Package 'HDtest'

December 30, 2016

Type Package

Title High Dimensional Hypothesis Testing for Mean Vectors, Covariance Matrices, and White Noise of Vector Time Series

Version 0.1

Date 2016-12-18

Author Meng Cao, Tong He, Wen Zhou

Maintainer Meng Cao <meng.cao@colostate.edu>

Description High dimensional testing procedures on mean, covariance and white noises.

Depends R (>= 3.2.2)

Imports checkmate (>= 1.6.0), MASS, stats, mvtnorm, foreach, doParallel, expm

License Apache License (== 2.0)

URL http://www.stat.colostate.edu/~riczw/SW.html

Repository CRAN

LazyData TRUE

RoxygenNote 5.0.1

NeedsCompilation yes

Date/Publication 2016-12-30 10:12:06

R topics documented:

`aeS` *Quantile of the absolute values of Gaussian vectors with long run covariance*

Description

This is an auxillary function to compute alpha-level quantile of absolute values of Gaussian vectors whose covariance matrices are specified by W.

Usage

```
aeS(ft, Sn, W, M, alpha)
```

Arguments

`ft`	The input multivariate time series
`Sn`	Long run covariance matrix
`W`	The inverse of covariance matrix at lag 0
`M`	Number of Gaussian vectors sampled for computing empirical quantile
`alpha`	level of significance (default 0.05)

Details

For input multivariate time series ε_t, derive

$$f_t = \{vec(\varepsilon_{t+1}\varepsilon_t^T), \ldots, vec(\varepsilon_{t+K}\varepsilon_t^T)\}^T$$

. Long run covariance matrix S_n is estimated separately using the method described in Section 2.3 in the reference below and inverse covariance matrix at lag 0 is estimated using ε_t. M independent Gaussian vectors with desired long run covariance are sampled to compute the α-level empirical quantiles for their absolute values.

Value

alpha-level quantiles for testing whether the input multivariate time series is a white noise or not, alpha is 0.05

Author(s)

Meng Cao, Wen Zhou

References

J. Chang, Q. Yao, and W. Zhou (2016) Testing for high-dimensional white noise using maximum cross correlations. *Biometrika*, to appear.

`autocovm` *Lag-k autocovariance matrix for multivariate time series*

Description

This is an auxillary function to compute the autocovariance matrix for multivariate time series at lag k.

Usage

```
autocovm(Y, k)
```

Arguments

`Y`	A multivariate time series.
`k`	The lag k.

Details

Compute the autocovariance matrix of mutlivariate time series Y at lag k.

Value

sm The autocovariace matrix at lag k.

Author(s)

Meng Cao, Wen Zhou

`CLX` *CLX-test for two sample means*

Description

Testing the equality of two high dimensional mean vectors using the testing procedure by Cai, Liu and Xia (2014).

Usage

```
CLX(X, Y, alpha, DNAME)
```

Arguments

X	The n x p data matrix from the sample 1
Y	The n x p data matrix from the sample 2.
alpha	The prescribed level of significance
DNAME	Default input.

Details

Implementing testing procedure proposed by Cai, Liu, and Xia (2014) to test the equality of two sample high dimensional mean vectors under the assumption of sparsity of signals.

Value

Value of testing statistic, p-value, alternative hypothesis, and the name of testing procedure.

Author(s)

Tong He

References

T. Cai, W. Liu, and Y. Xia (2014). Two-sample test of high dimensional means under dependence. J. R. Statist. Soc. B. 76, 349–372

CQ2 *CQ-test for two sample means*

Description

Testing the equality of two high dimensional mean vectors using the testing procedure by Chen and Qin (2010)

Usage

```
CQ2(X, Y, DNAME)
```

Arguments

X	The n x p data matrix from the sample 1
Y	The n x p data matrix from the sample 2.
DNAME	Default input.

Details

Implementing testing procedure proposed by Chen and Qin (2010) to test the equality of two sample high dimensional mean vectors under the assumption of sparsity of signals.

Value

res Value of testing statistic, alternative hypothesis, and the name of testing procedure.

Author(s)

Tong He

References

S. Chen and Y. Qin (2010). A two-sample test for high-dimensional data with applications to gene-set testing. Ann. Statist. 38, 808-835

---

equalCovs *LC-test for equality of high dimensional covariances*

Description

Testing the equality of two high dimensional covariance matrices using the testing procedure by Li and Chen (2012).

Usage

```
equalCovs(X, Y, alpha, DNAME)
```

Arguments

X	The n x p data matrix from the sample 1
Y	The n x p data matrix from the sample 2.
alpha	The prescribed level of significance
DNAME	Default input.

Details

Implementing testing procedure proposed by Li and Chen (2012) to test the equality of two sample high dimensional covariance matrices.

Value

Value of testing statistic, p-value, alternative hypothesis, and the name of testing procedure.

Author(s)

Tong He

References

J. Li and S. Chen (2012). Two sample tests for high-dimensional covariance matrices. Ann. Statist. 40, 908–940

GO26 *GO26*

Description

A list of two sample matrices sliced from GO:0016032, it is about viral reproduction

Usage

```
data(GO26)
```

Format

A list of two data objects `GO26$X` and `GO26$Y`.

GO54 *GO54*

Description

A list of two sample matrices sliced from GO:0034080, it is about CenH3-containing nucleosome assembly at centromere.

Usage

```
data(GO54)
```

Format

A list of two data objects `GO54$X` and `GO54$Y`.

oneMean *CZZZ-test for one sample mean vector*

Description

Testing the equality of high dimensional mean vector to zero using the method developed in arXiv:1406.1939 [math.ST]

Usage

```
oneMean(X, m = 2500, filter = TRUE, S = NULL, alpha = 0.05, DNAME)
```

Arguments

X	The nxp data matrix.
m	The number of Monte-Carlo samples in the test, default to be 2500
filter	A logical indicator of the filtering process, defaul to be TRUE
S	Covariance matrix of X, if not presented it will be estimated from the input sample.
alpha	The significant level of the test.
DNAME	Defaul input.

Details

Implement the method developed in arXiv:1406.1939 [math.ST] to test whether a high dimensional mean vector is zero or not, which is equivalent to test $H_0 : \mu = \mu_0$ for some prescribed value μ_0 which can be subtracted from the data. The procedure utilizes bootstrap concept and derive the critical values using independent Gaussian vectors whose covariance is estimated using sample covariance matrix.

Value

Value of testing statistics, p-values (the non-studentized statistic and the studentized statistic respectively), alternative hypothesis, and the name of testing procedure.

Author(s)

Tong He

References

J. Chang, W. Zhou and W.-X. Zhou, Simulation-Based Hypothesis Testing of High Dimensional Means Under Covariance Heterogeneity (2014), arXiv:1406.1939.

---

opbw *Optimal bandwidth to estimate long-run covariance*

Description

This is an auxillary function to estimate the bandwidth that used to estimate the long run covariance for testing multivariate white nosies.

Usage

```
opbw(X)
```

Arguments

X	pxn data matrix, with p time series of length n

Value

An optimal bandwidth.

Author(s)

Meng Cao, Wen Zhou

References

J. Chang, Q. Yao, and W. Zhou (2016) Testing for high-dimensional white noise using maximum cross correlations. *Biometrika*, to appear.

`segment` *segment*

Description

segment

Usage

```
segment(Y, mean_y, k, n, p)
```

Arguments

`Y`	pxn data matrix, with p time series of length n
`mean_y`	mean vector
`k`	lag
`n`	each time series has length n
`p`	p time series

Value

b An optimal bandwidth.

Author(s)

Meng Cao

References

J. Chang, Q. Yao, and W. Zhou (2016) Testing for high-dimensional white noise using maximum cross correlations. *Biometrika*, to appear.

testCov *Testing the equality of two sample covariance matrices in high dimension.*

Description

Testing the equality of two sample covariance matrices in high dimension using different methods.

Usage

```
testCov(X, Y, method = "ALL", J = 2500, alpha = 0.05, n.core = 1)
```

Arguments

X	the n x p training data, could be a matrix or a data.frame object.
Y	the n x p training data matrix, could be a matrix or a data.frame object.
method	a string incidating the method for the test. The current available methods are ALL, HD, LC, CLX, Scott.
J	the number of repetition in the test
alpha	the significant level of the test.
n.core	the number of cores to be used in parallel when HD is called.

Value

For any single method, the function returns an htest object.

For method ALL: A list of four htest objects.

HD refers to "Chang, J., Zhou, W., Zhou, W.-X., and Wang, L. (2016). Comparing large covariance matrices under weak conditions on the dependence structure and its application to gene clustering. Biometrics. To appear"#'

CLX refers to "Cai, T. T., Liu, W., and Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. Journal of the American Statistical Association 108, 265-277."

Sc refers to "Schott, J. R. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample size. Computational Statistics and Data Analysis 51, 6535-6542."

Author(s)

Tong He

Examples

```
data(GO54)
testCov(GO54$X, GO54$Y, method = "ALL", J = 100)
data(GO26)
testCov(GO26$X, GO26$Y, method = "ALL", J = 100)
```

testMean *Testing the equality of two sample mean vectors in high dimension.*

Description

Testing the equality of two sample mean vectors in high dimension using different methods.

Usage

```
testMean(X, Y = NULL, method = "HD", m = 2500, filter = TRUE,
  alpha = 0.05, SX = NULL, SY = NULL)
```

Arguments

X	the n x p training data matrix, could be a matrix or a data.frame object.
Y	the n x p training data matrix, if presented the method will perform a two-sample test of mean, one-sample test otherwise. Could be a matrix or a data.frame object.
method	a string incidating the method for the test. The current available methods are ALL, HD, CQ, CLX.
m	the number of repetition in the test
filter	a logical indicator of the filtering process
alpha	the significant level of the test.
SX	covariance matrix of X, if not presented it will be estimated from the input sample.
SY	covariance matrix of T, if not presented it will be estimated from the input sample.

Value

For method HD, the function returns two htest objects for non-studentized and studentized test respectively.

For method CLX and CQ, the function returns an htest object.

For method ALL: A list of four htest objects.

HD refers to arXiv:1406.1939 [math.ST]

Author(s)

Tong He

Examples

```
data(GO54)
testMean(GO54$X, m = 100, method = "HD")
testMean(GO54$X, GO54$Y, m = 100, method = "ALL")
```

`twoMeans` *CZZZ-test for two sample mean vectors*

Description

Testing the equality of two sample high dimensional mean vectors using the method developed in arXiv:1406.1939 [math.ST]

Usage

```
twoMeans(X, Y, m = 2500, filter = TRUE, SX = NULL, SY = NULL,
  alpha = 0.05, DNAME)
```

Arguments

`X`	The n x p training data matrix.
`Y`	The n x p training data matrix.
`m`	The number of repetition in the test, default to be 2500
`filter`	A logical indicator of the filtering process, default to be TRUE
`SX`	The covariance matrix of X, if not presented it will be estimated from the input sample.
`SY`	The covariance matrix of T, if not presented it will be estimated from the input sample.
`alpha`	The significant level of the test.
`DNAME`	Defaulf input.

Details

Implement the method developed in arXiv:1406.1939 [math.ST] to test whether a high dimensional mean vector is zero or not, which is equivalent to test $H_0 : \mu_1 = \mu_2$. The procedure utilizes bootstrap concept and derive the critical values using independent Gaussian vectors whose covariance is estimated using sample covariance matrix.

Value

Value of testing statistics, p-values (the non-studentized statistic and the studentized statistic respectively), alternative hypothesis, and the name of testing procedure.

Author(s)

Tong He

References

J. Chang, W. Zhou and W.-X. Zhou, Simulation-Based Hypothesis Testing of High Dimensional Means Under Covariance Heterogeneity (2014), arXiv:1406.1939.

wntest *Test_our_new*

Description

Our new method to test white noise

Usage

```
wntest(Y, M, k_max = 10, kk, type = 1, alpha = 0.05, k0 = 10,
  delta = 1.5, opt = 1)
```

Arguments

Y	is input multivaraite time series data
M	is a parameter, could be 1000, 2000 for example
k_max	is a parameter (for example default 10)
kk	is a vector of parameters, could be a scalar as well (kk = c(2:10))
type	1: wntest, 2: test_LM, 3: test_pre, 4: test_TB type = 1 need X, k_max, ,kk, M, bw, type = 2 need Y, k, type = 3: need Y, k_max, kk, type = 4: need Y
alpha	level of significance
k0	is parameter in time series PCA for transformation (default 10)
delta	is 2nd parameter in time series PCA for transformation (default 1.5)
opt	= 1, perform transformation, else do not perform transformation

Value

res white noise or not

Author(s)

Meng Cao

Examples

```
library(expm)
p = 15
n = 300
S1 = diag(1, p, p)
for(ii in c(1:p)){
for(jj in c(1:p)){
S1[ii, jj] = 0.995^(abs(ii-jj))
}
}
S11 = sqrtm(S1)
X = S11 %*% matrix(rt(n*p, df = 8), ncol = n)
```

```
k_max = 10
kk = seq(2, k_max, 2)
M = 2000
k0 = 10
delta = 1.5
alpha = 0.05
wntest(X, M, k_max, kk, type = 1, opt = 0)
```

Index