qingyang@email.sc.edu> (ORCID)

Authors:

- Xianzheng Huang <huang@stat.sc.edu> (ORCID)
- Ray Bai <rbai@mailbox.sc.edu> (ORCID)

See Also

Useful links:

- https://arxiv.org/pdf/2211.10776

---

modal_regression *Bayesian Modal Regression*

Description

Bayesian Modal Regression

Usage

```
modal_regression(formula, data, model, ...)
```

Arguments

`formula`	a formula.
`data`	a dataframe.
`model`	a description of the error distribution. Can be one of "FG", "DTP" and "TPSC".
`...`	Arguments passed to `rstan::sampling` (e.g. iter, chains).

Details

The Bayesian modal regression model based on the FG, DTP or TPSC distribution is defined as:

$$Y_i = \mathbf{X}_i \boldsymbol{\beta} + e_i,$$

where e_i follows the FG, DTP or TPSC distribution.

More details of the Bayesian modal regression model can be found at at Liu, Huang, and Bai (2022) https://arxiv.org/pdf/2211.10776.

Value

A `draw` object from the **posterior** package.

Examples

```
# Save current user's options.
old <- options()
# (Optional - Running Multiple Chains in Parallel)
options(mc.cores = 2)

# Need Boston housing data from MASS package.
require(MASS)

# Fit the modal regression based on the FG distribution to the Boston housing data.
FG_model <- modal_regression(formula = medv ~ .,
                             data = Boston,
                             model = "FG",
                             chains = 2,
                             iter = 2000)
summary(FG_model)

# Fit the modal regression based on the TPSC-Student-t distribution to the Boston housing data.
TPSC_model <- modal_regression(formula = medv ~ .,
                               data = Boston,
                               model = "TPSC",
                               chains = 2,
                               iter = 2000)
summary(TPSC_model)

# Fit the modal regression based on the DTP-Student-t distribution to the Boston housing data.
DTP_model <- modal_regression(formula = medv ~ .,
                              data = Boston,
```

```
                              model = "DTP",
                              chains = 2,
                              iter = 2000)
summary(DTP_model)

# reset (all) initial options
options(old)
```

Index