Package 'vasicekreg'

September 25, 2026

Type Package

Title Vasicek-Type Distributions and Regression Models

Version 1.3.0

Date 2026-09-20

Description Provides density, cumulative distribution, quantile, and random generation functions for Vasicek-type distributions with standard normal, standard logistic, and standard hyperbolic-secant kernels. The normal-kernel distribution is parameterized by either its mean or a fixed quantile, whereas the logistic- and hyperbolic-secant-kernel distributions use fixed-quantile parameterizations. Zero-augmented, one-augmented, and zero-and-one-augmented extensions of the normal-kernel mean parameterization are also provided for responses that include boundary values. The corresponding 'NVASIM', 'NVASIQ', 'LVASIQ', 'HVASIQ', 'ZANVASIM', 'OANVASIM', and 'ZOANVASIM' families are available for fitting Generalized Additive Models for Location, Scale and Shape (GAMLSS), as introduced by Rigby and Stasinopoulos (2005, <doi:10.1111/j.1467-9876.2005.00510.x>). Two-part random-intercept regression models for zero-augmented longitudinal responses are included with either a beta or a normal-kernel Vasicek positive component, extending the framework of Chen and Li (2016, <doi:10.1093/bioinformatics/btw308>). Some functions are written in 'C++' using 'Rcpp', developed by Eddelbuettel and Francois (2011, <doi:10.18637/jss.v040.i08>).

URL https://jmazucheli.github.io/vasicekreg/

License MIT + file LICENSE

Encoding UTF-8

ByteCompile yes

LazyData true

Depends R (>= 3.6)

Imports Rcpp, stats, utils, gamlss, gamlss.dist, mvtnorm, numDeriv, statmod

LinkingTo Rcpp

Suggests betareg, knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

NeedsCompilation yes

Config/testthat/edition 3

Config/roxygen2/version 8.1.0

Author Josmar Mazucheli [aut, cre],
Bruna Alves [ctb]

Maintainer Josmar Mazucheli <jmazucheli@gmail.com>

Repository CRAN

Date/Publication 2026-09-25 07:20:02 UTC

Contents

vasicekreg-package 2
aep 5
bodyfat 6
HVASIQ 8
LVASIQ 11
NVASIM 14
NVASIQ 17
OANVASIM 20
please_microbiome 22
transport 24
trees 25
vasicek_envelope 26
zabr 29
ZANVASIM 32
zavr 34
ZOANVASIM 37

Index **41**

`vasicekreg-package` *Overview of the vasicekreg package*

Description

The **vasicekreg** package provides distribution functions and GAMLSS families for Vasicek-type distributions on the unit interval. Four base families are available:

- `NVASIM`: normal kernel with mean parameterization, where $\mu = E(Y)$.
- `NVASIQ`: normal kernel with quantile parameterization, where $\mu = Q_Y(\tau)$ for a fixed $\tau \in (0, 1)$.
- `LVASIQ`: logistic kernel with quantile parameterization, where $\mu = Q_Y(\tau)$ for a fixed $\tau \in (0, 1)$.

- HVASIQ: hyperbolic-secant kernel with quantile parameterization, where $\mu = Q_Y(\tau)$ for a fixed $\tau \in (0, 1)$.

For responses observed at the boundaries, the normal-kernel mean model is also available as:

- ZANVASIM: point mass at zero and a continuous component on $(0, 1)$.
- OANVASIM: point mass at one and a continuous component on $(0, 1)$.
- ZOANVASIM: point masses at zero and one and a continuous component on $(0, 1)$.

The documentation uses *augmented* for these boundary mixtures and *Vasicek-type* for kernel-based constructions. The established family names are retained for backward compatibility and follow familiar GAMLSS abbreviations in which ZA and OA historically denote zero- and one-adjusted families. For repeated responses in $[0, 1)$, zabr and zavr provide two-part random-intercept regressions with, respectively, beta and normal-kernel Vasicek positive components. These dedicated maximum-likelihood functions are separate from the GAMLSS families. The shape parameter $\sigma \in (0, 1)$ controls dispersion in the continuous Vasicek component. The corresponding d, p, q, and r functions provide density or probability mass values, cumulative probabilities, quantiles, and random observations, respectively.

Details

Included datasets:

- bodyfat: body-fat proportions in $(0, 1)$ and demographic covariates for 298 individuals.
- aep: hospital-stay data from 1,383 patients; noinap / los contains observations at zero and one.
- transport: bicycle-trip proportions for 60 respondents, including observations at zero.
- trees: two-year tree-survival proportions for 26 parks, including observations at one.
- please_microbiome: longitudinal genus-level relative abundances from the pediatric study of Lewis et al. (2015), in long format (3186 rows, 59 subjects, 18 genera, baseline included as a subject-level covariate).

zabr: Zero-augmented beta regression for longitudinal responses. A logistic component models presence and a beta component models positive abundance; both include independent subject-specific Gaussian random intercepts.

zavr: Zero-augmented Vasicek regression with the same two-part random-intercept structure, replacing the positive beta distribution by NVASIM.

NVASIM: Normal-kernel mean parameterization and GAMLSS family. In regression models, covariates describe the conditional mean through μ.

NVASIQ: Normal-kernel quantile parameterization and GAMLSS family. For a fixed quantile level τ, covariates describe the conditional τ-th quantile through μ.

LVASIQ: Logistic-kernel quantile parameterization and GAMLSS family. For a fixed quantile level τ, covariates describe the conditional τ-th quantile through μ. A logistic-kernel mean-regression family is not provided because the mean has no closed-form expression and does not equal μ under this parameterization.

HVASIQ: Hyperbolic-secant-kernel quantile parameterization and GAMLSS family. For a fixed quantile level τ, covariates describe the conditional τ-th quantile through μ. Its conditional mean and variance are obtained by numerical quadrature and μ must not be interpreted as the mean.

ZANVASIM: Zero-augmented normal-kernel mean family. Here $\nu = P(Y = 0)$, $\mu = E(Y \mid Y > 0)$, and the marginal mean is $E(Y) = (1 - \nu)\mu$.

OANVASIM: One-augmented normal-kernel mean family. Here $\nu = P(Y = 1)$, $\mu = E(Y \mid Y < 1)$, and the marginal mean is $E(Y) = \nu + (1 - \nu)\mu$. The parameters μ and ν therefore have the same interpretations as their counterparts in the one-inflated beta family BEOI. The shape parameter σ is distribution-specific and should not be compared directly between these families.

ZOANVASIM: Zero-and-one-augmented normal-kernel mean family. Here $\nu = P(Y = 0)$, $\tau = P(Y = 1 \mid Y > 0)$, and $\mu = E(Y \mid 0 < Y < 1)$. Consequently, $P(Y = 1) = (1 - \nu)\tau$ and $E(Y) = (1 - \nu)[\tau + (1 - \tau)\mu]$.

The distribution functions dNVASIM, pNVASIM, qNVASIM, dNVASIQ, pNVASIQ, qNVASIQ, dLVASIQ, pLVASIQ, qLVASIQ, dHVASIQ, pHVASIQ, and qHVASIQ call compiled C++ routines through **Rcpp**. The boundary-augmented distribution functions are implemented in R and reuse the compiled NVASIM functions for their continuous component. Parameter validation, the GAMLSS family definitions, and all log-likelihood derivatives are implemented in R. The mean and variance components of the LVASIQ() and HVASIQ() family objects are obtained by numerical quadrature because these moments have no closed-form expressions. The longitudinal two-part models use non-adaptive Gauss–Hermite quadrature from **statmod**; Hessian-based covariance estimates are obtained with **numDeriv**.

For the distribution functions and GAMLSS constructors associated with NVASIQ, LVASIQ, and HVASIQ, the fixed quantile level is supplied through the quantile argument. It is stored in the family definition and embedded in the residual expression, so no global variable is required. This fixed quantile level is distinct from the parameter tau in ZOANVASIM(), which represents the conditional probability at one among nonzero observations.

Author(s)

Josmar Mazucheli <jmazucheli@gmail.com>

Bruna Alves <pg402900@uem.br>

References

Chen, E. Z. and Li, H. (2016). A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. *Bioinformatics*, **32**(17), 2611–2617. doi:10.1093/bioinformatics/btw308

Dunn, P. K. and Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**(3), 236–244. doi:10.1080/10618600.1996.10474708

Fischer, M. J., Hui, A. and Hösle, S. (2017). wHS-type distributions with application to finance. *Journal of Statistics and Management Systems*, **20**(1), 67–89. doi:10.1080/09720510.2016.1190575

Lewis, J. D., Chen, E. Z., Baldassano, R. N., et al. (2015). Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric Crohn's disease. *Cell Host & Microbe*, **18**(4), 489–500. doi:10.1016/j.chom.2015.09.008

Mazucheli, J., Alves, B., Korkmaz, M. Ç., and Leiva, V. (2022). Vasicek quantile and mean regression models for bounded data: New formulation, mathematical derivations, and numerical applications. *Mathematics*, **10**, 1389. doi:10.3390/math10091389

Mazucheli, J. (2026). A zero-augmented Vasicek mixed-effects regression model for longitudinal microbiome relative abundance data. Under review.

Moral, R. A., Hinde, J. and Demetrio, C. G. B. (2017). Half-normal plots and overdispersed models in R: The hnp package. *Journal of Statistical Software*, **81**(10), 1–23. doi:10.18637/jss.v081.i10

Ospina, R. and Ferrari, S. L. P. (2010). Inflated beta distributions. *Statistical Papers*, **51**, 111–126. doi:10.1007/s0036200801254

Ospina, R. and Ferrari, S. L. P. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, **56**(6), 1609–1623. doi:10.1016/j.csda.2011.10.005

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Applied Statistics*, **54**(3), 507–554. doi:10.1111/j.14679876.2005.00510.x

Segata, N., Waldron, L., Ballarini, A., et al. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods*, **9**(8), 811–814. doi:10.1038/nmeth.2066

Vasicek, O. A. (2002). The distribution of loan portfolio value. *Risk*, **15**(12), 160–162.

Witzany, J. (2013). A note on the Vasicek's model with the logistic distribution. *Ekonomický časopis (Journal of Economics)*, **61**(10), 1053–1066.

Zhao, Y., Lee, A. H., Yau, K. K. W. and McLachlan, G. J. (2011). Assessing the adequacy of Weibull survival models: A simulated envelope approach. *Journal of Applied Statistics*, **38**(10), 2089–2097. doi:10.1080/02664763.2010.545115

See Also

Useful links:

- https://jmazucheli.github.io/vasicekreg/

`aep` *Hospital-stay appropriateness data*

Description

Data on 1,383 patients admitted to Hospital del Mar, Barcelona, during 1988 and 1990. The data record the total length of stay and the number of days classified as inappropriate.

Format

A data frame with 1,383 observations and 8 variables:

los Total number of days spent in hospital.

noinap Number of hospital-stay days classified as inappropriate.

loglos Logarithm of length of stay divided by 10, `log(los / 10)`.

sex Patient's sex: factor with levels `1` (male) and `2` (female).

ward Hospital ward: factor with levels `1` (medical), `2` (surgical), and `3` (other).

year Admission year: factor with levels `88` and `90`.

age Patient's age minus 55 years.

y Two-column matrix response whose first column is `noinap` and whose second column is `los - noinap`.

Details

Gange et al. (1996) modeled the number of inappropriate days conditional on total length of stay using binomial and beta-binomial models. For augmented continuous-response models, the patient-level proportion can be constructed as `noinap / los`. The object distributed here retains the structure and values supplied by the **gamlss.data** package.

Source

Gange, S. J., Munoz, A., Saez, M. and Alonso, J. (1996). Use of the beta-binomial distribution to model the effect of policy changes on appropriateness of hospital stays. *Applied Statistics*, **45**(3), 371–382.

References

Stasinopoulos, M. and Rigby, R. (2025). **gamlss.data**: Data for Generalized Additive Models for Location Scale and Shape. R package version 6.0-7. https://CRAN.R-project.org/package=gamlss.data

Examples

```
data("aep", package = "vasicekreg")
str(aep)
inappropriate <- with(aep, noinap / los)
table(inappropriate == 0, inappropriate == 1)
```

---

bodyfat *Body Fat Proportions Dataset*

Description

Body fat proportions from individuals assisted in a public hospital in Curitiba, Paraná, Brazil.

Usage

```
bodyfat
```

Format

A data frame with 298 observations and 10 variables. The five body-fat responses are proportions in $(0, 1)$ (for example, 0.163 represents 16.3 percent):

- `ID`: individual identifier.
- `ARMS`: arms fat proportion.
- `LEGS`: legs fat proportion.
- `BODY`: body fat proportion.
- `ANDROID`: android fat proportion.

- GYNECOID: gynoid fat proportion.
- AGE: age of individuals.
- BMI: body mass index.
- SEX: 1 for female and 2 for male.
- IPAQ: physical activity level according to IPAQ (0 = sedentary, 1 = insufficiently active, 2 = active).

Author(s)

Josmar Mazucheli <jmazucheli@gmail.com>

Bruna Alves <pg402900@uem.br>

References

Mazucheli, J., Alves, B., Korkmaz, M. Ç., and Leiva, V. (2022). Vasicek quantile and mean regression models for bounded data: New formulation, mathematical derivations, and numerical applications. *Mathematics*, **10**, 1389.

Mazucheli, J., Leiva, V., Alves, B., and Menezes, A. F. B. (2021). A new quantile regression for modeling bounded data under a unit Birnbaum-Saunders distribution with applications in medicine and politics. *Symmetry*, **13**(4), 1–21.

Petterle, R. R., Bonat, W. H., Scarpin, C. T., Jonasson, T., and Borba, V. Z. C. (2020). Multivariate quasi-beta regression models for continuous bounded data. *The International Journal of Biostatistics*, **17**(1), 39–53.

Examples

```
data(bodyfat, package = "vasicekreg")

bodyfat$AGE <- bodyfat$AGE - 46.00
bodyfat$BMI <- bodyfat$BMI - 24.72
bodyfat$SEX <- as.factor(bodyfat$SEX)
bodyfat$IPAQ<- as.factor(bodyfat$IPAQ)

library(gamlss)

## Mean regression model
fitmean <- gamlss(
  ARMS ~ AGE + BMI + SEX + IPAQ,
  data = bodyfat,
  family = NVASIM(mu.link = "logit", sigma.link = "logit")
)

## Not run:
quantile_levels <- c(0.25, 0.50, 0.75)

## Quantile regression models with the normal kernel
fit_normal <- lapply(quantile_levels, function(level) {
  gamlss(
    ARMS ~ AGE + BMI + SEX + IPAQ,
```

```
    data = bodyfat,
    family = NVASIQ(
      quantile = level,
      mu.link = "logit",
      sigma.link = "logit"
    )
  )
})

## Quantile regression models with the logistic kernel
fit_logistic <- lapply(quantile_levels, function(level) {
  gamlss(
    ARMS ~ AGE + BMI + SEX + IPAQ,
    data = bodyfat,
    family = LVASIQ(
      quantile = level,
      mu.link = "logit",
      sigma.link = "logit"
    )
  )
})

## Quantile regression models with the Hyperbolic-secant-kernel
fit_hsk <- lapply(quantile_levels, function(level) {
  gamlss(
    ARMS ~ AGE + BMI + SEX + IPAQ,
    data = bodyfat,
    family = HVASIQ(
      quantile = level,
      mu.link = "logit",
      sigma.link = "logit"
    )
  )
})

lapply(fit_normal, summary)
lapply(fit_logistic, summary)
lapply(fit_hsk, summary)

## End(Not run)
```

HVASIQ *Hyperbolic-secant-kernel Vasicek-type quantile distribution*

Description

`HVASIQ()` defines the hyperbolic-secant-kernel Vasicek-type distribution as a `gamlss.family` object for conditional quantile regression. The functions `dHVASIQ()`, `pHVASIQ()`, `qHVASIQ()`, and `rHVASIQ()` provide the density, distribution function, quantile function, and random generation.

Usage

```
dHVASIQ(x, mu, sigma, quantile = 0.5, log = FALSE)

pHVASIQ(q, mu, sigma, quantile = 0.5, lower.tail = TRUE, log.p = FALSE)

qHVASIQ(p, mu, sigma, quantile = 0.5, lower.tail = TRUE, log.p = FALSE)

rHVASIQ(n, mu, sigma, quantile = 0.5)

HVASIQ(quantile = 0.5, mu.link = "logit", sigma.link = "logit")
```

Arguments

x	Vector of quantiles in $(0, 1)$.
mu	Vector of conditional τ-th quantiles, $0 < \mu < 1$.
sigma	Vector of shape parameter values, $0 < \sigma < 1$.
quantile	Fixed quantile level $\tau \in (0, 1)$ represented by μ, used in the distribution functions and in `HVASIQ()`.
log	Logical; if `TRUE`, returns the log-density.
q	Vector of values in $[0, 1]$ at which the cumulative distribution function is evaluated.
lower.tail	Logical; if `TRUE`, probabilities are $P(X \leq x)$.
log.p	Logical; if `TRUE`, probabilities are supplied or returned on the logarithmic scale.
p	Vector of probabilities in $[0, 1]$ on the probability scale.
n	Number of observations.
mu.link	Link function for μ.
sigma.link	Link function for σ.

Details

Let

$$H(w) = \frac{2}{\pi} \arctan\{\exp(w)\}, \qquad h(w) = \frac{1}{\pi \cosh(w)},$$

and

$$q(u) = H^{-1}(u) = \log\left\{\tan\left(\frac{\pi u}{2}\right)\right\}.$$

For fixed $\tau \in (0, 1)$, define

$$a = \sqrt{\frac{1-\sigma}{\sigma}}, \qquad z = a\{q(x) - q(\mu)\} + q(\tau).$$

The cumulative distribution function and density are

$$F(x \mid \mu, \sigma, \tau) = H(z)$$

and

$$f(x \mid \mu, \sigma, \tau) = a \frac{h(z)}{h\{q(x)\}}.$$

The quantile function is

$$Q(p \mid \mu, \sigma, \tau) = H\left[q(\mu) + \sqrt{\frac{\sigma}{1-\sigma}}\{q(p) - q(\tau)\}\right].$$

Consequently, $Q(\tau) = \mu$; μ is exactly the conditional τ-th quantile and is not the conditional mean in general.

The GAMLSS family uses analytical derivatives. For one observation, let

$$d = q(y) - q(\mu), \quad z = ad + q(\tau), \quad T = \tanh(z), \quad S = \operatorname{sech}^2(z),$$

$$b = \frac{1}{2\sigma(1-\sigma)}, \qquad g = \frac{\pi}{\sin(\pi\mu)}.$$

If ℓ is the individual log-likelihood contribution, then

$$\frac{\partial \ell}{\partial \mu} = agT,$$

$$\frac{\partial \ell}{\partial \sigma} = b(adT - 1),$$

$$\frac{\partial^2 \ell}{\partial \mu^2} = g^2\{-a\cos(\pi\mu)T - a^2 S\},$$

$$\frac{\partial^2 \ell}{\partial \mu\, \partial \sigma} = -abg(T + adS),$$

and

$$\frac{\partial^2 \ell}{\partial \sigma^2} = b^2\{2(1-2\sigma) - (3-4\sigma)adT - a^2 d^2 S\}.$$

The conditional mean and variance do not have elementary closed forms and are evaluated by numerical integration of the quantile function.

The fixed level is supplied through the `quantile` argument. It is stored in the family definition and embedded as a numeric literal in the family components used by GAMLSS; no global variable is required.

Value

`dHVASIQ()` returns the density, `pHVASIQ()` the distribution function, `qHVASIQ()` the quantile function, and `rHVASIQ()` random deviates. `HVASIQ()` returns a `gamlss.family` object.

Author(s)

Josmar Mazucheli <jmazucheli@gmail.com>

References

Dunn, P. K. and Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**(3), 236–244.

Fischer, M. J., Hui, A. and Hösle, S. (2017). wHS-type distributions with application to finance. *Journal of Statistics and Management Systems*, **20**(1), 67–89. doi:10.1080/09720510.2016.1190575

Mazucheli, J., Alves, B., Korkmaz, M. C. and Leiva, V. (2022). Vasicek quantile and mean regression models for bounded data: New formulation, mathematical derivations, and numerical applications. *Mathematics*, **10**, 1389. doi:10.3390/math10091389

Examples

```
set.seed(123)
y <- rHVASIQ(500, mu = 0.60, sigma = 0.30, quantile = 0.25)

fit <- gamlss::gamlss(
    y ~ 1,
    sigma.formula = ~ 1,
    family = HVASIQ(quantile = 0.25),
    control = gamlss::gamlss.control(trace = FALSE)
)
fitted(fit, what = "mu")[1]
```

`LVASIQ` *Logistic-kernel Vasicek-type distribution with quantile parameterization*

Description

The function `LVASIQ()` defines the logistic-kernel Vasicek-type distribution as a `gamlss.family` object for conditional quantile regression. The functions `dLVASIQ`, `pLVASIQ`, `qLVASIQ`, and `rLVASIQ` give the density, distribution function, quantile function, and random generation. The parameter μ is the conditional τ-th quantile ($0 < \mu < 1$), σ is a shape parameter ($0 < \sigma < 1$), and $\tau \in (0, 1)$ is fixed by the user. The fixed level is supplied through the `quantile` argument.

Usage

```
dLVASIQ(x, mu, sigma, quantile = 0.5, log = FALSE)

pLVASIQ(q, mu, sigma, quantile = 0.5, lower.tail = TRUE, log.p = FALSE)

qLVASIQ(p, mu, sigma, quantile = 0.5, lower.tail = TRUE, log.p = FALSE)

rLVASIQ(n, mu, sigma, quantile = 0.5)

LVASIQ(quantile = 0.5, mu.link = "logit", sigma.link = "logit")
```

Arguments

x	Vector of quantiles in $(0, 1)$.
mu	Vector of τ-quantiles, $0 < \mu < 1$.
sigma	Vector of shape parameter values, $0 < \sigma < 1$.
quantile	Fixed quantile level $\tau \in (0, 1)$ represented by μ, used in the distribution functions and in LVASIQ().
log	Logical; if TRUE, the log-density is returned.
q	Vector of values in $[0, 1]$ at which the cumulative distribution function is evaluated.
lower.tail	Logical; if TRUE, probabilities are $P(X \leq x)$.
log.p	Logical; if TRUE, probabilities p are given as log(p) or cumulative probabilities are returned on the log scale, as appropriate.
p	Vector of probabilities in $[0, 1]$ on the probability scale.
n	Number of observations.
mu.link	Link function for the μ parameter.
sigma.link	Link function for the σ parameter.

Details

Let $\mathrm{logit}(u) = \log\left(\frac{u}{1-u}\right)$ and $\Lambda(z) = \frac{1}{1+e^{-z}}$, with $\lambda(z) = \Lambda(z)\,[1 - \Lambda(z)]$. Define

$$z = \sqrt{\frac{1-\sigma}{\sigma}}\,[\mathrm{logit}(x) - \mathrm{logit}(\mu)] + \mathrm{logit}(\tau).$$

Cumulative distribution function

$$F(x \mid \mu, \sigma, \tau) = \Lambda(z).$$

Probability density function

$$f(x \mid \mu, \sigma, \tau) = \sqrt{\frac{1-\sigma}{\sigma}}\,\frac{\lambda(z)}{x(1-x)}.$$

Quantile function

$$Q(p \mid \mu, \sigma, \tau) = \Lambda\left\{\mathrm{logit}(\mu) + \sqrt{\frac{\sigma}{1-\sigma}}\,[\mathrm{logit}(p) - \mathrm{logit}(\tau)]\right\}.$$

By construction $Q(\tau) = \mu$, i.e. μ is the τ-th quantile. Note that, unlike the normal-kernel Vasicek distribution, the logistic kernel does **not** yield a closed-form mean; in particular $E(X) \neq \mu$ in general.

The GAMLSS family uses analytical derivatives. For one observation, let

$$a = \sqrt{\frac{1-\sigma}{\sigma}}, \qquad d = \mathrm{logit}(y) - \mathrm{logit}(\mu), \qquad P = \Lambda\left\{ad + \mathrm{logit}(\tau)\right\},$$

and define

$$V = P(1-P), \qquad b = \frac{1}{2\sigma(1-\sigma)}, \qquad g = \frac{1}{\mu(1-\mu)}.$$

If ℓ denotes the individual log-likelihood contribution, the first derivatives are

$$\frac{\partial \ell}{\partial \mu} = -ag(1-2P)$$

and

$$\frac{\partial \ell}{\partial \sigma} = -b\left\{1 + ad(1-2P)\right\}.$$

The second and cross derivatives are

$$\frac{\partial^2 \ell}{\partial \mu^2} = g^2\left\{a(1-2\mu)(1-2P) - 2a^2V\right\},$$

$$\frac{\partial^2 \ell}{\partial \mu\, \partial \sigma} = abg\left\{(1-2P) - 2adV\right\},$$

and

$$\frac{\partial^2 \ell}{\partial \sigma^2} = b^2\left\{2(1-2\sigma) + ad(3-4\sigma)(1-2P) - 2a^2d^2V\right\}.$$

These expressions are evaluated directly by `LVASIQ()`; numerical differentiation is not used. The `mean` and `variance` components of the family object use numerical quadrature because the corresponding moments do not have elementary closed forms.

Value

`dLVASIQ` gives the density, `pLVASIQ` the distribution function, `qLVASIQ` the quantile function, and `rLVASIQ` generates random deviates. `LVASIQ()` returns a `gamlss.family` object.

Note

The level supplied through `quantile` is stored in the family definition and embedded as a numeric literal in the family components used by GAMLSS; no global variable is required.

Author(s)

Josmar Mazucheli <jmazucheli@gmail.com>

References

Mazucheli, J., Alves, B., Korkmaz, M. C. and Leiva, V. (2022). Vasicek quantile and mean regression models for bounded data: New formulation, mathematical derivations, and numerical applications. *Mathematics*, **10**, 1389.

Vasicek, O. A. (2002). The distribution of loan portfolio value. *Risk*, **15**(12), 160–162.

Witzany, J. (2013). A note on the Vasicek's model with the logistic distribution. *Ekonomický časopis (Journal of Economics)*, **61**(10), 1053–1066.

Examples

```
set.seed(123)
x <- rLVASIQ(n = 1000, mu = 0.50, sigma = 0.25, quantile = 0.5)
S <- seq(min(x), max(x), length.out = 1000)

hist(x, prob = TRUE, main = "Logistic-kernel Vasicek-type")
lines(S, dLVASIQ(x = S, mu = 0.50, sigma = 0.25, quantile = 0.5), col = 2)

plot(ecdf(x))
lines(S, pLVASIQ(q = S, mu = 0.50, sigma = 0.25, quantile = 0.5), col = 2)

data <- data.frame(
  y = rLVASIQ(n = 100, mu = 0.50, sigma = 0.25, quantile = 0.50)
)
fit <- gamlss::gamlss(
    y ~ 1,
    data = data,
    family = LVASIQ(
      quantile = 0.50, mu.link = "logit", sigma.link = "logit"
    )
)
fitted(fit, what = "mu")[1:5]
```

NVASIM *Normal-kernel Vasicek-type distribution with mean parameterization*

Description

Defines the normal-kernel Vasicek distribution under a mean parameterization for use as a `gamlss.family`. The parameter μ represents the mean of the distribution, with $0 < \mu < 1$, and σ is a shape parameter.

The density, distribution function, quantile function and random number generation are provided by `dNVASIM()`, `pNVASIM()`, `qNVASIM()` and `rNVASIM()`, respectively.

Usage

```
dNVASIM(x, mu, sigma, log = FALSE)

pNVASIM(q, mu, sigma, lower.tail = TRUE, log.p = FALSE)

qNVASIM(p, mu, sigma, lower.tail = TRUE, log.p = FALSE)

rNVASIM(n, mu, sigma)

NVASIM(mu.link = "logit", sigma.link = "logit")
```

Arguments

x	Vector of quantiles in the interval $(0, 1)$.
mu	Vector of mean values.
sigma	Vector of shape parameter values.
log	Logical; if TRUE, the log-density is returned.
q	Vector of values in $[0, 1]$ at which the cumulative distribution function is evaluated.
lower.tail	Logical; if TRUE, probabilities $P(X \leq x)$ are returned.
log.p	Logical; if TRUE, probabilities p are given as log(p) or cumulative probabilities are returned on the log scale, as appropriate.
p	Vector of probabilities in $[0, 1]$ on the probability scale.
n	Number of observations. If length(n) > 1, the length is taken to be the number required.
mu.link	Link function for the μ parameter.
sigma.link	Link function for the σ parameter.

Details

Probability density function

$$f(x \mid \mu, \sigma) = \sqrt{\frac{1-\sigma}{\sigma}} \exp\left\{ \frac{1}{2} \left[\Phi^{-1}(x)^2 - \left(\frac{\Phi^{-1}(x)\sqrt{1-\sigma} - \Phi^{-1}(\mu)}{\sqrt{\sigma}} \right)^2 \right] \right\}$$

Cumulative distribution function

$$F(x \mid \mu, \sigma) = \Phi\left(\frac{\Phi^{-1}(x)\sqrt{1-\sigma} - \Phi^{-1}(\mu)}{\sqrt{\sigma}} \right)$$

Quantile function

$$Q(p \mid \mu, \sigma) = F^{-1}(p \mid \mu, \sigma) = \Phi\left(\frac{\Phi^{-1}(\mu) + \Phi^{-1}(p)\sqrt{\sigma}}{\sqrt{1-\sigma}} \right)$$

Expected value

$$E(X) = \mu$$

Variance

$$Var(X) = \Phi_2\left(\Phi^{-1}(\mu), \Phi^{-1}(\mu), \sigma\right) - \mu^2$$

where $(x, \mu, \sigma, p) \in (0, 1)$ and $\Phi_2(\cdot)$ is the cumulative distribution function for the standard bivariate normal distribution with correlation σ.

Value

NVASIM() returns a gamlss.family object.

Note

In the NVASIM() parameterization, μ corresponds to the mean of the distribution and σ is a shape parameter.

Author(s)

Josmar Mazucheli <jmazucheli@gmail.com> Bruna Alves <pg402900@uem.br>

References

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall, London.

Mazucheli, J., Alves, B., Korkmaz, M. C., and Leiva, V. (2022). Vasicek quantile and mean regression models for bounded data: New formulation, mathematical derivations, and numerical applications. *Mathematics*, **10**, 1389. doi:10.3390/math10091389

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape (with discussion). *Applied Statistics*, **54**(3), 507–554.

Rigby, R. A., Stasinopoulos, D. M., Heller, G. Z., and De Bastiani, F. (2019). *Distributions for Modeling Location, Scale, and Shape: Using GAMLSS in R*. Chapman and Hall/CRC.

Stasinopoulos, D. M. and Rigby, R. A. (2007). Generalized additive models for location, scale and shape (GAMLSS) in R. *Journal of Statistical Software*, **23**(7), 1–46.

Stasinopoulos, D. M., Rigby, R. A., Heller, G., Voudouris, V., and De Bastiani, F. (2017). *Flexible Regression and Smoothing: Using GAMLSS in R*. Chapman and Hall/CRC.

Vasicek, O. A. (1987). Probability of loss on loan portfolio. KMV Corporation.

Vasicek, O. A. (2002). The distribution of loan portfolio value. *Risk*, **15**(12), 160–162.

See Also

NVASIQ, pmvnorm

Examples

```
set.seed(123)
x <- rNVASIM(n = 1000, mu = 0.50, sigma = 0.69)
R <- range(x)
S <- seq(from = R[1], to = R[2], length.out = 1000)

hist(x, prob = TRUE, main = 'Vasicek')
lines(S, dNVASIM(x = S, mu = 0.50, sigma = 0.69), col = 2)

plot(ecdf(x))
lines(S, pNVASIM(q = S, mu = 0.50, sigma = 0.69), col = 2)

plot(quantile(x, probs = S), type = "l")
lines(qNVASIM(p = S, mu = 0.50, sigma = 0.69), col = 2)

library(gamlss)
set.seed(123)
data <- data.frame(y =  rNVASIM(n = 100, mu = 0.5, sigma = 0.69))
```

```
fit <- gamlss(y ~ 1, data = data, mu.link = 'logit', sigma.link = 'logit', family = NVASIM)
1 /(1 + exp(-fit$mu.coefficients))
1 /(1 + exp(-fit$sigma.coefficients))

## Not run:
library(gamlss)
set.seed(123)

n <- 1000
x <- rbinom(n, size = 1, prob = 0.5)
eta <- 0.5 + 1 * x;
mu <- 1 / (1 + exp(-eta));
sigma <- 0.5;
y <- rNVASIM(n, mu, sigma)
data <- data.frame(y, x)

fit <- gamlss(y ~ x, data = data, family = NVASIM, mu.link = 'logit', sigma.link = 'logit',
control = gamlss.control(n.cyc = 200))
summary(fit)

## End(Not run)
```

NVASIQ *Normal-kernel Vasicek-type distribution with quantile parameterization*

Description

The function NVASIQ() defines the normal-kernel Vasicek-type distribution as a gamlss.family object. In this parameterization, μ corresponds to the fixed τ-th quantile and σ is a shape parameter. The fixed level is supplied through the quantile argument. The functions dNVASIQ, pNVASIQ, qNVASIQ, and rNVASIQ define the density, distribution function, quantile function, and random generation for the Vasicek distribution, respectively.

Usage

```
dNVASIQ(x, mu, sigma, quantile = 0.5, log = FALSE)

pNVASIQ(q, mu, sigma, quantile = 0.5, lower.tail = TRUE, log.p = FALSE)

qNVASIQ(p, mu, sigma, quantile = 0.5, lower.tail = TRUE, log.p = FALSE)

rNVASIQ(n, mu, sigma, quantile = 0.5)

NVASIQ(quantile = 0.5, mu.link = "logit", sigma.link = "logit")
```

Arguments

x	Vector of quantiles in the interval $(0, 1)$.
mu	Vector of τ-th quantile parameter values.
sigma	Vector of shape parameter values.
quantile	Fixed quantile level $\tau \in (0, 1)$ used in the distribution functions and in the NVASIQ() GAMLSS family.
log, log.p	Logical; if TRUE, probabilities are returned on the log scale.
q	Vector of values in $[0, 1]$ at which the cumulative distribution function is evaluated.
lower.tail	Logical; if TRUE (default), $P(X \leq x)$ is returned; otherwise, $P(X > x)$.
p	Vector of probabilities in $[0, 1]$ on the probability scale.
n	Number of observations. If length(n) > 1, the length is taken to be the number required.
mu.link	Link function for the μ parameter.
sigma.link	Link function for the σ parameter.

Details

Probability density function:

$$f\left(x \mid \mu, \sigma, \tau\right) = \sqrt{\frac{1-\sigma}{\sigma}} \exp\left\{\frac{1}{2}\left[\Phi^{-1}(x)^2 - \left(\frac{\sqrt{1-\sigma}\left(\Phi^{-1}(x) - \Phi^{-1}(\mu)\right) - \sqrt{\sigma}\,\Phi^{-1}(\tau)}{\sqrt{\sigma}}\right)^2\right]\right\}.$$

Cumulative distribution function:

$$F\left(x \mid \mu, \sigma, \tau\right) = \Phi\left(\frac{\sqrt{1-\sigma}\left(\Phi^{-1}(x) - \Phi^{-1}(\mu)\right) - \sqrt{\sigma}\,\Phi^{-1}(\tau)}{\sqrt{\sigma}}\right).$$

where $0 < x < 1$, $0 < \mu < 1$, $0 < \sigma < 1$, and $0 < \tau < 1$; μ is the τ-th quantile and σ is the shape parameter.

Value

NVASIQ() returns a gamlss.family object that can be used to fit a Vasicek-type distribution using the gamlss function.

Note

For NVASIQ(), μ corresponds to the τ-th quantile and σ is a shape parameter. The level supplied through quantile is stored in the family definition and embedded as a numeric literal in the family components used by GAMLSS; no global variable is required.

Author(s)

Josmar Mazucheli <jmazucheli@gmail.com>

Bruna Alves <pg402900@uem.br>

References

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall, London.

Mazucheli, J., Alves, B., Korkmaz, M. Ç., and Leiva, V. (2022). Vasicek quantile and mean regression models for bounded data: New formulation, mathematical derivations, and numerical applications. *Mathematics*, **10**, 1389.

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape (with discussion). *Applied Statistics*, **54**(3), 507–554.

Rigby, R. A., Stasinopoulos, D. M., Heller, G. Z., and De Bastiani, F. (2019). *Distributions for Modeling Location, Scale, and Shape: Using GAMLSS in R*. Chapman and Hall/CRC.

Stasinopoulos, D. M. and Rigby, R. A. (2007). Generalized additive models for location, scale and shape (GAMLSS) in R. *Journal of Statistical Software*, **23**(7), 1–46.

Stasinopoulos, D. M., Rigby, R. A., Heller, G., Voudouris, V., and De Bastiani, F. (2017). *Flexible Regression and Smoothing: Using GAMLSS in R*. Chapman and Hall/CRC.

Vasicek, O. A. (1987). Probability of loss on loan portfolio. *KMV Corporation*.

Vasicek, O. A. (2002). The distribution of loan portfolio value. *Risk*, **15**(12), 160–162.

See Also

NVASIM

Examples

```
set.seed(123)
x <- rNVASIQ(n = 1000, mu = 0.50, sigma = 0.69, quantile = 0.50)
R <- range(x)
S <- seq(from = R[1], to = R[2], length.out = 1000)

hist(x, prob = TRUE, main = "Vasicek")
lines(S, dNVASIQ(x = S, mu = 0.50, sigma = 0.69, quantile = 0.50), col = 2)

plot(ecdf(x))
lines(S, pNVASIQ(q = S, mu = 0.50, sigma = 0.69, quantile = 0.50), col = 2)

plot(quantile(x, probs = S), type = "l")
lines(qNVASIQ(p = S, mu = 0.50, sigma = 0.69, quantile = 0.50), col = 2)

library(gamlss)
set.seed(123)
data <- data.frame(
  y = rNVASIQ(n = 100, mu = 0.50, sigma = 0.69, quantile = 0.50)
)

fit <- gamlss(y ~ 1, data = data,
              family = NVASIQ(quantile = 0.50,
                              mu.link = "logit",
                              sigma.link = "logit"))
1 / (1 + exp(-fit$mu.coefficients))
1 / (1 + exp(-fit$sigma.coefficients))
```

```
set.seed(123)
n <- 100
x <- rbinom(n, size = 1, prob = 0.5)
eta <- 0.5 + 1 * x
mu <- 1 / (1 + exp(-eta))
sigma <- 0.5
y <- rNVASIQ(n, mu, sigma, quantile = 0.5)
data <- data.frame(y, x)

fit <- gamlss(
  y ~ x, data = data, family = NVASIQ(quantile = 0.50)
)

fitquantiles <- lapply(c(0.10, 0.25, 0.50, 0.75, 0.90), function(level) {
  gamlss(y ~ x, data = data, family = NVASIQ(quantile = level))
})

lapply(fitquantiles, summary)
```

OANVASIM *One-augmented normal-kernel Vasicek-type distribution*

Description

Defines a one-augmented normal-kernel Vasicek-type distribution for responses in $(0, 1]$. The parameter ν is the probability at one. Conditional on an observation in $(0, 1)$, the distribution is NVASIM with mean μ and shape parameter σ.

Usage

```
d1NVASIM(x, mu = 0.5, sigma = 0.5, nu = 0.1, log = FALSE)

p1NVASIM(q, mu = 0.5, sigma = 0.5, nu = 0.1, lower.tail = TRUE, log.p = FALSE)

q1NVASIM(p, mu = 0.5, sigma = 0.5, nu = 0.1, lower.tail = TRUE, log.p = FALSE)

r1NVASIM(n, mu = 0.5, sigma = 0.5, nu = 0.1)

dOANVASIM(x, mu = 0.5, sigma = 0.5, nu = 0.1, log = FALSE)

pOANVASIM(q, mu = 0.5, sigma = 0.5, nu = 0.1, lower.tail = TRUE, log.p = FALSE)

qOANVASIM(p, mu = 0.5, sigma = 0.5, nu = 0.1, lower.tail = TRUE, log.p = FALSE)

rOANVASIM(n, mu = 0.5, sigma = 0.5, nu = 0.1)

OANVASIM(mu.link = "logit", sigma.link = "logit", nu.link = "logit")
```

Arguments

x	Vector of values in $[0, 1]$ at which the density or probability mass is evaluated. The distribution has support $(0, 1]$, and the returned value is zero at $x = 0$.
mu	Mean of the continuous Vasicek component, in $(0, 1)$.
sigma	Shape parameter of the continuous Vasicek component, in $(0, 1)$.
nu	Probability at one, in $(0, 1)$.
log	Logical; if TRUE, log probabilities or log densities are returned.
q	Vector of values in $[0, 1]$ at which the cumulative distribution function is evaluated.
lower.tail	Logical; if TRUE, probabilities are $P(Y \leq y)$; otherwise, they are $P(Y > y)$.
log.p	Logical; if TRUE, probabilities are supplied or returned on the log scale.
p	Vector of probabilities.
n	Number of observations. If length(n) > 1, its length is taken to be the number required.
mu.link	Link function for μ.
sigma.link	Link function for σ.
nu.link	Link function for ν.

Details

Let $Y_c \sim \mathrm{NVASIM}(\mu, \sigma)$ and let $0 < \nu < 1$. The BEOI-type one-augmented distribution is defined by

$$P(Y = 1) = \nu$$

and

$$f_Y(y) = (1 - \nu) f_{Y_c}(y \mid \mu, \sigma), \quad 0 < y < 1.$$

Consequently,

$$E(Y) = \nu + (1 - \nu)\mu$$

and

$$\mathrm{Var}(Y) = (1 - \nu)\mathrm{Var}(Y_c) + \nu(1 - \nu)(1 - \mu)^2.$$

Thus, $\mu = E(Y \mid 0 < Y < 1)$ is the mean of the continuous component, whereas $\nu + (1 - \nu)\mu$ is the marginal mean.

Value

OANVASIM() returns a gamlss.family object. The functions d1NVASIM(), p1NVASIM(), q1NVASIM(), and r1NVASIM() return probability mass or density values, cumulative probabilities, quantiles, and random observations, respectively. dOANVASIM(), pOANVASIM(), qOANVASIM(), and rOANVASIM() are equivalent names following the GAMLSS family-name convention.

References

Ospina, R. and Ferrari, S. L. P. (2010). Inflated beta distributions. *Statistical Papers*, **51**, 111–126.

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Applied Statistics*, **54**(3), 507–554.

See Also

NVASIM, ZANVASIM, BEOI

Examples

```
set.seed(123)
y <- r1NVASIM(1000, mu = 0.60, sigma = 0.30, nu = 0.20)
mean(y == 1)
mean(y)
0.20 + (1 - 0.20) * 0.60

## Not run:
library(gamlss)
fit <- gamlss(
  y ~ 1,
  sigma.formula = ~ 1,
  nu.formula = ~ 1,
  family = OANVASIM(),
  control = gamlss.control(trace = FALSE)
)

## End(Not run)
```

---

please_microbiome *Longitudinal microbiome abundances from the PLEASE study*

Description

A long-format data frame of genus-level relative abundances and associated clinical covariates from the pediatric study of Lewis et al. (2015). The analytic dataset was reconstructed from the public `chvlyl/PLEASE` repository and processed following the filtering and recoding conventions of the **ZIBR** package (Chen and Li, 2016). Genus-level relative abundances were originally quantified from shotgun metagenomic sequencing using MetaPhlAn 1.7.6 (Segata et al., 2012).

Usage

```
please_microbiome
```

Format

A data frame with 3186 rows and 7 variables. Each row is one post-baseline observation of a genus in a subject, so the 18 genera contribute 177 rows each (3 visits $\times$ 59 subjects).

Genus character. Genus name, e.g. `"g__Bacteroides"`.

Sample character. Sample identifier.

Subject character. Subject identifier (e.g. `"S5001"`).

Time numeric. Follow-up week: `1`, `4` or `8`. The baseline (week 0) is stored in `Baseline` and does not appear here.

Treat factor. Treatment arm with levels `"antiTNF"` (reference) and `"EEN"` (exclusive enteral nutrition).

Baseline numeric. Subject's relative abundance of the same genus at week 0, on its original proportion scale in $[0, 1]$.

Y numeric. Relative abundance of the genus at the given post-baseline visit, in $[0, 1]$. May contain zeros.

Details

The dataset covers 59 subjects (47 anti-TNF, 12 EEN) with observations at all four scheduled visits (baseline and weeks 1, 4 and 8) and 18 bacterial genera, for a total of 236 samples in the wide format used by Chen and Li (2016). The long format provided here contains the 177 post-baseline observations per genus, together with the baseline abundance replicated within subject and genus.

Processing followed the ZIBR workflow: samples with fewer than 10,000 non-human reads were removed; genera were retained if present in more than 40% of the remaining samples and if their 90th percentile of relative abundance, computed including zeros, exceeded 1%; and the abundances of the retained genera were renormalized to sum to one within each sample, so that each modeled abundance is relative to the retained genera rather than to the whole community. These steps preceded the restriction to the anti-TNF and EEN arms (the partial enteral nutrition arm was excluded) and the selection of subjects with observations at all four scheduled visits.

In the analysis reported in the companion paper, baseline abundance was used as a subject-level covariate on its original proportion scale, without centering or standardization, and both components of each two-part model included baseline abundance, week, and treatment. One genus, *Bacteroides*, had only three zeros among the post-baseline observations and was excluded from the Beta–Vasicek model comparison, leaving 17 genera; its observations are nonetheless retained here for completeness.

Source

PLEASE repository: https://github.com/chvlyl/PLEASE

Original study: Lewis, J. D., Chen, E. Z., Baldassano, R. N., et al. (2015). Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric Crohn's disease. *Cell Host & Microbe*, **18**(4), 489–500. doi:10.1016/j.chom.2015.09.008

References

Chen, E. Z. and Li, H. (2016). A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. *Bioinformatics*, **32**(17), 2611–2617. doi:10.1093/bioinformatics/btw308

Segata, N., Waldron, L., Ballarini, A., et al. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods*, **9**(8), 811–814. doi:10.1038/nmeth.2066

Examples

```
data(please_microbiome)
head(please_microbiome)

# Dimensions and structure
```

```
dim(please_microbiome)                     # 3186 x 7
length(unique(please_microbiome$Genus))    # 18 genera
length(unique(please_microbiome$Subject))  # 59 subjects
table(please_microbiome$Treat)             # 47 antiTNF, 12 EEN

# Balanced within genus: 177 rows each
table(please_microbiome$Genus)
```

`transport` *Transportation to campus*

Description

Data from a stratified sample of 60 respondents concerning their mode of transportation to campus. The sampling design oversampled respondents who sometimes traveled to campus by bicycle.

Format

A data frame with 60 observations and 7 variables:

ntrips Number of trips to campus during the preceding four weeks.

nbiked Number of those trips made by bicycle.

status Respondent's institutional status: faculty, staff, or student.

gender Respondent's gender, recorded as `"F"` or `"M"`.

parking Duration of the parking permit, in months.

distance Distance to campus. The source documentation does not specify the measurement unit.

propbiked Proportion of trips to campus made by bicycle, calculated as `nbiked / ntrips`.

Details

The data originate from a consulting study reported by Korosteleva (2019). Menezes, Mazucheli, and Bourguignon (2021) analyzed them using an inflated unit-Weibull quantile regression model. The object distributed here retains the variable names and values supplied by the **uwquantreg** package.

Source

Korosteleva, O. (2019). *Advanced Regression Models with SAS and R*. Boca Raton, FL: CRC Press.

https://github.com/AndrMenezes/uwquantreg

References

Menezes, A. F. B., Mazucheli, J. and Bourguignon, M. (2021). A parametric quantile regression approach for modeling zero- or one-inflated double bounded data. *Biometrical Journal*, **63**(4), 841–858. doi:10.1002/bimj.202000126

Menezes, A. F. B. (2026). **uwquantreg**: unit-Weibull quantile regression. R package version 0.1.0. https://github.com/AndrMenezes/uwquantreg

Examples

```
data("transport", package = "vasicekreg")
str(transport)
with(transport, all.equal(propbiked, nbiked / ntrips))
```

`trees` *Survival of young trees*

Description

Data from a study conducted by a parks and recreation department on the two-year survival of young trees planted in 26 parks.

Format

A data frame with 26 observations and 7 variables:

planted Number of trees planted.

survived Number of trees that survived for two years.

pest Frequency of pest-control treatment.

fertilization Frequency of soil fertilization.

precip Average annual precipitation, in inches.

wind Average annual wind speed, in miles per hour.

prop Proportion of planted trees that survived for two years, calculated as `survived / planted`.

Details

The data originate from a consulting study reported by Korosteleva (2019). Menezes, Mazucheli, and Bourguignon (2021) analyzed them using an inflated unit-Weibull quantile regression model. The object distributed here retains the variable names and values supplied by the **uwquantreg** package.

Source

Korosteleva, O. (2019). *Advanced Regression Models with SAS and R*. Boca Raton, FL: CRC Press.

https://github.com/AndrMenezes/uwquantreg

References

Menezes, A. F. B., Mazucheli, J. and Bourguignon, M. (2021). A parametric quantile regression approach for modeling zero- or one-inflated double bounded data. *Biometrical Journal*, **63**(4), 841–858. doi:10.1002/bimj.202000126

Menezes, A. F. B. (2026). **uwquantreg**: unit-Weibull quantile regression. R package version 0.1.0. https://github.com/AndrMenezes/uwquantreg

Examples

```
data("trees", package = "vasicekreg")
str(trees)
with(trees, all.equal(prop, survived / planted))
```

---

vasicek_envelope *Simulated envelopes for residuals from Vasicek GAMLSS models*

Description

Constructs pointwise simulated envelopes for normalized randomized quantile residuals and generalized Cox–Snell residuals from a Vasicek model fitted by `gamlss()`. Each simulated response is generated under the fitted model, the model is re-estimated, and the residuals from the re-estimated model are ordered before the envelope is computed.

Usage

```
vasicek_envelope(
  object,
  residual = c("quantile", "cox-snell"),
  nsim = 200L,
  level = 0.95,
  envelope = c("quantile", "minmax"),
  seed = NULL,
  data = NULL,
  max.attempts = 5L * nsim,
  refit = NULL,
  simulate = NULL,
  verbose = interactive()
)

## S3 method for class 'vasicek_envelope'
print(x, ...)

## S3 method for class 'vasicek_envelope'
plot(
  x,
  which = x$residual[1L],
  show.mean = TRUE,
  xlab = NULL,
  ylab = NULL,
  main = "",
  envelope.col = "grey85",
  mean.col = "blue",
  reference.col = "red",
  point.col = "black",
```

```
  ...
)
```

Arguments

object	A fitted object of class "gamlss" using one of the Vasicek families supplied by **vasicekreg**.
residual	Character vector selecting "quantile", "cox-snell", or both.
nsim	Number of successfully re-fitted simulated samples.
level	Pointwise coverage probability used when envelope = "quantile".
envelope	Either "quantile", for percentile limits, or "minmax", for the minimum and maximum at each order position.
seed	Optional integer seed. The previous random-number state is restored on exit.
data	Optional data frame used in the original fit. It is normally recovered from object$call$data.
max.attempts	Maximum number of simulation and re-fitting attempts. The default is five times nsim.
refit	Optional function with arguments object, y, and data. It must return a re-fitted "gamlss" object. This is useful for transformed responses or nonstandard fitting calls.
simulate	Optional function with the single argument object that returns one simulated response vector. By default, the random generator associated with the fitted Vasicek family is used.
verbose	Logical; if TRUE, reports progress and failed fits.
x	An object returned by vasicek_envelope().
...	Further arguments. For the plot method, they are passed to points().
which	Residual type to be plotted.
show.mean	Logical; if TRUE, draws the pointwise mean of the ordered simulated residuals.
xlab, ylab, main	Graphical labels.
envelope.col, mean.col, reference.col, point.col	Graphical colors.

Details

Let U_i denote the probability integral transform used by the normalized randomized quantile residual. The two residuals are

$$r_i^q = \Phi^{-1}(U_i)$$

and

$$r_i^{CS} = -\log(1 - U_i).$$

For augmented families, U_i is randomized over the relevant jump of the fitted distribution at zero or one, so the reported residuals inherit whatever randomization residuals() applies for those families; the Cox–Snell residual is then also randomized at boundary observations and has an $\mathrm{Exp}(1)$ reference distribution under a correctly specified model. The Cox–Snell residual is obtained from

the quantile residual on the log survival scale, $-\log\{\Pr(Z > r_i^q)\}$, which is numerically stable in the upper tail and preserves the ordering of the quantile residuals.

Each residual is displayed as a full quantile–quantile plot of the ordered residuals against the corresponding theoretical quantiles (normal or exponential), not as a half-normal plot. The envelope is pointwise, not simultaneous: even under a correctly specified model a fraction of points is expected to fall outside the band. With `envelope = "quantile"`, the two tail probabilities are equal and sum to $1 - \text{level}$. With `envelope = "minmax"`, the limits are the minimum and maximum at each order position, following the construction used by Zhao et al.

The pointwise mean of the ordered simulated residuals is the reference calibrated to the fitted model and to the sample size. The identity line drawn by the plot method is a theoretical idealization and, for Cox–Snell residuals, may separate from the mean curve in the upper tail at finite n; in that region the mean curve is the more reliable reference.

Failed or nonconverged re-fits are discarded and replaced until `nsim` successful samples are obtained or `max.attempts` is reached. The simulation uses fitted values for every distribution parameter, so covariate-dependent parameters are retained. Automatic simulation and refitting of `NVASIQ`, `LVASIQ`, and `HVASIQ` recover the fixed level embedded in the fitted object; no global quantile-level variable is consulted. Consequently, fitted models at different quantile levels can be used in the same R session safely.

Automatic refitting assumes that `data` contains exactly the observations used by the original fit; if rows were dropped for missing values or via `subset`, supply a matching `data` or a custom `refit`.

Value

An object of class `"vasicek_envelope"`. Its `results` component contains, for each requested residual, the theoretical order statistics, ordered observed residuals, pointwise lower, mean and upper curves, and the matrix of ordered simulated residuals. The object also records the number of successful simulations, attempts and failures.

References

Moral, R. A., Hinde, J., and Demetrio, C. G. B. (2017). Half-normal plots and overdispersed models in R: The hnp package. *Journal of Statistical Software*, **81**(10), 1–23. doi:10.18637/jss.v081.i10

Zhao, Y., Lee, A. H., Yau, K. K. W., and McLachlan, G. J. (2011). Assessing the adequacy of Weibull survival models: A simulated envelope approach. *Journal of Applied Statistics*, **38**, 2089–2097. doi:10.1080/02664763.2010.545115

Examples

```
## Not run:
library(gamlss)
set.seed(123)
dat <- data.frame(y = rNVASIM(100, mu = 0.55, sigma = 0.30))
fit <- gamlss(y ~ 1, sigma.formula = ~ 1, family = NVASIM(),
              data = dat, trace = FALSE)

env <- vasicek_envelope(fit, nsim = 200, seed = 123)
plot(env, which = "quantile")
plot(env, which = "cox-snell")
```

```
## End(Not run)
```

---

zabr *Zero-Augmented Beta random-intercept model*

Description

Fits a two-component model for responses in $[0, 1)$ with a point mass at zero (discrete component, parameterized by γ) and a Beta distribution for positive values (continuous component, parameterized by β).

Usage

```
zabr(
  data,
  y,
  formula_bin = NULL,
  formula_cont = NULL,
  random = NULL,
  logistic_cov = NULL,
  beta_cov = NULL,
  subject_ind = NULL,
  time_ind,
  component_wise_test = TRUE,
  quad_n = 30,
  verbose = FALSE,
  joint_test = NULL,
  sd_lower = 1e-05,
  start = NULL,
  control = list(),
  hessian = TRUE
)
```

Arguments

`data`	A data.frame.
`y`	Character name of the response column.
`formula_bin`	One-sided formula for the presence component.
`formula_cont`	One-sided formula for the Beta component.
`random`	Formula `~ 1 | Subject` for the random intercept.
`logistic_cov`	Deprecated: character vector of presence covariates.
`beta_cov`	Deprecated: character vector of Beta covariates.
`subject_ind`	Deprecated: subject column name.

`time_ind`	Character name of the time column.
`component_wise_test`	Logical; compute component-wise LRTs?
`quad_n`	Integer; number of Gauss-Hermite quadrature points.
`verbose`	Logical; print progress?
`joint_test`	NULL, TRUE, or FALSE; compute joint LRTs?
`sd_lower`	Lower bound for the random-effect SD.
`start`	Optional list of starting values.
`control`	List of control parameters.
`hessian`	Logical; compute Hessian-based standard errors?

Details

Formula and legacy arguments are mutually exclusive for each model component: use either `formula_bin` or `logistic_cov`, either `formula_cont` or `beta_cov`, and either `random` or `subject_ind`. Supplying both arguments in any pair is an error.

For a response $Y_{it} \in [0, 1)$ observed on subject i ($i = 1, \ldots, N$) at time t ($t = 1, \ldots, T$), the model places a point mass at zero and a continuous component on $(0, 1)$:

$$Y_{it} = 0 \quad \text{with probability } 1 - p_{it},$$

$$Y_{it} \sim \text{Beta}\left(\mu_{it}\phi,\ (1 - \mu_{it})\phi\right) \quad \text{with probability } p_{it},$$

where $0 < p_{it} < 1$, $0 < \mu_{it} < 1$, and $\phi > 0$. Let X_{it} and Z_{it} be the covariate vectors entering the discrete and continuous components, respectively; they may share columns or be disjoint. Both components are modeled on the logit scale:

$$\text{logit}(p_{it}) = \log\left(\frac{p_{it}}{1 - p_{it}}\right) = a_i + \gamma_0 + X_{it}^{\top}\gamma,$$

$$\text{logit}(\mu_{it}) = \log\left(\frac{\mu_{it}}{1 - \mu_{it}}\right) = b_i + \beta_0 + Z_{it}^{\top}\beta,$$

where a_i and b_i are subject-specific random intercepts that induce correlation across repeated measurements on the same subject,

$$a_i \sim N(0, \sigma_1^2), \qquad b_i \sim N(0, \sigma_2^2).$$

The fixed-effect coefficients are denoted by γ (discrete component) and β (continuous component); ϕ is the Beta precision parameter. The two random intercepts are independent, and the components have no shared parameters. Consequently, the marginal likelihood factorizes into discrete and continuous components. The two components are optimized separately; this is equivalent to maximizing their joint likelihood. Random effects are integrated out using non-adaptive Gauss–Hermite quadrature.

Value

An object of class "zabr" with components including logistic_est_table, beta_est_table, precision_table, random_effects, loglikelihood, joint_p, fit_statistics (joint model), fit_statistics_bin and fit_statistics_cont (per component; see section *Fit statistics*), and vcov. Fixed-effect names in coef() and vcov() use the prefix gamma_ for the discrete component and beta_ for the continuous component, including their intercepts.

Fit statistics

fit_statistics refers to the joint model and follows the "Fit Statistics" table of SAS PROC NLMIXED: $AIC = -2\ell + 2k$ and $BIC = -2\ell + k\log(s)$, where ℓ is the maximized log-likelihood, k the total number of parameters and s the number of subjects. BIC uses subjects, not observations. AIC() and BIC() return the same values as fit_statistics.

fit_statistics_bin and fit_statistics_cont are informational. The continuous component counts only observations with $Y > 0$ and subjects with at least one such observation, as if it were fitted alone to the positive responses. Consequently, AIC is additive across components, whereas BIC is additive only when every subject has at least one positive response. To compare models, use fit_statistics or BIC(); do not sum component-wise values.

logLik() sets the nobs attribute to the number of subjects, so that BIC() follows PROC NLMIXED.

The values coincide numerically with PROC NLMIXED only when the log-likelihood coincides: the same model with independent random intercepts and the same number of parameters, non-adaptive quadrature with the same number of points (NOAD and QPOINTS= equal to quad_n), and data sorted by subject.

References

Chen, E. Z. and Li, H. (2016). A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. *Bioinformatics*, **32**(17), 2611–2617. doi:10.1093/bioinformatics/btw308

See Also

zavr

Examples

```
data(please_microbiome)

d <- subset(please_microbiome, Genus == "g__Bifidobacterium")
d$Subject <- factor(d$Subject)

fit <- zabr(
  data         = d,
  y            = "Y",
  formula_bin  = ~ Baseline + Time + Treat,
  formula_cont = ~ Baseline + Time + Treat,
  random       = ~ 1 | Subject,
  time_ind     = "Time"
)
fit
```

ZANVASIM *Zero-augmented normal-kernel Vasicek-type distribution*

Description

Defines a zero-augmented normal-kernel Vasicek-type distribution for responses in $[0, 1)$. The parameter ν is the probability of a structural zero. Conditional on a positive response, the distribution is NVASIM with mean μ and shape parameter σ.

Usage

```
d0NVASIM(x, mu = 0.5, sigma = 0.5, nu = 0.1, log = FALSE)

p0NVASIM(q, mu = 0.5, sigma = 0.5, nu = 0.1, lower.tail = TRUE, log.p = FALSE)

q0NVASIM(p, mu = 0.5, sigma = 0.5, nu = 0.1, lower.tail = TRUE, log.p = FALSE)

r0NVASIM(n, mu = 0.5, sigma = 0.5, nu = 0.1)

dZANVASIM(x, mu = 0.5, sigma = 0.5, nu = 0.1, log = FALSE)

pZANVASIM(q, mu = 0.5, sigma = 0.5, nu = 0.1, lower.tail = TRUE, log.p = FALSE)

qZANVASIM(p, mu = 0.5, sigma = 0.5, nu = 0.1, lower.tail = TRUE, log.p = FALSE)

rZANVASIM(n, mu = 0.5, sigma = 0.5, nu = 0.1)

ZANVASIM(mu.link = "logit", sigma.link = "logit", nu.link = "logit")
```

Arguments

x	Vector of values in $[0, 1]$ at which the density or probability mass is evaluated. The distribution has support $[0, 1)$, and the returned value is zero at $x = 1$.
mu	Mean of the positive Vasicek component, in $(0, 1)$.
sigma	Shape parameter of the positive Vasicek component, in $(0, 1)$.
nu	Probability of a structural zero, in $(0, 1)$.
log	Logical; if TRUE, log probabilities or log densities are returned.
q	Vector of values in $[0, 1]$ at which the cumulative distribution function is evaluated.
lower.tail	Logical; if TRUE, probabilities are $P(Y \leq y)$; otherwise, they are $P(Y > y)$.
log.p	Logical; if TRUE, probabilities are supplied or returned on the log scale.
p	Vector of probabilities.

n	Number of observations. If length(n) > 1, its length is taken to be the number required.
mu.link	Link function for μ.
sigma.link	Link function for σ.
nu.link	Link function for ν.

Details

Let $Y_+ \sim \text{NVASIM}(\mu, \sigma)$ and let $0 < \nu < 1$. The zero-augmented distribution is defined by

$$P(Y = 0) = \nu$$

and

$$f_Y(y) = (1 - \nu) f_{Y_+}(y \mid \mu, \sigma), \quad 0 < y < 1.$$

Its cumulative distribution function is

$$F_Y(y) = \nu + (1 - \nu) F_{Y_+}(y \mid \mu, \sigma), \quad 0 < y < 1.$$

Consequently,

$$E(Y) = (1 - \nu)\mu$$

and

$$\text{Var}(Y) = (1 - \nu)\text{Var}(Y_+) + \nu(1 - \nu)\mu^2.$$

Thus, μ is the mean conditional on $Y > 0$; it is not the marginal mean when $\nu > 0$. The marginal mean is $(1 - \nu)\mu$.

Value

ZANVASIM() returns a gamlss.family object. The functions d0NVASIM(), p0NVASIM(), q0NVASIM(), and r0NVASIM() return density or probability mass values, cumulative probabilities, quantiles, and random observations, respectively. dZANVASIM(), pZANVASIM(), qZANVASIM(), and rZANVASIM() are equivalent names following the GAMLSS family-name convention.

References

Mazucheli, J., Alves, B., Korkmaz, M. C., and Leiva, V. (2022). Vasicek quantile and mean regression models for bounded data: New formulation, mathematical derivations, and numerical applications. *Mathematics*, **10**, 1389. doi:10.3390/math10091389

Ospina, R. and Ferrari, S. L. P. (2010). Inflated beta distributions. *Statistical Papers*, **51**, 111–126.

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Applied Statistics*, **54**(3), 507–554.

See Also

NVASIM, BEZI

Examples

```
set.seed(123)
y <- r0NVASIM(1000, mu = 0.60, sigma = 0.30, nu = 0.20)
mean(y == 0)
mean(y)
(1 - 0.20) * 0.60

library(gamlss)
fit <- gamlss(
  y ~ 1,
  sigma.formula = ~ 1,
  nu.formula = ~ 1,
  family = ZANVASIM(),
  control = gamlss.control(trace = FALSE)
)
fitted(fit, what = "mu")[1]
fitted(fit, what = "sigma")[1]
fitted(fit, what = "nu")[1]
```

---

zavr *Zero-Augmented Vasicek random-intercept model*

Description

Fits a two-component model for responses in $[0, 1)$ with a point mass at zero (discrete component, parameterized by γ) and a Vasicek distribution for positive values (continuous component, parameterized by β).

Usage

```
zavr(
  data,
  y,
  formula_bin = NULL,
  formula_cont = NULL,
  random = NULL,
  logistic_cov = NULL,
  vasicek_cov = NULL,
  subject_ind = NULL,
  time_ind,
  component_wise_test = TRUE,
  quad_n = 30,
  verbose = FALSE,
  joint_test = NULL,
  sd_lower = 1e-05,
  start = NULL,
  control = list(),
```

```
  hessian = TRUE
)
```

Arguments

`data`	A data.frame.
`y`	Character name of the response column.
`formula_bin`	One-sided formula for the presence component.
`formula_cont`	One-sided formula for the Vasicek component.
`random`	Formula `~ 1 | Subject` for the random intercept.
`logistic_cov`	Deprecated: character vector of presence covariates.
`vasicek_cov`	Deprecated: character vector of Vasicek covariates.
`subject_ind`	Deprecated: subject column name.
`time_ind`	Character name of the time column.
`component_wise_test`	Logical; compute component-wise LRTs?
`quad_n`	Integer; number of Gauss-Hermite quadrature points.
`verbose`	Logical; print progress?
`joint_test`	NULL, TRUE, or FALSE; compute joint LRTs?
`sd_lower`	Lower bound for the random-effect SD.
`start`	Optional list of starting values.
`control`	List of control parameters.
`hessian`	Logical; compute Hessian-based standard errors?

Details

Formula and legacy arguments are mutually exclusive for each model component: use either `formula_bin` or `logistic_cov`, either `formula_cont` or `vasicek_cov`, and either `random` or `subject_ind`. Supplying both arguments in any pair is an error.

For a response $Y_{it} \in [0, 1)$ observed on subject i ($i = 1, \ldots, N$) at time t ($t = 1, \ldots, T$), the model places a point mass at zero and a continuous component on $(0, 1)$:

$$Y_{it} = 0 \quad \text{with probability } 1 - p_{it},$$
$$Y_{it} \sim \text{NVASIM}\,(\mu_{it}, \sigma) \quad \text{with probability } p_{it},$$

where $0 < p_{it} < 1$, $0 < \mu_{it} < 1$, and $\sigma \in (0, 1)$. The continuous component is the normal-kernel Vasicek (mean-parameterized) distribution `NVASIM`, for which $\mu_{it} = E(Y_{it} \mid Y_{it} > 0)$. Let X_{it} and Z_{it} be the covariate vectors entering the discrete and continuous components, respectively; they may share columns or be disjoint. Both components are modeled on the logit scale:

$$\text{logit}(p_{it}) = \log\left(\frac{p_{it}}{1 - p_{it}}\right) = a_i + \gamma_0 + X_{it}^{\top}\gamma,$$

$$\text{logit}(\mu_{it}) = \log\left(\frac{\mu_{it}}{1-\mu_{it}}\right) = b_i + \beta_0 + Z_{it}^{\top}\beta,$$

where a_i and b_i are subject-specific random intercepts that induce correlation across repeated measurements on the same subject,

$$a_i \sim N(0, \sigma_1^2), \qquad b_i \sim N(0, \sigma_2^2).$$

The fixed-effect coefficients are denoted by γ (discrete component) and β (continuous component); σ is the shape parameter of the Vasicek component. The two random intercepts are independent, and the components have no shared parameters. Consequently, the marginal likelihood factorizes into discrete and continuous components. The two components are optimized separately; this is equivalent to maximizing their joint likelihood. Random effects are integrated out using non-adaptive Gauss–Hermite quadrature.

This function differs from `zabr` only in the continuous component: `zabr` uses a Beta distribution with precision ϕ, whereas `zavr` uses the Vasicek `NVASIM` distribution with shape σ. The two models share the same discrete component and the same random-intercept structure.

Value

An object of class `"zavr"` with components including `logistic_est_table`, `vasicek_est_table`, `shape_table`, `random_effects`, `loglikelihood`, `joint_p`, `fit_statistics` (joint model), `fit_statistics_bin` and `fit_statistics_cont` (per component; see section *Fit statistics*), and `vcov`. Fixed-effect names in `coef()` and `vcov()` use the prefix `gamma_` for the discrete component and `beta_` for the continuous component, including their intercepts.

Fit statistics

`fit_statistics` refers to the joint model and follows the "Fit Statistics" table of SAS PROC NLMIXED: $AIC = -2\ell + 2k$ and $BIC = -2\ell + k\log(s)$, where ℓ is the maximized log-likelihood, k the total number of parameters and s the number of subjects. BIC uses subjects, not observations. `AIC()` and `BIC()` return the same values as `fit_statistics`.

`fit_statistics_bin` and `fit_statistics_cont` are informational. The continuous component counts only observations with $Y > 0$ and subjects with at least one such observation, as if it were fitted alone to the positive responses. Consequently, AIC is additive across components, whereas BIC is additive only when every subject has at least one positive response. To compare models, use `fit_statistics` or `BIC()`; do not sum component-wise values.

`logLik()` sets the `nobs` attribute to the number of subjects, so that `BIC()` follows PROC NLMIXED.

The values coincide numerically with PROC NLMIXED only when the log-likelihood coincides: the same model with independent random intercepts and the same number of parameters, non-adaptive quadrature with the same number of points (`NOAD` and `QPOINTS=` equal to `quad_n`), and data sorted by subject.

References

Chen, E. Z. and Li, H. (2016). A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. *Bioinformatics*, **32**(17), 2611–2617. doi:10.1093/bioinformatics/btw308

Mazucheli, J., Alves, B., Korkmaz, M. Ç., and Leiva, V. (2022). Vasicek quantile and mean regression models for bounded data: New formulation, mathematical derivations, and numerical applications. *Mathematics*, **10**, 1389. doi:10.3390/math10091389

See Also

zabr, NVASIM

Examples

```
data(please_microbiome)

d <- subset(please_microbiome, Genus == "g__Bifidobacterium")
d$Subject <- factor(d$Subject)

fit <- zavr(
  data         = d,
  y            = "Y",
  formula_bin  = ~ Baseline + Time + Treat,
  formula_cont = ~ Baseline + Time + Treat,
  random       = ~ 1 | Subject,
  time_ind     = "Time"
)
fit
```

ZOANVASIM *Zero-and-one-augmented normal-kernel Vasicek-type distribution*

Description

Defines a normal-kernel Vasicek-type distribution augmented by point masses at zero and one. Conditional on an observation in $(0, 1)$, the continuous component is NVASIM with mean μ and shape parameter σ.

Usage

```
d01NVASIM(x, mu = 0.5, sigma = 0.5, nu = 0.1, tau = 0.1, log = FALSE)

p01NVASIM(
  q,
  mu = 0.5,
  sigma = 0.5,
  nu = 0.1,
  tau = 0.1,
  lower.tail = TRUE,
  log.p = FALSE
)
```

```
q01NVASIM(
  p,
  mu = 0.5,
  sigma = 0.5,
  nu = 0.1,
  tau = 0.1,
  lower.tail = TRUE,
  log.p = FALSE
)

r01NVASIM(n, mu = 0.5, sigma = 0.5, nu = 0.1, tau = 0.1)

dZOANVASIM(x, mu = 0.5, sigma = 0.5, nu = 0.1, tau = 0.1, log = FALSE)

pZOANVASIM(
  q,
  mu = 0.5,
  sigma = 0.5,
  nu = 0.1,
  tau = 0.1,
  lower.tail = TRUE,
  log.p = FALSE
)

qZOANVASIM(
  p,
  mu = 0.5,
  sigma = 0.5,
  nu = 0.1,
  tau = 0.1,
  lower.tail = TRUE,
  log.p = FALSE
)

rZOANVASIM(n, mu = 0.5, sigma = 0.5, nu = 0.1, tau = 0.1)

ZOANVASIM(
  mu.link = "logit",
  sigma.link = "logit",
  nu.link = "logit",
  tau.link = "logit"
)
```

Arguments

x	Vector of values in $[0, 1]$ at which the density or probability mass is evaluated. The distribution has support $[0, 1]$.

mu	Mean of the continuous Vasicek component, in $(0, 1)$.
sigma	Shape parameter of the continuous Vasicek component, in $(0, 1)$.
nu	Probability at zero, $\nu = P(Y = 0)$, in $(0, 1)$.
tau	Conditional probability at one among nonzero observations, $\tau = P(Y = 1 \mid Y > 0)$, in $(0, 1)$. This parameter is unrelated to the fixed quantile level used by `NVASIQ()`, `LVASIQ()`, and `HVASIQ()`.
log	Logical; if `TRUE`, log probabilities or log densities are returned.
q	Vector of values in $[0, 1]$ at which the cumulative distribution function is evaluated.
lower.tail	Logical; if `TRUE`, probabilities are $P(Y \leq y)$; otherwise, they are $P(Y > y)$.
log.p	Logical; if `TRUE`, probabilities are supplied or returned on the log scale.
p	Vector of probabilities.
n	Number of observations. If `length(n) > 1`, its length is taken to be the number required.
mu.link	Link function for μ.
sigma.link	Link function for σ.
nu.link	Link function for ν.
tau.link	Link function for τ.

Details

Let $Y_c \sim \mathrm{NVASIM}(\mu, \sigma)$. Write p_0, p_1, and p_c for the probabilities of zero, one, and the continuous component. The sequential BEOI-type parameterization is

$$\nu = P(Y = 0), \qquad \tau = P(Y = 1 \mid Y > 0).$$

Hence,

$$p_0 = \nu, \qquad p_1 = (1 - \nu)\tau, \qquad p_c = (1 - \nu)(1 - \tau).$$

The distribution is

$$P(Y = 0) = p_0, \qquad P(Y = 1) = p_1$$

and

$$f_Y(y) = p_c f_{Y_c}(y \mid \mu, \sigma), \quad 0 < y < 1.$$

Its marginal mean and variance are

$$E(Y) = (1 - \nu)\left[\tau + (1 - \tau)\mu\right]$$

and

$$\mathrm{Var}(Y) = (1 - \nu)\left[(1 - \tau)\left\{\mathrm{Var}(Y_c) + \mu^2\right\} + \tau\right] - \left\{(1 - \nu)\left[\tau + (1 - \tau)\mu\right]\right\}^2.$$

Logit links for ν and τ guarantee valid probabilities. In the closure of the parameter space, $\nu = 0$ gives the BEOI-type one-augmented model and $\tau = 0$ gives the zero-augmented model. The implemented distribution functions and GAMLSS family use the open parameter space $0 < \nu < 1$ and $0 < \tau < 1$; these nested models are therefore limiting cases rather than admissible interior parameter values.

Value

ZOANVASIM() returns a four-parameter gamlss.family object. The functions d01NVASIM(), p01NVASIM(), q01NVASIM(), and r01NVASIM() return probability mass or density values, cumulative probabilities, quantiles, and random observations, respectively. dZOANVASIM(), pZOANVASIM(), qZOANVASIM(), and rZOANVASIM() are equivalent names following the GAMLSS family-name convention.

References

Ospina, R. and Ferrari, S. L. P. (2010). Inflated beta distributions. *Statistical Papers*, **51**, 111–126.

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Applied Statistics*, **54**(3), 507–554.

See Also

NVASIM, ZANVASIM, OANVASIM, BEOI

Examples

```
set.seed(123)
y <- r01NVASIM(
  1000, mu = 0.60, sigma = 0.30, nu = 0.20, tau = 0.25
)
c(zero = mean(y == 0), one = mean(y == 1))
mean(y)
(1 - 0.20) * (0.25 + (1 - 0.25) * 0.60)

## Not run:
library(gamlss)
fit <- gamlss(
  y ~ 1,
  sigma.formula = ~ 1,
  nu.formula = ~ 1,
  tau.formula = ~ 1,
  family = ZOANVASIM(),
  control = gamlss.control(trace = FALSE)
)

## Real-data example with observations at both boundaries
data("aep", package = "vasicekreg")
aep$inappropriate <- with(aep, noinap / los)
fit_aep <- gamlss(
  inappropriate ~ sex + ward + year + age + loglos,
  sigma.formula = ~ loglos,
  nu.formula = ~ loglos,
  tau.formula = ~ loglos,
  family = ZOANVASIM(),
  data = aep,
  control = gamlss.control(n.cyc = 200, trace = FALSE)
)

## End(Not run)
```

Index

∗ **datasets**
aep, 5
bodyfat, 6
please_microbiome, 22
transport, 24
trees, 25

aep, *3*, 5

BEOI, *4*, *22*, *40*
BEZI, *33*
bodyfat, *3*, 6

d01NVASIM *(*ZOANVASIM*)*, 37
d0NVASIM *(*ZANVASIM*)*, 32
d1NVASIM *(*OANVASIM*)*, 20
dHVASIQ *(*HVASIQ*)*, 8
dLVASIQ *(*LVASIQ*)*, 11
dNVASIM *(*NVASIM*)*, 14
dNVASIQ *(*NVASIQ*)*, 17
dOANVASIM *(*OANVASIM*)*, 20
dZANVASIM *(*ZANVASIM*)*, 32
dZOANVASIM *(*ZOANVASIM*)*, 37

gamlss, *18*

HVASIQ, *3*, 8

LVASIQ, *3*, 11

NVASIM, *3*, 14, *19*, *22*, *33*, *37*, *40*
NVASIQ, *3*, *16*, 17

OANVASIM, *4*, 20, *40*

p01NVASIM *(*ZOANVASIM*)*, 37
p0NVASIM *(*ZANVASIM*)*, 32
p1NVASIM *(*OANVASIM*)*, 20
pHVASIQ *(*HVASIQ*)*, 8
please_microbiome, *3*, 22
plot.vasicek_envelope *(*vasicek_envelope*)*, 26
pLVASIQ *(*LVASIQ*)*, 11
pmvnorm, *16*
pNVASIM *(*NVASIM*)*, 14
pNVASIQ *(*NVASIQ*)*, 17
pOANVASIM *(*OANVASIM*)*, 20
print.vasicek_envelope *(*vasicek_envelope*)*, 26
pZANVASIM *(*ZANVASIM*)*, 32
pZOANVASIM *(*ZOANVASIM*)*, 37

q01NVASIM *(*ZOANVASIM*)*, 37
q0NVASIM *(*ZANVASIM*)*, 32
q1NVASIM *(*OANVASIM*)*, 20
qHVASIQ *(*HVASIQ*)*, 8
qLVASIQ *(*LVASIQ*)*, 11
qNVASIM *(*NVASIM*)*, 14
qNVASIQ *(*NVASIQ*)*, 17
qOANVASIM *(*OANVASIM*)*, 20
qZANVASIM *(*ZANVASIM*)*, 32
qZOANVASIM *(*ZOANVASIM*)*, 37

r01NVASIM *(*ZOANVASIM*)*, 37
r0NVASIM *(*ZANVASIM*)*, 32
r1NVASIM *(*OANVASIM*)*, 20
rHVASIQ *(*HVASIQ*)*, 8
rLVASIQ *(*LVASIQ*)*, 11
rNVASIM *(*NVASIM*)*, 14
rNVASIQ *(*NVASIQ*)*, 17
rOANVASIM *(*OANVASIM*)*, 20
rZANVASIM *(*ZANVASIM*)*, 32
rZOANVASIM *(*ZOANVASIM*)*, 37

transport, *3*, 24
trees, *3*, 25

vasicek_envelope, 26
vasicekreg *(*vasicekreg-package*)*, 2
vasicekreg-package, 2

zabr, *3*, 29, *36*, *37*
ZANVASIM, *4*, *22*, 32, *40*

zavr, *3*, *31*, 34
ZOANVASIM, *4*, 37