Package ‘histmdl’

May 23, 2018

Type Package
Title A Most Informative Histogram-Like Model
Version 0.7-1
Date 2018-05-23
Author Jouke Witteveen
Maintainer Jouke Witteveen <j.witteveen@gmail.com>
Description Using the MDL principle, it is possible to estimate parameters for a histogram-like model. The package contains the implementation of such an estimation method.
Imports graphics
License GPL (>= 2)
ByteCompile yes
NeedsCompilation yes
Repository CRAN
Date/Publication 2018-05-23 15:39:13 UTC

R topics documented:

histmdl ................................................................. 1

Index

histmdl  Most Informative Histograms

Description

Compute an estimate of the maximum likelihood parameter of a histogram-like model. If plot=TRUE, the resulting object of class "histogram" is plotted according to plot.histogram, before it is returned.
Usage

```
histmdl(x, model = "Witteveen", gain = 0, precision = 0, support = 4,
        plot = TRUE, main = paste("Histogram of", xname),
        xlab = xname, ylab = "Density", ...)
```

Arguments

- **x**: a vector of values for which the histogram is desired.
- **model**: a character string naming the desired histogram-like model. Currently, only "Witteveen" is implemented.
- **gain**: minimum required complexity reduction before an additional interval is accepted.
- **precision**: a value giving the minimum resolution of the data. When computing complexities, boundary values are blurred by this amount to mitigate the effects of, for example, rounding.
- **support**: minimum number of data points desired per interval.
- **plot**: logical. If TRUE (default), a histogram is plotted, otherwise a list of breaks and densities is returned.
- **main**, **xlab**, **ylab**: these arguments to `title` have useful defaults here.
- **...**: further arguments and graphical parameters passed to `plot.histogram` and thence to `title` and `axis` (if plot=TRUE).

Value

an object of class "histogram" which is a list of components:

- **breaks**: the boundaries of intervals. Note that consecutive values are not the same as the boundaries of the intervals that, recursively, define the model instance.
- **density**: densities of the data inside the model-interval that a section is part of.
- **xname**: a character string with the actual x argument name.

Author(s)

Jouke Witteveen and Richard Gill

See Also

`hist`

Examples

```
set.seed(28011988)

x <- c(rnorm(1000, -6), rnorm(1000, 6))
histmdl(x, gain=2, col="peru")
hist(x, freq=FALSE, add=TRUE, col="#80808080")
```
```r
x <- c(runif(50), runif(50, max=3))
histmdl(x, col="peru", ylim=0:1)
hist(x, freq=FALSE, add=TRUE, col="#00080800")
```
Index

*Topic distribution
  histmdl, 1
*Topic dplot
  histmdl, 1
*Topic histogram
  histmdl, 1
*Topic hplot
  histmdl, 1

axis, 2

class, 1

graphical parameters, 2

hist, 2
histmdl, 1

plot.histogram, 1, 2

title, 2