Package ‘weightedGCM’

October 12, 2022

Type Package

Title Weighted Generalised Covariance Measure Conditional Independence Test

Version 0.1.0

Description A conditional independence test that can be applied both to univariate and multivariate random variables. The test is based on a weighted form of the sample covariance of the residuals after a nonlinear regression on the conditioning variables. Details are described in Scheidegger, Hoerrmann and Buehlmann (2021) "The Weighted Generalised Covariance Measure" <arXiv:2111.04361>. The test is a generalisation of the Generalised Covariance Measure (GCM) implemented in the R package 'GeneralisedCovarianceMeasure' by Jonas Peters and Rajen D. Shah based on Shah and Peters (2020) "The Hardness of Conditional Independence Testing and the Generalised Covariance Measure" <arXiv:1804.07203>.

License GPL-2

Imports GeneralisedCovarianceMeasure, methods, mgcv, stats, xgboost

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

Encoding UTF-8

RoxygenNote 7.1.1

NeedsCompilation no

Author Cyrill Scheidegger [aut, cre], Julia Hoerrmann [ths], Peter Buehlmann [ths], Jonas Peters [ctb, cph] (The code in 'trainFunctions.R' is copied (with small modifications) from the R package 'GeneralisedCovarianceMeasure' by Jonas Peters and Rajen D. Shah), Rajen D. Shah [ctb, cph] (The code in 'trainFunctions.R' is copied (with small modifications) from the R package 'GeneralisedCovarianceMeasure' by Jonas Peters and Rajen D. Shah)

Maintainer Cyrill Scheidegger <cyrill.scheidegger@stat.math.ethz.ch>
Description

The Weighted Generalised Covariance Measure (WGCM) with Estimated Weight Function is a test for conditional independence. It is a generalisation of the Generalised Covariance Measure implemented in the R package GeneralisedCovarianceMeasure.

Usage

wgcm.est(X, Y, Z, beta = 0.3, regr.meth, regr.pars = list(), nsim = 499)

Arguments

X
A (n x d_X) numeric matrix with n observations of d_X variables.

Y
A (n x d_Y) numeric matrix with n observations of d_Y variables.

Z
A (n x d_Z) numeric matrix with n observations of d_Z variables.

beta
A real number between 0 and 1 indicating the fraction of the sample used to estimate the weight function.

regr.meth
One of "gam" and "xgboost" indicating the regression method used to estimate the conditional expectations E[X|Z] and E[Y|Z] and the weight function sign(E[(X-E[X|Z])(Y-E[Y|Z])|Z]).

regr.pars
Optional additional regression parameters according to GeneralisedCovarianceMeasure::comp.resids()

nsim
Number of samples used to calculate the p-value using simulation. Only used if max(d_X, d_Y) > 1.

Value

A p-value for the null hypothesis of conditional independence of X and Y given Z.
References

Please cite the following papers. Cyrill Scheidegger, Julia Hoerrmann, Peter Buehlmann: "The Weighted Generalised Covariance Measure" [https://arxiv.org/abs/2111.04361](https://arxiv.org/abs/2111.04361)


Examples

```r
set.seed(1)
n <- 200
Z <- rnorm(n)
X <- Z + 0.3*rnorm(n)
Y1 <- Z + 0.3*rnorm(n)
Y2 <- Z + 0.3*rnorm(n) + 0.3*X
Y3 <- Z + 0.3*rnorm(n) + 0.15*X^2
wgcm.est(X, Y1, Z, beta = 0.3, regr.meth = "gam")
wgcm.est(X, Y2, Z, beta = 0.3, regr.meth = "gam")
wgcm.est(X, Y3, Z, beta = 0.3, regr.meth = "gam")
```

---

wgcm.fix

Weighted Generalised Covariance Measure (WGCM) With Fixed Weight Functions Conditional Independence Test

Description

The Weighted Generalised Covariance Measure (WGCM) with Fixed Weight Functions is a test for conditional independence. It is a generalisation of the Generalised Covariance Measure implemented in the R package GeneralisedCovarianceMeasure.

Usage

```r
wgcm.fix(
  X,
  Y,
  Z,
  regr.meth,
  regr.pars = list(),
  weight.num,
  weight.meth = "sign",
  nsim = 499
)
```

Arguments

- `X` A (n x d_X) numeric matrix with n observations of d_X variables.
- `Y` A (n x d_Y) numeric matrix with n observations of d_Y variables.
**wgcm.fix**

**Z**
A \((n \times d_Z)\) numeric matrix with \(n\) observations of \(d_Z\) variables.

**regr.meth**
One of "gam" and "xgboost" indicating the regression method used to estimate the conditional expectations \(E[X|Z]\) and \(E[Y|Z]\).

**regr.pars**
Optional additional regression parameters according to GeneralisedCovarianceMeasure::comp.resids().

**weight.num**
Number \(k_0\) of weight functions per dimension of \(Z\) to be used additionally to the constant weight function \(w(z) = 1\). The total number of weight functions will be \(1 + k_0 \times d_Z\). In case of \(\max(d_X, d_Y) > 1\), the same \(1 + k_0 \times d_Z\) weight functions are used for every combination of the components of \(X\) and \(Y\).

**weight.meth**
String indicating the method to choose the weight functions. Currently, only "sign" is implemented.

**nsim**
Number of samples used to calculate the p-value using simulation.

**Value**
A p-value for the null hypothesis of conditional independence of \(X\) and \(Y\) given \(Z\).

**References**
Please cite the following papers. Cyrill Scheidegger, Julia Hoerrmann, Peter Bühlmann: "The Weighted Generalised Covariance Measure" [https://arxiv.org/abs/2111.04361](https://arxiv.org/abs/2111.04361)


**Examples**

```r
define_seed
set.seed(1)
n <- 200
Z <- rnorm(n)
X <- Z + 0.3*rnorm(n)
Y1 <- Z + 0.3*rnorm(n)
Y2 <- Z + 0.3*rnorm(n) + 0.3*X
Y3 <- Z + 0.3*rnorm(n) + 0.15*X^2
wgcm.fix(X, Y1, Z, regr.meth = "gam", weight.num = 7, weight.meth = "sign")
wgcm.fix(X, Y2, Z, regr.meth = "gam", weight.num = 7, weight.meth = "sign")
wgcm.fix(X, Y3, Z, regr.meth = "gam", weight.num = 7, weight.meth = "sign")
```
Index

wgcm.est, 2
wgcm.fix, 3