Package ‘Conake’

October 12, 2022

Encoding UTF-8

Type Package

Title Continuous Associated Kernel Estimation

Version 1.0.1

Date 2015-02-27

Author W. E. Wansouwé, F. G. Libengué and C. C. Kokonendji

Maintainer W. E. Wansouwé <ericwansouwe@gmail.com>

Description

Continuous smoothing of probability density function on a compact or semi-infinite support is performed using four continuous associated kernels: extended beta, gamma, lognormal and reciprocal inverse Gaussian. The cross-validation technique is also implemented for bandwidth selection.

License GPL (>= 2)

LazyLoad yes

URL www.r-project.org

NeedsCompilation no

Repository CRAN

Date/Publication 2022-06-13 15:03:41 UTC

R topics documented:

`Conake-package` *Continuous Associated Kernel Estimation*

Description

Continuous smoothing of probability density function defined on a compact $T = [a, b]$ or semi-infinite support $T = [0, \infty)$ is performed using four continuous associated kernels: extended beta, gamma, lognormal and reciprocal inverse Gaussian. The cross-validation technique is also implemented to select the smoothing parameter.

Details

The estimated density: The kernel estimator $\widehat{f}_n$ of f is defined as

$$\widehat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_{x,h}(X_i),$$

where $K_{x,h}$ is one of the kernels defined below. In practice, we first calculate the normalizing constant

$$C_n = \int_{x \in T} \widehat{f}_n(x)dx,$$

where T is the support of the density function. This normalizing constant is not generally equal to 1. The estimated density is then $\tilde{f}_n = \widehat{f}_n / C_n$.

Given a data sample, the **Conake** package allows to compute the density `dke` using one of the four kernel functions: extended beta, gamma, lognormal and reciprocal inverse Gaussian. The bandwidth parameter is calculated using the cross-validation technique `cvbw`.The kernel functions `kef` are defined below.

Extended beta kernel : The extended beta kernel is defined on $S_{x,h,a,b} = [a, b] = T$ with $a < b < \infty$, $x \in T$ and $h > 0$:

$$BE_{x,h,a,b}(y) = \frac{(y-a)^{(x-a)/\{(b-a)h\}}(b-y)^{(b-x)/\{(b-a)h\}}}{(b-a)^{1+h^{-1}}B\left(1+(x-a)/(b-a)h, 1+(b-x)/(b-a)h\right)} 1_{S_{x,h,a,b}}(y),$$

where $B(r, s) = \int_0^1 t^{r-1}(1-t)^{s-1}dt$ is the usual beta function with $r > 0$, $s > 0$ and 1_A denotes the indicator function of A. For $a = 0$ and $b = 1$, the extended beta kernel corresponds to the beta kernel which is the probability density function of the beta distribution with shape parameters $1 + x/h$ and $(1 - x)/h$; see Libengué (2013).

Gamma kernel: The gamma kernel is defined on $S_{x,h} = [0, +\infty) = T$ with $x \in T$ and $h > 0$:

$$GA_{x,h}(y) = \frac{y^{x/h}}{\Gamma(1 + x/h)h^{1+x/h}} exp\left(-\frac{y}{h}\right) 1_{S_{x,h}}(y),$$

where $\Gamma(.)$ is the classical gamma function. It is the probability density function of the gamma distribution with scale parameter $1 + x/h$ and shape parameter h; see Chen (2000) and also Libengué (2013).

Lognormal kernel : The lognormal kernel is defined on $S_{x,h} = [0,\infty) = T$ with $x \in T$ and $h > 0$:

$$LN_{x,h}(y) = \frac{1}{yh\sqrt{2\pi}} exp\left\{-\frac{1}{2}\left(\frac{1}{h}log(\frac{y}{x}) - h\right)^2\right\} 1_{S_{x,h}}(y).$$

It is the probability densiy function of the classical lognormal distribution with mean $log(x) + h^2$ and standard deviation h; see Igarashi and Kakizawa (2015) and also Libengué (2013).

Reciprocal inverse Gaussian kernel: The reciprocal inverse Gaussian kernel is defined on $S_{x,h} =]0,\infty) = T$ with $x \in T$ and $h > 0$:

$$RIG_{x,h}(y) = \frac{1}{\sqrt{2\pi hy}} exp\left\{-\frac{\zeta(x,h)}{2h}\left(\frac{y}{\zeta(x,h)} - 2 + \frac{\zeta(x,h)}{y}\right)\right\} 1_{S_{x,h}}(y),$$

where $\zeta(x,h) = (x^2 + xh)^{1/2}$. It is the probability densiy function of the classical reciprocal inverse Gaussian distribution with mean $1/\sqrt{x^2 + xh}$ and standard deviation $1/h$; see Igarashi and Kakizawa (2015) and also Libengué (2013).

The bandwidth selection: The cross-validation technique `cvbw` is used for the bandwidth selection. The optimal parameter is the one which minimizes the cross-validation function defined by:

$$CV(h) = \int_{x\in T} \{\widehat{f}_n(x)\}^2 dx - \frac{2}{n}\sum_{i=1}^{n} \widehat{f}_{n,-i}(X_i),$$

where $\widehat{f}_{n,-i}(X_i) = (n-1)^{-1}\sum_{j\neq i}^{n} K_{X_i,h}(X_j)$ is the density estimator computed without the observation X_i.

Author(s)

W. E. Wansouwé, F.G. Libengué and C. C. Kokonendji

Maintainer: W. E. Wansouwé <ericwansouwe@gmail.com>

References

Chen, S. X. (1999). Beta kernels estimators for density functions, *Computational Statistics and Data Analysis* **31**, 131 - 145.

Chen, S. X. (2000). Gamma kernels estimators for density functions, *Annals of the Institute of Statistical Mathematics* **52**, 471 - 480.

Libengué, F.G. (2013). *Méthode Non-Paramétrique par Noyaux Associés Mixtes et Applications*, Ph.D. Thesis Manuscript (in French) to Université de Franche-Comté, Besançon, France and Université de Ouagadougou, Burkina Faso, June 2013, **LMB no. 14334**, Besançon.

Igarashi, G. and Kakizawa, Y. (2015). Bias correction for some asymmetric kernel estimators, *Journal of Statistical Planning and Inference* **159**, 37 - 63.

Conakereport *A brief summary of the results*

Description

For a sample, the function gives automatically the result of computations of the normalizing constant and the smoothing parameter. One can then plot the histogram.

Usage

```
Conakereport(Vec, ker, h = NULL, a = 0, b = 1)
```

Arguments

Vec	The sample of data.
ker	The kernel function:
h	The bandwidth or smoothing parameter.
a	The left bound of the support used for extended beta kernel. Default value is 0 for beta kernel.
b	The right bound of the support used for extended beta kernel. Default value is 1 for beta kernel.

Value

Returns a list containing:

h_n	The bandwith parameter used to compute f_n
C_n	The normalizing constant

Author(s)

W. E. Wansouwé, F.G. Libengué and C. C. Kokonendji

References

Libengué, F.G. (2013). *Méthode Non-Paramétrique par Noyaux Associés Mixtes et Applications*, Ph.D. Thesis Manuscript (in French) to Université de Franche-Comté, Besançon, France and Université de Ouagadougou, Burkina Faso, June 2013, **LMB no. 14334**, Besançon.

Examples

```
## Data can be simulated data or real data
## We use simulate data
Vec<-rgamma(100,1.5,2.6)
## Not run:
Conakereport(V,ker="GA")

## End(Not run)
```

cvbw *Cross-validation function for bandwidth selection*

Description

The function allows to calculate the optimal bandwidth using the cross-validation method. Four kernels are available: extended beta, gamma, lognormal and reciprocal inverse Gaussian kernels.

Usage

```
cvbw(Vec, bw = NULL, ker,a=0,b=1)
```

Arguments

`Vec`	The sample data.
`bw`	The sequence of bandwidths where the cross-validation is computed. If NULL, the procedure defines a sequence of bandwidths.
`ker`	The associated kernel: "BE" extended beta, "GA" gamma, "LN" lognormal and "RIG" reciprocal inverse Gaussian.
`a`	The left bound of the support used for extended beta kernel. Default value is 0 for beta kernel.
`b`	The right bound of the support used for extended beta kernel. Default value is 1 for beta kernel.

Details

The selection of the bandwidth parameter is crucial. If the bandwidth is small, we will obtain an undersmoothed estimator, with high variability. On the contrary, if the value is large, the resulting estimator will be very smoothed and farther from the function that we are trying to estimate. See Libengué (2013).

Value

Returns a list containing:

`hcv`	The optimal bandwidth obtained by cross-validation.
`CV`	The values of the cross-validation function in the sequence of bandwidths.
`bw`	The sequence of bandwidths used.

Author(s)

W. E. Wansouwé, F.G. Libengué and C. C. Kokonendji

References

Libengué, F.G. (2013). *Méthode Non-Paramétrique par Noyaux Associés Mixtes et Applications*, Ph.D. Thesis Manuscript (in French) to Université de Franche-Comté, Besançon, France and Université de Ouagadougou, Burkina Faso, June 2013, **LMB no. 14334**, Besançon.

Examples

```
## Data can be simulated data or real data
## We use simulate data
## and then compute the cross validation.
Vec<-rgamma(100,1.5,2.6)
## Not run:
CV<-cvbw(Vec,ker="GA")
CV$hcv

## End(Not run)
```

---

dke — *Function for probability density estimation*

Description

The function estimates the density in a single value or in a grid using discrete associated kernels. Four different associated kernels are available: extended beta, gamma, lognormal and reciprocal inverse Gaussian.

Usage

```
dke(vec_data, ker, bw, x = NULL,a=0,b=1)
```

Arguments

`vec_data`	The data sample.
`ker`	The associated kernel: "BE" extended beta, "GA" gamma, "LN" lognormal and "RIG" reciprocal inverse Gaussian.
`bw`	The bandwidth or smoothing parameter.
`x`	The single value or grid where estimation is computed
`a`	The left bound of the support used for extended beta kernel. Default value is 0 for beta kernel.
`b`	The right bound of the support used for extended beta kernel. Default value is 1 for beta kernel.

Details

The kernel estimator $\widehat{f}_n$ of f is defined in the above sections. We recall that in general, the sum of the estimated values on the support is not equal to 1. In practice, we calculate the normalizing constant C_n before computing the estimated density $\tilde{f}_n$; see Libengué (2013) .

The bandwidth parameter in the function is obtained using the cross-validation technique for the four kernels.

Value

Returns a list containing:

C_n	The normalizing constant.
f_n	The values of the estimated function

Author(s)

W. E. Wansouwé, F.G. Libengué and C. C. Kokonendji

References

Libengué, F.G. (2013). *Méthode Non-Paramétrique par Noyaux Associés Mixtes et Applications*, Ph.D. Thesis Manuscript (in French) to Université de Franche-Comté, Besançon, France and Université de Ouagadougou, Burkina Faso, June 2013, **LMB no. 14334**, Besançon.

Examples

```
## A sample data with n=100.
V<-rgamma(100,1.5,2.6)


##The bandwidth can be the one obtained by cross validation.
h<-0.052
## We choose Gamma kernel.

est<-dke(V,"GA",h)
est$f_n
```

`kef` *Continuous associated kernel function*

Description

This function computes the discrete associated kernel function; see Chen (1999) and also Chen (2000).

Usage

```
kef(x, t, h, ker, a = 0, b = 1)
```

Arguments

`x`	The target.
`t`	A single value or the grid where the continuous associated kernel function is computed.
`h`	The bandwidth or smoothing parameter.
`ker`	The associated kernel: "BE" extended beta, "GA" gamma, "LN" lognormal and "RIG" reciprocal inverse Gaussian.
`a`	The left bound of the support used for extended beta kernel. Default value is 0 for beta kernel.
`b`	The right bound of the support used for extended beta kernel. Default value is 1 for beta kernel.

Details

The associated kernel is one of the four which have been defined in the sections above : extended beta, gamma, lognormal and reciprocal inverse Gaussian; see Igarashi and Kakizawa (2015) and also Libengué (2013).

Value

Returns the value of the discrete associated kernel function at t according to the target and the bandwidth.

Author(s)

W. E. Wansouwé, F.G. Libengué and C. C. Kokonendji

References

Chen, S. X. (1999). Beta kernels estimators for density functions, *Computational Statistics and Data Analysis* **31**, 131 - 145.

Chen, S. X. (2000). Gamma kernels estimators for density functions, *Annals of the Institute of Statistical Mathematics* **52**, 471 - 480.

Libengué, F.G. (2013).*Méthode Non-Paramétrique par Noyaux Associés Mixtes et Applications*, Ph.D. Thesis Manuscript (in French) to Université de Franche-Comté, Besançon, France and Université de Ouagadougou, Burkina Faso, June 2013, **LMB no. 14334**, Besançon.

Igarashi, G. and Kakizawa, Y. (2015). Bias correction for some asymmetric kernel estimators, *Journal of Statistical Planning and Inference* **159**, 37 - 63.

Examples

```
x<-4
h<-0.1
t<-0:10
kef(x,t,h,"GA")
```

simp_int *The Simpson method to compute integral*

Description

This function computes the Simpson method to calculate an integral.

Usage

```
simp_int(x, fx, n.pts = 256, ret = FALSE)
```

Arguments

x	The vector where the integral is computed
fx	The function to integrate
n.pts	The number of points used to compute the integral through the Simpson technique.
ret	A boolean control parameter. Default value is FALSE.

Value

Returns the value of the integral.

Author(s)

W. E. Wansouwé, F.G. Libengué and C. C. Kokonendji

Examples

```
Vec=rgamma(100,1.5,2.6)
x=seq(min(Vec),max(Vec),length.out=100)
simp_int(x,dgamma(x,1.5,2.6))
```

Index