

giniVarCI: Gini Indices, Variances and Confidence Intervals for Finite and Infinite Populations

R Package version 0.0.1-3

<https://CRAN.R-project.org/package=giniVarCI>

Full citation to this publication:

Muñoz, J.F., Pavía, J.M. & Álvarez-Verdejo, E. (2023). giniVarCI: Gini Indices, Variances and Confidence Intervals for Finite and Infinite Populations. R packages version 0.0.1-3. <https://CRAN.R-project.org/package=giniVarCI>

Thank for your interest in this publication.

Package ‘giniVarCI’

January 8, 2024

Type Package

Title Gini Indices, Variances and Confidence Intervals for Finite and Infinite Populations

Version 0.0.1-3

Description Estimates the Gini index and computes variances and confidence intervals for finite and infinite populations, using different methods; also computes Gini index for continuous probability distributions, draws samples from continuous probability distributions with Gini indices set by the user; uses 'Rcpp'.

References:

Muñoz et al. (2023) <[doi:10.1177/00491241231176847](https://doi.org/10.1177/00491241231176847)>.

Álvarez et al. (2021) <[doi:10.3390/math9243252](https://doi.org/10.3390/math9243252)>.

Giorgi and Gigliarano (2017) <[doi:10.1111/joes.12185](https://doi.org/10.1111/joes.12185)>.

Langel and Tillé (2013) <[doi:10.1111/j.1467-985X.2012.01048.x](https://doi.org/10.1111/j.1467-985X.2012.01048.x)>.

License GPL

Imports Rcpp (>= 1.0.10), stats

Depends R (>= 3.5.0)

Suggests knitr, rmarkdown, VGAM, utils, microbenchmark, laeken, REAT, DescTools, ineq, ggplot2

Encoding UTF-8

RoxygenNote 7.1.1

LinkingTo Rcpp

VignetteBuilder knitr

NeedsCompilation yes

Author Juan Francisco Muñoz [aut, cre]

(<<https://orcid.org/0000-0001-7427-6630>>),

Jose M. Pavía [aut] (<<https://orcid.org/0000-0002-0129-726X>>),

Encarnación Álvarez Verdejo [aut]

(<<https://orcid.org/0000-0002-0473-6037>>),

MCIN-AEI and ERDF. Reference PID2022-136235NB-I00 [fnd]

Maintainer Juan Francisco Muñoz <jfmunoz@ugr.es>

Repository CRAN

Date/Publication 2024-01-08 10:30:02 UTC

R topics documented:

fcompareCI	2
fgini	8
fginindex	14
gbeta	17
gburr	19
gchisq	21
gdagum	22
gf	24
gfisk	26
gfrechet	27
ggamma	29
ggompertz	30
glnorm	32
gpareto	33
gparetoI	35
gparetoII	36
gparetoIII	38
gparetoIV	40
gsample	42
gunif	44
gweibull	46
icompareCI	47
igini	51
iginindex	54

Index **59**

fcompareCI	<i>Comparisons of variance estimates and confidence intervals for the Gini index in finite populations</i>
------------	--

Description

Compares variance estimates and confidence intervals for the Gini index in finite populations.

Usage

```
fcompareCI(
  y,
  w,
  Pi = NULL,
  Pij = NULL,
  PiU,
  alpha = 0.05,
  B = 1000L,
  digitsgini = 2L,
```

```

digitsvar = 4L,
na.rm = TRUE,
plotCI = TRUE,
line.types = c(1L, 2L, 4L),
colors = c("red", "green", "blue"),
shapes = c(8L, 4L, 3L),
save.plot = FALSE,
large.sample = FALSE)

```

Arguments

y	A vector with the non-negative real numbers to be used for estimating the Gini index.
w	A numeric vector with the survey weights to be used for estimating the Gini index, the variance estimation and the confidence interval. This argument can be missing if argument Pi is provided.
Pi	A numeric vector with the (sample) first inclusion probabilities to be used for estimating the Gini index, the variance estimation and the confidence interval. This argument can be NULL if argument w is provided. The default value is Pi = NULL.
Pij	A numeric square matrix with the (sample) second (joint) inclusion probabilities to be used for the variance estimation and the confidence interval. The Hajek approximation is used when Pij = NULL. This argument is used by the intervals "zjackknife", "zlinearization" and "zblinearization". The default value is Pij = NULL.
PiU	A numeric vector with the (population) first inclusion probabilities. The Hartley-Rao (HR) expression for the variance estimation is also computed if this argument is provided.
alpha	A single numeric value between 0 and 1 specifying the confidence level 1-alpha to be used for computing the confidence interval for the Gini index. Some authors call alpha the significance level. The default value is alpha = 0.05.
B	A single integer specifying the number of bootstrap replicates. The default value is B = 1000L.
digitsgini	A single integer specifying the number of decimals used in the estimation of the Gini index and confidence intervals. The default value is digitsgini = 2L.
digitsvar	A single integer specifying the number of decimals used in the variance estimation of the Gini index. The default value is digitsvar = 4L.
na.rm	A 'TRUE/FALSE' logical value indicating whether NA values should be removed before the computation proceeds. The default value is na.rm = TRUE.
plotCI	A 'TRUE/FALSE' logical value indicating whether confidence intervals are compared using a plot. The default value is plotCI = TRUE.
line.types	A numeric vector of length 3 specifying the line types. See the function plot for the different line types. The default value is line.types = c(1L, 2L, 4L).
colors	A vector of length 3 specifying the colors for lines of the plot. The default value is colors = c("red", "green", "blue").

shapes	A numeric vector specifying the point shapes for the limits of intervals. If PiU is missing, the function uses the two first components of shapes, i.e., it must have at least length 2. If PiU is provided, shapes must have at least length 3. See the function plot for the different point shapes. The default value is shapes = c(8L, 4L, 3L).
save.plot	A 'TRUE/FALSE' logical value indicating whether the ggplot object of the plot comparing the confidence intervals should be saved in the output. The default value is save.plot = FALSE.
large.sample	A 'TRUE/FALSE' logical value indicating whether the sample is large to apply a faster algorithm to sort the sample values in the computation of the Gini index. The default value is large.sample = FALSE.

Details

For a sample S , with size n and inclusion probabilities $\pi_i = P(i \in S)$ (argument Pi), derived from a finite population U , with size N , different formulations of the Gini index have been proposed in the literature. This function estimates the Gini index, variances and confidence intervals using various formulations. The different methods for estimating the Gini index are (see also Muñoz et al., 2023):

\ Gini Index formulae.

Method 1 (Langel and Tillé, 2013)

$$\hat{G}_{w1} = \frac{1}{2\hat{N}^2\bar{y}_w} \sum_{i \in S} \sum_{j \in S} w_i w_j |y_i - y_j|,$$

where $\hat{N} = \sum_{i \in S} w_i$, $\bar{y}_w = \hat{N}^{-1} \sum_{i \in S} w_i y_i$, and w_i are the survey weights. For example, the survey weights can be $w_i = \pi_i^{-1}$. w or Pi must be provided, but not both. It is required that $w_i = \pi_i^{-1}$, for $i \in S$, when both w and Pi are provided.

Method 2 (Alfons and Templ, 2012; Langel and Tillé, 2013)

$$\hat{G}_{w2} = \frac{2 \sum_{i \in S} w_{(i)}^+ \hat{N}_{(i)} y_{(i)} - \sum_{i \in S} w_i^2 y_i}{\hat{N}^2 \bar{y}_w} - 1,$$

where $y_{(i)}$ are the values y_i sorted in increasing order, $w_{(i)}^+$ are the values w_i sorted according to the increasing order of the values y_i , and $\hat{N}_{(i)} = \sum_{j=1}^i w_{(j)}^+$. Langel and Tillé (2013) show that $\hat{G}_{w1} = \hat{G}_{w2}$, so the computation of \hat{G}_{w1} is omitted in results.

Method 3 (Berger, 2008)

$$\hat{G}_{w3} = \frac{2}{\hat{N} \bar{y}_w} \sum_{i \in S} w_i y_i \hat{F}_w^*(y_i) - 1,$$

where

$$\hat{F}_w^*(t) = \frac{1}{\hat{N}} \sum_{i \in S} w_i [\delta(y_i < t) + 0.5\delta(y_i = t)]$$

is the smooth (mid-point) distribution function, and $\delta(\cdot)$ is the indicator variable that takes the value 1 when its argument is true, and 0 otherwise. It can be seen that $\widehat{G}_{w2} = \widehat{G}_{w3}$, so the computation of \widehat{G}_{w3} is omitted in results.

Method 4 (Berger and Gedik-Balay, 2020)

$$\widehat{G}_{w4} = 1 - \frac{\bar{v}_w}{\bar{y}_w},$$

where $\bar{v}_w = \widehat{N}^{-1} \sum_{i \in S} w_i v_i$ and

$$v_i = \frac{1}{\widehat{N} - w_i} \sum_{\substack{j \in S \\ j \neq i}} \min(y_i, y_j).$$

Method 5 (Lerman and Yitzhaki, 1989)

$$\widehat{G}_{w5} = \frac{2}{\widehat{N} \bar{y}_w} \sum_{i \in S} w_{(i)}^+ [y_{(i)} - \bar{y}_w] \left[\widehat{F}_w^{LY}(y_{(i)}) - \bar{F}_w^{LY} \right],$$

where

$$\widehat{F}_w^{LY}(y_{(i)}) = \frac{1}{\widehat{N}} \left(\widehat{N}_{(i-1)} + \frac{w_{(i)}^+}{2} \right)$$

and $\bar{F}_w^{LY} = \widehat{N}^{-1} \sum_{i \in S} w_{(i)}^+ \widehat{F}_w^{LY}(y_{(i)})$.

\ Variances and confidence intervals.

For a given estimator \widehat{G}_w and variable z , the Horvitz-Thompson type variance estimator (Hortvitz and Thompson, 1952) is given by

$$\widehat{V}_{HT}(\widehat{G}_w) = \sum_{i \in S} \sum_{j \in S} \check{\Delta}_{ij} w_i w_j z_i z_j,$$

where

$$\check{\Delta}_{ij} = \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}}$$

and π_{ij} is the second (joint) inclusion probability of the individuals i and j , i.e., $\pi_{ij} = P\{(i, j) \in S\}$ (argument $\text{Pi } j$).

The Sen-Yates-Grundy type variance estimator (Sen, 1953; Yates and Grundy, 1953) is defined as

$$\widehat{V}_{SYG}(\widehat{G}_w) = -\frac{1}{2} \sum_{i \in S} \sum_{j \in S} \check{\Delta}_{ij} (w_i z_i - w_j z_j)^2$$

.

The Hartley-Rao type variance estimator (Hartley and Rao, 1962) is given by

$$\widehat{V}_{HR}(\widehat{G}_w) = \frac{1}{n-1} \sum_{i \in S} \sum_{\substack{j \in S \\ j < i}} \left(1 - \pi_i - \pi_j + \frac{1}{n} \sum_{k \in U} \pi_k^2 \right) (w_i z_i - w_j z_j)^2.$$

Note that the The Horvitz-Thompson variance estimator can give negative values. We observe that both Horvitz-Thompson and Sen-Yates-Grundy variance estimators depend on second (joint) inclusion probabilities (argument π_{ij}). The Hajek (1964) approximation

$$\pi_{ij} \cong \pi_i \pi_j \left[1 - \frac{(1 - \pi_i)(1 - \pi_j)}{\sum_{i \in S} (1 - \pi_i)} \right]$$

is used when the second (joint) inclusion probabilities are not available ($\pi_{ij} = \text{NULL}$). Note that the Hajek approximation is suggested for large-entropy sampling designs, large samples, and large populations (see Tille 2006; Berger and Tillé, 2009; Haziza et al., 2008; Berger, 2011). For instance, this approximation is not recommended for highly-stratified samples (Berger, 2005). The Hartley-Rao variance estimator requires the first inclusion probabilities at the population level (argument π_{iU}). `zjackknife` computes the confidence interval based on the jackknife technique with critical values based on the Normal approximation. `zlinearization` and `zblinearization` compute the confidence intervals based on the linearization technique applied to the estimators

$$\widehat{G}_w^a = \widehat{G}_{w1}$$

and

$$\widehat{G}_w^b = \frac{2}{\widehat{N}\bar{y}_w} \sum_{i \in S} w_i y_i \widehat{F}_w(y_i) - 1,$$

respectively, where

$$\widehat{F}_w(t) = \frac{1}{\widehat{N}} \sum_{i \in S} w_i \delta(y_i \leq t).$$

Critical values are also based on the Normal approximation. `pbootstrap` computes the variance using the rescaled bootstrap, and the confidence interval is constructed using the percentile method. The **vignette** `vignette("GiniVarInterval")` contains a detailed description of the various methods for variance estimation and confidence intervals for the Gini index.

The following table summarises the various types of variances and confidence intervals that the function `fcompareCI` computes.

Interval	Variance	Critical values	References
<code>zjackknife</code>	Jackknife	Normal	Berger (2008)
<code>zlinearization</code>	Linearization	Normal	Langel and Tille (2013)
<code>zblinearization</code>	Linearization	Normal	Berger (2008)
<code>pBootstrap</code>	Rescaled bootstrap	Percentile bootstrap	Berger and Gedik-Balay (2020)

Value

If `save.plot = FALSE`, a data frame with columns:

1. `interval`. The method used to construct the confidence interval.
2. `method`. The method used to estimate the Gini index.
3. `varformula`. The type of formula for the variance estimator. Possible values are HT and SYG if argument `PiU` is missing, and HT, SYG and HR if argument `PiU` is provided.
4. `gini`. The estimation of the Gini index.

5. `lowerlimit`. The lower limit of the confidence interval.
6. `upperlimit`. The upper limit of the confidence interval.
7. `var.gini`. The variance estimation for the estimator of the Gini index.

If `save.plot = TRUE`, a list with two components: (i) ‘base.CI’ a data frame of seven columns as just described and (ii) ‘plot’ a (ggplot) description of the plot, which is a list with components that contain the plot itself, the data, information about the scales, panels, etc. As a side-effect, a plot that compares the various methods for constructing confidence intervals for the Gini index is displayed. `**ggplot2**` is needed to be installed for this option to work.

If `plotCI = TRUE`, as a side-effect, a plot that compares the various methods for constructing confidence intervals for the Gini index is displayed. `**ggplot2**` is needed to be installed for this option to work.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

- Alfons, A., and Templ, M. (2012). Estimation of social exclusion indicators from complex surveys: The R package `laeken`. KU Leuven, Faculty of Business and Economics Working Paper.
- Berger, Y. G. (2005). Variance estimation with highly stratified sampling designs with unequal probabilities. *Australian & New Zealand Journal of Statistics*, 47, 365–373.
- Berger, Y. G. (2008). A note on the asymptotic equivalence of jackknife and linearization variance estimation for the Gini Coefficient. *Journal of Official Statistics*, 24(4), 541-555.
- Berger, Y. G. (2011). Asymptotic consistency under large entropy sampling designs with unequal probabilities. *Pakistan Journal of Statistics*, 27, 407–426.
- Berger, Y., and Gedik-Balay, İ. (2020). Confidence intervals of Gini coefficient under unequal probability sampling. *Journal of Official Statistics*, 36(2), 237-249.
- Berger, Y. G. and Tillé, Y. (2009). Sampling with unequal probabilities. In *Sample Surveys: Design, Methods and Applications* (eds. D. Pfeiffermann and C. R. Rao), 39–54. Elsevier, Amsterdam.
- Hajek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *The Annals of Mathematical Statistics*, 35, 4, 1491–1523.
- Hartley, H. O., and Rao, J. N. K. (1962). Sampling with unequal probabilities and without replacement. *The Annals of Mathematical Statistics*, 350-374.
- Haziza, D., Mecatti, F. and Rao, J. N. K. (2008). Evaluation of some approximate variance estimators under the Rao-Sampford unequal probability sampling design. *Metron*, LXVI, 91–108.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663–685.
- Langel, M., and Tillé, Y. (2013). Variance estimation of the Gini index: revisiting a result several times published. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2), 521-540.

Lerman, R. I., and Yitzhaki, S. (1989). Improving the accuracy of estimates of Gini coefficients. *Journal of econometrics*, 42(1), 43-47.

Muñoz, J. F., Moya-Fernández, P. J., and Álvarez-Verdejo, E. (2023). Exploring and Correcting the Bias in the Estimation of the Gini Measure of Inequality. *Sociological Methods & Research*. <https://doi.org/10.1177/00491241231176847>

Sen, A. R. (1953). On the estimate of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, 5, 119–127.

Tillé, Y. (2006). *Sampling Algorithms*. Springer, New York.

Yates, F., and Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society B*, 15, 253–261.

See Also

[fgini](#), [fginindex](#)

Examples

```
# Income and weights (region 'Burgenland') from the 2006 Austrian EU-SILC (Package 'laeken').
data(eusilc, package="laeken")
y <- eusilc$eqIncome[eusilc$db040 == "Burgenland"]
w <- eusilc$rb050[eusilc$db040 == "Burgenland"]

# Estimation of the Gini index and confidence intervals using different methods.
fcompareCI(y, w)

y <- c(30428.83, 14976.54, 18094.09, 29476.79, 20381.93, 6876.17,
      10360.96, 8239.82, 29476.79, 32230.71)
w <- c(357.86, 480.99, 480.99, 476.01, 498.58, 498.58, 476, 498.58, 476.01, 476.01)
fcompareCI(y, w, plotCI = FALSE)
```

fgini

Gini index, variances and confidence intervals in finite populations

Description

Estimates the Gini index and computes variances and confidence intervals in finite populations.

Usage

```
fgini(
  y,
  w,
  method = 2L,
  interval = NULL,
  Pi = NULL,
  Pij = NULL,
  PiU,
```

```

alpha = 0.05,
B = 1000L,
na.rm = TRUE,
varformula = "SYG",
large.sample = FALSE
)

```

Arguments

y	A vector with the non-negative real numbers to be used for estimating the Gini index.
w	A numeric vector with the survey weights to be used for estimating the Gini index, the variance and the confidence interval. This argument can be missing if argument Pi is provided.
method	An integer between 1 and 5 selecting one of the 5 methods detailed below for estimating the Gini index in finite populations. The default method is method = 2L.
interval	A character string specifying the type of variance estimation and confidence interval to be used. Possible values are "zjackknife", "zlinearization", "zblinearization" and "pbootstrap". interval = NULL omits the computation of both variance and confidence interval. The default value is interval = NULL.
Pi	A numeric vector with the (sample) first inclusion probabilities to be used for estimating the Gini index, the variance and the confidence interval. This argument can be NULL if argument w is provided. The default value is Pi = NULL.
Pij	A numeric square matrix with the (sample) second (joint) inclusion probabilities to be used for the variance estimation and the confidence interval. The Hajek approximation is used when Pij = NULL. This argument is used when interval = {"zjackknife", "zlinearization", "zblinearization"}. The default value is Pij = NULL.
PiU	A numeric vector with the (population) first inclusion probabilities. This argument is only required when the Hartley-Rao expression for the variance estimation is selected (varformula = "HR").
alpha	A single numeric value between 0 and 1. If interval is not NULL, the confidence level to be used for computing the confidence interval for the Gini is 1-alpha. Some authors call alpha the significance level. The default value is alpha = 0.05.
B	A single integer specifying the number of bootstrap replicates. This argument is required when interval = "pbootstrap". The default value is B = 1000L.
na.rm	A 'TRUE/FALSE' logical value indicating whether NA's should be removed before the computation proceeds. The default value is na.rm = TRUE.
varformula	A character string specifying the type of formula to be used for the variance estimator when interval = {"zjackknife", "zlinearization", "zblinearization"}. Possible values are "HT" (Hortvitz-Thompson), "SYG" (Sen-Yates-Grundy) and "HR" (Hartley-Rao). The default value is varformula = "SYG".

large.sample A ‘TRUE/FALSE’ logical value indicating whether the sample is large to apply a faster algorithm to sort the sample values in the computation of the Gini index. The default value is large.sample = FALSE.

Details

For a sample S , with size n and inclusion probabilities $\pi_i = P(i \in S)$ (argument `Pi`), derived from a finite population U , with size N , different formulations of the Gini index have been proposed in the literature. This function estimates the Gini index, variances and confidence intervals using various formulations. The different methods for estimating the Gini index are (see also Muñoz et al., 2023):

\Gini Index formulae.

method = 1 (Langel and Tillé, 2013)

$$\hat{G}_{w1} = \frac{1}{2\hat{N}^2\bar{y}_w} \sum_{i \in S} \sum_{j \in S} w_i w_j |y_i - y_j|,$$

where $\hat{N} = \sum_{i \in S} w_i$, $\bar{y}_w = \hat{N}^{-1} \sum_{i \in S} w_i y_i$, and w_i are the survey weights. For example, the survey weights can be $w_i = \pi_i^{-1}$. `w` or `Pi` must be provided, but not both. It is required that $w_i = \pi_i^{-1}$, for $i \in S$, when both `w` and `Pi` are provided.

method = 2 (Alfons and Templ, 2012; Langel and Tillé, 2013)

$$\hat{G}_{w2} = \frac{2 \sum_{i \in S} w_{(i)}^+ \hat{N}_{(i)} y_{(i)} - \sum_{i \in S} w_i^2 y_i}{\hat{N}^2 \bar{y}_w} - 1,$$

where $y_{(i)}$ are the values y_i sorted in increasing order, $w_{(i)}^+$ are the values w_i sorted according to the increasing order of the values y_i , and $\hat{N}_{(i)} = \sum_{j=1}^i w_{(j)}^+$. Langel and Tillé (2013) show that $\hat{G}_{w1} = \hat{G}_{w2}$.

method = 3 (Berger, 2008)

$$\hat{G}_{w3} = \frac{2}{\hat{N}\bar{y}_w} \sum_{i \in S} w_i y_i \hat{F}_w^*(y_i) - 1,$$

where

$$\hat{F}_w^*(t) = \frac{1}{\hat{N}} \sum_{i \in S} w_i [\delta(y_i < t) + 0.5\delta(y_i = t)]$$

is the smooth (mid-point) distribution function, and $\delta(\cdot)$ is the indicator variable that takes the value 1 when its argument is true, and the value 0 otherwise. It can be seen that $\hat{G}_{w2} = \hat{G}_{w3}$.

method = 4 (Berger and Gedik-Balay, 2020)

$$\hat{G}_{w4} = 1 - \frac{\bar{v}_w}{\bar{y}_w},$$

where $\bar{v}_w = \hat{N}^{-1} \sum_{i \in S} w_i v_i$ and

$$v_i = \frac{1}{\hat{N} - w_i} \sum_{\substack{j \in S \\ j \neq i}} \min(y_i, y_j).$$

method = 5 (Lerman and Yitzhaki, 1989)

$$\hat{G}_{w5} = \frac{2}{\hat{N}\bar{y}_w} \sum_{i \in S} w_{(i)}^+ [y_{(i)} - \bar{y}_w] \left[\hat{F}_w^{LY}(y_{(i)}) - \bar{F}_w^{LY} \right],$$

where

$$\hat{F}_w^{LY}(y_{(i)}) = \frac{1}{\hat{N}} \left(\hat{N}_{(i-1)} + \frac{w_{(i)}^+}{2} \right)$$

and $\bar{F}_w^{LY} = \hat{N}^{-1} \sum_{i \in S} w_{(i)}^+ \hat{F}_w^{LY}(y_{(i)})$.

\ Variances and confidence intervals.

For a given estimator \hat{G}_w and variable z , the Horvitz-Thompson type variance estimator (Horvitz and Thompson, 1952)

$$\hat{V}_{HT}(\hat{G}_w) = \sum_{i \in S} \sum_{j \in S} \check{\Delta}_{ij} w_i w_j z_i z_j$$

is computed when varformula = "HT", where

$$\check{\Delta}_{ij} = \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}}$$

and π_{ij} is the second (joint) inclusion probability of the individuals i and j , i.e., $\pi_{ij} = P\{(i, j) \in S\}$ (argument Pi j).

The Sen-Yates-Grundy type variance estimator (Sen, 1953; Yates and Grundy, 1953)

$$\hat{V}_{SYG}(\hat{G}_w) = -\frac{1}{2} \sum_{i \in S} \sum_{j \in S} \check{\Delta}_{ij} (w_i z_i - w_j z_j)^2$$

is computed when varformula = "SYG", and the Hartley-Rao type variance estimator (Hartley and Rao, 1962)

$$\hat{V}_{HR}(\hat{G}_w) = \frac{1}{n-1} \sum_{i \in S} \sum_{\substack{j \in S \\ j < i}} \left(1 - \pi_i - \pi_j + \frac{1}{n} \sum_{k \in U} \pi_k^2 \right) (w_i z_i - w_j z_j)^2$$

is computed when varformula = "HR". Note that the The Horvitz-Thompson variance estimator can give negative values. We observe that both Horvitz-Thompson and Sen-Yates-Grundy variance estimators depend on second (joint) inclusion probabilities (argument Pi j). The Hajek (1964) approximation

$$\pi_{ij} \cong \pi_i \pi_j \left[1 - \frac{(1 - \pi_i)(1 - \pi_j)}{\sum_{i \in S} (1 - \pi_i)} \right]$$

is used when the second (joint) inclusion probabilities are not available (Pi j = NULL). Note that the Hajek approximation is suggested for large-entropy sampling designs, large samples, and large populations (see Tille 2006; Berger and Tille, 2009; Haziza et al., 2008; Berger, 2011). For instance, this approximation is not recommended for highly-stratified samples (Berger, 2005). The Hartley-Rao variance estimator requires the first inclusion probabilities at the population level (argument PiU). zjackknife computes the confidence interval based on the jackknife technique with critical values based on the Normal approximation. zalinearization and zblinearization compute the confidence intervals based on the linearization technique applied to the estimators

$$\hat{G}_w^a = \hat{G}_{w1}$$

and

$$\widehat{G}_w^b = \frac{2}{\widehat{N}\bar{y}_w} \sum_{i \in S} w_i y_i \widehat{F}_w(y_i) - 1,$$

respectively, where

$$\widehat{F}_w(t) = \frac{1}{\widehat{N}} \sum_{i \in S} w_i \delta(y_i \leq t).$$

Critical values are also based on the Normal approximation. `pbootstrap` computes the variance using the rescaled bootstrap, and the confidence interval is constructed using the percentile method. The **vignette vignette("GiniVarInterval")** contains a detailed description of the various methods for variance estimation and confidence intervals for the Gini index.

The following table summarises the various types of variances and confidence intervals that the function `fgini` computes. The argument `varformula` only applies for the jackknife and linearization techniques (see Berger, 2008; Langel and Tillé, 2013).

Interval	Variance	Critical values	References
<code>zjackknife</code>	Jackknife	Normal	Berger (2008)
<code>zlinearization</code>	Linearization	Normal	Langel and Tillé (2013)
<code>zblinearization</code>	Linearization	Normal	Berger (2008)
<code>pBootstrap</code>	Rescaled bootstrap	Percentile bootstrap	Berger and Gedik-Balay (2020)

Value

When `interval = NULL`, the function returns a single numeric value between 0 and 1 informing about the estimation of the Gini index. When `interval` is not `NULL`, the function returns a list with 3 components: a single numeric value with the estimation of the Gini index; a single numeric value with the variance estimation of the Gini index; and a vector of length two containing the lower and upper limits of the confidence interval for the Gini index.

Author(s)

Juan F Muñoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

- Alfons, A., and Templ, M. (2012). Estimation of social exclusion indicators from complex surveys: The R package `laeken`. KU Leuven, Faculty of Business and Economics Working Paper.
- Berger, Y. G. (2005). Variance estimation with highly stratified sampling designs with unequal probabilities. *Australian & New Zealand Journal of Statistics*, 47, 365–373.
- Berger, Y. G. (2008). A note on the asymptotic equivalence of jackknife and linearization variance estimation for the Gini Coefficient. *Journal of Official Statistics*, 24(4), 541-555.
- Berger, Y. G. (2011). Asymptotic consistency under large entropy sampling designs with unequal probabilities. *Pakistan Journal of Statistics*, 27, 407–426.

- Berger, Y. G. and Tillé, Y. (2009). Sampling with unequal probabilities. In *Sample Surveys: Design, Methods and Applications* (eds. D. Pfeffermann and C. R. Rao), 39–54. Elsevier, Amsterdam
- Berger, Y., and Gedik-Balay, I. (2020). Confidence intervals of Gini coefficient under unequal probability sampling. *Journal of Official Statistics*, 36(2), 237-249.
- Hajek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *The Annals of Mathematical Statistics*, 35, 4, 1491–1523.
- Hartley, H. O., and Rao, J. N. K. (1962). Sampling with unequal probabilities and without replacement. *The Annals of Mathematical Statistics*, 350-374.
- Haziza, D., Mecatti, F. and Rao, J. N. K. (2008). Evaluation of some approximate variance estimators under the Rao-Sampford unequal probability sampling design. *Metron*, LXVI, 91–108.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663–685.
- Langel, M., and Tille, Y. (2013). Variance estimation of the Gini index: revisiting a result several times published. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2), 521-540.
- Lerman, R. I., and Yitzhaki, S. (1989). Improving the accuracy of estimates of Gini coefficients. *Journal of econometrics*, 42(1), 43-47.
- Muñoz, J. F., Moya-Fernández, P. J., and Álvarez-Verdejo, E. (2023). Exploring and Correcting the Bias in the Estimation of the Gini Measure of Inequality. *Sociological Methods & Research*. <https://doi.org/10.1177/00491241231176847>
- Sen, A. R. (1953). On the estimate of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, 5, 119–127.
- Tillé, Y. (2006). *Sampling Algorithms*. Springer, New York.
- Yates, F., and Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society B*, 15, 253–261.

See Also

[fginindex](#), [fcompareCI](#)

Examples

```
# Income and weights (region 'Burgenland') from the 2006 Austrian EU-SILC (Package 'laeken').
data(eusilc, package="laeken")
y <- eusilc$eqIncome[eusilc$db040 == "Burgenland"]
w <- eusilc$rb050[eusilc$db040 == "Burgenland"]

# Estimation of the Gini index using 'method = 2' .
fgini(y, w)

y <- c(30428.83, 14976.54, 18094.09, 29476.79, 20381.93, 6876.17,
      10360.96, 8239.82, 29476.79, 32230.71)
w <- c(357.86, 480.99, 480.99, 476.01, 498.58, 498.58, 476, 498.58, 476.01, 476.01)

# Gini index estimation and confidence interval using:
## a: The method 2 for point estimation.
```

```

## b: The method 'zjackknife' for variance estimation.
## c: The Sen-Yates-Grundy type variance estimator.
## d: The Hajek approximation for the joint inclusion probabilities.
fgini(y, w, interval = "zjackknife")

# Gini index estimation and confidence interval using:
## a: The method 2 for point estimation.
## b: The method 'zlinearization' for variance estimation.
## c: The Sen-Yates-Grundy type variance estimator.
## d: The Hajek approximation for the joint inclusion probabilities.
fgini(y, w, interval = "zlinearization")

# Gini index estimation and confidence interval using:
## a: The method 3 for point estimation.
## b: The method 'zblinearization' for variance estimation.
## c: The Sen-Yates-Grundy type variance estimator.
## d: The Hajek approximation for the joint inclusion probabilities.
fgini(y, w, method = 3L, interval = "zblinearization")

# Gini index estimation and confidence interval using:
## a: The method 2 for point estimation.
## b: The method 'pbootstrap' for variance estimation.
## c: The percentile bootstrap method for the confidence interval.
fgini(y, w, interval = "pbootstrap")

```

fginindex

Gini index for finite populations and different estimation methods.

Description

Estimates the Gini index in finite populations, using different methods.

Usage

```

fginindex(
  y,
  w,
  method = 2L,
  Pi = NULL,
  na.rm = TRUE,
  useRcpp = TRUE
)

```

Arguments

y	A vector with the non-negative real numbers to be used for estimating the Gini index.
w	A numeric vector with the survey weights to be used for estimating the Gini index. This argument can be missing if argument Pi is provided.

method	An integer between 1 and 5 selecting one of the 5 methods detailed below for estimating the Gini index in finite populations. The default method is method = 2L.
Pi	A numeric vector with the (sample) first inclusion probabilities to be used for estimating the Gini index. This argument can be NULL if argument w is provided. The default value is Pi = NULL.
na.rm	A 'TRUE/FALSE' logical value indicating whether NA's should be removed before the computation proceeds. The default value is na.rm = TRUE.
useRcpp	A 'TRUE/FALSE' logical value indicating whether Rcpp (useRcpp = TRUE), or R (useRcpp = FALSE), is used for computation. The default value is UseRcpp = TRUE.

Details

For a sample S , with size n and inclusion probabilities $\pi_i = P(i \in S)$ (argument Pi), derived from a finite population U , with size N , different formulations of the Gini index have been proposed in the literature. This function estimates the Gini index using various formulations, and both R and C++ codes are implemented. This can be useful for research purposes, and speed comparisons can be made. The different methods for estimating the Gini index are (see also Muñoz et al., 2023):

method = 1 (Langel and Tillé, 2013)

$$\hat{G}_{w1} = \frac{1}{2\hat{N}^2\bar{y}_w} \sum_{i \in S} \sum_{j \in S} w_i w_j |y_i - y_j|,$$

where $\hat{N} = \sum_{i \in S} w_i$, $\bar{y}_w = \hat{N}^{-1} \sum_{i \in S} w_i y_i$, and w_i are the survey weights. For example, the survey weights can be $w_i = \pi_i^{-1}$. w or Pi must be provided, but not both. It is required that $w_i = \pi_i^{-1}$, for $i \in S$, when both w and Pi are provided.

method = 2 (Alfons and Templ, 2012; Langel and Tillé, 2013)

$$\hat{G}_{w2} = \frac{2 \sum_{i \in S} w_{(i)}^* \hat{N}_{(i)} y_{(i)} - \sum_{i \in S} w_i^2 y_i}{\hat{N}^2 \bar{y}_w} - 1,$$

where $y_{(i)}$ are the values y_i sorted in increasing order, $w_{(i)}^*$ are the values w_i sorted according to the increasing order of the values y_i , and $\hat{N}_{(i)} = \sum_{j=1}^i w_{(j)}^*$. Langel and Tillé (2013) show that $\hat{G}_{w1} = \hat{G}_{w2}$.

method = 3 (Berger, 2008)

$$\hat{G}_{w3} = \frac{2}{\hat{N}\bar{y}_w} \sum_{i \in S} w_i y_i \hat{F}_w^*(y_i) - 1,$$

where

$$\hat{F}_w^*(t) = \frac{1}{\hat{N}} \sum_{i \in S} w_i [\delta(y_i < t) + 0.5\delta(y_i = t)]$$

is the smooth (mid-point) distribution function, and $\delta(\cdot)$ is the indicator variable that takes the value 1 when its argument is true, and the value 0 otherwise. It can be seen that $\hat{G}_{w2} = \hat{G}_{w3}$.

method = 4 (Berger and Gedik-Balay, 2020)

$$\hat{G}_{w4} = 1 - \frac{\bar{z}_w}{\bar{y}_w},$$

where $\bar{z}_w = \hat{N}^{-1} \sum_{i \in S} w_i z_i$ and

$$z_i = \frac{1}{\hat{N} - w_i} \sum_{\substack{j \in S \\ j \neq i}} \min(y_i, y_j).$$

method = 5 (Lerman and Yitzhaki, 1989)

$$\hat{G}_{w5} = \frac{2}{\hat{N} \bar{y}_w} \sum_{i \in S} w_i [y_i - \bar{y}_w] \left[\hat{F}_w^{LY}(y_i) - \bar{F}_w^{LY} \right],$$

where

$$\hat{F}_w^{LY}(y_i) = \frac{1}{\hat{N}} \left(\hat{N}_{(i-1)} + \frac{w_{(i)}^*}{2} \right)$$

and $\bar{F}_w^{LY} = \hat{N}^{-1} \sum_{i \in S} w_i \hat{F}_w^{LY}(y_i)$.

Value

A single numeric value between 0 and 1. The estimation of the Gini index.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

- Alfons, A., and Templ, M. (2012). Estimation of social exclusion indicators from complex surveys: The R package laeken. KU Leuven, Faculty of Business and Economics Working Paper.
- Berger, Y. G. (2008). A note on the asymptotic equivalence of jackknife and linearization variance estimation for the Gini Coefficient. *Journal of Official Statistics*, 24(4), 541-555.
- Berger, Y. G., and Gedik-Balay, İ. (2020). Confidence intervals of Gini coefficient under unequal probability sampling. *Journal of official statistics*, 36(2), 237-249.
- Langel, M., and Tillé, Y. (2013). Variance estimation of the Gini index: revisiting a result several times published. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2), 521-540.
- Lerman, R. I., and Yitzhaki, S. (1989). Improving the accuracy of estimates of Gini coefficients. *Journal of econometrics*, 42(1), 43-47.
- Muñoz, J. F., Moya-Fernández, P. J., and Álvarez-Verdejo, E. (2023). Exploring and Correcting the Bias in the Estimation of the Gini Measure of Inequality. *Sociological Methods & Research*. <https://doi.org/10.1177/00491241231176847>

See Also

[fgini](#), [fcompareCI](#)

Examples

```
# Income and weights (region "Burgenland") from the 2006 Austrian EU-SILC (Package 'laeken').
data(eusilc, package="laeken")
y <- eusilc$eqIncome[eusilc$db040 == "Burgenland"]
w <- eusilc$rb050[eusilc$db040 == "Burgenland"]

#Comparing the computation time for the various estimation methods and using R
microbenchmark::microbenchmark(
  fginindex(y, w, method = 1L, useRcpp = FALSE),
  fginindex(y, w, method = 2L, useRcpp = FALSE),
  fginindex(y, w, method = 3L, useRcpp = FALSE),
  fginindex(y, w, method = 4L, useRcpp = FALSE),
  fginindex(y, w, method = 5L, useRcpp = FALSE)
)

# Comparing the computation time for the various estimation methods and using Rcpp
microbenchmark::microbenchmark(
  fginindex(y, w, method = 1L),
  fginindex(y, w, method = 2L),
  fginindex(y, w, method = 3L),
  fginindex(y, w, method = 4L),
  fginindex(y, w, method = 5L)
)

# Estimation of the Gini index using 'method = 4'.
y <- c(30428.83, 14976.54, 18094.09, 29476.79, 20381.93, 6876.17,
      10360.96, 8239.82, 29476.79, 32230.71)
w <- c(357.86, 480.99, 480.99, 476.01, 498.58, 498.58, 476, 498.58, 476.01, 476.01)
fginindex(y, w, method = 4L)
```

gbeta

Gini index for the Beta distribution with user-defined shape parameters

Description

Calculates the Gini index for the Beta distribution with shape parameters a (shape1) and b (shape2).

Usage

```
gbeta(shape1, shape2)
```

Arguments

shape1	A positive real number specifying the shape1 parameter a of the Beta distribution.
shape2	A positive real number specifying the shape2 parameter b of the Beta distribution.

Details

The Beta distribution with shape parameters a (argument shape1) and b (argument shape2) and denoted as $Beta(a, b)$, where $a > 0$ and $b > 0$, has a probability density function given by (Kleiber and Kotz, 2003; Johnson et al., 1995; Yee, 2022)

$$f(y) = \frac{1}{B(a, b)} y^{a-1} (1-y)^{b-1},$$

and a cumulative distribution function given by

$$F(y) = \frac{B(y; a, b)}{B(a, b)}$$

where $0 \leq y \leq 1$,

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

is the beta function,

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$$

is the gamma function, and

$$B(y; a, b) = \int_0^y t^{a-1} (1-t)^{b-1} dt$$

is the incomplete beta function.

The Gini index can be computed as

$$G = \frac{2}{a} \frac{B(a+b, a+b)}{B(a, a)B(b, b)}.$$

Value

A numeric value with the Gini index. A NA is returned when a shape parameter is non-numeric or non-positive.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

- Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*, Hoboken, NJ, USA: Wiley-Interscience.
- Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, volume 1, chapter 14. Wiley, New York.
- Yee, T. W. (2022). *VGAM: Vector Generalized Linear and Additive Models*. R package version 1.1-7, <https://CRAN.R-project.org/package=VGAM>.

See Also

[gf](#), [gunif](#), [gweibull](#), [ggamma](#), [gchisq](#)

Examples

```
# Gini index for the Beta distribution with shape parameters 'a = 2' and 'b = 1'.
gbeta(shape1 = 2, shape2 = 1)

# Gini index for the Beta distribution with shape parameters 'a = 1' and 'b = 2'.
gbeta(shape1 = 1, shape2 = 2)
```

gburr

Gini index for the Burr Type XII (Singh-Maddala) distribution with user-defined scale and shape parameters

Description

Calculates the Gini index for the Burr Type XII (Singh-Maddala) distribution with scale parameter b and shape parameters g (shape.g) and s (shape.s).

Usage

```
gburr(
  scale = 1,
  shape.g = 1,
  shape.s = 1
)
```

Arguments

- | | |
|---------|--|
| scale | A positive real number specifying the scale parameter b of the Burr Type XII (Singh-Maddala) distribution. The default value is <code>scale = 1</code> . |
| shape.g | A positive real number specifying the shape parameter g of the Burr Type XII (Singh-Maddala) distribution. The default value is <code>shape.g = 1</code> . |
| shape.s | A positive real number specifying the shape parameter s of the Burr Type XII (Singh-Maddala) distribution. The default value is <code>shape.s = 1</code> . |

Details

The Burr Type XII (Singh-Maddala) distribution with scale parameter b , shape parameters g (argument shape .g) and s (argument shape .s) and denoted as $BurrXII(b, g, s)$, where $b > 0$, $g > 0$ and $s > 0$, has a probability density function given by (Kleiber and Kotz, 2003; Johnson et al., 1995; Rodriguez, 1977; Yee, 2022)

$$f(y) = \frac{gs}{b} \left(\frac{y}{b}\right)^{g-1} \left[1 + \left(\frac{y}{b}\right)^g\right]^{-(s+1)},$$

and a cumulative distribution function given by

$$F(y) = 1 - \left[1 + \left(\frac{y}{b}\right)^g\right]^{-s},$$

where $y > 0$.

The Gini index can be computed as

$$G = 2 \left(0.5 - \frac{1}{E[y]} \int_0^1 \int_0^{Q(y)} y f(y) dy \right),$$

where $Q(y)$ is the quantile function of the Burr Type XII (Singh-Maddala) distribution, and $E[y]$ is the expectation of the distribution. The Burr Type XII (Singh-Maddala) distribution is related to the Pareto (IV) distribution: $BurrXII(b, g, s) = ParetoIV(0, b, 1/g, s)$.

Value

A numeric value with the Gini index. A NA is returned when any of the parameter is non-numeric or non-positive.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, NJ, USA: Wiley-Interscience.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). Continuous Univariate Distributions, volume 1, chapter 14. Wiley, New York.

Rodriguez, R. N. (1977). A guide to the Burr type XII distributions. Biometrika, 64(1), 129-134.

Yee, T. W. (2022). VGAM: Vector Generalized Linear and Additive Models. R package version 1.1-7, <https://CRAN.R-project.org/package=VGAM>.

See Also

[gparetoIV](#), [gpareto](#), [gparetoI](#), [gparetoII](#), [gparetoIII](#), [gfisk](#)

Examples

```
# Gini index for the Burr Type XII distribution with 'scale = 1', 'shape.g = 2', 'shape.s = 1'.
gburr(scale = 1, shape.g = 2, shape.s = 1)
```

```
# Gini index for the Burr Type XII distribution with 'scale = 1', 'shape.g = 5', 'shape.s = 3'.
gburr(scale = 1, shape.g = 5, shape.s = 3)
```

gchisq	<i>Gini index for the Chi-Squared distribution with user-defined degrees of freedom</i>
--------	---

Description

Calculates Gini indices for the Chi-Squared distribution with degrees of freedom n (df).

Usage

```
gchisq(df)
```

Arguments

df A vector of positive real numbers specifying degrees of freedom of the Chi-Squared distribution.

Details

The Chi-Squared distribution with degrees of freedom n (argument df) and denoted as χ_n^2 , where $n > 0$, has a probability density function given by (Kleiber and Kotz, 2003; Johnson et al., 1995)

$$f(y) = \frac{1}{2^{n/2} \Gamma(\frac{n}{2})} y^{n/2-1} e^{-y/2},$$

and a cumulative distribution function given by

$$F(y) = \frac{\gamma(\frac{n}{2}, \frac{y}{2})}{\Gamma(\frac{n}{2})},$$

where $y \geq 0$, the gamma function is defined by

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt,$$

and the lower incomplete gamma function is given by

$$\gamma(\alpha, y) = \int_0^y t^{\alpha-1} e^{-t} dt.$$

The Gini index can be computed as

$$G = \frac{2\Gamma(\frac{1+n}{2})}{n\Gamma(\frac{n}{2})\sqrt{\pi}}.$$

The Chi-Squared distribution is related to the Gamma distribution: $\chi_n^2 = \text{Gamma}(n/2, 2)$.

Value

A numeric vector with the Gini indices. A NA is returned when degrees of freedom are non-numeric or non-positive.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, NJ, USA: Wiley-Interscience.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) Continuous Univariate Distributions, volume 1, chapter 14. Wiley, New York.

See Also

[ggamma](#), [gf](#), [gbeta](#), [glnorm](#)

Examples

```
# Gini index for the Chi-Squared distribution with degrees of freedom equal to 2.
gchisq(df = 2)
```

```
# Gini indices for the Chi-Squared distribution and different degrees of freedom.
gchisq(df = 5:10)
```

gdagum

Gini index for the Dagum distribution with user-defined shape parameters

Description

Calculates the Gini index for the Dagum distribution with shape parameters a (shape1.a) and p (shape2.p).

Usage

```
gdagum(shape1.a, shape2.p)
```

Arguments

shape1.a A positive real number specifying the shape1 parameter a of the Dagum distribution.

shape2.p A positive real number specifying the shape parameter p of the Dagum distribution.

Details

The Dagum distribution with scale parameter b , shape parameters a (argument shape1 .a) and p (argument shape2.p) and denoted as $Dagum(b, a, p)$, where $b > 0$, $a > 0$ and $p > 0$, has a probability density function given by (Kleiber and Kotz, 2003; Johnson et al., 1995; Rodriguez, 1977; Yee, 2022)

$$f(y) = \frac{ap}{y} \frac{\left(\frac{y}{b}\right)^{ap}}{\left[\left(\frac{y}{b}\right)^a + 1\right]^{p+1}},$$

and a cumulative distribution function given by

$$F(y) = \left[1 + \left(\frac{y}{b}\right)^{-a}\right]^{-p},$$

where $y > 0$.

The Gini index can be computed as

$$G = \frac{\Gamma(p)\Gamma(2p + 1/a)}{\Gamma(2p)\Gamma(p + 1/a)} - 1,$$

where the gamma function is defined as

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt.$$

The Dagum distribution is also known the Burr III, inverse Burr, beta-K, or 3-parameter kappa distribution. The Dagum distribution is related to the Fisk (Log Logistic) distribution: $Dagum(b, a, 1) = Fisk(b, a)$. The Dagum distribution is also related to the inverse Lomax distribution and the inverse paralogistic distribution (see Kleiber and Kotz, 2003; Johnson et al., 1995; Yee, 2022).

Value

A numeric value with the Gini index. A NA is returned when a shape parameter is non-numeric or non-positive.

Note

The Gini index of the Dagum distribution does not depend on its scale parameter.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, NJ, USA: Wiley-Interscience.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) Continuous Univariate Distributions, volume 1, chapter 14. Wiley, New York.

Yee, T. W. (2022). VGAM: Vector Generalized Linear and Additive Models. R package version 1.1-7, <https://CRAN.R-project.org/package=VGAM>.

See Also

[gburr](#), [gpareto](#), [gfisk](#), [ggompertz](#), [gfrechet](#)

Examples

```
# Gini index for the Dagum distribution with shape parameters 'a = 2' and 'p = 20'.
gdagum(shape1.a = 2, shape2.p = 20)
```

 gf

Gini index for the F distribution with user-defined degrees of freedom

Description

Calculates the Gini index for the F distribution with degrees of freedom ν_1 (df1) and ν_2 (df2).

Usage

```
gf(df1, df2)
```

Arguments

df1 A positive real number specifying the degrees of freedom ν_1 of the F distribution.

df2 A positive real number higher or equal than two specifying the degrees of freedom ν_2 of the F distribution.

Details

The F distribution with ν_1 (argument df1) and ν_2 (argument df2) degrees of freedom and denoted as F_{ν_1, ν_2} , where $\nu_1 > 0$ and $\nu_2 > 0$, has a probability density function given by (Kleiber and Kotz, 2003; Johnson et al., 1995)

$$f(y) = \frac{\Gamma\left(\frac{\nu_1}{2} + \frac{\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} y^{\nu_1/2-1} \left(1 + \frac{\nu_1 y}{\nu_2}\right)^{-(\nu_1+\nu_2)/2},$$

and a cumulative distribution function given by

$$F(y) = I_{\nu_1 y / (\nu_1 y + \nu_2)}\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right),$$

where $y \geq 0$,

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$$

is the gamma function,

$$I_y(a, b) = \frac{B(y; a, b)}{B(a, b)}$$

is the regularized incomplete beta function,

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

is the beta function, and

$$B(y; a, b) = \int_0^y t^{a-1}(1-t)^{b-1} dt$$

is the incomplete beta function.

The Gini index, for $\nu_2 \geq 2$, can be computed as

$$G = 2 \left(0.5 - \frac{\nu_2 - 2}{\nu_2} \int_0^1 \int_0^{Q(y)} y f(y) dy \right),$$

where $Q(y)$ is the quantile function of the F distribution.

Value

A numeric value with the Gini index. A NA is returned when degrees of freedom are non-numeric or $df1 \leq 0$ or $df2 < 2$.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, NJ, USA: Wiley-Interscience.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) Continuous Univariate Distributions, volume 1, chapter 14. Wiley, New York.

See Also

[gchisq](#), [ggamma](#), [ggompertz](#), [glnorm](#)

Examples

```
# Gini index for the F distribution with 'df1 = 10' and 'df2 = 20' degrees of freedom.
gf(df1 = 10, df2 = 20)
```

gfisk	<i>Gini index for the Fisk (Log Logistic) distribution with user-defined shape parameters</i>
-------	---

Description

Calculates the Gini indices for the Fisk (Log Logistic) distribution with shape parameters a (shape1 . a).

Usage

```
gfisk(shape1 . a)
```

Arguments

shape1 . a A vector of positive real numbers specifying shape parameters a of the Fisk (Log Logistic) distribution.

Details

The Fisk (Log Logistic) distribution with scale parameter b , shape parameter a (argument shape1 . a) and denoted as $Fisk(b, a)$, where $b > 0$ and $a > 0$, has a probability density function given by (Kleiber and Kotz, 2003; Johnson et al., 1995; Yee, 2022)

$$f(y) = \frac{a}{y} \frac{\left(\frac{y}{b}\right)^a}{\left[\left(\frac{y}{b}\right)^a + 1\right]^2},$$

and a cumulative distribution function given by

$$F(y) = 1 - \left[1 + \left(\frac{y}{b}\right)^a\right]^{-1},$$

where $y \geq 0$.

The Gini index can be computed as

$$G = \begin{cases} 1, & 0 < a < 1; \\ \frac{1}{a}, & a \geq 1. \end{cases}$$

The Fisk (Log Logistic) distribution is related to the Dagum distribution: $Fisk(b, a) = Dagum(b, a, 1)$.

Value

A numeric vector with the Gini indices. A NA is returned when a shape parameter is non-numeric or non-positive.

Note

The Gini index of the Fisk (Log Logistic) distribution does not depend on its scale parameter.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>
 Jose M Pavia <pavia@uv.es>
 Encarnacion Alvarez <encarniav@ugr.es>

References

Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, NJ, USA: Wiley-Interscience.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) Continuous Univariate Distributions, volume 1, chapter 14. Wiley, New York.

Yee, T. W. (2022). VGAM: Vector Generalized Linear and Additive Models. R package version 1.1-7, <https://CRAN.R-project.org/package=VGAM>.

See Also

[gdagum](#), [gburr](#), [gpareto](#), [ggompertz](#)

Examples

```
# Gini index for the Fisk distribution with a shape parameter 'a = 2'.
gfisk(shape1.a = 2)

# Gini indices for the Fisk distribution and different shape parameters.
gfisk(shape1.a = 1:10)
```

gfrechet

Gini index for the Frechet distribution with user-defined shape parameters

Description

Calculates the Gini indices for the Frechet distribution with shape parameters s .

Usage

```
gfrechet(shape)
```

Arguments

shape A vector of positive real numbers higher or equal than 1 specifying shape parameters s of the Frechet distribution.

Details

The Frechet distribution with location parameter a , scale parameter b , shape parameter s and denoted as $Frechet(a, b, s)$, where $a > 0$, $b > 0$ and $s > 0$, has a probability density function given by (Kleiber and Kotz, 2003; Johnson et al., 1995)

$$f(y) = \frac{sb}{(y-a)^2} \left(\frac{b}{y-a}\right)^{s-1} \exp\left[-\left(\frac{b}{y-a}\right)^s\right],$$

and a cumulative distribution function given by

$$F(y) = \exp\left[-\left(\frac{b}{y-a}\right)^s\right],$$

where $y > a$.

The Gini index, for $s \geq 1$, can be computed as

$$G = 2^{1/s} - 1.$$

Value

A numeric vector with the Gini indices. A NA is returned when a shape parameter is non-numeric or smaller than 1.

Note

The Gini index of the Frechet distribution does not depend on its location and scale parameters and only is defined when its shape parameter is at least 1.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, NJ, USA: Wiley-Interscience.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) Continuous Univariate Distributions, volume 1, chapter 14. Wiley, New York.

See Also

[gdagum](#), [gburr](#), [gfisk](#), [gpareto](#), [ggompertz](#)

Examples

```
# Gini index for the Frechet distribution with a shape parameter 's = 1'.
gfrechet(shape = 1)

# Gini indices for the Frechet distribution and different shape parameters.
gfrechet(shape = 1:10)
```

ggamma	<i>Gini index for the Gamma distribution with user-defined shape parameter</i>
--------	--

Description

Calculates the Gini indices for the Gamma distribution with shape parameters α .

Usage

```
ggamma(shape)
```

Arguments

shape A vector of positive real numbers specifying the shape parameters α of the Gamma distribution.

Details

The Gamma distribution with shape parameter α , scale parameter σ and denoted as $Gamma(\alpha, \sigma)$, where $\alpha > 0$ and $\sigma > 0$, has a probability density function given by (Kleiber and Kotz, 2003; Johnson et al., 1995)

$$f(y) = \frac{1}{\sigma^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-y/\sigma},$$

and a cumulative distribution function given by

$$F(y) = \frac{\gamma\left(\alpha, \frac{y}{\sigma}\right)}{\Gamma(\alpha)},$$

where $y \geq 0$, the gamma function is defined by

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt,$$

and the lower incomplete gamma function is given by

$$\gamma(\alpha, y) = \int_0^y t^{\alpha-1} e^{-t} dt.$$

The Gini index can be computed as

$$G = \frac{\Gamma\left(\frac{2\alpha+1}{2}\right)}{\alpha \Gamma(\alpha) \sqrt{\pi}}.$$

The Gamma distribution is related to the Chi-squared distribution: $Gamma(n/2, 2) = \chi_n^2$.

Value

A numeric vector with the Gini indices. A NA is returned when a shape parameter is non-numeric or non-positive.

Note

The Gini index of the Gamma distribution does not depend on its scale parameter.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, NJ, USA: Wiley-Interscience.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) Continuous Univariate Distributions, volume 1, chapter 14. Wiley, New York.

See Also

[gchisq](#), [gf](#), [gbeta](#), [gweibull](#), [glnorm](#)

Examples

```
# Gini index for the Gamma distribution with 'shape = 1'.  
ggamma(shape = 1)
```

```
# Gini indices for the Gamma distribution and different shape parameters.  
ggamma(shape = 1:10)
```

ggompertz

Gini index for the Gompertz distribution with user-defined scale and shape parameters

Description

Calculate the Gini index for the Gompertz distribution with scale parameter β and shape parameter α .

Usage

```
ggompertz(  
  scale = 1,  
  shape  
)
```

Arguments

scale	A positive real number specifying the scale parameter β of the Gompertz distribution. The default value is <code>scale = 1</code> .
shape	A positive real number specifying the shape parameter α of the Gompertz distribution.

Details

The Gompertz distribution with scale parameter β , shape parameter α and denoted as $Gompertz(\beta, \alpha)$, where $\beta > 0$ and $\alpha > 0$, has a probability density function given by (Kleiber and Kotz, 2003; Johnson et al., 1995; Rodriguez, 1977; Yee, 2022)

$$f(y) = \alpha e^{\beta y} \exp \left[-\frac{\alpha}{\beta} (e^{\beta y} - 1) \right],$$

and a cumulative distribution function given by

$$F(y) = 1 - \exp \left[-\frac{\alpha}{\beta} (e^{\beta y} - 1) \right],$$

where $y \geq 0$.

The Gini index can be computed as

$$G = 2 \left(0.5 - \frac{1}{E[y]} \int_0^1 \int_0^{Q(y)} y f(y) dy \right),$$

where $Q(y)$ is the quantile function of the Gompertz distribution, and $E[y]$ is the expectation of the distribution. If `scale` is not specified it assumes the default value of 1.

Value

A numeric value with the Gini index. A NA is returned when a parameter is non-numeric or non-positive.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, NJ, USA: Wiley-Interscience.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) Continuous Univariate Distributions, volume 1, chapter 14. Wiley, New York.

Yee, T. W. (2022). VGAM: Vector Generalized Linear and Additive Models. R package version 1.1-7, <https://CRAN.R-project.org/package=VGAM>.

See Also

[ggamma](#), [gbeta](#), [gchisq](#), [gpareto](#)

Examples

```
# Gini index for the Gompertz distribution with 'scale = 1' and 'shape = 3'.
ggompertz(scale = 1, shape = 3)
```

glnorm	<i>Gini index for the Log Normal distribution with user-defined standard deviations</i>
--------	---

Description

Calculates the Gini indices for the Log Normal distribution with standard deviations σ (sdlog).

Usage

```
glnorm(sdlog)
```

Arguments

sdlog A vector of positive real numbers specifying standard deviations σ of the Log Normal distribution.

Details

The Log Normal distribution with mean μ , standard deviation σ on the log scale (argument sdlog) and denoted as $\text{logNormal}(\mu, \sigma)$, has a probability density function given by (Kleiber and Kotz, 2003; Johnson et al., 1995)

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

and a cumulative distribution function given by

$$F(y) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right),$$

where $y > 0$ and

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-t^2/2} dt$$

is the cumulative distribution function of a standard Normal distribution.

The Gini index can be computed as

$$G = 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1.$$

Value

A numeric vector with the Gini indices. A NA is returned when a standard deviation is non-numeric or non-positive.

Note

The Gini index of the logNormal distribution does not depend on the mean parameter.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, NJ, USA: Wiley-Interscience.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) Continuous Univariate Distributions, volume 1, chapter 14. Wiley, New York.

See Also

[ggamma](#), [gpareto](#), [gchisq](#), [gweibull](#)

Examples

```
# Gini index for the Log Normal distribution with standard deviation 'sdlog = 2'.
glnorm(sdlog = 2)
```

```
# Gini indices for the Log Normal distribution with different standard deviations.
glnorm(sdlog = c(0.2, 0.5, 1:3))
```

gpareto

Gini index for the Pareto distribution with user-defined shape parameters

Description

Calculates the Gini indices for the Pareto distribution with shape parameters α .

Usage

```
gpareto(shape)
```

Arguments

shape A vector of positive real numbers specifying shape parameters α of the Pareto distribution.

Details

The Pareto distribution with scale parameter k , shape parameter α and denoted as $Pareto(k, \alpha)$, where $k > 0$ and $\alpha > 0$, has a probability density function given by (Kleiber and Kotz, 2003; Johnson et al., 1995; Yee, 2022)

$$f(y) = \frac{\alpha k^\alpha}{y^{\alpha+1}},$$

and a cumulative distribution function given by

$$F(y) = 1 - \left(\frac{k}{y}\right)^\alpha,$$

where $y \geq k$.

The Gini index can be computed as

$$G = \begin{cases} 1, & 0 < \alpha < 1; \\ \frac{1}{2\alpha - 1}, & \alpha \geq 1. \end{cases}$$

Value

A numeric vector with the Gini indices. A NA is returned when a shape parameter is non-numeric or non-positive.

Note

The Gini index of the Pareto distribution does not depend on the shape parameter.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, NJ, USA: Wiley-Interscience.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) Continuous Univariate Distributions, volume 1, chapter 14. Wiley, New York.

Yee, T. W. (2022). VGAM: Vector Generalized Linear and Additive Models. R package version 1.1-7, <https://CRAN.R-project.org/package=VGAM>.

See Also

[gparetoI](#), [gparetoII](#), [gparetoIII](#), [gparetoIV](#), [gdagum](#), [gburr](#), [gfisk](#)

Examples

```
# Gini index for the Pareto distribution with 'shape = 2'.
gpareto(shape = 2)

# Gini indices for the Pareto distribution and different shape parameters.
gpareto(shape = 1:5)
```

gparetoI	<i>Gini index for the Pareto (I) distribution with user-defined scale and shape parameters</i>
----------	--

Description

Calculate the Gini index for the Pareto (I) distribution with scale parameter b and shape parameter s .

Usage

```
gparetoI(
  scale = 1,
  shape = 1
)
```

Arguments

scale	A positive real number specifying the scale parameter b of the Pareto (I) distribution. The default value is <code>scale = 1</code> .
shape	A positive real number specifying the shape parameter s of the Pareto (I) distribution. The default value is <code>shape = 1</code> .

Details

The Pareto (I) distribution with scale parameter b , shape parameter s and denoted as `ParetoI(b, s)`, where $b > 0$ and $s > 0$, has a probability density function given by (Kleiber and Kotz, 2003; Johnson et al., 1995; Yee, 2022)

$$f(y) = \frac{s}{b} \left(\frac{y}{b}\right)^{-(s+1)},$$

and a cumulative distribution function given by

$$F(y) = 1 - \left(\frac{y}{b}\right)^{-s},$$

where $y > b$.

The Gini index can be computed as

$$G = 2 \left(0.5 - \frac{1}{E[y]} \int_0^1 \int_0^{Q(y)} y f(y) dy \right),$$

where $Q(y)$ is the quantile function of the Pareto (I) distribution, and $E[y]$ is the expectation of the distribution. If scale or shape are not specified they assume the default value of 1. The Pareto (I) distribution is related to the Pareto (IV) distribution: $ParetoI(b, s) = ParetoIV(b, b, 1, s)$

Value

A numeric value with the Gini index. A NA is returned when a parameter is non-numeric or non-positive.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, NJ, USA: Wiley-Interscience.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) Continuous Univariate Distributions, volume 1, chapter 14. Wiley, New York.

Yee, T. W. (2022). VGAM: Vector Generalized Linear and Additive Models. R package version 1.1-7, <https://CRAN.R-project.org/package=VGAM>.

See Also

[gpareto](#), [gparetoII](#), [gparetoIII](#), [gparetoIV](#), [gdagum](#), [gburr](#), [gfisk](#)

Examples

```
# Gini index for the Pareto (I) distribution with scale 'b = 1' and shape 's = 3'.
gparetoI(scale = 1, shape = 3)
```

gparetoII

Gini index for the Pareto (II) distribution with user-defined location, scale and shape parameters

Description

Calculates the Gini index for the Pareto (II) distribution with location parameter a , scale parameter b and shape parameter s .

Usage

```
gparetoII(
  location = 0,
  scale = 1,
  shape = 1
)
```

Arguments

location A positive real number specifying the location parameter a of the Pareto (II) distribution. The default value is `location = 0`.

scale A positive real number specifying the scale parameter b of the Pareto (II) distribution. The default value is `scale = 1`.

shape A positive real number specifying the shape parameter s of the Pareto (II) distribution. The default value is `shape = 1`.

Details

The Pareto (II) distribution with location parameter a , scale parameter b , shape parameter s and denoted as $ParetoII(a, b, s)$, where $a \geq 0$, $b > 0$ and $s > 0$, has a probability density function given by (Kleiber and Kotz, 2003; Johnson et al., 1995; Yee, 2022)

$$f(y) = \frac{s}{b} \left[1 + \left(\frac{y-a}{b} \right) \right]^{-(s+1)},$$

and a cumulative distribution function given by

$$F(y) = 1 - \left(1 + \frac{y-a}{b} \right)^{-s},$$

where $y > a$.

The Gini index can be computed as

$$G = 2 \left(0.5 - \frac{1}{E[y]} \int_0^1 \int_0^{Q(y)} y f(y) dy \right),$$

where $Q(y)$ is the quantile function of the Pareto (II) distribution, and $E[y]$ is the expectation of the distribution. If `location` is not specified it assumes the default value of 0, and `scale` and `shape` assume the default value of 1. The Pareto (II) distribution is related to the Pareto (IV) distribution: $ParetoII(a, b, s) = ParetoIV(a, b, 1, s)$.

Value

A numeric value with the Gini index. A NA is returned when a parameter is non-numeric or positive, except the location parameter that can be equal to 0.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, NJ, USA: Wiley-Interscience.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) Continuous Univariate Distributions, volume 1, chapter 14. Wiley, New York.

Yee, T. W. (2022). VGAM: Vector Generalized Linear and Additive Models. R package version 1.1-7, <https://CRAN.R-project.org/package=VGAM>.

See Also

[gpareto](#), [gparetoI](#), [gparetoIII](#), [gparetoIV](#), [gdagum](#), [gburr](#), [gfisk](#)

Examples

```
# Gini index for the Pareto (II) distribution with parameters 'a = 1', 'b = 1' and 's = 3'.
gparetoII(location = 1, scale = 1, shape = 3)
```

gparetoIII	<i>Gini index for the Pareto (III) distribution with user-defined inequality parameters</i>
------------	---

Description

Calculate the Gini index for the Pareto (III) distribution with inequality parameters g .

Usage

```
gparetoIII(
  inequality = 1
)
```

Arguments

`inequality` A vector of positive numbers in the $[0, 1]$ interval specifying inequality parameters g of the Pareto (III) distribution. The default value is `inequality = 1`.

Details

The Pareto (III) distribution with location parameter a , scale parameter b , inequality parameter g and denoted as $ParetoIII(a, b, g)$, where $a > 0$, $b > 0$, and $g \in [0, 1]$, has a probability density function given by (Kleiber and Kotz, 2003; Johnson et al., 1995; Yee, 2022)

$$f(y) = \frac{1}{bg} \left(\frac{y-a}{b} \right)^{1/g-1} \left[1 + \left(\frac{y-a}{b} \right)^{1/g} \right]^{-2},$$

and a cumulative distribution function given by

$$F(y) = 1 - \left[1 + \left(\frac{y-a}{b} \right)^{1/g} \right]^{-1},$$

where $y > a$.

The Gini index is $G = g$.

If inequality is not specified it assumes the default value of 1. The Pareto (III) distribution is related to the Pareto (IV) distribution: $ParetoIII(a, b, g) = ParetoIV(a, b, g, 1)$.

Value

A numeric vector with the Gini indices. A NA is returned when a inequality parameter is non-numeric or it is out of the interval $[0, 1]$.

Note

The Gini index of the Pareto (III) distribution does not depend on its location and scale parameters.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, NJ, USA: Wiley-Interscience.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) Continuous Univariate Distributions, volume 1, chapter 14. Wiley, New York.

Yee, T. W. (2022). VGAM: Vector Generalized Linear and Additive Models. R package version 1.1-7, <https://CRAN.R-project.org/package=VGAM>.

See Also

[gpareto](#), [gparetoI](#), [gparetoII](#), [gparetoIV](#), [gdagum](#), [gburr](#), [gfisk](#)

Examples

```
# Gini index for the Pareto (III) distribution with inequality parameter 'g = 0.3'.
gparetoIII(inequality = 0.3)

# Gini indices for the Pareto (III) distribution with different inequality parameters.
gparetoIII(inequality = seq(0.1, 0.9, by=0.1))
```

gparetoIV	<i>Gini index for the Pareto (IV) distribution with user-defined location, scale, inequality and shape parameters</i>
-----------	---

Description

Calculates the Gini index for the Pareto (IV) distribution with location parameter a , scale parameter b , inequality parameter g and shape parameter s .

Usage

```
gparetoIV(
  location = 0,
  scale = 1,
  inequality = 1,
  shape = 1
)
```

Arguments

location	A non-negative real number specifying the location parameter a of the Pareto (IV) distribution. The default value is <code>location = 0</code> .
scale	A positive real number specifying the scale parameter b of the Pareto (IV) distribution. The default value is <code>scale = 1</code> .
inequality	A positive real number specifying the inequality parameter g of the Pareto (IV) distribution. The default value is <code>inequality = 1</code> .
shape	A positive real number specifying the shape parameter s of the Pareto (IV) distribution. The default value is <code>shape = 1</code> .

Details

The Pareto (IV) distribution with location parameter a , scale parameter b , inequality parameter g , shape parameter s and denoted as $\text{ParetoIV}(a, b, g, s)$, where $a \geq 0$, $b > 0$, $g > 0$ and $s > 0$, has a probability density function given by (Kleiber and Kotz, 2003; Johnson et al., 1995; Yee, 2022)

$$f(y) = \frac{s}{bg} \left(\frac{y-a}{b} \right)^{1/g-1} \left[1 + \left(\frac{y-a}{b} \right)^{1/g} \right]^{-(s+1)},$$

and a cumulative distribution function given by

$$F(y) = 1 - \left[1 + \left(\frac{y-a}{b} \right)^{1/g} \right]^{-s},$$

where $y > a$.

The Gini index can be computed as

$$G = 2 \left(0.5 - \frac{1}{E[y]} \int_0^1 \int_0^{Q(y)} y f(y) dy \right),$$

where $Q(y)$ is the quantile function of the Pareto (IV) distribution, and $E[y]$ is the expectation of the distribution. If location is not specified it assumes the default value of 0, and the remaining parameters assume the default value of 1. The Pareto (IV) distribution is related to:

1. The Burr distribution: $ParetoIV(0, b, g, s) = BurrXII(b, 1/g, s)$.
2. The Pareto (I) distribution: $ParetoIV(b, b, 1, s) = ParetoI(b, s)$.
3. The Pareto (II) distribution: $ParetoIV(a, b, 1, s) = ParetoII(a, b, s)$.
4. The Pareto (III) distribution: $ParetoIV(a, b, g, 1) = ParetoIII(a, b, g)$.

Value

A numeric value with the Gini index. A NA is returned when a parameter is non-numeric or positive, except for the location parameter that can be equal to 0.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, NJ, USA: Wiley-Interscience.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) Continuous Univariate Distributions, volume 1, chapter 14. Wiley, New York.

Yee, T. W. (2022). VGAM: Vector Generalized Linear and Additive Models. R package version 1.1-7, <https://CRAN.R-project.org/package=VGAM>.

See Also

[gpareto](#), [gparetoI](#), [gparetoII](#), [gparetoIII](#), [gdagum](#), [gburr](#), [gfisk](#)

Examples

```
# Gini index for the Pareto (IV) distribution with 'a = 1', 'b = 1', 'g = 0.5', 's = 1'.
gparetoIV(location = 1, scale = 1, inequality = 0.5, shape = 1)

# Gini index for the Pareto (IV) distribution with 'a = 1', 'b = 1', 'g = 2', 's = 3'.
gparetoIV(location = 1, scale = 1, inequality = 2, shape = 3)
```

gsample	<i>Samples from a set of continuous probability distributions with user-defined Gini indices</i>
---------	--

Description

Draws samples from a continuous probability distribution with Gini indices set by the user.

Usage

```
gsample(
  n,
  gini,
  distribution = c("pareto", "dagum", "lognormal", "fisk", "weibull", "gamma",
    "chisq", "frechet"),
  scale = 1,
  meanlog = 0,
  shape2.p = 1,
  location = 0
)
```

Arguments

n	An integer specifying the sample(s) size.
gini	A numeric vector of values between 0 and 1, indicating the Gini indices for the continuous distribution from which samples are generated.
distribution	A character string specifying the continuous probability distribution to be used to generate the sample. Possible values are "pareto", "dagum", "lognormal", "fisk", "weibull", "gamma", "chisq" and "frechet" for the Pareto, Dagum, logNormal, Fisk (Log-logistic), Weibull, Gamma, Chi-Squared and Frechet distributions, respectively.
scale	The scale parameter for the Pareto, Dagum, Fisk, Weibull, Gamma and Frechet distributions. The default value is scale = 1.
meanlog	The mean for the logNormal distribution on the log scale. The default value is meanlog = 0.
shape2.p	The scale parameter p for the Dagum distribution. The default value is shape2.p = 1.
location	The location parameter for the Frechet distribution. The default value is location = 0.

Details

For each continuous probability distribution, parameters involved in the theoretical formulation of the Gini index (G) are selected such that G takes the values set in the argument `gini`. Additional parameters required in the distribution can be set by the user, and default values are provided. `scale` is the scale parameter for the Pareto, Dagum, Fisk, Weibull, Gamma and Frechet distributions, `meanlog` is the mean for the Lognormal distribution on the log scale, `shape2.p` is the scale parameter `p` for the Dagum distribution, and `location` is the location parameter for the Frechet distribution. Additional information for the continuous probability distributions used by this function can be seen in Kleiber and Kotz (2003), Johnson et al. (1995) and Yee (2022).

Value

A numeric vector (or matrix of order $n \times \text{size}(\text{gini})$) with the samples by columns extracted from the continuous probability distribution stated in `distribution` and the Gini indices corresponding to the vector `gini`.

Note

Underestimation problems may appear for large heavy-tailed distributions (Pareto, Dagum, Lognormal, Fisk and Frechet) and large values of `gini`. A larger sample size may solve/minimize this problem.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>
 Jose M Pavia <pavia@uv.es>
 Encarnacion Alvarez <encarniav@ugr.es>

References

Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, NJ, USA: Wiley-Interscience.
 Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) Continuous Univariate Distributions, volume 1, chapter 14. Wiley, New York.
 Yee, T. W. (2022). VGAM: Vector Generalized Linear and Additive Models. R package version 1.1-7, <https://CRAN.R-project.org/package=VGAM>.

See Also

[gpareto](#), [gdagum](#), [glnorm](#), [gfisk](#), [gweibull](#), [ggamma](#), [gchisq](#), [gfrechet](#)

Examples

```
# Sample from the Pareto distribution and parameter selected such that the Gini index is 0.3.
gsample(n = 10, gini = 0.3, "pareto")

# Samples from the Pareto distribution and gini indices 0.2 and 0.5.
gsample(n = 10, gini = c(0.2,0.5), "par", scale = 2)
```

```

# Samples from the Lognormal distribution and gini indices 0.2 and 0.5.
gsample(n = 10, gini = c(0.2,0.5), "lognormal", meanlog = 5)

# Samples from the Dagum distribution and gini indices 0.2 and 0.5.
gsample(n = 10, gini = c(0.2,0.5), "dagum")

# Samples from the Fisk (Log-logistic) distribution and gini indices 0.3 and 0.6.
gsample(n = 10, gini = c(0.3,0.6), "fisk")

# Sample from the Weibull distribution and parameter selected such that the Gini index is 0.2.
gsample(n = 10, gini = 0.2, "weibull")

# Sample from the Gamma distribution and parameter selected such that the Gini index is 0.3.
gsample(n = 10, gini = 0.2, "gamma")

# Samples from the Chi-Squared distribution and gini indices 0.3 and 0.6..
gsample(n = 10, gini = c(0.3,0.6), "chi")

# Samples from the Frechet distribution and gini indices 0.3 and 0.6.
gsample(n = 10, gini = c(0.3,0.6), "fre")

```

gunif	<i>Gini index for the Uniform distribution with user-defined lower and upper limits</i>
-------	---

Description

Calculates the Gini index for the Uniform distribution with lower limit `min` and upper limit `max`.

Usage

```

gunif(
  min = 0,
  max = 1
)

```

Arguments

<code>min</code>	A non-negative real number specifying the lower limit of the Uniform distribution. The default value is <code>min = 0</code> .
<code>max</code>	A positive real number higher than <code>min</code> specifying the upper limit of the Uniform distribution. The default value is <code>max = 1</code> .

Details

The Uniform distribution with lower and upper limits `min` and `max`, and denoted as $U(\min, \max)$, where $\min \geq 0$, $\max > 0$, $\min < \max$ and both must be finite, has a probability density function

given by (Kleiber and Kotz, 2003; Johnson et al., 1995; Yee, 2022)

$$f(y) = \frac{1}{\max - \min},$$

where $y \in [\min, \max]$. The cumulative distribution function is given by

$$F(y) = \begin{cases} 0, & y < \min; \\ \frac{y - \min}{\max - \min}, & y \in [\min, \max]; \\ 1, & y > \max. \end{cases}$$

The Gini index can be computed as

$$G = \frac{\max - \min}{3(\min + \max)}.$$

If min or max are not specified they assume the default values of 0 and 1, respectively.

Value

A numeric value with the Gini index. A NA value is returned when a limit is non-numeric or non-negative, or $\min \geq \max$.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, NJ, USA: Wiley-Interscience.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) Continuous Univariate Distributions, volume 1, chapter 14. Wiley, New York.

Yee, T. W. (2022). VGAM: Vector Generalized Linear and Additive Models. R package version 1.1-7, <https://CRAN.R-project.org/package=VGAM>.

See Also

[gbeta](#), [ggamma](#), [gchisq](#), [gf](#)

Examples

```
# Gini index for the Uniform distribution with lower limit 0 and upper limit 1.
gunif()
```

```
# Gini index for the Uniform distribution with lower limit 10 and upper limit 190.
gunif(min = 10, max = 190)
```

gweibull	<i>Gini index for the Weibull distribution with user-defined shape parameters</i>
----------	---

Description

Calculate the Gini indices for the Weibull distribution with shape parameters a .

Usage

```
gweibull(shape)
```

Arguments

shape	A vector of positive real numbers specifying shape parameters a of the Weibull distribution.
-------	--

Details

The Weibull distribution with scale parameter σ , shape parameter a , and denoted as $Weibull(\sigma, a)$, where $\sigma > 0$ and $a > 0$, has a probability density function given by (Kleiber and Kotz, 2003; Johnson et al., 1995; Yee, 2022)

$$f(y) = \frac{a}{\sigma} \left(\frac{y}{\sigma}\right)^{a-1} e^{-(y/\sigma)^a},$$

and a cumulative distribution function given by

$$F(y) = 1 - e^{-(y/\sigma)^a},$$

where $y \geq 0$.

The Gini index can be computed as

$$G = 1 - 2^{-1/a}.$$

Value

A numeric vector with the Gini indices. A NA is returned when a shape parameter is non-numeric or non-positive.

Note

The Gini index of the Weibull distribution does not depend on its scale parameter.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*, Hoboken, NJ, USA: Wiley-Interscience.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) *Continuous Univariate Distributions*, volume 1, chapter 14. Wiley, New York.

See Also

[gbeta](#), [ggamma](#), [gchisq](#), [gunif](#)

Examples

```
# Gini index for the Weibull distribution with 'shape = 1'.
gweibull(shape = 1)

# Gini indices for the Weibull distribution and different shape parameters.
gweibull(shape = 1:10)
```

icompareCI	<i>Comparisons of variance estimators and confidence intervals for the Gini index in infinite populations</i>
------------	---

Description

Compares variance estimates and confidence intervals for the Gini index in infinite populations.

Usage

```
icompareCI(
  y,
  B = 1000L,
  alpha = 0.05,
  plotCI = TRUE,
  digitsgini = 2L,
  digitsvar = 4L,
  cum.sums = NULL,
  na.rm = TRUE,
  precisionEL = 1e-4,
  maxiterEL = 100L,
  line.types = c(1L, 2L),
  colors = c("red", "green"),
  save.plot = FALSE
)
```


Arguments

<code>y</code>	A vector with the non-negative real numbers to be used for estimating the Gini index. This argument can be missing if argument <code>cum.sums</code> is provided.
<code>B</code>	A single integer specifying the number of bootstrap replicates. The default value is <code>B = 1000L</code> .
<code>alpha</code>	A single numeric value between 0 and 1 specifying the confidence level $1-\alpha$ to be used for computing the confidence interval for the Gini. Some authors call <code>alpha</code> the significance level. The default value is <code>alpha = 0.05</code> .
<code>plotCI</code>	A 'TRUE/FALSE' logical value indicating whether confidence intervals are compared using a plot. The default value is <code>plotCI = TRUE</code> .
<code>digitsgini</code>	A single integer specifying the number of decimals used in the estimation of the Gini index and confidence intervals. The default value is <code>digitsgini = 2L</code> .
<code>digitsvar</code>	A single integer specifying the number of decimals used in the variance estimation of the Gini index. The default value is <code>digitsvar = 4L</code> .
<code>cum.sums</code>	A numeric vector of non-negative real numbers specifying the cumulative sums of the variable used to estimate the Gini index. This argument can be <code>NULL</code> if argument <code>y</code> is provided. The default value is <code>cum.sums = NULL</code> .
<code>na.rm</code>	A 'TRUE/FALSE' logical value indicating whether the NA should be removed before the computation proceeds. The default value is <code>na.rm = TRUE</code> .
<code>precisionEL</code>	A single numeric value specifying the precision for the confidence interval based on the empirical likelihood method. The default value is <code>precisionEL = 1e-4</code> , i.e., limits of the confidence interval have a total of 4 decimal places.
<code>maxiterEL</code>	A single integer specifying the maximum number of iterations allowed for the convergence in the empirical likelihood method. The default value is <code>maxiterEL = 100L</code> .
<code>line.types</code>	A numeric vector with length equal 2 specifying the line types. See the function <code>plot</code> for the different line types. The default value is <code>lty = c(1L, 2L)</code> .
<code>colors</code>	A numeric vector with length equal 2 specifying the colors for lines of the plot. The default value is <code>colors = c("red", "green")</code> .
<code>save.plot</code>	A 'TRUE/FALSE' logical value indicating whether the ggplot object of the plot comparing the confidence intervals should be saved in the output. The default value is <code>save.plot = FALSE</code> .

Details

For a sample S , with size n , derived from an infinite population, the Gini index is estimated by two different versions (see Muñoz et al., 2023 for more details):

$$\hat{G} = \frac{2}{\bar{y}n^2} \sum_{i \in S} iy_{(i)} - \frac{n+1}{n};$$

$$\hat{G}^{bc} = \frac{2}{\bar{y}n(n-1)} \sum_{i \in S} iy_{(i)} - \frac{n+1}{n-1},$$

where the label *bc* indicates that the bias correction is applied. The table below summarises the various types of variances and confidence intervals that computes this function. Methods based on the jackknife technique use the fast algorithm suggested by Ogwang (2000). The linearization technique for variance estimation (Deville, 1999) has been applied to the following estimators of the Gini index (Berger, 2008; Langel and Tille, 2013):

$$\widehat{G}^a = \frac{1}{2\bar{y}n^2} \sum_{i \in S} \sum_{j \in S} |y_i - y_j|$$

and

$$\widehat{G}^b = \frac{2}{\bar{y}n} \sum_{i \in S} y_i \widehat{F}_n(y_i) - 1,$$

where

$$\widehat{F}_n(y_i) = \frac{1}{n} \sum_{j \in S} \delta(y_j \leq y_i).$$

zlinearization and zblinearization linearize, respectively, the estimators \widehat{G}^a and \widehat{G}^b . The percentile bootstrap (see Qin et al., 2010) is computed using pbootstrap. bca is the bias corrected bootstrap confidence interval (Efron and Tibshirani, 1993). ELchisq and ELboot are the confidence intervals based on the empirical likelihood method. The **vignette** vignette("GiniVarInterval") contains a detailed description of the various methods for variance estimation and confidence intervals for the Gini index.

Interval	Variance	Critical values	References
zjackknife	Jackknife	Normal	Berger (2008)
tjackknife	Jackknife	Studentized bootstrap	Biewen (2002); Berger (2008)
zlinearization	Linearization	Normal	Langel and Tille (2013)
zblinearization	Linearization	Normal	Berger (2008)
tlinearization	Linearization	Studentized bootstrap	Langel and Tille (2013)
tblinearization	Linearization	Studentized bootstrap	Biewen (2002); Berger (2008)
pBootstrap	Bootstrap	Percentile bootstrap	Qin et al. (2010)
BCa	Bootstrap	BCa bootstrap	Davison and Hinkley (1997)
ELchisq	Linearization	Chi-Squared	Qin et al. (2010)
ELboot	Bootstrap	Percentile bootstrap	Qin et al. (2010)

Value

If `save.plot = FALSE`, a data frame with columns:

1. `interval`. The method used to construct the confidence interval.
2. `bc`. A 'TRUE/FALSE' logical value indicating whether the bias correction is applied.
3. `gini`. The estimation of the Gini index.
4. `lowerlimit`. The lower limit of the confidence interval.
5. `upperlimit`. The upper limit of the confidence interval.
6. `var.gini`. The variance estimation for the estimator of the Gini index.

If `save.plot = TRUE`, a list with two components: (i) 'base.CI' a data frame of six columns as just described and (ii) 'plot' a (ggplot) description of the plot, which is a list with components that contain the plot itself, the data, information about the scales, panels, etc. As a side-effect, a plot that compares the various methods for constructing confidence intervals for the Gini index is displayed. ****ggplot2**** is needed to be installed for this option to work.

If `plotCI = TRUE`, as a side-effect, a plot that compares the various methods for constructing confidence intervals for the Gini index is displayed. ****ggplot2**** is needed to be installed for this option to work.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

Berger, Y. G. (2008). A note on the asymptotic equivalence of jackknife and linearization variance estimation for the Gini Coefficient. *Journal of Official Statistics*, 24(4), 541-555.

Biewen, M. (2002). Bootstrap inference for inequality, mobility and poverty measurement. *Journal of Econometrics*, 108(2), 317-342.

Davison, A. C., and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application* (Cambridge Series in Statistical and Probabilistic Mathematics, No 1)—Cambridge University Press.

Deville, J.C. (1999). Variance Estimation for Complex Statistics and Estimators: Linearization and Residual Techniques. *Survey Methodology*, 25, 193–203.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York, London.

Langel, M., and Tille, Y. (2013). Variance estimation of the Gini index: revisiting a result several times published. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2), 521-540.

Muñoz, J. F., Moya-Fernández, P. J., and Álvarez-Verdejo, E. (2023). Exploring and Correcting the Bias in the Estimation of the Gini Measure of Inequality. *Sociological Methods & Research*. <https://doi.org/10.1177/00491241231176847>

Ogwang, T. (2000). A convenient method of computing the Gini index and its standard error. *Oxford Bulletin of Economics and Statistics*, 62(1), 123-123.

Qin, Y., Rao, J. N. K., and Wu, C. (2010). Empirical likelihood confidence intervals for the Gini measure of income inequality. *Economic Modelling*, 27(6), 1429-1435.

See Also

[igini](#), [iginindex](#)

Examples

```
# Sample, with size 50, from a Lognormal distribution. The true Gini index is 0.5.
set.seed(123)
y <- gsample(n = 50, gini = 0.5, distribution = "lognormal")

# Estimation of the Gini index and confidence intervals using different methods.
icompareCI(y)
```

igini	<i>Gini index, variances and confidence intervals in infinite populations</i>
-------	---

Description

Estimation of the Gini index and computation of variances and confidence interval for infinite populations.

Usage

```
igini(
  y,
  bias.correction = TRUE,
  interval = NULL,
  B = 1000L,
  alpha = 0.05,
  cum.sums = NULL,
  na.rm = TRUE,
  precisionEL = 1e-04,
  maxiterEL = 100L,
  large.sample = FALSE
)
```

Arguments

y	A vector with the non-negative real numbers to be used for estimating the Gini index. This argument can be missing if argument <code>cum.sums</code> is provided.
bias.correction	A 'TRUE/FALSE' logical value indicating whether the bias correction should be applied to the estimation of the Gini index. The default value is <code>bias.correction = TRUE</code> .
interval	A character string specifying the type of variance estimation and confidence interval to be used, or NULL (the default value) to omit the computation of both variance and confidence interval. Possible values are "zjackknife", "tjackknife", "zlinearization", "zblinearization", "tlinearization", "tblinearization", "pbootstrap", "BCa", "ELchisq" and "ELboot". The default value is <code>interval = NULL</code> .
B	A single integer specifying the number of bootstrap replicates. The default value is <code>B = 1000L</code> .

alpha	A single numeric value between 0 and 1. If interval is not NULL, the confidence level to be used for computing the confidence interval for the Gini is 1-alpha. Some authors call alpha the significance level. The default value is alpha = 0.05.
cum.sums	A vector with the non-negative real numbers specifying the cumulative sums of the variable used to estimate the Gini index. This argument can be NULL if argument y is provided. The default value is cum.sums = NULL.
na.rm	A 'TRUE/FALSE' logical value indicating whether NA's should be removed before the computation proceeds. The default value is na.rm = TRUE.
precisionEL	A single numeric value specifying the precision for the confidence interval based on the empirical likelihood method. The default value is precisionEL = 1e-4, i.e., limits of the confidence interval have a total of 4 decimal places.
maxiterEL	A single integer specifying the maximal number of iterations allowed for the convergence of the empirical likelihood method. The default value is maxiterEL = 100L.
large.sample	A 'TRUE/FALSE' logical value indicating whether the sample is large to apply a faster algorithm to sort the sample values. The default value is large.sample = FALSE.

Details

For a sample S , with size n , derived from an infinite population, the Gini index is estimated by

$$\hat{G} = \frac{2}{\bar{y}n^2} \sum_{i \in S} iy_{(i)} - \frac{n+1}{n}$$

when `bias.correction = FALSE`, and by

$$\hat{G}^{bc} = \frac{2}{\bar{y}n(n-1)} \sum_{i \in S} iy_{(i)} - \frac{n+1}{n-1}$$

when `bias.correction = TRUE`. For more details, see Muñoz et al. (2023). The table below summarises the various types of variances and confidence intervals that computes this function. Methods based on the jackknife technique use the fast algorithm suggested by Ogwang (2000). The linearization technique for variance estimation (Deville, 1999) has been applied to the following estimators of the Gini index (Berger, 2008; Langel and Tille, 2013):

$$\hat{G}^a = \frac{1}{2\bar{y}n^2} \sum_{i \in S} \sum_{j \in S} |y_i - y_j|$$

and

$$\hat{G}^b = \frac{2}{\bar{y}n} \sum_{i \in S} y_i \hat{F}_n(y_i) - 1,$$

where

$$\hat{F}_n(y_i) = \frac{1}{n} \sum_{j \in S} \delta(y_j \leq y_i).$$

zlinearization and zblinearization linearize, respectively, the estimators \hat{G}^a and \hat{G}^b . The percentile bootstrap (see Qin et al., 2010) is computed using pbootstrap. Bca is the bias corrected bootstrap confidence interval (Efron and Tibshirani, 1993). ELchisq and ELboot are the confidence intervals based on the empirical likelihood method. The **vignette** vignette("GiniVarInterval") contains a detailed description of the various methods for variance estimation and confidence intervals for the Gini index.

Interval	Variance	Critical values	References
zjackknife	Jackknife	Normal	Berger (2008)
tjackknife	Jackknife	Studentized bootstrap	Biewen (2002); Berger (2008)
zlinearization	Linearization	Normal	Langel and Tille (2013)
zblinearization	Linearization	Normal	Berger (2008)
tlinearization	Linearization	Studentized bootstrap	Langel and Tille (2013)
tblinearization	Linearization	Studentized bootstrap	Biewen (2002); Berger (2008)
pBootstrap	Bootstrap	Percentile bootstrap	Qin et al. (2010)
BCa	Bootstrap	BCa bootstrap	Davison and Hinkley (1997)
ELchisq	Linearization	Chi-Squared	Qin et al. (2010)
ELboot	Bootstrap	Percentile bootstrap	Qin et al. (2010)

Value

When interval = NULL, a single numeric value between 0 and 1, containing the estimation of the Gini index based on the vector y or the vector cum. sums. When interval is not NULL, a list of 3 components: a single numeric value with the estimation of the Gini index; a single numeric value with the variance estimation of the Gini index; and a numeric matrix with 1 row and 2 columns containing the lower and upper limits of the confidence intervals for the Gini index.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

- Berger, Y. G. (2008). A note on the asymptotic equivalence of jackknife and linearization variance estimation for the Gini Coefficient. *Journal of Official Statistics*, 24(4), 541-555.
- Biewen, M. (2002). Bootstrap inference for inequality, mobility and poverty measurement. *Journal of Econometrics*, 108(2), 317-342.
- Davison, A. C., and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application* (Cambridge Series in Statistical and Probabilistic Mathematics, No 1)—Cambridge University Press.
- Deville, J.C. (1999). Variance Estimation for Complex Statistics and Estimators: Linearization and Residual Techniques. *Survey Methodology*, 25, 193–203.
- Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York, London.

Langel, M., and Tille, Y. (2013). Variance estimation of the Gini index: revisiting a result several times published. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2), 521-540.

Muñoz, J. F., Moya-Fernández, P. J., and Álvarez-Verdejo, E. (2023). Exploring and Correcting the Bias in the Estimation of the Gini Measure of Inequality. *Sociological Methods & Research*. <https://doi.org/10.1177/00491241231176847>

Ogwang, T. (2000). A convenient method of computing the Gini index and its standard error. *Oxford Bulletin of Economics and Statistics*, 62(1), 123-123.

Qin, Y., Rao, J. N. K., and Wu, C. (2010). Empirical likelihood confidence intervals for the Gini measure of income inequality. *Economic Modelling*, 27(6), 1429-1435.

See Also

[icompareCI](#), [iginindex](#)

Examples

```
# Sample, with size 50, from a Lognormal distribution. The true Gini index is 0.5.
set.seed(123)
y <- gsample(n = 50, gini = 0.5, distribution = "lognormal")

# Bias corrected estimation of the Gini index.
igini(y)

# Estimation of the Gini index and confidence interval based on jackknife and studentized bootstrap.
igini(y, interval = "tjackknife")
```

iginindex

Gini index for infinite populations and different estimation methods.

Description

Estimates the Gini index in infinite populations, using different methods.

Usage

```
iginindex(
  y,
  method = 5L,
  bias.correction = TRUE,
  cum.sums = NULL,
  na.rm = TRUE,
  useRcpp = TRUE
)
```

Arguments

y	A vector with the non-negative real numbers to be used for estimating the Gini index. This argument can be missing if argument cum.sums is provided.
method	An integer between 1 and 10 selecting one of the 10 methods detailed below for estimating the Gini index in infinite populations. The default method is method = 5L.
bias.correction	A ‘TRUE/FALSE’ logical value indicating whether the bias correction should be applied to the estimation of the Gini index. The default value is bias.correction = TRUE.
cum.sums	A vector with the non-negative real numbers specifying the cumulative sums of the variable used to estimate the Gini index. This argument can be NULL if argument y is provided. The default value is cum.sums = NULL.
na.rm	A ‘TRUE/FALSE’ logical value indicating whether NA’s should be removed before the computation proceeds. The default value is na.rm = TRUE.
useRcpp	A ‘TRUE/FALSE’ logical value indicating whether Rcpp (useRcpp = TRUE) or R (useRcpp = FALSE) is used for computation. The default value is UseRcpp = TRUE.

Details

For a sample S , with size n , derived from an infinite population, different formulations of the Gini index have been proposed in the literature, but they only provide two different outputs.

This function estimates the Gini index using the various formulations, and both R and C++ codes are implemented. This can be useful for research purposes, and speed comparisons can be made. The argument cum.sums does not require that the cumulative sums are based on the non-decreasing order of the variable y .

The different methods for estimating the Gini index are (see Wang et al., 2016; Giorgi and Gigliarano, 2017; Mukhopadhyay and Sengupta, 2021; Muñoz et al., 2023):

method = 1

$$\hat{G}_1 = \frac{1}{2\bar{y}n^2} \sum_{i \in S} \sum_{j \in S} |y_i - y_j|;$$

$$\hat{G}_1^{bc} = \frac{1}{2\bar{y}n(n-1)} \sum_{i \in S} \sum_{j \in S} |y_i - y_j|,$$

where $\bar{y} = n^{-1} \sum_{i \in S} y_i$ is the sample mean and the label bc indicates that the bias correction is applied to the estimation of the Gini index.

method = 2

$$\hat{G}_2 = \frac{n-1}{n} \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum_{i=1}^{n-1} p_i},$$

$$\hat{G}_2^{bc} = \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum_{i=1}^{n-1} p_i},$$

where

$$p_i = \frac{i}{n}; \quad q_i = \frac{y_i^+}{y_n^+},$$

and $y_i^+ = \sum_{j=1}^i y_{(j)}$, with $i = \{1, \dots, n\}$, are the cumulative sums of the ordered values $y_{(i)}$ (in non-decreasing order) of the variable of interest y .

method = 3

$$\hat{G}_3 = \frac{n-1}{n} - \frac{2}{n} \sum_{i=1}^{n-1} q_i;$$

$$\hat{G}_3^{bc} = 1 - \frac{2}{n-1} \sum_{i=1}^{n-1} q_i.$$

method = 4

$$\hat{G}_4 = 1 - \sum_{i=0}^{n-1} (q_{i+1} + q_i)(p_{i+1} - p_i);$$

$$\hat{G}_4^{bc} = \frac{n}{n-1} \left[1 - \sum_{i=0}^{n-1} (q_{i+1} + q_i)(p_{i+1} - p_i) \right],$$

where $p_0 = q_0 = 0$.

method = 5

$$\hat{G}_5 = \frac{2}{\bar{y}n^2} \sum_{i \in S} iy_{(i)} - \frac{n+1}{n};$$

$$\hat{G}_5^{bc} = \frac{2}{\bar{y}n(n-1)} \sum_{i \in S} iy_{(i)} - \frac{n+1}{n-1}.$$

method = 6

$$\hat{G}_6 = \frac{2}{\bar{y}n} \text{cov}(i, y_{(i)});$$

$$\hat{G}_6^{bc} = \frac{2}{\bar{y}(n-1)} \text{cov}(i, y_{(i)}).$$

method = 7

$$\hat{G}_7 = \frac{1}{\bar{y}n^2} \sum_{i \in S} \sum_{j \in S} |y_i - y_j| \cdot |\hat{F}_n^*(y_i) - \hat{F}_n^*(y_j)|;$$

$$\hat{G}_7^{bc} = \frac{1}{\bar{y}n(n-1)} \sum_{i \in S} \sum_{j \in S} |y_i - y_j| \cdot |\hat{F}_n^*(y_i) - \hat{F}_n^*(y_j)|,$$

where

$$\hat{F}_n^*(t) = \frac{1}{n} \sum_{i \in S} [\delta(y_i < t) + 0.5\delta(y_i = t)]$$

is the smooth (mid-point) distribution function.

method = 8

$$\hat{G}_8 = 1 - \frac{1}{\bar{y}n^2} \sum_{i \in S} \sum_{j \in S} \min(y_i, y_j);$$

$$\hat{G}_8^{bc} = 1 - \frac{1}{\bar{y}n(n-1)} \sum_{i \in S} \sum_{\substack{j \in S \\ j \neq i}} \min(y_i, y_j).$$

method = 9

$$\hat{G}_9 = \frac{2}{\bar{y}n} \sum_{i \in S} y_i \hat{F}_n^*(y_i) - 1;$$

$$\hat{G}_9^{bc} = \frac{2}{\bar{y}(n-1)} \sum_{i \in S} y_i \hat{F}_n^*(y_i) - \frac{n}{n-1}.$$

method = 10

$$\hat{G}_{10} = \frac{n-1}{2\bar{y}n} \binom{n}{2}^{-1} \sum_{i \leq i_1 < i_2 \leq n} |y_{i_1} - y_{i_2}|;$$

$$\hat{G}_{10}^{bc} = \frac{1}{2\bar{y}} \binom{n}{2}^{-1} \sum_{i \leq i_1 < i_2 \leq n} |y_{i_1} - y_{i_2}|.$$

Value

A single numeric value between 0 and 1 containing the estimation of the Gini index based on the vector `y` or the vector `cum.sums`.

Author(s)

Juan F Munoz <jfmunoz@ugr.es>

Jose M Pavia <pavia@uv.es>

Encarnacion Alvarez <encarniav@ugr.es>

References

Giorgi, G. M., and Gigliarano, C. (2017). The Gini concentration index: a review of the inference literature. *Journal of Economic Surveys*, 31(4), 1130-1148.

Mukhopadhyay, N., and Sengupta, P. P. (Eds.). (2021). *Gini inequality index: Methods and applications*. CRC press.

Muñoz, J. F., Moya-Fernández, P. J., and Álvarez-Verdejo, E. (2023). Exploring and Correcting the Bias in the Estimation of the Gini Measure of Inequality. *Sociological Methods & Research*. <https://doi.org/10.1177/00491241231176847>

Wang, D., Zhao, Y., and Gilmore, D. W. (2016). Jackknife empirical likelihood confidence interval for the Gini index. *Statistics & Probability Letters*, 110, 289-295.

See Also

[igini](#), [icompareCI](#)

Examples

```
# Sample, with size 50, from a Lognormal distribution. The true Gini index is 0.5.
set.seed(123)
y <- gsample(n = 50, gini = 0.5, meanlog = 5)

# Estimation of the Gini index using the method = 5, bias correction, and Rcpp.
```

```
iginindex(y)

# Estimation of the Gini index using the method = 5, bias correction, and R.
iginindex(y, useRcpp = FALSE)

#Comparing the computation time for the various estimation methods and using R
microbenchmark::microbenchmark(
  iginindex(y, method = 1, useRcpp = FALSE),
  iginindex(y, method = 2, useRcpp = FALSE),
  iginindex(y, method = 3, useRcpp = FALSE),
  iginindex(y, method = 4, useRcpp = FALSE),
  iginindex(y, method = 5, useRcpp = FALSE),
  iginindex(y, method = 6, useRcpp = FALSE),
  iginindex(y, method = 7, useRcpp = FALSE),
  iginindex(y, method = 8, useRcpp = FALSE),
  iginindex(y, method = 9, useRcpp = FALSE),
  iginindex(y, method = 10, useRcpp = FALSE)
)

# Comparing the computation time for the various estimation methods and using Rcpp
microbenchmark::microbenchmark(
  iginindex(y, method = 1),
  iginindex(y, method = 2),
  iginindex(y, method = 3),
  iginindex(y, method = 4),
  iginindex(y, method = 5),
  iginindex(y, method = 6),
  iginindex(y, method = 7),
  iginindex(y, method = 8),
  iginindex(y, method = 9),
  iginindex(y, method = 10) )
```

Index

fcompareCI, [2](#), [13](#), [17](#)
fgini, [8](#), [8](#), [17](#)
fginindex, [8](#), [13](#), [14](#)

gbeta, [17](#), [22](#), [30](#), [32](#), [45](#), [47](#)
gburr, [19](#), [24](#), [27](#), [28](#), [35](#), [36](#), [38](#), [39](#), [41](#)
gchisq, [19](#), [21](#), [25](#), [30](#), [32](#), [33](#), [43](#), [45](#), [47](#)
gdagum, [22](#), [27](#), [28](#), [35](#), [36](#), [38](#), [39](#), [41](#), [43](#)
gf, [19](#), [22](#), [24](#), [30](#), [45](#)
gfisk, [20](#), [24](#), [26](#), [28](#), [35](#), [36](#), [38](#), [39](#), [41](#), [43](#)
gfrechet, [24](#), [27](#), [43](#)
ggamma, [19](#), [22](#), [25](#), [29](#), [32](#), [33](#), [43](#), [45](#), [47](#)
ggompertz, [24](#), [25](#), [27](#), [28](#), [30](#)
glnorm, [22](#), [25](#), [30](#), [32](#), [43](#)
gpareto, [20](#), [24](#), [27](#), [28](#), [32](#), [33](#), [33](#), [36](#), [38](#), [39](#),
[41](#), [43](#)
gparetoI, [20](#), [35](#), [35](#), [38](#), [39](#), [41](#)
gparetoII, [20](#), [35](#), [36](#), [36](#), [39](#), [41](#)
gparetoIII, [20](#), [35](#), [36](#), [38](#), [38](#), [41](#)
gparetoIV, [20](#), [35](#), [36](#), [38](#), [39](#), [40](#)
gsample, [42](#)
gunif, [19](#), [44](#), [47](#)
gweibull, [19](#), [30](#), [33](#), [43](#), [46](#)

icompareCI, [47](#), [54](#), [57](#)
igini, [50](#), [51](#), [57](#)
iginindex, [50](#), [54](#), [54](#)