Package ‘L2DensityGoFtest’

October 7, 2021

Encoding UTF-8

Type Package

Title Density Goodness-of-Fit Test

Version 0.2.0

Author Dimitrios Bagkavos [aut, cre]

Maintainer Dimitrios Bagkavos <dimitrios.bagkavos@gmail.com>

Description Provides functions for the implementation of a density goodness-of-fit test, based on piecewise approximation of the L2 distance.

Imports fGarch, kedd, nor1mix

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2021-10-07 19:30:02 UTC

R topics documented:

`cutoff.asymptotic` *Asymptotic critical value for goodness-of-fit testing*

Description

Implements an asymptotic based critical value for goodness-of-fit testing of densities.

Usage

```
cutoff.asymptotic(dist,  p1, p2, sig.lev)
```

Arguments

`dist`	The null distribution.
`p1`	Argument 1 (vector or object) for the null distribution.
`p2`	Argument 2 (vector or object) for the null distribution.
`sig.lev`	Significance level of the hypothesis test.

Details

Implements the asymptotic critical value in Remark 1, Bagkavos, Patil and Wood (2021), equal to $z_\alpha \sigma_{0,\theta_0}$ where z_α is the $1-\alpha$ quantile of the normal distribution and

$$\sigma^2_{0,\theta_0} = 2\left(\int K^2(u)\,du\right)\left(\int f_0^2(x;\theta_0)\,dx\right).$$

Value

A scalar, the estimate the critical value at the given significance level

Author(s)

Dimitrios Bagkavos

R implementation and documentation: Dimitrios Bagkavos <dimitrios.bagkavos@gmail.com>

References

Bagkavos, Patil and Wood: Nonparametric goodness-of-fit testing for a continuous multivariate parametric model, (2021), under review.

cutoff.edgeworth *Edgeworth based critical value for goodness-of-fit testing*

Description

Implements the Edgeworth based critical value for goodness-of-fit testing of densities.

Usage

```
cutoff.edgeworth(xin, dist, kfun, p1, p2, sig.lev)
```

Arguments

xin	A vector of data points - the available sample size.
dist	The null distribution.
kfun	The kernel to use in the density estimates used in the bandwidth expression.
p1	Argument 1 (vector or object) for the null distribution.
p2	Argument 2 (vector or object) for the null distribution.
sig.lev	Significance level of the hypothesis test.

Details

Implements the Edgwworth based finite sample critical value, given by

$$l_\alpha = z_\alpha + d_0\sqrt{h} + d_2(n\sqrt{h})^{-1}$$

where z_α is the $1-\alpha$ quantile of the normal distribution and $d_0 = d_1 - C_{H_0}$ and $d_j = (z_\alpha^2 - 1)c_j, j = 1, 2$ with

$$c_1 = \frac{4K^{(3)}(0)\mu_2^3\nu_3}{3\sigma^3},\ c_2 = \frac{\mu_3^2 K^2(0)}{\sigma^3},\ \mu_i = \int K^i(x)\,dx, i = 1, \ldots.$$

and

$$C_{H_0} = 2\left(Ef_0'(\theta_0)\right)^2\Delta^{-1},\ \nu_i = E\left\{f^i(x)\right\} = \int f^{i+1}(x)\,dx, i = 1, \ldots$$

This is similar in nature to the critical value estimate obtained for the closely relatated regression setting in Gao and Gijbels (2008).

Value

A scalar, the estimate the critical value at the given significance level

Author(s)

Dimitrios Bagkavos

R implementation and documentation: Dimitrios Bagkavos <dimitrios.bagkavos@gmail.com>

References

Bagkavos, Patil and Wood: Nonparametric goodness-of-fit testing for a continuous multivariate parametric model, (2021), under review.

Gao and Gijbels, Bandwidth selection in nonparametric kernel testing, pp. 1584-1594, JASA (2008)

`hopt.be` *Power-optimal bandwidth and critical value*

Description

Implements the power-optimal bandwidth for density goodness-of-fit testing based on a Berry Essen bound.

Usage

```
hopt.be(xin)
```

Arguments

`xin`	A vector of data points - the available sample size.

Details

Implements the power-optimal bandwidth for the test statistic `S.n` given by

$$h = n^{-1/2}\sqrt{\frac{\hat{\nu}_p R_4(K)}{\rho_*^2 \hat{\nu}_4 I_0(K)}},$$

where

$$\hat{\nu}_p = n^{-1}\sum_{j=1}^{n} \hat{f}(X_j; \hat{h}_a),$$

and $\hat{h}_a$ is the density optimal bandwidth calculated by a reference to a prametric distribution, $\rho_\star = 1$ and

$$R_4(K) = \int K^4(x)\, dx.$$

Value

The estimate of the power-optimal bandwidth

Author(s)

Dimitrios Bagkavos

R implementation and documentation: Dimitrios Bagkavos <dimitrios.bagkavos@gmail.com>

References

Bagkavos, Patil and Wood: Nonparametric goodness-of-fit testing for a continuous multivariate parametric model, (2021), under review.

`hopt.edgeworth` *Power-optimal bandwidth for density goodness-of-fit testing*

Description

Implements the Edgeworth based power-optimal bandwidth for density goodness-of-fit testing.

Usage

```
hopt.edgeworth(xin, dist, kfun, p1, p2, sig.lev)
```

Arguments

`xin`	A vector of data points - the available sample size.
`dist`	The null distribution.
`kfun`	The kernel to use in the density estimates used in the bandwidth expression.
`p1`	Argument 1 (vector or object) for the null distribution.
`p2`	Argument 2 (vector or object) for the null distribution.
`sig.lev`	Significance level of the hypothesis test.

Details

Implements: the power-optimal bandwidth for the test statistic `S.n` given by

$$h = \left\{ \frac{\sqrt{2}K^{(3)}(0)}{3R(K)^{3/2}} \frac{\nu_2}{R(f)^{3/2}} \right\}^{-1/2} \left\{ \frac{n \int \Delta_n^2(x) f^2(x)\, dx}{\sigma^2 \{2\nu_2 R(K)\}^{1/2}} \right\}^{-3/2},$$

which is similar in nature to the corresponding optimal bandwidth value obtained for the closely relatated regression setting in Gao and Gijbels (2008).

Value

A scalar, the estimate the power-optimal bandwidth at the given significance level α.

Author(s)

Dimitrios Bagkavos

R implementation and documentation: Dimitrios Bagkavos <dimitrios.bagkavos@gmail.com>

References

Gao and Gijbels, Bandwidth selection in nonparametric kernel testing, pp. 1584-1594, JASA (2008)

kde *Kernel Density Estimation*

Description

Implements the (classical) kernel density estimator, see (2.2a) in Silverman (1986).

Usage

```
kde(xin, xout, h, kfun)
```

Arguments

xin	A vector of data points. Missing values not allowed.
xout	A vector of grid points at which the estimate will be calculated.
h	A scalar, the bandwidth to use in the estimate, e.g. bw.nrd(xin)
kfun	Kernel function to use. Supported kernels: Epanechnikov, Biweight, Gaussian, Rectangular, Triangular.

Details

The classical kernel density estimator is given by

$$\hat{f}(x;h) = n^{-1} \sum_{i=1}^{n} K_h(x - X_i)$$

h is determined by a bandwidth selector such as Silverman's default plug-in rule.

Value

A vector with the density estimates at the designated points xout.

Author(s)

R implementation and documentation: Dimitrios Bagkavos <dimitrios.bagkavos@gmail.com>

References

Silverman (1986), Density Estimation for Statistics and Data Analysis, Chapman and Hall, London.

Examples

```
x<-seq(-5, 5,length=100)          #design points where the estimate will be calculated
plot(x, dnorm(x),  type="l", xlab = "x", ylab="density") #plot true density function
SampleSize <- 100
ti<- rnorm(SampleSize)            #draw a random sample from the actual distribution

huse<-bw.nrd(ti)
arg2<-kde(ti, x, huse, Epanechnikov) #Calculate the estimate
lines(x, arg2, lty=2)             #draw the result on the graphics device.
```

`Kernels` *Kernel functions*

Description

Implements various kernel functions, including boundary, integrated and discrete kernels for use in the definition of the nonparametric estimates

Usage

```
Biweight(x, ...)
Epanechnikov(x, ...)
Triangular(x, ...)
Gaussian(x, ...)
Rectangular(x, ...)
```

Arguments

`x`	A vector of data points where the kernel will be evaluated.
`...`	Further arguments.

Details

Implements the Biweight, Triangular, Guassian, Rectangular and Epanechnikov kernels.

Value

The value of the kernel at x

References

Wand and Jones, (1996), Kernel Smoothing, Chapman and Hall, London

---

`NDistDens` *Select null distribution*

Description

Implements the selection of null distribution; to be used within the implementation of the test statistic S.n

Usage

```
NDistDens(x, dist, p1, p2)
```

Arguments

`x`	A vector of data points - the available sample size.
`dist`	The null distribution.
`p1`	Argument 1 (vector or object) for the null distribution.
`p2`	Argument 2 (vector or object) for the null distribution.

Details

Implements the null distribution evaluation at designated points, given the parameters p1 and p2.

Value

A vector containing the density values of the designated distribution

Author(s)

Dimitrios Bagkavos

R implementation and documentation: Dimitrios Bagkavos <dimitrios.bagkavos@gmail.com>

References

Bagkavos, Patil and Wood: Nonparametric goodness-of-fit testing for a continuous multivariate parametric model, (2021), under review.

S.n *Goodness-of-Fit test statistic based on discretized L2 distance*

Description

Implements a density goodness of fit test based on a discretized approximation of the L2 distance.

Usage

```
S.n(xin, h, drate, dist, p1, p2)
```

Arguments

`xin`	A vector of data points - the available sample size.
`h`	The bandwidth to use, typically the output of `hopt.edgeworth`.
`drate`	power to use for the binning, the default is 3/4.
`dist`	The null distribution.
`p1`	Argument 1 (vector or object) for the null distribution.
`p2`	Argument 2 (vector or object) for the null distribution.

Details

Implements the test statistic used for testing the hypothesis

$$H_0 : f(x) = f_0(x, p1, p2) \;\; vs \;\; H_a : f(x) \neq f_0(x, p1, p2).$$

This density goodness-of-fit test is based on a discretized approximation of the L2 distance. Assuming that n is the number of observations and $g = (max(xin) - min(xin))/n^{-drate}$ is the number of bins in which the range of the data is split, the test statistic is:

$$S_n(h) = n\Delta^2 h^{-1/2} \sum \sum_{i \neq j} K\{(X_i - X_j)h^{-1}\}\{Y_i - f_0(X_i)\}\{Y_j - f_0(X_j)\}$$

where K is the Epanechnikov kernel implemented in this package with the `Epanechnikov` function. The null model f_0 is specified through the `dist` argument with parameters passed through the `p1` and `p2` arguments. The test is implemented either with bandwidth `hopt.edgeworth` or with bandwidth `hopt.be` which provide the value of h needed for calculation of $S_n(h)$ and the critical value used to determine acceptance or rejection of the null hypothesis. See the example below for an application to a real world dataset.

Value

A vector with the value of the test statistic as well as the bandwidth used for its calculation

Author(s)

R implementation and documentation: Dimitrios Bagkavos <dimitrios.bagkavos@gmail.com>

References

Bagkavos, Patil and Wood: Nonparametric goodness-of-fit testing for a continuous multivariate parametric model, (2021), under review.

See Also

hopt.edgeworth

Examples

```
library(fGarch)
library(kedd)
data(EuStockMarkets)
DAX <- as.ts(EuStockMarkets[,"DAX"])
dax <-  diff(log(DAX))#[,"DAX"]

# Fit a GARCH(1,1) model to dax returns:
lll<-garchFit(~ garch(1,1), data = as.ts(dax), trace = FALSE, cond.dist ="std")
# define the model innovations, to be used as input to the test statistic
xin<-lll@residuals /lll@sigma.t
# exclude smallest value - only for uniform presentation of results
#(this step can be excluded):
xin = xin[xin!= min(xin)]

#inputs for the test statistic:
#kernel function to use in implementing the statistic
#and functional estimates for optimal h:
kfun<-"epanechnikov"
a.sig<-0.05 #define the significance level
#null hypothesis is that the innovations are normaly distributed:
Nulldist<-"normal"

p1<-mean(xin)
p2<- sd(xin)
#Power optimal bandwidth:
h<-hopt.edgeworth(xin,   Nulldist, kfun, p1, p2, a.sig )
h.be <- hopt.be(xin)
# Edgeworth based cutoff point:
cutoff<-cutoff.edgeworth(xin,   Nulldist, kfun, p1, p2, a.sig )

TestStatistic<-S.n(xin, h, 1/3, Nulldist, p1, p2)
TestStatistic.be<-S.n(xin, h.be, 1/3, Nulldist, p1, p2)

cat("L2 test statistic value Edgeworth:", TestStatistic[1],
"\nL2 test statistic value Barry-Essen bandwidth:", TestStatistic.be[1],
"\ncritical value :", cutoff, "\n")
#L2 test statistic value Edgeworth: 21.49267
#L2 test statistic value Barry-Essen bandwidth: 52.66799
# critical value : 2.140446
# L2 test statistic >  critical value, hence normality is rejected
```

Index