

Package ‘DeBoinR’

December 19, 2023

Type Package

Title Box-Plots and Outlier Detection for Probability Density Functions

Version 1.0

Date 2023-12-15

Maintainer Alexander C. Murph <murph@lanl.gov>

Description Orders a data-set consisting of an ensemble of probability density functions on the same x-grid. Visualizes a box-plot of these functions based on the notion of distance determined by the user. Reports outliers based on the distance chosen and the scaling factor for an interquartile range rule. For further details, see: Alexander C. Murph et al. (2023). “Visualization and Outlier Detection for Probability Density Function Ensembles.” <<https://sirmurphalot.github.io/publications>>.

Copyright file COPYRIGHT

License MIT + file LICENSE

Imports parallel (>= 3.6.2), KernSmooth, ggplot2, gridExtra, pracma, stats, dplyr, graphics

Depends R (>= 3.5.0)

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

NeedsCompilation no

Author Alexander C. Murph [aut, cre] (<<https://orcid.org/0000-0001-7170-867X>>), Justin D. Strait [aut] (<<https://orcid.org/0000-0003-4356-9443>>)

Repository CRAN

Date/Publication 2023-12-19 17:00:11 UTC

R topics documented:

deboinr	2
pdf_data	3
print.DeBoinR	4

deboinr	<i>Orders a data-set consisting of probability density functions on the same x-grid. Visualizes a boxplot of these functions based on the notion of distance determined by the user. Reports outliers based on the distance chosen and k value.</i>
---------	---

Description

Orders a data-set consisting of probability density functions on the same x-grid. Visualizes a boxplot of these functions based on the notion of distance determined by the user. Reports outliers based on the distance chosen and k value.

Usage

```
deboinr(
  x_grid,
  densities_matrix,
  distance = c("hellinger", "nLQD", "fisher_rao", "TV_dist", "CLR", "wasserstein",
    "BD_fboxplot", "MBD_fboxplot", "user_defined"),
  median_type = c("cross", "geometric"),
  center_PDFs = FALSE,
  user_dist = NULL,
  k = 1.5,
  num_cores = 1
)
```

Arguments

x_grid	Vector. X values of the PDF
densities_matrix	Matrix. A n x p matrix where rows are individual PDFs and p matches the length of x_grid.
distance	Character. The distance metric to use for the pairwise distances, or one of the two band depth options.
median_type	Character. Whether the cross-median or the geometric median should be used.
center_PDFs	Logical. Whether or not the modes of all the PDFs should be aligned prior to performing any calculations.
user_dist	R Function. User-defined function that takes in two PDFs as vectors and returns a non-negative float corresponding to a distance between them.
k	Float. The factor by which to expand the IQR when calculating outliers.
num_cores	Integer. The number of cores to use if parallelizing the distance matrix calculations.

Value

An `deboinr` object containing the following:

- `density_order`. Vector of indices corresponding to rows of `densities_matrix` that sort from closest to furthest from the median PDF.
- `outliers`. Vector of indices corresponding to rows of `densities_matrix` that are determined to be outliers.
- `box_plot`. `ggplot` object of graphic output by calling this method.

Examples

```
example_data = DeBoinR::pdf_data[1:100,]
xx = deboinr(DeBoinR::x_grid,
             as.matrix(example_data),
             distance = "hellinger",
             median_type = 'cross',
             center_PDFs = TRUE,
             num_cores = 1
            )

print("about to print DeBoinR object...")
print(xx)
```

pdf_data

Simulated PDF data.

Description

Data simulated using the the `dfnWorks` suite.

Usage

pdf_data

x_grid

Format

‘pdf_data’ is an $n \times p$ matrix, where n is the number of PDFs and p matches the length of `x_grid`. `x_grid` contains the points at which the PDFs are evaluated (assumed equally spaced apart).

Details

‘pdf_data’ is a data frame with 1,000 rows and 5 columns. ‘x_grid’; is a timestamp of each of ‘full_data’'s 1,000 rows.

Examples

```
pdf_data  
x_grid
```

print.DeBoinR	<i>Print function for a DeBoinR object. Prints ggplot graphs and other output values.</i>
---------------	---

Description

Print function for a DeBoinR object. Prints ggplot graphs and other output values.

Arguments

x	deboinr object. Fit from DeBoinR main method.
...	Additional plotting arguments.

Index

* **datasets**

pdf_data, 3

deboinr, 2

pdf_data, 3

print.DeBoinR, 4

x_grid(pdf_data), 3