Package 'good'

October 13, 2022

Type Package

Title Good Regression

Version 1.0.1

Description

Fit Good regression models to count data (Tur et al., 2021) <arXiv:2105.01557>. The package provides functions for model estimation and model prediction. Density, distribution function, quantile function and random generation for the Good distribution are also provided.

License GPL (>= 2)

Imports copula, maxLik, plyr

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

NeedsCompilation no

Author Jordi Tur [aut, cre],
David Moriña [ctb],
Pere Puig [ctb],
Argimiro Arratia [ctb],
Alejandra Cabaña [ctb],
Amanda Fernández-Fontelo [aut]

Maintainer Jordi Tur <jjtur@crm.cat>

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2021-05-10 13:50:02 UTC

R topics documented:

dgood *Probability mass function for the Good distribution*

Description

Probability mass function for the Good distribution with parameters z and s.

Usage

```
dgood ( x , z , s )
```

Arguments

x	vector of non-negative integer quantiles.
z	vector of first parameter for the Good distribution.
s	vector of second parameter for the Good distribution.

Details

The Good distribution has the probability mass function (pmf):

$$P(X = x) = (1/F(z, s)) \cdot (z^{(x+1)}/(x + 1)^s),$$

where $x = 0, 1, 2 \ldots$. Parameter z should be within the interval $(0, 1)$, and parameter s in the reals. $F(z, s)$ is the polylogarithm function:

$$F(z, s) = \sum_{i=1}^{\infty} z^n/n^s,$$

and acts in the pmf as the normalizing constant.

If $F(z, s)$ does not converge (e.g., for large negative values of the parameter s), the following approximation is used instead:

$$F(z, s) \approx \Gamma(1 - s) \cdot (-\log(z))^{(s-1)},$$

and dgood returns approximated probabilities:

$$P(X = x) \approx \exp((x + 1) \cdot \log(z) - s \cdot \log(x + 1) - \log(\Gamma(1 - s)) - (s - 1) \cdot \log(-\log(z))).$$

Value

dgood gives the probability mass function for the Good distribution with parameters z and s. x should be a vector of non-negative integer quantiles. If x is non-integer and/or negative, dgood returns 0 with a warning. z and s can be vectors with values within the interval $(0, 1)$ and the reals respectively. If vector z has negative values and/or outside the interval $(0, 1)$, dgood returns NaN with a warning.

If function `polylog` from package **copula** returns Inf (e.g., for large negative values of parameter s), dgood uses the approximation described above for probabilities, and additionally returns an informative warning.

Author(s)

Jordi Tur, David Moriña, Pere Puig, Alejandra Cabaña, Argimiro Arratia, Amanda Fernández-Fontelo

References

Good, J. (1953). The population frequencies of species and the estimation of population parameters. Biometrika, 40: 237–264.

Zörnig, P. and Altmann, G. (1995). Unified representation of zipf distributions. Computational Statistics & Data Analysis, 19: 461–473.

Kulasekera, K.B. and Tonkyn, D. (1992). A new distribution with applications to survival dispersal anddispersion. Communication in Statistics - Simulation and Computation, 21: 499–518.

Doray, L.G. and Luong, A. (1997). Efficient estimators for the good family. Communications in Statistics - Simulation and Computation, 26: 1075–1088.

Johnson, N.L., Kemp, A.W. and Kotz, S. Univariate Discrete Distributions. Wiley, Hoboken, 2005.

Kemp. A.W. (2010). Families of power series distributions, with particular reference to the lerch family. Journal of Statistical Planning and Inference, 140:2255–2259.

Wood, D.C. (1992). The Computation of Polylogarithms. Technical report. UKC, University of Kent, Canterbury, UK (KAR id:21052).

See Also

See also `polylog` from **copula**, `pgood`, and `qgood` and `rgood` from **good**.

Examples

```
# if x is not a non-negative integer, dgood returns 0 with a warning
dgood ( x = -3 , z = c ( 0.6 , 0.5 ) , s = -3 )
dgood ( x = 4.5 , z = c ( 0.6 , 0.5 ) , s = -3 )

# if z is not within 0 and 1, dgood returns NaN with a warning
dgood ( x = 4 , z = c ( 0.6 , 0.5 , -0.9 ) , s = -3 )

# if the approximation is used, dgood returns a warning
dgood ( x = 330 : 331 , z = c ( 0.6 , 0.5 ) , s = -170 )

dgood ( x = 4 , z = 0.6 , s = -3 )
dgood ( x = 4 , z = c ( 0.6 , 0.5 ) , s = -3 )
dgood ( x = 4 : 5 , z = c ( 0.6 , 0.5 ) , s = c ( -3 , -10 ) )
dgood ( x = 4 : 6 , z = c ( 0.6 , 0.5 ) , s = c ( -3 , -10 ) )
dgood ( x = 3 : 5 ,  z = c ( 0.6 , 0.5 , 0.9 , 0.4 ) , s = c ( -3 , -10 ) )
```

glm.good *Maximum Likelihood Estimation and Good Regression*

Description

glm.good is used to fit generalized linear models with a response variable following a Good distribution with parameters z and s. glm.good allows incorporating predictors in the model with a link function (log, logit and identity) that relates parameter z and predictors. A summary method over an object of class glm.good provides essential information regarding the fitted model such as parameters estimates, standard errors, and some goodness-of-fit measures. A prediction method over an object of class glm.good provides the fitted values with the estimated model and optionally standard errors and predictions for a new data set.

Usage

```
glm.good ( formula , data , link = "log" , start = NULL )
```

Arguments

formula	symbolic description of the model to be fitted. A typical predictor has the form response ~ terms where the response is the integer-valued response vector following a Good distribution with parameters s and z, and terms is a series of predictors.
data	an optional data frame with the variables in the model.
link	character specification of link function: "logit", "log" or "identity". By default link="log".
start	a vector with the starting values for the model parameters. Used for numerically maximize the likelihood function for parameters estimation. By default start = NULL.

Value

glm.good returns an object of class glm.good that is a list including:

coefs	The vector of coefficients.
loglik	Log-likelihood of the fitted model.
vcov	Variance-covariance matrix of all model parameters (derived from the Hessian matrix returned by nlm()).
hess	Hessian matrix, returned by nlm().
fitted.values	The fitted mean values. These are obtained by transforming the linear predictors by the link function inverse.

Author(s)

Jordi Tur, David Moriña, Pere Puig, Alejandra Cabaña, Argimiro Arratia, Amanda Fernández-Fontelo

References

Good, J. (1953). The population frequencies of species and the estimation of population parameters. Biometrika, 40: 237–264.

Zörnig, P. and Altmann, G. (1995). Unified representation of zipf distributions. Computational Statistics & Data Analysis, 19: 461–473.

Kulasekera, K.B. and Tonkyn, D. (1992). A new distribution with applications to survival dispersal anddispersion. Communication in Statistics - Simulation and Computation, 21: 499–518.

Doray, L.G. and Luong, A. (1997). Efficient estimators for the good family. Communications in Statistics - Simulation and Computation, 26: 1075–1088.

Johnson, N.L., Kemp, A.W. and Kotz, S. Univariate Discrete Distributions. Wiley, Hoboken, 2005.

Kemp. A.W. (2010). Families of power series distributions, with particular reference to the lerch family. Journal of Statistical Planning and Inference, 140:2255–2259.

Wood, D.C. (1992). The Computation of Polylogarithms. Technical report. UKC, University of Kent, Canterbury, UK (KAR id:21052).

See Also

See also `polylog` from **copula**, `dgood`, and `pgood`, `qgood` and `rgood` from **good**, and `maxLik` from **maxLik**.

Examples

```
strikes <- c ( rep ( 0, 46 ) , rep ( 1, 76 ) , rep ( 2, 24 ) , rep ( 3, 9 ) , rep ( 4, 1 ) )
mle <- glm.good ( strikes ~ 1 , link = "log" )
names ( mle )
mle$coefficients
mle$fitted.values
mean ( strikes )
summary ( mle )
predict ( mle , newdata = NULL , se.fit = TRUE )
```

---

pgood *Distribution function for the Good distribution*

Description

Distribution function for the Good distribution with parameters z and s.

Usage

```
pgood ( q , z , s , lower.tail = TRUE )
```

Arguments

q	vector of non-negative integer quantiles.
z	vector of first parameter for the Good distribution.
s	vector of second parameter for the Good distribution.
lower.tail	logical; if TRUE (default), probabilities are $P(X \leq x)$. Otherwise, $P(X > x)$.

Value

pgood returns the cumulative distribution function (cdf) for the Good distribution with parameters z and s. Parameter z should be within the interval $(0, 1)$, and parameter s in the reals. If q is non-integer, pgood returns the cdf of `floor(q)` with a warning. If q is negative, pgood returns 0 with a warning. pgood calls dgood from package **good**.

Author(s)

Jordi Tur, David Moriña, Pere Puig, Alejandra Cabaña, Argimiro Arratia, Amanda Fernández-Fontelo

References

Good, J. (1953). The population frequencies of species and the estimation of population parameters. Biometrika, 40: 237–264.

Zörnig, P. and Altmann, G. (1995). Unified representation of zipf distributions. Computational Statistics & Data Analysis, 19: 461–473.

Kulasekera, K.B. and Tonkyn, D. (1992). A new distribution with applications to survival dispersal anddispersion. Communication in Statistics - Simulation and Computation, 21: 499–518.

Doray, L.G. and Luong, A. (1997). Efficient estimators for the good family. Communications in Statistics - Simulation and Computation, 26: 1075–1088.

Johnson, N.L., Kemp, A.W. and Kotz, S. Univariate Discrete Distributions. Wiley, Hoboken, 2005.

Kemp. A.W. (2010). Families of power series distributions, with particular reference to the lerch family. Journal of Statistical Planning and Inference, 140:2255–2259.

Wood, D.C. (1992). The Computation of Polylogarithms. Technical report. UKC, University of Kent, Canterbury, UK (KAR id:21052).

See Also

See also `polylog` from **copula**, `dgood`, and `qgood` and `rgood` from **good**.

Examples

```
# if q < 0, pgood returns NaN with a warning
pgood ( q = -3 , z = 0.6 , s = -3  )

# if q is non-integer, pgood returns the cdf of floor(q) with a warning
pgood ( q = 3.4 , z = 0.6 , s = -3 )

# if z is not within 0 and 1, pgood returns returns NaN with a warning
```

```
pgood ( q = 3.4 , z = c( -0.6 , 0.6) , s = -3 )

pgood ( q = 0 : 2 , z = 0.6 , s = -3 )
pgood ( q = 0 : 1 , z = c ( 0.6 , 0.9 ) , s = -3 )
pgood ( q = 0 : 1 , z = c ( 0.6 , 0.9 ) , s = -3 , lower.tail = FALSE )
pgood ( q = 0 : 2 , z = c ( 0.6 , 0.9 ) , s = c ( -3 , -4 , -5 ) )
```

`polar` *Polar bear litter size data set*

Description

This data set corresponds to live-captured polar bears from late March 1992 to beginning of May 2017 at Svalbard, Norway.

Usage

```
data(polar)
```

Format

A data frame with 231 rows and 7 columns.

year Catch year

days Number of the day of the catch year

id Unique specimen id

age Age of the specimen, estimated using premolar tooth

agecat Categorized age of the specimen

length Body straight length (cm)

cubnumber Litter size

Source

Folio, Dorinda Marie et al. (2019), Data from: How many cubs can a mum nurse? Maternal age and size influence litter size in polar bears, Dryad, Dataset.

References

Folio D. M., Aars J., Gimenez O., Derocher A. E., Wiig O. and Cubaynes S. (2019) How many cubs can a mum nurse? Maternal age and size influence litter size in polar bears, Biology letters, 15.

Examples

```
data(polar)
head(polar)
```

qgood *Quantile function for the Good distribution*

Description

Quantile function for the Good distribution with parameters z and s.

Usage

```
qgood ( p , z , s , lower.tail = TRUE )
```

Arguments

p	vector of non-negative integer quantiles.
z	vector of first parameter for the Good distribution.
s	vector of second parameter for the Good distribution.
lower.tail	logical; if TRUE (default), probabilities are $P(X \leq x)$. Otherwise, $P(X > x)$.

Value

The smallest integer x such that $P(X \leq x) \geq p$ (or such that $P(X \leq x) \geq 1 - p$ if lower.tail is FALSE), where X is a random variable following a Good distribution with parameters z and s. Parameter z should be within the interval $(0, 1)$, and parameter s in the reals. Vector p should have values between 0 and 1. If vector p has negative values and/or outside the interval $(0, 1)$, qgood returns NaN with a warning. If vector p contains 1, qgood returns Inf. qgood calls dgood from package **good**.

Author(s)

Jordi Tur, David Moriña, Pere Puig, Alejandra Cabaña, Argimiro Arratia, Amanda Fernández-Fontelo

References

Good, J. (1953). The population frequencies of species and the estimation of population parameters. Biometrika, 40: 237–264.

Zörnig, P. and Altmann, G. (1995). Unified representation of zipf distributions. Computational Statistics & Data Analysis, 19: 461–473.

Kulasekera, K.B. and Tonkyn, D. (1992). A new distribution with applications to survival dispersal anddispersion. Communication in Statistics - Simulation and Computation, 21: 499–518.

Doray, L.G. and Luong, A. (1997). Efficient estimators for the good family. Communications in Statistics - Simulation and Computation, 26: 1075–1088.

Johnson, N.L., Kemp, A.W. and Kotz, S. Univariate Discrete Distributions. Wiley, Hoboken, 2005.

Kemp. A.W. (2010). Families of power series distributions, with particular reference to the lerch family. Journal of Statistical Planning and Inference, 140:2255–2259.

Wood, D.C. (1992). The Computation of Polylogarithms. Technical report. UKC, University of Kent, Canterbury, UK (KAR id:21052).

See Also

See also `polylog` from **copula**, `dgood`, and `pgood` and `rgood` from **good**.

Examples

```
# if p is not within [0, 1], NaN is returned with a warning
qgood ( p = c ( -0.6 , 1.3 ) , z = 0.5 , s = -3 )

# if z is not within 0 and 1, NaN is returned with a warning
qgood ( p = 0.5 , z = c(-0.6, -9, 0.5) , s = -3 )

qgood ( p = 0.5 , z = 0.6 , s = -3 )
qgood ( p = c ( 0.025 , 0.5 , 0.975 ) , z = 0.6 , s = -3 )
qgood ( p = c ( 0.025 , 0.5 , 0.975 ) , z = c ( 0.6 , 0.3 , 0.1 ) , s = -5 )
qgood ( p = c ( 0.025 , 0.5 , 0.975 ) , z = c ( 0.6 , 0.3 , 0.5 ) , s = -3 , lower.tail = FALSE )
qgood ( p = c ( 0.025 , 0.5 , 0.975 ) , z = c ( 0.6 , 0.3 ) , s = -3 )
```

`rgood` *Random generation for the Good distribution*

Description

Random generation for the Good distribution with parameters z and s.

Usage

```
rgood ( n , z , s , th = 10^-6 )
```

Arguments

`n`	vector of number of observations to be generated
`z`	vector of first parameter for the Good distribution
`s`	vector of second parameter for the Good distribution
`th`	defines the lower (q_1) and upper (q_2) quantiles such that $P(X \leq q_1) = th$ and $P(X \leq q_2) = 1 - th$ respectively.

Value

A vector containing n random deviates from a Good distribution with parameters z and s. Parameter z should be within the interval $(0, 1)$, and parameter s in the reals. `rgood` returns NaN if either arguments n or th are negative. `rgood` calls `qgood` and `pgood` from package **good**.

Author(s)

Jordi Tur, David Moriña, Pere Puig, Alejandra Cabaña, Argimiro Arratia, Amanda Fernández-Fontelo

References

Good, J. (1953). The population frequencies of species and the estimation of population parameters. Biometrika, 40: 237–264.

Zörnig, P. and Altmann, G. (1995). Unified representation of zipf distributions. Computational Statistics & Data Analysis, 19: 461–473.

Kulasekera, K.B. and Tonkyn, D. (1992). A new distribution with applications to survival dispersal anddispersion. Communication in Statistics - Simulation and Computation, 21: 499–518.

Doray, L.G. and Luong, A. (1997). Efficient estimators for the good family. Communications in Statistics - Simulation and Computation, 26: 1075–1088.

Johnson, N.L., Kemp, A.W. and Kotz, S. Univariate Discrete Distributions. Wiley, Hoboken, 2005.

Kemp. A.W. (2010). Families of power series distributions, with particular reference to the lerch family. Journal of Statistical Planning and Inference, 140:2255–2259.

Wood, D.C. (1992). The Computation of Polylogarithms. Technical report. UKC, University of Kent, Canterbury, UK (KAR id:21052).

See Also

See also `polylog` from **copula**, `dgood`, and `pgood` and `qgood` from **good**.

Examples

```
# if n is not a non-negative interger, function returns NaN with a warning
rgood ( n = -100 , z = 0.5 , s = -3 )

# if th is not positive, th is replaced by 1e-06 and a warning is provided
rgood ( n = 1 , z = 0.5 , s = -3 , th = -9 )

# if z is not within 0 and 1, NaN is returned with a warning
rgood ( n = 2 , z = c( -0.5, 0.5 ) , s = -3 )

rgood ( n = 10 , z = 0.6 , s = -3 )
rgood ( n = 1000 , z = 0.6 , s = -3 )
rgood ( n = c ( 3 , 10 ) , z = 0.6  , s = -3 )
rgood ( n = c ( 3 , 10 ) , z = c ( 0.2 , 0.8 ) , s = - 3 )
rgood ( n = c ( 3 , 10 , 6 ) , z = c ( 0.2 , 0.8 ) , s = c ( - 3 , -2 ) )
rgood ( n = 1000 , z = 0.3 , s = - 170 )
```

Index