

Package: benchden (via r-universe)

August 31, 2024

Type Package

Title 28 Benchmark Densities from Berlinet/Devroye (1994)

Version 1.0.8

Date 2023-08-31

Maintainer Thoralf Mildenerger <mild@zhaw.ch>

Description Full implementation of the 28 distributions introduced as benchmarks for nonparametric density estimation by Berlinet and Devroye (1994) <<https://hal.science/hal-03659919>>. Includes densities, cdfs, quantile functions and generators for samples as well as additional information on features of the densities. Also contains the 4 histogram densities used in Rozenholc/Mildenerger/Gather (2010) <[doi:10.1016/j.csda.2010.04.021](https://doi.org/10.1016/j.csda.2010.04.021)>.

License GPL (>= 2)

URL <https://github.com/thmild/benchden>

Repository <https://thmild.r-universe.dev>

RemoteUrl <https://github.com/thmild/benchden>

RemoteRef HEAD

RemoteSha 4334cbcac30e2502ee40191cb546f13ebe474dc1

Contents

| | |
|-------------------|----|
| bberdev | 2 |
| berdev | 3 |
| bhisto | 5 |
| dberdev | 7 |
| dhisto | 9 |
| histo | 11 |

| | |
|--------------|-----------|
| Index | 13 |
|--------------|-----------|

bberdev

Some properties of 28 benchmark densities

Description

Names and points of nonsmoothness for the 28 distributions from Berlinet/Devroye (1994).

Usage

```
bberdev(dnum = 1)
nberdev(dnum = 1)
```

Arguments

dnum number of distribution as in Berlinet/Devroye (1994), Section 3.2.

Details

These functions implement the 28 distributions from Berlinet and Devroye (1994), Section 3.2, which are:

dnum == 1 "uniform" on [0,1] as in stats-package

dnum == 2 "exponential" as in stats-package

dnum == 3 "Maxwell"

dnum == 4 "double exponential"

dnum == 5 "logistic" as in stats-package

dnum == 6 "Cauchy" as in stats-package

dnum == 7 "extreme value"

dnum == 8 "infinite peak"

dnum == 9 "Pareto"

dnum == 10 "symmetric Pareto"

dnum == 11 "normal" as in stats-package

dnum == 12 "lognormal"

dnum == 13 "uniform scale mixture"

dnum == 14 "Matterhorn"

dnum == 15 "logarithmic peak"

dnum == 16 "isosceles triangle"

dnum == 17 "beta 2,2" as in stats-package

dnum == 18 "chi-square 1" as in stats-package

dnum == 19 "normal cubed"

dnum == 20 "inverse exponential"

```

dnum == 21 "Marronite"
dnum == 22 "skewed bimodal"
dnum == 23 "claw"
dnum == 24 "smooth comb"
dnum == 25 "caliper"
dnum == 26 "trimodal uniform"
dnum == 27 "sawtooth"
dnum == 28 "bilogarithmic peak"

```

Value

| | |
|---------|---|
| nberdev | gives the name of the distribution (the same as name in berdev). |
| bberdev | Since evaluation of loss functions in nonparametric density estimation often requires numerical integration, bberdev returns a vector of points you should generally take care not to integrate over, e.g. points where the density is not continuous or not differentiable (gives the same as breaks in berdev). |

Author(s)

Thoralf Mildenerger, Henrike Weinert and Sebastian Tiemeyer

References

A. Berlinet and L. Devroye, "A comparison of kernel density estimates", *Publications de l'Institut de Statistique de l'Universite de Paris*, vol. 38(3), pp. 3-59, 1994. <https://hal.science/hal-03659919>

T. Mildenerger and H. Weinert, "The benchden Package: Benchmark Densities for Nonparametric Density Estimation", *Journal of Statistical Software*, vol. 46(14), 1-14, 2012. <https://www.jstatsoft.org/v46/i14/>

Examples

```

# name of "Claw"-distribution
nberdev(dnum=23)

```

| | |
|--------|--|
| berdev | <i>Some Properties of 28 benchmark densities</i> |
|--------|--|

Description

Name, position of modes, support and points of nonsmoothness for the 28 distributions from Berlinet/Devroye (1994).

Usage

```
berdev(dnum = 1)
```

Arguments

dnum number of distribution as in Berlinet/Devroye (1994), Section 3.2.

Details

These functions implement the 28 distributions from Berlinet and Devroye (1994), Section 3.2, which are:

dnum == 1 "uniform" on [0,1] as in stats-package

dnum == 2 "exponential" as in stats-package

dnum == 3 "Maxwell"

dnum == 4 "double exponential"

dnum == 5 "logistic" as in stats-package

dnum == 6 "Cauchy" as in stats-package

dnum == 7 "extreme value"

dnum == 8 "infinite peak"

dnum == 9 "Pareto"

dnum == 10 "symmetric Pareto"

dnum == 11 "normal" as in stats-package

dnum == 12 "lognormal"

dnum == 13 "uniform scale mixture"

dnum == 14 "Matterhorn"

dnum == 15 "logarithmic peak"

dnum == 16 "isosceles triangle"

dnum == 17 "beta 2,2" as in stats-package

dnum == 18 "chi-square 1" as in stats-package

dnum == 19 "normal cubed"

dnum == 20 "inverse exponential"

dnum == 21 "Marronite"

dnum == 22 "skewed bimodal"

dnum == 23 "claw"

dnum == 24 "smooth comb"

dnum == 25 "caliper"

dnum == 26 "trimodal uniform"

dnum == 27 "sawtooth"

dnum == 28 "bilogarithmic peak"

Value

berdev returns a list with components

| | |
|---------|--|
| name | gives the name of the distribution, |
| peaks | gives a vector of the positions of peaks or modes of the density, and |
| support | gives a matrix as follows: in each row an interval is defined (with the first column giving the left and the second column the right end of the interval). Together the intervals give the support of the distribution (for most distributions only one interval). |
| breaks | Since evaluation of loss functions in nonparametric density estimation often requires numerical integration, bberdev returns a vector of points you should generally take care not to integrate over, e.g. points where the density is not continuous or not differentiable. |

Author(s)

Thoralf Mildenerger, Henrike Weinert and Sebastian Tiemeyer

References

A. Berlinet and L. Devroye, "A comparison of kernel density estimates", Publications de l'Institut de Statistique de l'Universite de Paris, vol. 38(3), pp. 3-59, 1994. <https://hal.science/hal-03659919>

T. Mildenerger and H. Weinert, "The benchden Package: Benchmark Densities for Nonparametric Density Estimation", Journal of Statistical Software, vol. 46(14), 1-14, 2012. <https://www.jstatsoft.org/v46/i14/>

Examples

```
# position of peaks of "Claw"-distribution
berdev(dnum=23)$peaks
```

```
# support of the "Trimodal uniform"
berdev(dnum=26)$support
```

bhisto

Some properties of 4 histogram benchmark densities

Description

Names and breakpoints for the 4 histogram benchmark distributions from Rozenholc/Mildenerger/Gather (2010).

Usage

```
bhisto(dnum = 1)
nhisto(dnum = 1)
```

Arguments

dnum number of distribution.

Details

These functions implement the 4 histogram benchmark distributions from Rozenholc/Mildenberger/Gather (2010). Defined as the following mixtures of uniform distributions:

dnum == 1 5 bin regular histogram:

$$0.15 * U[0, 0.2] + 0.35 * U(0.2, 0.4) + 0.2 * U(0.4, 0.6) + 0.1 * U(0.6, 0.8) + 0.2 * U(0.8, 1.0]$$

dnum == 2 5 bin irregular histogram:

$$0.15 * U[0, 0.13] + 0.35 * U(0.13, 0.34] + 0.2 * U(0.34, 0.61] + 0.1 * U(0.61, 0.65] + 0.2 * U(0.65, 1.0]$$

dnum == 3 10 bin regular histogram:

$$\begin{aligned} &0.01 * U[0, 0.1] + 0.18 * U(0.1, 0.2] + 0.16 * U(0.2, 0.3] \\ &+ 0.07 * U(0.3, 0.4] + 0.06 * U(0.4, 0.5] + 0.01 * U(0.5, 0.6] \\ &+ 0.06 * U(0.6, 0.7] + 0.37 * U(0.7, 0.8] + 0.06 * U(0.8, 0.9] \\ &\quad + 0.02 * U(0.9, 1.0] \end{aligned}$$

dnum == 4 10 bin irregular histogram:

$$\begin{aligned} &0.01 * U[0, 0.02] + 0.18 * U(0.02, 0.07] + 0.16 * U(0.07, 0.14] \\ &+ 0.07 * U(0.14, 0.44] + 0.06 * U(0.44, 0.53] + 0.01 * U(0.53, 0.56] \\ &+ 0.06 * U(0.56, 0.67] + 0.37 * U(0.67, 0.77] + 0.06 * U(0.77, 0.91] \\ &\quad + 0.02 * U(0.91, 1.0] \end{aligned}$$

where $U[a, b]$ denotes the uniform distribution on $[a, b]$.

Value

nhisto gives the name of the distribution (the same as name in histo).
bhisto gives the vector of break points (the same as breaks in histo).

Author(s)

Thoralf Mildenberger

References

- T. Mildenberger and H. Weinert, "The benchden Package: Benchmark Densities for Nonparametric Density Estimation", Journal of Statistical Software, vol. 46(14), 1-14, 2012. <https://www.jstatsoft.org/v46/i14/>
- Y. Rozenholc, T. Mildenberger and U. Gather (2010), "Combining Regular and Irregular Histograms by Penalized Likelihood", Computational Statistics and Data Analysis, 54, 3313-3323. doi:10.1016/j.csda.2010.04.021 Earlier version including explicit definition of the densities: doi:10.17877/DE290R15901

Examples

```
# name string of 5 bin regular histogram  
nhisto(dnum=1)
```

| | |
|---------|--|
| dberdev | <i>28 benchmark densities from Berlinet/Devroye (1994)</i> |
|---------|--|

Description

Density, distribution function, quantile function and random variate generation for the 28 distributions from Berlinet/Devroye (1994).

Usage

```
dberdev(x, dnum = 1)  
pberdev(q, dnum = 1)  
qberdev(p, dnum = 1)  
rberdev(n, dnum = 1)
```

Arguments

| | |
|------|--|
| dnum | number of distribution as in Berlinet/Devroye (1994), Section 3.2. |
| x, q | vector of quantiles. |
| p | vector of probabilities. |
| n | number of observations. |

Details

These functions implement the 28 distributions from Berlinet and Devroye (1994), Section 3.2, which are:

```
dnum == 1 "uniform" on [0,1] as in stats-package  
dnum == 2 "exponential" as in stats-package  
dnum == 3 "Maxwell"  
dnum == 4 "double exponential"  
dnum == 5 "logistic" as in stats-package  
dnum == 6 "Cauchy" as in stats-package  
dnum == 7 "extreme value"  
dnum == 8 "infinite peak"  
dnum == 9 "Pareto"  
dnum == 10 "symmetric Pareto"  
dnum == 11 "normal" as in stats-package
```

dnum == 12 "lognormal"
dnum == 13 "uniform scale mixture"
dnum == 14 "Matterhorn"
dnum == 15 "logarithmic peak"
dnum == 16 "isosceles triangle"
dnum == 17 "beta 2,2" as in stats-package
dnum == 18 "chi-square 1" as in stats-package
dnum == 19 "normal cubed"
dnum == 20 "inverse exponential"
dnum == 21 "Marronite"
dnum == 22 "skewed bimodal"
dnum == 23 "claw"
dnum == 24 "smooth comb"
dnum == 25 "caliper"
dnum == 26 "trimodal uniform"
dnum == 27 "sawtooth"
dnum == 28 "bilogarithmic peak"

Value

| | |
|---------|----------------------------------|
| dberdev | gives the density, |
| pberdev | gives the distribution function, |
| qberdev | gives the quantile function, and |
| rberdev | generates random deviates. |

Acknowledgement

The authors thank Luc Devroye for providing his original implementation for testing purposes.

Author(s)

Thoralf Mildenerger, Henrike Weinert and Sebastian Tiemeyer

References

- A. Berlinet and L. Devroye, "A comparison of kernel density estimates," Publications de l'Institut de Statistique de l'Universite de Paris, vol. 38(3), pp. 3-59, 1994. <https://hal.science/hal-03659919>
- T. Mildenerger and H. Weinert, "The benchden Package: Benchmark Densities for Nonparametric Density Estimation", Journal of Statistical Software, vol. 46(14), 1-14, 2012. <https://www.jstatsoft.org/v46/i14/>

Examples

```
# histogram and true density of "Claw"-distribution
hist(rberdev(1000,dnum=23),breaks=100, main = " ",freq=FALSE)
lines(seq(-3,3,0.01),dberdev(seq(-3,3,0.01),dnum=23),col="blue",lwd=2)
title(paste(nberdev(dnum=23)))

# plot cdf of simulated data and the df of "Matterhorn"-distribution
plot.stepfun(rberdev(100,dnum=14),do.points=TRUE,main="")
lines(seq(-1,1,0.001),pberdev(seq(-1,1,0.001),dnum=14),col="blue")
title(paste(nberdev(dnum=14)))

# plot quantiles of "smooth comb"-distribution
plot(qberdev(seq(0,1,0.01),dnum=24),t="l")
title(paste(nberdev(dnum=24)))
```

dhisto

4 histogram benchmark densities

Description

Density, distribution function, quantile function and random variate generation for the 4 histogram benchmark distributions from Rozenholc/Mildenberger/Gather (2010).

Usage

```
dhisto(x,dnum = 1)
phisto(q,dnum = 1)
qhisto(p,dnum = 1)
rhisto(n,dnum = 1)
```

Arguments

| | |
|------|---|
| dnum | number of distribution as in Rozenholc/Mildenberger/Gather (2010) |
| x, q | vector of quantiles. |
| p | vector of probabilities. |
| n | number of observations. |

Details

These functions implement the 4 histogram benchmark distributions from Rozenholc/Mildenberger/Gather (2010). Defined as the following mixtures of uniform distributions:

dnum == 1 5 bin regular histogram:

$$0.15 * U[0, 0.2] + 0.35 * U(0.2, 0.4] + 0.2 * U(0.4, 0.6] + 0.1 * U(0.6, 0.8] + 0.2 * U(0.8, 1.0]$$

dnum == 2 5 bin irregular histogram:

$$0.15 * U[0, 0.13] + 0.35 * U(0.13, 0.34] + 0.2 * U(0.34, 0.61] + 0.1 * U(0.61, 0.65] + 0.2 * U(0.65, 1.0]$$

dnum == 3 10 bin regular histogram:

$$\begin{aligned} &0.01 * U[0, 0.1] + 0.18 * U(0.1, 0.2] + 0.16 * U(0.2, 0.3] \\ &+ 0.07 * U(0.3, 0.4] + 0.06 * U(0.4, 0.5] + 0.01 * U(0.5, 0.6] \\ &+ 0.06 * U(0.6, 0.7] + 0.37 * U(0.7, 0.8] + 0.06 * U(0.8, 0.9] \\ &+ 0.02 * U(0.9, 1.0] \end{aligned}$$

dnum == 4 10 bin irregular histogram:

$$\begin{aligned} &0.01 * U[0, 0.02] + 0.18 * U(0.02, 0.07] + 0.16 * U(0.07, 0.14] \\ &+ 0.07 * U(0.14, 0.44] + 0.06 * U(0.44, 0.53] + 0.01 * U(0.53, 0.56] \\ &+ 0.06 * U(0.56, 0.67] + 0.37 * U(0.67, 0.77] + 0.06 * U(0.77, 0.91] \\ &+ 0.02 * U(0.91, 1.0] \end{aligned}$$

where $U[a, b]$ denotes the uniform distribution on $[a, b]$.

Value

| | |
|--------|----------------------------------|
| dhisto | gives the density, |
| phisto | gives the distribution function, |
| qhisto | gives the quantile function, and |
| rhisto | generates random deviates. |

Author(s)

Thoralf Mildenerger

References

- T. Mildenerger and H. Weinert, "The benchden Package: Benchmark Densities for Nonparametric Density Estimation", Journal of Statistical Software, vol. 46(14), 1-14, 2012. <https://www.jstatsoft.org/v46/i14/>
- Y. Rozenholc, T. Mildenerger and U. Gather (2010), "Combining Regular and Irregular Histograms by Penalized Likelihood", Computational Statistics and Data Analysis, 54, 3313-3323. [doi:10.1016/j.csda.2010.04.021](https://doi.org/10.1016/j.csda.2010.04.021) Earlier version including explicit definition of the densities: [doi:10.17877/DE290R15901](https://doi.org/10.17877/DE290R15901)

Examples

```
# histogram and true density of "5 bin irregular"-distribution
hist(rhisto(2000,dnum=2),breaks=250, main = " ",freq=FALSE)
lines(seq(0,1,0.01),dhisto(seq(0,1,0.01),dnum=2),col="blue",lwd=1)
title(paste("sample from",nhisto(dnum=2),"density"))
```

histo

Some properties of 4 histogram benchmark densities

Description

Name, position of modes, support and break points for the 4 histogram benchmark distributions from Rozenholc/Mildenberger/Gather (2010).

Usage

```
histo(dnum = 1)
```

Arguments

dnum number of distribution.

Details

These functions implement the 4 histogram benchmark distributions from Rozenholc/Mildenberger/Gather (2010). Defined as the following mixtures of uniform distributions:

dnum == 1 5 bin regular histogram:

$$0.15 * U[0, 0.2] + 0.35 * U(0.2, 0.4] + 0.2 * U(0.4, 0.6] + 0.1 * U(0.6, 0.8] + 0.2 * U(0.8, 1.0]$$

dnum == 2 5 bin irregular histogram:

$$0.15 * U[0, 0.13] + 0.35 * U(0.13, 0.34] + 0.2 * U(0.34, 0.61] + 0.1 * U(0.61, 0.65] + 0.2 * U(0.65, 1.0]$$

dnum == 3 10 bin regular histogram:

$$\begin{aligned} &0.01 * U[0, 0.1] + 0.18 * U(0.1, 0.2] + 0.16 * U(0.2, 0.3] \\ &+ 0.07 * U(0.3, 0.4] + 0.06 * U(0.4, 0.5] + 0.01 * U(0.5, 0.6] \\ &+ 0.06 * U(0.6, 0.7] + 0.37 * U(0.7, 0.8] + 0.06 * U(0.8, 0.9] \\ &+ 0.02 * U(0.9, 1.0] \end{aligned}$$

dnum == 4 10 bin irregular histogram:

$$\begin{aligned} &0.01 * U[0, 0.02] + 0.18 * U(0.02, 0.07] + 0.16 * U(0.07, 0.14] \\ &+ 0.07 * U(0.14, 0.44] + 0.06 * U(0.44, 0.53] + 0.01 * U(0.53, 0.56] \\ &+ 0.06 * U(0.56, 0.67] + 0.37 * U(0.67, 0.77] + 0.06 * U(0.77, 0.91] \\ &+ 0.02 * U(0.91, 1.0] \end{aligned}$$

where $U[a, b]$ denotes the uniform distribution on $[a, b]$.

Value

histo returns a list with the following components:

| | |
|---------|--|
| name | gives the name of the distribution. |
| peaks | gives a vector of the positions of peaks of the density, defined here as mid points of maximal intervals. |
| support | gives a matrix with one row with the endpoints of the support, which is $[0, 1]$ for all four histogram densities. |
| breaks | gives the vector of break points. |

Author(s)

Thoralf Mildenerger

References

- T. Mildenerger and H. Weinert, "The benchden Package: Benchmark Densities for Nonparametric Density Estimation", Journal of Statistical Software, vol. 46(14), 1-14, 2012. <https://www.jstatsoft.org/v46/i14/>
- Y. Rozenholc, T. Mildenerger and U. Gather (2010), "Combining Regular and Irregular Histograms by Penalized Likelihood", Computational Statistics and Data Analysis, 54, 3313-3323. [doi:10.1016/j.csda.2010.04.021](https://doi.org/10.1016/j.csda.2010.04.021) Earlier version including explicit definition of the densities: [doi:10.17877/DE290R15901](https://doi.org/10.17877/DE290R15901)

Examples

```
# position of peaks of the 5 bin irregular histogram density
histo(dnum=2)$peaks

# support of the 10 bin regular histogram density
histo(dnum=3)$support
```

Index

* datagen

bberdev, 2
berdev, 3
bhisto, 5
dberdev, 7
dhisto, 9
histo, 11

* distribution

bberdev, 2
berdev, 3
bhisto, 5
dberdev, 7
dhisto, 9
histo, 11

* nonparametric

bberdev, 2
berdev, 3
bhisto, 5
dberdev, 7
dhisto, 9
histo, 11

bberdev, 2
berdev, 3
bhisto, 5

dberdev, 7
dhisto, 9

histo, 11

nberdev (bberdev), 2
nhisto (bhisto), 5

pberdev (dberdev), 7
phisto (dhisto), 9

qberdev (dberdev), 7
qhisto (dhisto), 9

rberdev (dberdev), 7
rhisto (dhisto), 9