Package ‘SDNNtests’

June 10, 2021

Title SHAPE CONSTRAINED TESTS OF STOCHASTIC DOMINANCE

Version 0.0.0.9000

Description This package implements the methods discussed in Laha et al. (2020). In particular, this package performs nonparametric and shape-constrained (unimodality and log-concavity) test of one sided stochastic dominance against the null of non-dominance. This package also gives an estimator of the Hellinger distance between two densities under the shape restriction of unimodality or log-concavity.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Imports logcondens,
pracma,
lattice,
gridExtra,
ggpubr,
grid,
ggplot2,
ks

RoxygenNote 7.1.0

R topics documented:

calc_mode *Unimodal density estimator when the mode is unknown*

Description

Estimates the density of a given sample under the assumption that the underlying density is unimodal. The mode is estimated from the data. The method is based on the unimodal regularization of Birge (1997).

Usage

```
calc_mode(x, t)
```

Arguments

x	Vector of independent and identically distributed random variables; must be sorted.
t	A positive real number. Default value is one.

Details

Birge(1997)'s estimator gives a a pieciewise constant unimodal density. The discontinuity points of the respective density are called the knots, which belong to the set of datapoints. The density estimator is constant between two consecutive knots.

`t`: The parameter t corresponds to the parameter τ in Birge (1997). Higher values of t leads to more accurate estimation of the unimodal densities. This value ontrols the accuracy in unimodal density estimation upto the term n^{-t} where n is the sample size. We recommend a value greater than or equal to one. See Birge (1997) for more details.

Value

- mode - The estimator of the mode
- x.knots - The vector of the knots of the estimated density.
- F.knots - A vector consisting the values of the estimated distribution function evaluated at the knots.
- f.knots - A vector whose i-th element gives the value of the estimated density on the segment joining the i-th and (i+1)-th knot. Recall that the estimated density is piecewise constant between two knots.

Author(s)

Nilanjana Laha (maintainer), <nlaha@hsph.harvard.edu>, Alex Luedtke, <aluedtke@uw.edu>.

References

Birge, L. (1997). *Estimation of unimodal densities without smoothness assumptions*, Ann. Statist., 25, 970–981.

Examples

```
x <- sort(rnorm(100)); calc_mode(x, 1)
```

hd.lc *Hellinger distance between log-concave densities*

Description

Provides an estimate of the squared hellinger distance between two log-concave densities. This function uses the log-concave density estimator of Dumbgen and Rufibatch (2009), given by logConDens of logcondens package.

Usage

```
hd.lc(x, y)
```

Arguments

x	Vector of m independent and identically distributed random variables; corresponds to the first sample.
y	Vector of n independent and identically distributed random variables; corresponds to the second sample.

Value

A point estimator of the Hellinger distance.

Author(s)

Nilanjana Laha (maintainer), <nlaha@hsph.harvard.edu>,

Alex Luedtke, <aluedtke@uw.edu>.

References

Laha, N., Moodie, Z., Huang, Y., and Luedtke (2021), A. *Improved inference for vaccine-induced immune responses via shape-constrained methods*. Submitted.

Dumbgen, L. and Rufibatch, K. (2009). *Maximum likelihood estimation of a logconcave density and its distribution function: Basic properties and uniform consistency*, Bernoulli, 15, 40–68.

See Also

hd.uni, hd.lc.sm, hell.ci

Examples

```
x <- sort(rnorm(100)); y <- sort(rgamma(50, shape=1));
hd.lc(x,y)
```

hd.lc.sm *Hellinger distance between log-concave densities*

Description

Provides an estimate of the squared Hellinger distance between two log-concave densities. This function uses the smoothed log-concave density estimator of Chen and Samworth (2013), given by logConDens of logcondens package.

Usage

```
hd.lc.sm(x, y)
```

Arguments

x	Vector of m independent and identically distributed random variables; corresponds to the first sample.
y	Vector of n independent and identically distributed random variables; corresponds to the second sample.

Value

A point estimator of the Hellinger distance.

Author(s)

Nilanjana Laha (maintainer), <nlaha@hsph.harvard.edu>,

Alex Luedtke, <aluedtke@uw.edu>.

References

Laha, N., Moodie, Z., Huang, Y., and Luedtke (2021), A. *Improved inference for vaccine-induced immune responses via shape-constrained methods.* Submitted.

Chen, Y. and Samworth, R. J. (2013). *Smoothed log-concave maximum likelihood estimation with applications*, Statistica Sinica, 23, 1303-1398.

See Also

hd.lc, hd.uni, hell.ci

Examples

```
x <- sort(rnorm(100)); y <- sort(rgamma(50, shape=1));
hd.lc.sm(x,y)
```

hd.uni — *Hellinger distance between unimodal densities*

Description

Provides an estimate of the squared hellinger distance between the densities underlying the (two) observed samples. This estimator assumes that both underlying densities are unimodal. . This function uses the density estimator of Birges (1997), given by `calc_mode`. See Laha et al. (2021) for more details.

Usage

```
hd.uni(x, y)
```

Arguments

x	Vector of m independent and identically distributed random variables; corresponds to the first sample.
y	Vector of n independent and identically distributed random variables; corresponds to the second sample.

Details

This function calls `link{calc_mode}` where the parameter t is taken to be one.

Value

An estimate of the Hellinger distance between the densities of x and y.

Author(s)

Nilanjana Laha (maintainer), `<nlaha@hsph.harvard.edu>`,

Alex Luedtke, `<aluedtke@uw.edu>`.

References

Laha, N., Moodie, Z., Huang, Y., and Luedtke, A. (2021). *Improved inference for vaccine-induced immune responses via shape-constrained methods*. Submitted.

Birge, L. (1997). *Estimation of unimodal densities without smoothness assumptions*, Ann. Statist., 25, 970–981.

See Also

`calc_mode`, `hd.lc`, `hd.lc.sm`, `hell.ci`

Examples

```
x <- sort(rnorm(100)); y <- sort(rgamma(50, shape=1));
hd.uni(x,y)
```

hell.ci *Confidence interval for Hellinger distance estimators*

Description

Calculates the confidence interval of the squared Hellinger distance, as discussed in Laha et al. (2021). This function employs shape constrained methods to estimate the underlying densities. See below for more details. which include the log-concave density estimator of Dumbgen and Rufibatch (2009), and the smoothed log-concave density estimator given by ,

Usage

```
hell.ci(x, y, alpha = 0.05)
```

Arguments

x	Vector of m independent and identically distributed random variables; corresponds to the first sample.
y	Vector of n independent and identically distributed random variables; corresponds to the second sample.
alpha	Level of significance; returns a $1 - \alpha\%$ confidence interval.

Details

The confidence intervals follow the construnction of Laha et al. (2021). One confidence interval is based on the unimodal density estimator of Birges (1997), and provably requires the underlying densities to be unimodal to work. The other confidence intervals are based on the log-concave density estimators, i.e. the log-concave MLE of Dumbgen and Rufibatch (2009), and the smoothed log-concave MLE of Chen and Samworth (2013) . Both are computed using the function logConDens of logcondens package.

Value

- lc.ci - Confidence intervals based on the log-concave MLE.
- sm.lc.ci - Confidence intervals based on the smoothed log-concave MLE.
- um.ci - Confidence intervals based on unimodal density estimator.

Author(s)

Nilanjana Laha (maintainer), <nlaha@hsph.harvard.edu>,

Alex Luedtke, <aluedtke@uw.edu>.

References

Laha, N., Moodie, Z., Huang, Y., and Luedtke (2020), A. *Improved inference for vaccine-induced immune responses via shape-constrained methods*. Submitted.

Dumbgen, L. and Rufibatch, K. (2009). *Maximum likelihood estimation of a logconcave density and its distribution function: Basic properties and uniform consistency*, Bernoulli, 15, 40–68.

Chen, Y. and Samworth, R. J. (2013). *Smoothed log-concave maximum likelihood estimation with applications*, Statistica Sinica, 23, 1303-1398.

Birge, L. (1997). *Estimation of unimodal densities without smoothness assumptions*, Ann. Statist., 25, 970–981.

See Also

hd.uni, hd.lc.sm, hd.uni

Examples

```
x <- sort(rnorm(100)); y <- sort(rgamma(50, shape=1));
hell.ci(x,y)$lc.ci
```

SDNN — *One-sided test of stochastic dominance against the null of non-dominance*

Description

Calculates the p-values of one-sided tests of restricted stochastic dominance against the null of non-dominance. The concerned tests are the minimum t-statistic test and the two sample empirical process (TSEP) test of *Laha et al. (2021).* Each test can be either nonparametric, or semiparametric, ie. using unimodality or log-concavity assumption on the underlying densities.

Usage

```
SDNN(x, y, Method, t = 1, p1 = 0.05, p2 = 0.05)
```

Arguments

`x`	Vector of m independent and identically distributed random variables; corresponds to the first sample.
`y`	Vector of n independent and identically distributed random variables; corresponds to the second sample.
`Method`	Must be one among "NP" (nonparametric), "UM" (unimodal), and "LC" (log-concave). See 'Details'.
`t`	A positive real number. Only required when Method="UM", default value is 1. See Details.
`p1`	The proportion of combined data to be trimmed from the left prior to testing, should take value in $[0, 0.50)$, the default is set to 0.05.
`p2`	The proportion of combined data to be trimmed from the right prior to testing, should take value in $[0, 0.50)$, the default is set to 0.05.

Details

Suppose $X_1, \ldots, X_m$ and $Y_1, \ldots, y_n$ are independent random variables with distribution F and G, respectively. Denote by $D_{p,m,n}$ the set

$$[H_n^{-1}(p), H_n^{-1}(1-p)]$$

where $p \in [0, 0.5)$ and H_n is the empirical distribution function of the combined sample

$$\{X_1, \ldots, x_m, Y_1, \ldots, Y_n\}.$$

The function SDNN tests $H_0 : F(x) \geq G(x)$ for some $x \in D_{p,m,n}$ vs $H_1 : F(x) < G(x)$ for all $x \in D_{p,m,n}$. For more details, see Laha et al., 2021.

`Method`: "NP" corresponds to the nonparametric tests. "UM" corresponds to tests which use the function `calc_mode` to estimate the densities of X_i's and Y_i's. This function estimates the unimodal density estimator of Birge (1997). "LC" corresponds to tests which use the log-concave MLE (given by logConDens of R package "logcondens") of Dumbgen and Rufibatch (2009) to estimate the latter densities. For more detail, see Laha et al. (2021).

`t`: The parameter t corresponds to the parameter τ in Birge (1997). Higher values of t leads to more accurate estimation of the unimodal densities. This value ontrols the accuracy in unimodal density estimation upto the term $(m + n)^{-t}$. A value greater than or equal to one is recommended. See Birge (1997) for more details.

`p`: p corresponds to the set $D_{p,m,n}$ in H_0 and H_1. To overcome the difficulties arising from the tail region, $100p$ percent of data is trimmed from both sides of the combined sample.

Value

A list of two numbers.

- T1 - The p-value of the test based on minimum T-statistic of Laha *et al.*, (2021)
- T2 - The p-value of the TSEP test of Laha *et al.*, (2021).

Author(s)

Nilanjana Laha (maintainer), `<nlaha@hsph.harvard.edu>`, Alex Luedtke, `<aluedtke@uw.edu>`.

References

Laha, N., Moodie, Z., Huang, Y., and Luedtke, A. (2021). *Improved inference for vaccine-induced immune responses via shape-constrained methods*. Submitted.

Dumbgen, L. and Rufibatch, K. (2009). *Maximum likelihood estimation of a logconcave density and its distribution function: Basic properties and uniform consistency*, Bernoulli, 15, 40–68.

Birge, L. (1997). *Estimation of unimodal densities without smoothness assumptions*, Ann. Statist., 25, 970–981.

See Also

`calc_mode`, logConDens

Examples

```
x <- rnorm(100); y <- rgamma(50, shape=1);
 SDNN(x, y, Method="UM", t=1, p1=0.01, p2=0.05)
```

Index