Package ‘CMDMeasure’

February 12, 2018

Title Conditional Mean Dependence Measures via Energy Statistics

Version 1.0.0

Date 2018-01-30

Description Implementation of conditional mean dependence measures and conditional mean independence tests in Shao, X., and Zhang, J. (2014) <doi:10.1080/01621459.2014.887012> and Park, T., et al. (2015) <doi:10.1214/15-EJS1047>.

Depends R (>= 3.4.0)

Suggests testthat (>= 2.0.0),
energy (>= 1.7-0)

License GPL (>= 2)

LazyData true

RoxygenNote 6.0.1

Collate 'CMDMeasure-package.R'
'functions.R'
'pmdd.R'
'mdd.R'
'cmdm_test.R'
'mdc.R'
'pmdc.R'

R topics documented:

CMDMeasure-package *Conditional Mean Dependence Measures via Energy Statistics*

Description

CMDMeasure: A package for conditional mean dependence measures via energy statistics

Details

The CMDMeasure package provides measures of conditional mean dependence and tests of conditional mean independence.

Measuring conditional mean dependence

The conditional mean dependence measures include:

- conditional mean dependence of Y given X
  - martingale difference divergence
  - martingale difference correlation
- conditional mean dependence of Y given X conditioning on Z
  - partial martingale difference divergence
  - partial martingale difference correlation

Testing conditional mean independence

The conditional mean independence tests include:

- conditional mean independence of Y given X conditioning on Z
  - martingale difference divergence under a linear assumption
  - partial martingale difference divergence

The conditional mean independence tests based on the conditional mean dependence measures are implemented as permutation tests.

Author(s)

Ze Jin <zj58@cornell.edu>, Shun Yao <shunyao2@illinois.edu>,
Xiaofeng Shao <xshao@illinois.edu>, David S. Matteson <matteson@cornell.edu>

cmdm_test *Conditional Mean Independence Tests*

Description

cmdm_test tests conditional mean independence of Y given X conditioning on Z, where each contains one variable (univariate) or more variables (multivariate). All tests are implemented as permutation tests.

Usage

```
cmdm_test(X, Y, Z, num_perm = 500, type = "linmdd", compute = "C",
  center = "U")
```

Arguments

X	A vector, matrix or data frame, where rows represent samples, and columns represent variables.
Y	A vector, matrix or data frame, where rows represent samples, and columns represent variables.
Z	A vector, matrix or data frame, where rows represent samples, and columns represent variables.
num_perm	The number of permutation samples drawn to approximate the asymptotic distributions of mutual dependence measures.
type	The type of conditional mean dependence measures, including • linmdd: martingale difference divergence under a linear assumption; • pmdd: partial martingale difference divergence.
compute	The computation method for martingale difference divergence, including • C: computation implemented in C code; • R: computation implemented in R code.
center	The centering approach for martingale difference divergence, including • U: U-centering which leads to an unbiased estimator; • D: double-centering which leads to a biased estimator.

Value

cmdm_test returns a list including the following components:

stat	The value of the conditional mean dependence measure.
dist	The p-value of the conditional mean independence test.

References

Shao, X., and Zhang, J. (2014). Martingale difference correlation and its use in high-dimensional variable screening. Journal of the American Statistical Association, 109(507), 1302-1318. http://dx.doi.org/10.1080/01621459.2014.887012.

Park, T., Shao, X., and Yao, S. (2015). Partial martingale difference correlation. Electronic Journal of Statistics, 9(1), 1492-1517. http://dx.doi.org/10.1214/15-EJS1047.

Examples

```
## Not run:
# X, Y, Z are vectors with 10 samples and 1 variable
X <- rnorm(10)
Y <- rnorm(10)
Z <- rnorm(10)

cmdm_test(X, Y, Z, type = "linmdd")

# X, Y, Z are 10 x 2 matrices with 10 samples and 2 variables
X <- matrix(rnorm(10 * 2), 10, 2)
Y <- matrix(rnorm(10 * 2), 10, 2)
Z <- matrix(rnorm(10 * 2), 10, 2)

cmdm_test(X, Y, Z, type = "pmdd")

## End(Not run)
```

`mdc` *Martingale Difference Correlation*

Description

`mdc` measures conditional mean dependence of `Y` given `X`, where each contains one variable (univariate) or more variables (multivariate).

Usage

```
mdc(X, Y, center = "U")
```

Arguments

`X`	A vector, matrix or data frame, where rows represent samples, and columns represent variables.
`Y`	A vector, matrix or data frame, where rows represent samples, and columns represent variables.
`center`	The approach for centering, including • `U`: U-centering which leads to an unbiased estimator; • `D`: double-centering which leads to a biased estimator.

Value

`mdc` returns the value of squared martingale difference correlation.

References

Shao, X., and Zhang, J. (2014). Martingale difference correlation and its use in high-dimensional variable screening. Journal of the American Statistical Association, 109(507), 1302-1318. http://dx.doi.org/10.1080/01621459.2014.887012.

Park, T., Shao, X., and Yao, S. (2015). Partial martingale difference correlation. Electronic Journal of Statistics, 9(1), 1492-1517. http://dx.doi.org/10.1214/15-EJS1047.

Examples

```
# X, Y are 10 x 2 matrices with 10 samples and 2 variables
X <- matrix(rnorm(10 * 2), 10, 2)
Y <- matrix(rnorm(10 * 2), 10, 2)

mdc(X, Y, center = "U")
mdc(X, Y, center = "D")
```

mdd — *Martingale Difference Divergence*

Description

`mdd` measures conditional mean dependence of `Y` given `X`, where each contains one variable (univariate) or more variables (multivariate).

Usage

```
mdd(X, Y, compute = "C", center = "U")
```

Arguments

`X`	A vector, matrix or data frame, where rows represent samples, and columns represent variables.
`Y`	A vector, matrix or data frame, where rows represent samples, and columns represent variables.
`compute`	The method for computation, including • `C`: computation implemented in C code; • `R`: computation implemented in R code.
`center`	The approach for centering, including • `U`: U-centering which leads to an unbiased estimator; • `D`: double-centering which leads to a biased estimator.

Value

`mdd` returns the value of squared martingale difference divergence.

References

Shao, X., and Zhang, J. (2014). Martingale difference correlation and its use in high-dimensional variable screening. Journal of the American Statistical Association, 109(507), 1302-1318. http://dx.doi.org/10.1080/01621459.2014.887012.

Park, T., Shao, X., and Yao, S. (2015). Partial martingale difference correlation. Electronic Journal of Statistics, 9(1), 1492-1517. http://dx.doi.org/10.1214/15-EJS1047.

Examples

```
# X, Y are vectors with 10 samples and 1 variable
X <- rnorm(10)
Y <- rnorm(10)

mdd(X, Y, compute = "C")
mdd(X, Y, compute = "R")

# X, Y are 10 x 2 matrices with 10 samples and 2 variables
X <- matrix(rnorm(10 * 2), 10, 2)
Y <- matrix(rnorm(10 * 2), 10, 2)

mdd(X, Y, center = "U")
mdd(X, Y, center = "D")
```

---

pmdc *Partial Martingale Difference Correlation*

Description

pmdc measures conditional mean dependence of Y given X conditioning on Z, where each contains one variable (univariate) or more variables (multivariate).

Usage

```
pmdc(X, Y, Z)
```

Arguments

X	A vector, matrix or data frame, where rows represent samples, and columns represent variables.
Y	A vector, matrix or data frame, where rows represent samples, and columns represent variables.
Z	A vector, matrix or data frame, where rows represent samples, and columns represent variables.

Value

pmdc returns the value of squared partial martingale difference correlation.

References

Park, T., Shao, X., and Yao, S. (2015). Partial martingale difference correlation. Electronic Journal of Statistics, 9(1), 1492-1517. http://dx.doi.org/10.1214/15-EJS1047.

Examples

```
# X, Y, Z are 10 x 2 matrices with 10 samples and 2 variables
X <- matrix(rnorm(10 * 2), 10, 2)
Y <- matrix(rnorm(10 * 2), 10, 2)
Z <- matrix(rnorm(10 * 2), 10, 2)

pmdc(X, Y, Z)
```

pmdd *Partial Martingale Difference Divergence*

Description

pmdd measures conditional mean dependence of Y given X conditioning on Z, where each contains one variable (univariate) or more variables (multivariate).

Usage

```
pmdd(X, Y, Z)
```

Arguments

X	A vector, matrix or data frame, where rows represent samples, and columns represent variables.
Y	A vector, matrix or data frame, where rows represent samples, and columns represent variables.
Z	A vector, matrix or data frame, where rows represent samples, and columns represent variables.

Value

pmdd returns the value of squared partial martingale difference divergence.

References

Park, T., Shao, X., and Yao, S. (2015). Partial martingale difference correlation. Electronic Journal of Statistics, 9(1), 1492-1517. http://dx.doi.org/10.1214/15-EJS1047.

Examples

```
# X, Y, Z are vectors with 10 samples and 1 variable
X <- rnorm(10)
Y <- rnorm(10)
Z <- rnorm(10)

pmdd(X, Y, Z)

# X, Y, Z are 10 x 2 matrices with 10 samples and 2 variables
X <- matrix(rnorm(10 * 2), 10, 2)
Y <- matrix(rnorm(10 * 2), 10, 2)
Z <- matrix(rnorm(10 * 2), 10, 2)

pmdd(X, Y, Z)
```

Index