Package ‘tukeyGH’

April 10, 2021

Type Package

Title Tukey's g-and-h Probability Distribution

Version 1.0.0

Author Flavio Santi [cre, aut] (<https://orcid.org/0000-0002-2014-1981>),
Marco Bee [aut] (<https://orcid.org/0000-0002-9579-3650>),
Julien Hambuckers [aut] (<https://orcid.org/0000-0003-4567-5477>),
Luca Trapin [aut] (<https://orcid.org/0000-0001-8613-0962>)

Maintainer Flavio Santi <flavio.santi@univr.it>

URL https://github.com/f-santi/tukeyGH

BugReports https://github.com/f-santi/tukeyGH/issues

Description It provides distribution, density and quantile functions of the Tukey's g-and-h probability distribution, as well as functions for random number generation, parameter estimation and testing.

Depends R (>= 4.0)

Imports magrittr (>= 2.0.0), minqa, Rdpack, rootSolve (>= 1.8.2.1)

Suggests testthat (>= 3.0.0)

License GPL (>= 3)

LazyData true

NeedsCompilation no

Encoding UTF-8

RoxygenNote 7.1.1

Repository CRAN

RdMacros Rdpack

Config/testthat/edition 3

Date/Publication 2021-04-10 08:10:02 UTC

R topics documented:

`tukeyGH-package` *Tukey's g-and-h probability distribution*

Description

It provides distribution, density and quantile functions of the Tukey's g-and-h probability distribution, as well as functions for random number generation, parameter estimation and testing.

Author(s)

Maintainer: Flavio Santi `<flavio.santi@univr.it>` (ORCID)

Authors:

- Marco Bee `<marco.bee@unitn.it>` (ORCID)
- Julien Hambuckers `<jhambuckers@uliege.be>` (ORCID)
- Luca Trapin `<luca.trapin@unibo.it>` (ORCID)

See Also

Useful links:

- `https://github.com/f-santi/tukeyGH`
- Report bugs at `https://github.com/f-santi/tukeyGH/issues`

BDSF — *Losses on Business Disruption and System Failures*

Description

Operational losses from the *business disruption and system failures* of the Italian bank UniCredit between 2005 and 2014. The data, which are scaled by an unknown factor for anonymity reasons, have been provided by the UniCredit's Operational Risk Department. More information related to these data can be found in Hambuckers et al. (2018) and Bee et al. (2021).

Format

`numeric` with 152 operational losses.

References

Bee M, Hambuckers J, Trapin L (2021). "Estimating large losses in insurance analytics and operational risk using the g-and-h distribution." *Quantitative Finance*. doi: 10.1080/14697688.2020.1849778, `https://doi.org/10.1080/14697688.2020.1849778`.

Hambuckers J, Groll A, Kneib T (2018). "Understanding the economic determinants of the severity of operational losses: a regularized Generalized Pareto regression approach." *Journal of Applied Econometrics*, **33**(6), 898–935. doi: 10.1002/jae.2638, `https://doi.org/10.1002/jae.2638`.

See Also

Other data: `CPBP2014`, `EDPM2014`, `EPWS2014`

CPBP2014 — *Losses on Clients, Products, and Business Practices*

Description

Operational losses from the *clients, products, and business practices* of the Italian bank UniCredit in 2014. The data, which are scaled by an unknown factor for anonymity reasons, have been provided by the UniCredit's Operational Risk Department. Losses are adjusted for inflation by means of the monthly Italian consumer price index (base 100: December 2013). Around one-fourth of the losses are related to financial instruments and derivative products. More information related to these data can be found in Hambuckers et al. (2018) and Bee et al. (2019).

Format

`numeric` with 583 operational losses.

References

Bee M, Hambuckers J, Trapin L (2019). "Estimating value-at-risk for the g-and-h distribution: an indirect inference approach." *Quantitative Finance*, **19**(8), 1255–1266. doi: 10.1080/14697688.2019.1580762, https://doi.org/10.1080/14697688.2019.1580762.

Hambuckers J, Groll A, Kneib T (2018). "Understanding the economic determinants of the severity of operational losses: a regularized Generalized Pareto regression approach." *Journal of Applied Econometrics*, **33**(6), 898–935. doi: 10.1002/jae.2638, https://doi.org/10.1002/jae.2638.

See Also

Other data: BDSF, EDPM2014, EPWS2014

---

distr-g — *The g distribution*

Description

Density (dg), distribution function (pg), quantile function (qg), random generation (rg), and bounds of the support (infg and supg) of the g distribution (Tukey 1977). All functions with the exception of rg are vectorized with respect to all arguments on the g distribution (x, q, p, a, b, g). The functions are wrappers of the g-and-h family with h = 0.

Usage

```
dg(x, a = 0, b = 1, g = 0, log = FALSE, ...)

pg(q, a = 0, b = 1, g = 0, lower.tail = TRUE, log.p = FALSE, ...)

qg(p, a = 0, b = 1, g = 0, lower.tail = TRUE, log.p = FALSE)

rg(n, a = 0, b = 1, g = 0)

infg(a = 0, b = 1, g = 0)

supg(a = 0, b = 1, g = 0)
```

Arguments

x, q	vector of quantiles.
a	location parameter(s).
b	scale parameter(s).
g	skewness parameter(s).
log, log.p	logical; if TRUE, probabilities p are given as log(p).
...	arguments passed to rootSolve::uniroot.all().

`lower.tail`	logical; if TRUE (default), probabilities are $P[X \leq x]$ otherwise, $P[X > x]$.
`p`	vector of probabilities.
`n`	number of observations. If `length(n) > 1`, the length is taken to be the number required.

Value

`dg` gives the density, `pg` gives the distribution function, `qg` gives the quantile function, and `rg` generates random numbers.

The length of the result is determined by `n` for `rg`, and is the maximum of the lengths of the numerical arguments for the other functions.

The numerical arguments other than `n` are recycled to the length of the result. Only the first elements of the logical arguments are used.

References

Tukey JW (1977). “Modern techniques in data analysis.” NSF-sponsored regional research conference at Southern Massachusetts University.

`distr-gh` *The Tukey's g-and-h distribution*

Description

Density (`dgh`), distribution function (`pgh`), quantile function (`qgh`), random generation (`rgh`), and bounds of the support (`infgh` and `supgh`) of the Tukey's g-and-h distribution (Tukey 1977). All functions with the exception of `rgh` are vectorized with respect to all arguments on the Tukey's distribution (`x`, `q`, `p`, `a`, `b`, `g`, `h`).

Usage

```
dgh(x, a = 0, b = 1, g = 0, h = 0.2, log = FALSE, ...)

pgh(q, a = 0, b = 1, g = 0, h = 0.2, lower.tail = TRUE, log.p = FALSE, ...)

qgh(p, a = 0, b = 1, g = 0, h = 0.2, lower.tail = TRUE, log.p = FALSE)

rgh(n, a = 0, b = 1, g = 0, h = 0.2)

infgh(a = 0, b = 1, g = 0, h = 0.2)

supgh(a = 0, b = 1, g = 0, h = 0.2)
```

Arguments

`x, q`	vector of quantiles.
`a`	location parameter(s).
`b`	scale parameter(s).
`g`	skewness parameter(s).
`h`	heavy-taildness parameter(s). Only non-negative values will be accepted (see *Details*).
`log, log.p`	logical; if TRUE, probabilities p are given as log(p).
`...`	arguments passed to `rootSolve::uniroot.all()`.
`lower.tail`	logical; if TRUE (default), probabilities are $P[X \le x]$ otherwise, $P[X > x]$.
`p`	vector of probabilities.
`n`	number of observations. If `length(n) > 1`, the length is taken to be the number required.

Details

Given a Gaussian random variable $Z \sim \mathcal{N}(0,1)$, the following transformation:

$$X = a + b\,\frac{e^{gZ} - 1}{g}\,e^{\frac{hZ^2}{2}}$$

defines the Tukey's g-and-h distribution. Hence $X \sim gh(a,b,g,h)$ denotes a random variable distributed according to the Tukey's g-and-h distribution function, where $a \in \mathbf{R}$ is the location parameter, $b \in \mathbf{R}^+$ is the scale parameter, $g \in \mathbf{R}$ is the skewness parameter, and $h \in \mathbf{R}^+$ is the shape parameter.

In principle, the shape parameter h may also take negative values, however, in such a case, the above transformation is not monotone. All functions on this page require that $h \ge 0$.

Note that, when $g = 0$, the limit for $g \to 0$ of the previous transformation is considered:

$$X = \lim_{g \to 0}\left(a + b\,\frac{e^{gZ} - 1}{g}\,e^{\frac{hZ^2}{2}}\right) = a + b\,Z\,e^{\frac{hZ^2}{2}}$$

so that $X \sim gh(a,b,0,h)$.

Value

`dgh` gives the density, `pgh` gives the distribution function, `qgh` gives the quantile function, and `rgh` generates random numbers.

The length of the result is determined by `n` for `rgh`, and is the maximum of the lengths of the numerical arguments for the other functions.

The numerical arguments other than `n` are recycled to the length of the result. Only the first elements of the logical arguments are used.

References

Tukey JW (1977). "Modern techniques in data analysis." NSF-sponsored regional research conference at Southern Massachusetts University.

EDPM2014 *Losses on Execution, Delivery, and Process Management*

Description

Operational losses from the *execution, delivery, and process management* of the Italian bank UniCredit in 2014. The data, which are scaled by an unknown factor for anonymity reasons, have been provided by the UniCredit's Operational Risk Department. Losses are adjusted for inflation by means of the monthly Italian consumer price index (base 100: December 2013). Around one-fourth of the losses are related to financial instruments and derivative products. More information related to these data can be found in Hambuckers et al. (2018) and Bee et al. (2019).

Format

`numeric` with 417 operational losses.

References

Bee M, Hambuckers J, Trapin L (2019). "Estimating value-at-risk for the g-and-h distribution: an indirect inference approach." *Quantitative Finance*, **19**(8), 1255–1266. doi: 10.1080/14697688.2019.1580762, `https://doi.org/10.1080/14697688.2019.1580762`.

Hambuckers J, Groll A, Kneib T (2018). "Understanding the economic determinants of the severity of operational losses: a regularized Generalized Pareto regression approach." *Journal of Applied Econometrics*, **33**(6), 898–935. doi: 10.1002/jae.2638, `https://doi.org/10.1002/jae.2638`.

See Also

Other data: `BDSF`, `CPBP2014`, `EPWS2014`

EPWS2014 *Losses on Employment Practices and Workplace Safety*

Description

Operational losses from the *employment practices and workplace safety* of the Italian bank UniCredit in 2014. The data, which are scaled by an unknown factor for anonymity reasons, have been provided by the UniCredit's Operational Risk Department. More information related to these data can be in Hambuckers et al. (2018) and Bee et al. (2019).

Format

`numeric` with 97 operational losses.

References

Bee M, Hambuckers J, Trapin L (2019). "Estimating value-at-risk for the g-and-h distribution: an indirect inference approach." *Quantitative Finance*, **19**(8), 1255–1266. doi: 10.1080/14697688.2019.1580762, https://doi.org/10.1080/14697688.2019.1580762.

Hambuckers J, Groll A, Kneib T (2018). "Understanding the economic determinants of the severity of operational losses: a regularized Generalized Pareto regression approach." *Journal of Applied Econometrics*, **33**(6), 898–935. doi: 10.1002/jae.2638, https://doi.org/10.1002/jae.2638.

See Also

Other data: `BDSF`, `CPBP2014`, `EDPM2014`

`fitG` *Fit g distribution*

Description

Fit the g distribution on a dataset through maximum likelihood (Bee et al. 2019).

Usage

```
fitG(x, verbose = "v")
```

Arguments

`x`	data as a `numeric`.
`verbose`	function verbosity. Values `v`, `vv` and `vvv` are admitted, whereas other values (such as `""` or `FALSE`) will make the function silent.

Value

Object of class `fitGH`. Useful methods include:

- `coef()` point estimates of parameters
- `print()` short information about the object
- `summary()` summary information about the estimation process

References

Bee M, Hambuckers J, Trapin L (2019). "An improved approach for estimating large losses in insurance analytics and operational risk using the g-and-h distribution." Technical Report 2019/11, Department of Economics and Management, University of Trento.

Examples

```
data("EPWS2014")

# Fit to EPWS2014 data
modG <- fitG(EPWS2014)
summary(modG)
```

`fitGH` *Fit the Tukey's g-and-h distribution*

Description

Fit the Tukey's g-and-h distribution on a dataset through various methods: quantile estimator by Hoaglin (1985), indirect inference (Bee et al. 2019), and maximum likelihood (Bee et al. 2019).

Usage

```
fitGH(x, method = c("iinference", "quantile", "mle"), verbose = "vv")
```

Arguments

`x`	data as a `numeric`.
`method`	estimation method (partial string matching is allowed). Indirect inference is adopted as default.
`verbose`	function verbosity. Values `v`, `vv` and `vvv` are admitted, whereas other values (such as `""` or `FALSE`) will make the function silent.

Value

Object of class `fitGH`. Useful methods include:

- `coef()` point estimates of parameters
- `print()` short information about the object
- `summary()` summary information about the estimation process

References

Bee M, Hambuckers J, Trapin L (2019). "Estimating value-at-risk for the g-and-h distribution: an indirect inference approach." *Quantitative Finance*, **19**(8), 1255–1266. doi: 10.1080/14697688.2019.1580762, `https://doi.org/10.1080/14697688.2019.1580762`.

Bee M, Hambuckers J, Trapin L (2019). "An improved approach for estimating large losses in insurance analytics and operational risk using the g-and-h distribution." Technical Report 2019/11, Department of Economics and Management, University of Trento.

Hoaglin DC (1985). "Exploring Data Tables, Trends, and Shapes." In Hoaglin DC, Mosteller F, Tukey JW (eds.), chapter Summarizing Shape Numerically: The g-and-h Distributions, 461–513. Wiley. ISBN 978-0-470-04005-8.

Examples

```
data("EPWS2014")

# Fit to EPWS2014 data through indirect inference
modII <- fitGH(EPWS2014)
summary(modII)

# Fit to EPWS2014 data through the quantile estimator
modQ <- fitGH(EPWS2014, method = "quantile")
summary(modQ)

## Not run:

# Fit to EPWS2014 data through MLE (the computation time is much longer)
modMLE <- fitGH(EPWS2014, method = "mle")
summary(modMLE)

## End(Not run)
```

---

qqgh — *Draw a Q-Q plot based on g-and-h distribution*

Description

Draw a quantile-quantile plot based on the Tukey's g-and-h distribution.

Usage

```
qqgh(x, theta = NULL, qqline = TRUE, grid = TRUE, ...)
```

Arguments

`x`	either data as a `numeric` vector, or an object of class `fitGH`, as returned by `fitGH()`.
`theta`	parameters of the g-and-h distribution as a `numeric` vector of length four: (a, b, g, h). If argument `x` is a `fitGH` object, and argument `theta` is `NULL`, `theta` will be initialised with `coef(x)`.
`qqline`	if `TRUE` (default) a Q-Q line will be added to the graph by means of `qqline()`
`grid`	if `TRUE` (default) a Q-Q line will be added to the graph by means of `grid()`.
`...`	other arguments passed to `qqplot()`.

Value

A named list with the following components:

- `teo_quantile`: theoretical quantile function (with argument `p`)
- `qqplot`: output of function `qqplot()`

Examples

```
data("EPWS2014")
modII <- fitGH(EPWS2014)
qqgh(modII)
```

`testGvsGH` *Compute simulation-based* p*-value of log-likelihood ratio test*

Description

Compute simulation-based *p*-value of log-likelihood ratio test (Bee et al. 2021).

Usage

```
testGvsGH(x, nsim, verbose = "vv")
```

Arguments

`x`	data.
`nsim`	number of Monte Carlo simulations
`verbose`	function verbosity. Values `v`, `vv` and `vvv` are admitted, whereas other values (such as `""` or `FALSE`) will make the function silent.

Value

Object of class `testGvsGH`.

References

Bee M, Hambuckers J, Santi F, Trapin L (2021). "Testing a parameter restriction on the boundary for the g-and-h distribution: a simulated approach." *Computational Statistics*. doi: 10.1007/s00180-021010783, `https://doi.org/10.1007/s00180-021-01078-3`.

Examples

```
## Not run:

data(EPWS2014)
# Warning: the following code may take up to 30 mins to be run
testGvsGH(EPWS2014, 30)

## End(Not run)
```

Index