Package 'ghcm'

September 27, 2021

Type Package

Title Functional Conditional Independence Testing with the GHCM

Version 2.0.0

Description A statistical hypothesis test for conditional independence.
Given residuals from a sufficiently powerful regression, it tests whether
the covariance of the residuals is vanishing. It can be applied to both
discretely-observed functional data and multivariate data.
Details of the method can be found in Anton Rask Lundborg, Rajen D. Shah and Jonas
Peters (2021) <arXiv:2101.07108>.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Imports graphics, MASS, refund, stats, utils, CompQuadForm

Depends R (>= 4.0.0)

RoxygenNote 7.1.1

Suggests testthat, knitr, rmarkdown, bookdown,
GeneralisedCovarianceMeasure, ggplot2, reshape2

URL https://github.com/arlundborg/ghcm

BugReports https://github.com/arlundborg/ghcm/issues

VignetteBuilder knitr

NeedsCompilation no

Author Anton Rask Lundborg [aut, cre],
Rajen D. Shah [aut],
Jonas Peters [aut]

Maintainer Anton Rask Lundborg <a.lundborg@statslab.cam.ac.uk>

Repository CRAN

Date/Publication 2021-09-27 18:20:02 UTC

R topics documented:

`ghcm_sim_data` *GHCM simulated data*

Description

A simulated dataset containing a combination of functional and scalar variables. The functional variables each consists of 101 observations on an equidistant grid on [0, 1].

Usage

```
ghcm_sim_data
```

Format

A data frame with 500 rows of 5 variables:

X 500 x 101 matrix.

Z 500 x 101 matrix.

W 500 x 101 matrix.

Y_1 Numeric vector.

Y_2 Numeric vector.

Details

Y_1 and Y_2 are scalar random variables and are both functions of Z. X, Z and W are functional, Z is a function of X and W is a function of Z.

Source

The generation script can be found in the `data-raw` folder of the package.

ghcm_test *Conditional Independence Test using the GHCM*

Description

Testing X independent of Y given Z using the Generalised Hilbertian Covariance Measure. The function is applied to residuals from regressing X on Z and regressing Y on Z and its validity is contingent on the performance of the regression methods.

Usage

```
ghcm_test(resid_X_on_Z, resid_Y_on_Z, alpha = 0.05)
```

Arguments

resid_X_on_Z, resid_Y_on_Z	Numeric vectors or matrices. Residuals when regressing X (Y) on Z with a suitable regression method.
alpha	Numeric in the unit interval. Significance level of the test.

Value

An object of class ghcm containing:

test_statistic Numeric, test statistic of the test.

p Numeric in the unit interval, estimated p-value of the test.

cov matrix, estimated covariance of the truncated limiting Gaussian.

alpha Numeric in the unit interval, significance level of the test.

References

Please cite the following paper: Anton Rask Lundborg, Rajen D. Shah and Jonas Peters: "Conditional Independence Testing in Hilbert Spaces with Applications to Functional Data Analysis" https://arxiv.org/abs/2101.07108

Examples

```
library(refund)
set.seed(1)
data(ghcm_sim_data)
grid <- seq(0, 1, length.out = 101)

# Test independence of two scalars given a functional variable

m_1 <- pfr(Y_1 ~ lf(Z), data=ghcm_sim_data)
m_2 <- pfr(Y_2 ~ lf(Z), data=ghcm_sim_data)
ghcm_test(resid(m_1), resid(m_2))
```

```
# Test independence of a functional variable and a scalar variable given a
# functional variable

m_X <- pffr(X ~ ff(Z), data=ghcm_sim_data, chunk.size=31000)
ghcm_test(resid(m_X), resid(m_1))

# Test independence of two functional variables given a functional variable

m_W <- pffr(W ~ ff(Z), data=ghcm_sim_data, chunk.size=31000)
ghcm_test(resid(m_X), resid(m_W))
```

Index