Package 'ghcm'

January 25, 2021

Type Package

Title Functional Conditional Independence Testing with the GHCM

Version 1.0.0

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 (2020) <arXiv:2101.07108>.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Imports graphics, MASS, refund, stats, utils

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 2021-01-25 13:10:05 UTC

R topics documented:

ghcm_sim_data 2
ghcm_test 3
plot.ghcm 4

Index **6**

`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,
  X_grid = NULL,
  Y_grid = NULL,
  fpca_method = "fpca.sc",
  b = 10000,
  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.
`X_grid, Y_grid`	Numeric vectors or NA. The grid of values that X (Y) is observed on. When set to NULL (the default) an equidistant grid on the unit interval is created. If NA, X (Y) is assumed to not be a functional random variable.
`fpca_method`	String or function. If a string is given, will search the refund package for a function with the given name. If a function is given it must take a data matrix and a grid as input and return a matrix with the same number of rows as the input and the coordinates of the input in its FPCA basis as each row. Extra arguments to the fpca function are supplied with `....` Currently supported refund fpca functions are `fpca.sc` (the default), `fpca.ssvd` and `fpca.face`.
`b`	Positive integer. The number of Monte Carlo realisations from the estimated limiting distribution to estimate the p-value.
`alpha`	Numeric in the unit interval. Significance level of the test.
`...`	Additional arguments to be passed to the fpca_method.

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.

dim Positive integer, the dimension of the truncated limiting Gaussian.

cov dim x dim matrix, estimated covariance of the truncated limiting Gaussian.

samples Numeric vector, samples of the Hilbert-Schmidt norm of the estimated 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), X_grid = NA, Y_grid = NA )

# 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), X_grid = grid, Y_grid = NA )

# 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), X_grid = grid, Y_grid = grid)
```

---

plot.ghcm *Plotting function for the* ghcm*-class.*

Description

Plots the observed test statistic of a performed GHCM test together with a density estimate of the estimated asymptotic distribution of the test statistic under the null.

Usage

```
## S3 method for class 'ghcm'
plot(x, bw = "SJ", ...)
```

Arguments

x	ghcm-object. The result of running a ghcm_test.
bw	a string or numeric. If bw is a string, it should specify a bandwidth method for the density function in the **stats** package. If bw is a numeric, it will be used as the bandwidth in the call to density.
...	additional plotting parameters.

Value

None.

Examples

```
set.seed(1)
library(refund)
data(ghcm_sim_data)

# 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)
test <- ghcm_test(resid(m_1), resid(m_2), X_grid = NA, Y_grid = NA )
plot(test)
```

Index

∗ **datasets**
ghcm_sim_data, 2

ghcm_sim_data, 2
ghcm_test, 3

plot.ghcm, 4