

Package ‘kerdiest’

February 20, 2015

Type Package

Title Nonparametric kernel estimation of the distribution function.
Bandwidth selection and estimation of related functions.

Version 1.2

Date 2012-08-11

Author Alejandro Quintela del Rio and Graciela Estevez Perez

Maintainer Alejandro Quintela del Rio <aquintela@udc.es>

Description Nonparametric kernel distribution function estimation is performed. Three automatic bandwidth selection methods for nonparametric kernel distribution function estimation are implemented: the plug-in of Altman and Leger, the plug-in of Polansky and Baker, and the modified cross-validation of Bowman, Hall and Prvan. The exceedance function, the mean return period and the return level are also computed.

License GPL (>= 2)

LazyLoad yes

Depends date, chron, evir

Repository CRAN

Date/Publication 2012-08-13 07:57:06

NeedsCompilation no

R topics documented:

kerdiest-package	2
ALbw	2
CVbw	4
ef	5
kde	7
mrp	9
nwip	11
PBbw	12
rl	14
saltriver	15

Index**17**

kerdiest-package	<i>Nonparametric kernel estimation of the distribution. Bandwidth selection methods and estimation of related functions.</i>
------------------	--

Description

Nonparametric kernel distribution function estimation for continuous random variables is performed. Three automatic bandwidth selection procedures for nonparametric kernel distribution function estimation are implemented: the plug-in method of Altman and Leger, the plug-in method of Polansky and Baker, and the modified cross-validation method of Bowman, Hall and Prvan. The exceedance function, the mean return period and the return level are also computed.

Author(s)

Graciela Estevez Perez <graci@udc.es> and Alejandro Quintela del Rio <aquintela@udc.es>
 Maintainer: Alejandro Quintela del Rio <aquintela@udc.es>

References

Quintela-del-Rio, A. and Estevez-Perez, G. (2012) Nonparametric Kernel Distribution Function Estimation with kerdiest: An R Package for Bandwidth Choice and Applications, *Journal of Statistical Software* **50(8)**, pp. 1-21. URL <http://www.jstatsoft.org/v50/i08/>.

ALbw	<i>Computes the plug-in bandwidth of Altman and Leger.</i>
------	--

Description

The bandwidth parameter for the distribution function kernel estimator is calculated, using the plug-in method of Altman and Leger (1995). Four possible kernel functions can be used for the kernel estimator: "e" Epanechnikov, "n" Normal, "b" Biweight and "t" Triweight.

Usage

```
ALbw(type_kernel = "n", vec_data)
```

Arguments

type_kernel	The kernel function. You can use four types: "e" Epanechnikov, "n" Normal, "b" Biweight and "t" Triweight. The Normal kernel is used by default.
vec_data	The data sample.

Details

Altman and Leger (1995) recommend the use of the Epanechnikov kernel, because in this case the rate of convergence for the kernel derivative estimator is improved. For the sake of uniformity along the package, the gaussian kernel is used by default, but the user can obviously choose the Epanechnikov function.

Value

A real value for the bandwidth parameter.

Author(s)

Graciela Estevez Perez <graci@udc.es> and Alejandro Quintela del Rio <aquintela@udc.es>

References

Altman, N., Leger, C. (1995) Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference* **46**, pp. 195–214.

Quintela-del-Rio, A. and Estevez-Perez, G. (2012) Nonparametric Kernel Distribution Function Estimation with kerdier: An R Package for Bandwidth Choice and Applications, *Journal of Statistical Software* **50(8)**, pp. 1-21. URL <http://www.jstatsoft.org/v50/i08/>.

Examples

```
# Compute the plug-in bandwidth for a sample of 100 random N(0,1) data
x<-rnorm(100,0,1)
h_AL<- ALbw(type_kernel="e",vec_data=x)
h_AL
## Not run:
## A Quick plot of a distribution function estimate
x<-rnorm(1000)
h_AL<-ALbw(vec_data=x)
F_AL<-kde(vec_data=x, bw=h_AL)
plot(F_AL$grid,F_AL$Estimated_values,type="l")

## End(Not run)
## Not run:
## Plotting the distribution function estimate controlling the grid points
# and the kernel function
ss <- quantile(x, c(0.05, 0.95))
# number of points to be used in the representation of estimated distribution
# function
n_pts <- 100
y <- seq(ss[1],ss[2],length.out=n_pts)
F_AL <- kde(type_kernel="e", x, y, h_AL)$Estimated_values
## plot of the theoretical and estimated distribution functions
require(graphics)
plot(y,F_AL, type="l", lty=2)
lines(y, pnorm(y),type="l", lty=1)
legend(-1,0.8,c("real","nonparametric"),lty=1:2)
```

```
## End(Not run)
```

CVbw

Computes the cross-validation bandwidth of Bowman et al (1998).

Description

The bandwidth parameter for the distribution function kernel estimator is calculated, using the modified cross-validation method of Bowman, Hall and Prvan (1998). Four possible kernel functions can be used: "e" Epanechnikov, "n" Normal, "b" Biweight and "t" Triweight. The cross-validation function involves an integral term, that is calculated using the Simpson's rule.

Usage

```
CVbw(type_kernel = "n", vec_data, n_pts = 100, seq_bws =NULL)
```

Arguments

type_kernel	The kernel function. You can use four types: "e" Epanechnikov, "n" Normal, "b" Biweight and "t" Triweight. The normal kernel is used by default.
vec_data	The data sample.
n_pts	The number of points used to approximate, by the Simpson's rule, the integral term, in the cross-validation function. Because this numeric method enlarge the computing time, you can check different numbers, depending on your sample size. 100 points are used by default.
seq_bws	The sequence of bandwidths in which to compute the cross-validation function. By default, the procedure defines a sequence of 50 points, from the range of the data divided by 200 to the range divided by 2.

Value

A list consisting of

seq_bws	The sequence of bandwidths.
CV function	The values of the cross-validation function in the bandwidths grid.
bw	A real value for the cross-validation bandwidth.

Author(s)

Graciela Estevez Perez <graci@udc.es> and Alejandro Quintela del Rio <aquintela@udc.es>

References

Bowman, A.W., Hall, P. and Prvan,T. (1998) Cross-validation for the smoothing of distribution functions, *Biometrika* **85**, pp. 799-808.

Quintela-del-Rio, A. and Estevez-Perez, G. (2012) Nonparametric Kernel Distribution Function Estimation with kerdieat: An R Package for Bandwidth Choice and Applications, *Journal of Statistical Software* **50(8)**, pp. 1-21. URL <http://www.jstatsoft.org/v50/i08/>.

Examples

```
## Compute the cross-validation bandwidth for a sample of 100 random N(0,1) data
x<-rnorm(100,0,1)
num_bws <- 50
seq_bws <- seq(((max(x)-min(x))/2)/50, (max(x)-min(x))/2, length=num_bws)
hCV <- CVbw(type_kernel="e", vec_data=x, n_pts=200, seq_bws=seq_bws)
hCV
## The CV function is plotted
h_CV<-CVbw(vec_data=x)
h_CV$bw
plot(h_CV$seq_bws, h_CV$CVfunction, type="l")
## Not run:
## Plotting the distribution function estimate controlling the grid points
## and the kernel function
ss <- quantile(x, c(0.05, 0.95))
# number of points to be used in the representation of estimated distribution
# function
n_pts <- 100
y <- seq(ss[1],ss[2],length.out=n_pts)
F_CV<-kde(type_kernel="e", x, y, h_CV$bw)$Estimated_values
## plot of the theoretical and estimated distribution functions
require(graphics)
plot(y,F_CV, type="l", lty=2)
lines(y, pnorm(y),type="l", lty=1)
legend(-1,0.8,c("real", "nonparametric"),lty=1:2)

## End(Not run)
```

 ef

Exceedance function estimation

Description

We compute the exceedance probability, that is, the probability that a specified value c (a magnitude of a seismic event, a flow level...) will be exceeded in D time units.

Usage

```
ef(type_kernel = "n", vec_data, c,
   bw = PBbw(type_kernel = "n", vec_data, 2),
   Dmin = 0, Dmax = 15, size_grid = 50, lambda)
```

Arguments

type_kernel	The kernel function. You can use four types: "e" Epanechnikov, "n" Normal, "b" Biweight and "t" Triweight. The Normal kernel is used by default.
vec_data	The data sample (earthquake magnitudes, flow levels, wind speed...)
c	The concrete level in which we want to compute the exceedance probability.

bw	The bandwidth parameter. The plug-in method of Polansky and Baker (2000) is used by default.
Dmin	Minimum value for D time units (years, days...). Default is Dmin=0.
Dmax	Maximum value for D time units (years, days...). Default is Dmax=15.
size_grid	Length of a grid in which we compute the exceedance function. The size is 50 by default.
lambda	The mean activity rate.

Details

The exceedance function is usually calculated assuming that the occurrence process of events follows a Poisson one. In this case, the exceedance function, that is, the probability of an specific value c is calculated as

$$R(c, D) = 1 - \exp(-\lambda D(1 - F_h(c))).$$

See, for example, Orlecka-Sikora (2008) or Quintela del Rio (2010) for earthquake data applications.

Value

Returns a list containing:

Estimated_values	Vector containing the estimated function.
grid	The used grid.
bw	Value of the bandwidth.

Author(s)

Graciela Estevez Perez <graci@udc.es> and Alejandro Quintela del Rio <aquintela@udc.es>

References

Orlecka-Sikora, B. (2008) Resampling methods for evaluating the uncertainty of the nonparametric magnitude distribution estimation in the probabilistic seismic hazard analysis. *Tectonophysics* **456**, 38–51.

Quintela-del-Rio, A. (2010) On non-parametric techniques for area-characteristic seismic hazard parameters. *Geophysical Journal International* **180**, pp. 339–346.

Quintela-del-Rio, A. and Estevez-Perez, G. (2012) Nonparametric Kernel Distribution Function Estimation with kerdieft: An R Package for Bandwidth Choice and Applications, *Journal of Statistical Software* **50(8)**, pp. 1-21. URL <http://www.jstatsoft.org/v50/i08/>.

Examples

```
# Working with earthquake data. We use the catalogue of the National
# Geographic Institute (IGN) of Spain and select the data of the Northwest
# of the Iberian Peninsula.
data(nwip)
```

```

require(chron)
require(date)
# we consider the data with magnitude greater than 3
mg<-nwip$magnitude[nwip$magnitude>3.0]
x1<-nwip$year
x2<-nwip$month
x3<-nwip$day
ys<-paste(x1,x2,x3)
earthquake_date<-as.character(ys)
y1s<-as.date(earthquake_date, order = "ymd")
# we compute the total number of years
y2s<-as.POSIXct(y1s)
z<-years(y2s)
n.years<-length(levels(z))
# the mean rate of earthquakes per year
lambda<-length(mg)/n.years
## Not run:
# we estimate the exceedance probability for a value of the
# the magnitude = 4
est<-ef(vec_data=mg, m_c=4, lambda=lambda)
plot(est$grid, est$Estimated_values, type="l",
      xlab="years", ylab="Probability of Exceedance")

## End(Not run)

```

kde

Kernel estimator of the distribution function

Description

Computes the value of the kernel estimator of the distribution function, in a single value or in a grid. Four possibilities for the kernel function are implemented, and the bandwidth parameter can be directly calculated by the plug-in method of Polansky and Baker (2000).

Usage

```
kde(type_kernel = "n", vec_data, y = NULL, bw = PBbw(type_kernel = "n",
vec_data, 2))
```

Arguments

type_kernel	The kernel function. You can use four types: "e" Epanechnikov, "n" Normal, "b" Biweight and "t" Triweight. The Normal kernel is used by default.
vec_data	The data sample.
y	The single value or the grid vector where the distribution function is estimated. By default, a grid of 100 equidistant points from the minimum to the maximum of the data sample is selected.
bw	The bandwidth used. If it is not provided, the Plug-in bandwidth of Polansky and Baker (2000) is computed.

Value

Returns a list containing:

Estimated_values	Vector containing the estimated function in the grid values.
grid	The used grid.
bw	Value of the bandwidth.

Author(s)

Graciela Estevez Perez <graci@udc.es> and Alejandro Quintela del Rio <aquintela@udc.es>

References

Reiss, R.D. (1981) Nonparametric estimation of smooth distribution functions, *Scandinavian Journal of Statistics* **8**, pp:116-119.

Simonoff, J. (1996) *Smoothing Methods in Statistics*, Springer, New York.

Polansky, A.M. and Baker, E.R. (2000) Multistage plug-in bandwidth selection for kernel distribution function estimates, *Journal of Statistical Computation and Simulation* **65**, pp. 63-80.

Quintela-del-Rio, A. and Estevez-Perez, G. (2012) Nonparametric Kernel Distribution Function Estimation with kerdie: An R Package for Bandwidth Choice and Applications, *Journal of Statistical Software* **50(8)**, pp. 1-21. URL <http://www.jstatsoft.org/v50/i08/>.

Examples

```
# Comparison of three bandwidth selection methods

x<-rnorm(100)
# The bandwidths by cross-validation, plug-in of Altman and Leger
# and plug-in of Polansky and Baker are calculated, using a normal kernel and a
# standard setting of parameters, in each case
h_CV<-CVbw(vec_data=x)$bw
# plug-in of Altman and Leger
h_AL<- ALbw(vec_data=x)
# plug-in of Polansky and Baker
h_PB<- PBbw(vec_data=x)
## Not run: print(h_CV); print(h_AL); print(h_PB)
# plot of the three estimates together with the real distribution
F_CV<-kde(vec_data=x, bw= h_CV)
F_AL<-kde(vec_data=x, bw= h_AL)
F_PB<-kde(vec_data=x, bw= h_PB)
y<-F_CV$grid
Ft<-pnorm(y)
require(graphics)
plot(y,Ft, ylab="Distribution", xlab="data", type="l", lty=1)
lines(y,F_CV$Estimated_values, type="l",lty=2)
lines(y,F_AL$Estimated_values, type="l",lty=3)
lines(y,F_PB$Estimated_values, type="l",lty=4)

legend(1,0.4,c("real", "F_CV", "F_AL", "F_PB"),lty=1:4)
```



```
## End(Not run)
```

mrp *Mean return period estimation*

Description

This functions computes an estimate of the time between two values of a concrete level (size of an earthquake, flow level, wind speed...).

Usage

```
mrp(type_kernel = "n", vec_data, y = NULL,
     bw = PBbw(type_kernel = "n", vec_data, 2), lambda)
```

Arguments

type_kernel	The kernel function. You can use four types: "e" Epanechnikov, "n" Normal, "b" Biweight and "t" Triweight. The Normal kernel is used by default.
vec_data	The data sample (earthquake magnitudes, flow levels, wind speed...).
y	A grid or a singular value where the estimator is computed. By default, a grid of 50 values between the minimum and the maximum of the data is computed.
bw	The bandwidth parameter. The plug-in method of Polansky and Baker (2000) is used by default.
lambda	The mean activity rate.

Details

The mean return period is usually calculated assuming that the occurrence process of the events follows a Poisson one. In this case, the mean return period of events of size c is calculated as

$$T(c) = \frac{1}{\lambda(1 - F_h(c))}.$$

In Orlecka-Sikora (2008) or Quintela del Rio (2010) an application to earthquake data is made. In hydrological applications, if we work with annual maxima data, the parameter of the Poisson variable is 1 (one maximum per year). The mean return period between flow levels of value c is calculated as

$$T(c) = \frac{1}{1 - F_h(c)}.$$

See, for instance, Quintela del Rio (2011), for an application to data of Salt River near Roosevelt, AZ, USA (saltriver data).

Value

Returns a list containing:

Estimated_values	Vector containing the estimated function.
grid	The used grid.
bw	Value of the bandwidth.

Author(s)

Graciela Estevez Perez <graci@udc.es> and Alejandro Quintela del Rio <aquintela@udc.es>

References

- Orlecka-Sikora, B. (2008) Resampling methods for evaluating the uncertainty of the nonparametric magnitude distribution estimation in the probabilistic seismic hazard analysis. *Tectonophysics* **456**, 38–51.
- Quintela-del-Rio, A. (2010) On non-parametric techniques for area-characteristic seismic hazard parameters. *Geophysical Journal International* **180**, pp. 339–346.
- Quintela-del-Rio, A. (2011) On bandwidth selection for nonparametric estimation in flood frequency analysis. *Hydrological Processes* **25**, pp. 671–678.
- Quintela-del-Rio, A. and Estevez-Perez, G. (2012) Nonparametric Kernel Distribution Function Estimation with kerdie: An R Package for Bandwidth Choice and Applications, *Journal of Statistical Software* **50(8)**, pp. 1-21. URL <http://www.jstatsoft.org/v50/i08/>.

Examples

```
# Working with earthquake data. We use the catalogue of the National
# Geographic Institute (IGN) of Spain and select the data of the Northwest
# of the Iberian Peninsula.
data(nwip)
require(chron)
require(date)
# we consider the data with magnitude greater than 3
mg<-nwip$magnitude[nwip$magnitude>3.0]
x1<-nwip$year
x2<-nwip$month
x3<-nwip$day
ys<-paste(x1,x2,x3)
earthquake_date<-as.character(ys)
y1s<-as.date(earthquake_date, order = "ymd")
# we compute the total number of years
y2s<-as.POSIXct(y1s)
z<-years(y2s)
n.years<-length(levels(z))
# the mean rate of earthquakes per year
lambda<-length(mg)/n.years
## Not run:
# we estimate the mean return period (in years) between earthquakes of
```

```

# the same magnitude
est2<-mrp(vec_data=mg, lambda=lambda)
plot(est2$grid, est2$Estimated_values, type="l",
      xlab="magnitude", ylab="Mean return period (years)")

## End(Not run)
## Working with hydrological data: annual peak instantaneous flow of the
# Salt River near Roosevelt, AZ, USA, for 1924 - 2006.
data(saltriver)
peak<-saltriver$peakflow
## Not run:
year<-saltriver$year
plot(year,peak, type="l",ylab="Annual peak flow")

## End(Not run)
# mean return period for the Saltriver data
rp<-mrp(type_kernel="n", vec_data=peak, lambda=1)
## Not run:
plot(rp$grid, rp$Estimated_values, type="l",
      xlab="Flow level", ylab="Years ", main="Mean return period")

## End(Not run)

```

nwip

Earthquakes of the Norwesth of the Iberian Peninsula

Description

This data set corresponds to the earthquakes occurred in the Norwesth of the Iberian Peninsula, from 25/November/1924 to 31/July/2010. The area is limited by the coordinates 41 N – 44 N and 6 W – 10 W, involving the autonomic region of Galicia (Spain) and northern Portugal.

Usage

```
data(nwip)
```

Format

A data frame with 3491 observations on the following 10 variables, corresponding to the earthquake epicenters and time of occurrence.

day Day

month Month

year Year

hour Hour

minute Minute

second second

latitude Latitude in degrees
 longitude Longitude in degrees
 depth Depth in km
 magnitude Magnitude in Richter Scale

Source

The data catalogue has been obtained from the National Geographic Institute (IGN) of Spain. The web page is www.ign.es.

References

Rueda, J., and J. Mezcuca (2001) Sismicidad, sismotectonica y peligrosidad sismica en Galicia, *IGN Technical Publication* **35**.

Quintela-del-Rio, A. and Estevez-Perez, G. (2012) Nonparametric Kernel Distribution Function Estimation with kerdienst: An R Package for Bandwidth Choice and Applications, *Journal of Statistical Software* **50(8)**, pp. 1-21. URL <http://www.jstatsoft.org/v50/i08/>.

Examples

```
data(nwip)
```

PBbw

Computes the plug-in bandwidth of Polansky and Baker.

Description

The bandwidth parameter for the distribution function kernel estimator is calculated, using the plug-in method of Polansky and Baker (2000). Four possible kernel functions can be used for the kernel estimator: "e" Epanechnikov, "n" Normal, "b" Biweight and "t" Triweight. Because kernel estimators of derivatives of order bigger than two are required, only the normal kernel is used in this case.

Usage

```
PBbw(type_kernel = "n", vec_data, num_stage = 2)
```

Arguments

type_kernel	The kernel function used. You can use four types: "e" Epanechnikov, "n" Normal, "b" Biweight and "t" Triweight. The kernel normal is used by default.
vec_data	The data sample.
num_stage	The number of iterations in the Polansky and Baker's method. b=2 is usually a good option, and this is the value by default. b=3 or b=4 are also allowed.

Value

A real value for the bandwidth parameter.

Author(s)

Graciela Estevez Perez <graci@udc.es> and Alejandro Quintela del Rio <aquintela@udc.es>

References

Polansky, A.M. and Baker, E.R. (2000) Multistage plug-in bandwidth selection for kernel distribution function estimates, *Journal of Statistical Computation and Simulation* **65**, pp. 63-80.

Quintela-del-Rio, A. and Estevez-Perez, G. (2012) Nonparametric Kernel Distribution Function Estimation with kerdie: An R Package for Bandwidth Choice and Applications, *Journal of Statistical Software* **50(8)**, pp. 1-21. URL <http://www.jstatsoft.org/v50/i08/>.

Examples

```
# Compute the plug-in bandwidth for a sample of 100 random N(0,1) data
x<-rnorm(100,0,1)
h_PB<-PBbw(vec_data=x,num_stage=4)
h_PB
## Not run:
## A Quick plot of a distribution function estimate
x<-rnorm(1000)
h_PB<-PBbw(vec_data=x)
F_PB<-kde(vec_data=x, bw=h_PB)
plot(F_PB$grid, F_PB$Estimated_values, type="l")

## End(Not run)
## Not run:
## Plotting the distribution function estimate controlling the grid points and
## the kernel function
ss <- quantile(x, c(0.05, 0.95))
# number of points to be used in the representation of the estimated
# distribution function
n_pts <- 100
y <- seq(ss[1],ss[2],length.out=n_pts)
F_PB <- kde(type_kernel="e", x, y, h_PB)$Estimated_values
## plot of the theoretical and estimated distribution functions
require(graphics)
plot(y,F_PB, type="l", lty=2)
lines(y, pnorm(y),type="l", lty=1)
legend(-1.2,0.8,c("real", "nonparametric"),lty=1:2)

## End(Not run)
```

rl *Return level estimation*

Description

The T-return level is defined as the value of the observed variable that can be expected to be once exceeded during a T-period of time. This is computed as the quantile of the distribution, corresponding to the value $F^{-1}(1 - \frac{1}{T})$.

Usage

```
rl(type_kernel = "n", vec_data, T, bw = PBbw(type_kernel = "n",
vec_data, 2))
```

Arguments

type_kernel	The kernel function used. You can use four types: "e" Epanechnikov, "n" Normal, "b" Biweight and "t" Triweight. The Normal kernel is used by default.
vec_data	The data sample (earthquake magnitudes, flow levels, wind speeds...).
T	A particular value of time, or an array of time values.
bw	The bandwidth parameter. The plug-in method of Polansky and Baker (2000) is used by default.

Details

In several scientific fields results of interest to estimate quantiles corresponding to a probability of exceedance. For example, in hydrology, the T-return level x_T is defined as the value of the observed flow that can be expected to be once exceeded during a T-period of time; that is, the quantile

$$x_T = F^{-1}(1 - \frac{1}{T}).$$

We can estimate it directly by

$$\hat{x}_T = F_h^{-1}(1 - \frac{1}{T}).$$

See, for instance, Quintela del Rio (2011), for an application to data of Salt River near Roosevelt, AZ, USA.

Value

A single value or an array for the estimated quantiles.

Author(s)

Graciela Estevez Perez <graci@udc.es> and Alejandro Quintela del Rio <aquintela@udc.es>

References

Quintela-del-Rio, A. (2011) On bandwidth selection for nonparametric estimation in flood frequency analysis. *Hydrological Processes* **25**, pp. 671–678.

Quintela-del-Rio, A. and Estevez-Perez, G. (2012) Nonparametric Kernel Distribution Function Estimation with kerdienst: An R Package for Bandwidth Choice and Applications, *Journal of Statistical Software* **50(8)**, pp. 1-21. URL <http://www.jstatsoft.org/v50/i08/>.

Examples

```
data(saltriver)
peak<-saltriver$peakflow
## Not run:
year<-saltriver$year
plot(year,peak, type="l",ylab="Annual peak flow")

## End(Not run)
# Calculating the return values for a period from 2 to 100 years
T<-seq(2,100, length.out=100)
ret.lev<-r1(vec_data=peak, T=T)
## Not run:
plot(T, ret.lev, type="l", xlab="years", ylab="Flow (cumecs)",
main="Return level Plot")

## End(Not run)
```

saltriver

Data from the Salt River near Roosevelt, AZ, USA, for 1924-2006.

Description

The annual peak instantaneous flow of the Salt River near Roosevelt, AZ, USA, for 1924-2009. Data are in cfs (0.028317 m³/s); water year October-September. The data were examined in several papers related with extreme values in hydrology. Among others, they were analyzed by Anderson and Meerschaert (1998) and Dettinger and Diaz (2000), where they were fitted to a GEV and a GPD distribution. In Quintela del Rio (2011), a nonparametric analysis for this data set is made.

Usage

```
data(saltriver)
```

Format

A data frame with 85 observations on the following 2 variables.

year Year

peakflow The annual observed maximum peak flow

Source

US Geological Survey <http://water.usgs.gov/nwis/peak>.

References

Anderson, P.L. and Meerschaert, M.M. (1998) Modeling river flows with heavy tails, *Water Resources Research* **34**, pp. 2271–2280.

Dettinger, M.D. and Diaz, H.F. (2000) Global characteristics of stream flow seasonality and variability, *Journal of Hydrometeorology* **1**, pp. 289–310.

Quintela-del-Rio, A. (2011) On bandwidth selection for nonparametric estimation in flood frequency analysis. *Hydrological Processes* **25**, pp. 671–678.

Quintela-del-Rio, A. and Estevez-Perez, G. (2012) Nonparametric Kernel Distribution Function Estimation with kerdie: An R Package for Bandwidth Choice and Applications, *Journal of Statistical Software* **50(8)**, pp. 1-21. URL <http://www.jstatsoft.org/v50/i08/>.

Examples

```
data(saltriver)
peak<-saltriver$peakflow
year<-saltriver$year
plot(year,peak, type="l",ylab="Annual peak flow")
```


Index

*Topic **datasets**

nwip, 11
saltriver, 15

*Topic **nonparametric**

ALbw, 2
CVbw, 4
ef, 5
kde, 7
mrp, 9
PBbw, 12
r1, 14

*Topic **smooth**

ALbw, 2
CVbw, 4
ef, 5
kde, 7
mrp, 9
PBbw, 12
r1, 14

ALbw, 2

CVbw, 4

ef, 5

kde, 7

kerdiest (kerdiest-package), 2

kerdiest-package, 2

mrp, 9

nwip, 11

PBbw, 12

r1, 14

saltriver, 15