Package ‘overlapping’

May 28, 2019

Type Package

Title Estimation of Overlapping in Empirical Distributions

Version 1.5.3

Date 2019-05-28

Author Massimiliano Pastore

Maintainer Massimiliano Pastore <massimiliano.pastore@unipd.it>

Description Functions for estimating the overlapping area of two or more kernel density estimations from empirical data.

Depends R (>= 3.0.0), ggplot2, testthat

License GPL-2

Encoding UTF-8

NeedsCompilation no

Repository CRAN

Date/Publication 2019-05-28 11:20:02 UTC

R topics documented:

boot.overlap .................................................... 2
cutnumeric ..................................................... 3
final.plot ....................................................... 4
overlap ......................................................... 5

Index 7
Description

Bootstrap the estimated overlapping area of two or more kernel density estimations from empirical data.

Usage

`boot.overlap(x, B = 1000, ...)`

Arguments

- `x`: list of numerical vectors to be compared; each vector is an element of the list
- `B`: integer, number of bootstrap draws
- `...`: options, see function `overlap` for details

Details

If the list `x` contains more than two elements (i.e. more than two distributions) it computes bootstrap overlapping between all `q` number of paired distributions. For example, if `x` contains three elements, then `q = 3`; if `x` contains four elements, then `q = 6`, and so on.

Value

It returns a list containing the following components:

- `OVboot_stats`: Data frame `q x 3`; each row containing the following statistics: `estOV`, estimated overlapping area, \( \hat{\eta} \); `bias`, difference between the bootstrap’s expected value and the observed value of the overlapping area: \( E(\hat{\eta}^*) - \hat{\eta} \); `se`, bootstrap standard error \( \sigma_{\hat{\eta}} \).
- `OVboot_dist`: Matrix `B x q`, `B` rows (bootstrap replicates) and `q` columns (depending on the number of elements of `x`); each column is a bootstrap distribution of an overlapping index.

Note

Call function `overlap`.

Author(s)

Massimiliano Pastore
References

Examples
set.seed(20150605)
x <- list(X1=rnorm(100), X2=rt(50,8), X3=rchisq(80,2))

## bootstrapping
out <- boot.overlap( x, B = 10 )
out$O$boot_stats

# bootstrap quantile intervals
apply( out$O$boot_dist, 2, quantile, probs = c(.05, .9) )

# plot of bootstrap distributions
Y <- stack( data.frame( out$O$boot_dist ) )
ggplot( Y, aes( values ) ) + facet_wrap( ~ind ) + geom_density()

cutnumeric

Description
It divides a numerical variable x in classes, and returns for each class the central value.

Internal function, generally not to be called by the user.

Usage
cutnumeric( x, n = 1000 )

Arguments
x numeric vector
n number of classes

Details
It calls the cut function, and then converts factor classes in numeric classes, returning for each class its central value.

Value
It returns a numerical vector. The values are the central points of classes obtained by the function cut.
Author(s)

Massimiliano Pastore

See Also

cut

Examples

```r
x <- rnorm(50)
cutnumeric(x, 5)
```

Description

Graphical representation of estimated densities and overlapping area.

Usage

```r
final.plot( x, OV = NULL )
```

Arguments

- `x` list of numerical vectors to be compared; each vector is an element of the list, see `overlap`
- `OV` Optional vector of overlapping areas obtained by `overlap`

Details

It requires the package `ggplot2`.

Author(s)

Massimiliano Pastore

Examples

```r
set.seed(20150605)
x <- list(x1=rnorm(100), x2=rt(50, 8), x3=rchisq(80, 2))
out <- overlap(x)
final.plot(x, out$OV)
```
**Description**

It gives the overlapped estimated area of two or more kernel density estimations from empirical data.

**Usage**

```r
overlap( x, nbins = 1024, plot = FALSE, partial.plot = FALSE, boundaries = NULL, ... )
```

**Arguments**

- `x` list of numerical vectors to be compared; each vector is an element of the list
- `nbins` number of equally spaced points at which the overlapping density is evaluated; see `density` for details
- `plot` logical, if TRUE, final plot of estimated densities and overlapped areas is produced
- `partial.plot` logical, if TRUE, partial paired distributions are plotted
- `boundaries` an optional list for bounded distributions, see Details
- `...` optional arguments to be passed to function `density`

**Details**

If the list `x` contains more than two elements (i.e. more than two distributions) it computes overlapping between all paired distributions. Partial plots refer to these paired distributions.

If `plot=TRUE`, all overlapped areas are plotted. It requires `ggplot2`.

The optional list `boundaries` must contain two elements: `from` and `to`, indicating the empirical limits of input variables. Each element must be of length equal to the input data list `x` or, at least, length one when all boundaries are equal for all distributions. See examples below.

**Value**

It returns a list containing the following components:

- **DD** Data frame with information used for computing overlapping, containing the following variables: `x`, coordinates of the points where the density is estimated; `y1` and `y2`, densities; `ovy`, density for estimating overlapping area (i.e. `min(y1,y2)`); `ally`, density for estimating whole area (i.e. `max(y1,y2)`); `dominance`, indicates which distribution has the highest density; `k`, label indicating which distributions are compared.
- **OV** Estimates of overlapped areas relative to each pair of distributions.
- **xpoints** List of abscissas of intersection points among the density curves.
Note

Call function `final.plot`.

Author(s)

Massimiliano Pastore

References


Examples

```r
set.seed(20150605)
x <- list(x1=rnorm(100), x2=rt(50,8), x3=rchisq(80,2))
out <- overlap(x, plot=TRUE)
out$OV

# including boundaries
x <- list(x1=runif(100), x2=runif(100,.5,1))
boundaries <- list( from = c(0,.5), to = c(1,1) )
out <- overlap(x, plot=TRUE, boundaries=boundaries)
out$OV

# equal boundaries
x <- list(x1=runif(100), x2=runif(100), x3=runif(100))
boundaries <- list( from = 0, to = 1 )
out <- overlap(x, plot=TRUE, boundaries=boundaries)
out$OV

# changing kernel
out <- overlap(x, plot=TRUE, kernel="rectangular")
out$OV
```
Index

*Topic utility
  boot.overlap, 2
  cutnumeric, 3
  final.plot, 4
  overlap, 5

boot.overlap, 2

cut, 4
cutnumeric, 3
density, 5
final.plot, 4, 6
overlap, 2, 4, 5