Package ‘HDtest’

October 12, 2022

Type Package

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

Version 2.1

Date 2018-9-10

Author Meng Cao, Tong He, Wen Zhou

Maintainer Wen Zhou <riczw@stat.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, fastclime, clime

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

Repository CRAN

LazyData TRUE

RoxygenNote 6.1.0

NeedsCompilation yes

Date/Publication 2018-09-13 17:00:03 UTC

R topics documented:

`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.

---

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 *Testing for multivariate or high dimensional white noise*

Description

A variety of methods to test multivariate or high-dimensional white noise, including classical methods such as the multivariate portmanteau tests, Lagrange multiplier test, as well as the new method proposed by Chang, Yao and Zhou (2017) based on maximum cross correlations.

Usage

```
wntest(Y, M, k_max = 10, kk, type = 1, alpha = 0.05, k0 = 10,
  delta = 1.5, opt = 1, lambda = 0.01, lambda_search = seq(1e-04,
  0.01, length.out = 50), fold = 5, S1 = NULL, cv_opt = NULL)
```

Arguments

`Y`	A p by n data matrix with p time series of length n.
`M`	Number of bootstrap replicates, ex. 2000.
`k_max`	The largest time lag to be tested for white noise (default is 10).
`kk`	A vector of time lags using for test (ex. $kk = \texttt{seq}(2, 10, \texttt{by} = 2)$), scalar is allowed and the largest kk must be less than k_max.
`type`	Tests to be performed: 1 is coded for the newly proposed maximum cross-correlation-based test for high-dimensional white noise by Chang, Yao and Zhou (2017); 2 is coded for the Lagrange multiplier test; 3 is coded for the three portmanteau tests, where results for both χ^2 and normal approximations are reported; and 4 is coded for the Tiao-Box likelihoood ratio-based test.
`alpha`	Level of significance (default is 0.05).
`k0`	A parameter in time series PCA for pre-transformation (default is 10).
`delta`	The thresholding parameter in time series PCA for pre-transformation (default is 1.5).
`opt`	Options for pre-transformation of time series. That is, one considers a transformation matrix $A_{n\times p}$ and corresponding pre-transformed data AY. For parameter 'opt', 1 is coded for performing the transformation using package 'fastclime' and user-specific tuning parameter λ for estimating the contempaneous correlations; 2 is coded for performing the transformation using the sample covariance; 3 is coded for performing the transformation using package 'clime' with build-in cross validation on the tuning parameter for estimating the contempaneous correlations; 4 is coded for performing the transformation using 'fastclime' with cross validation on the tuning parameter for estimating the contempaneous correlations; and else do not perform the transformation.
`lambda`	The tuning parameter used in package 'fastclime', which is required for 'opt=1'. The default value is 0.1.

lambda_search	The tuning parameters search for package 'fastclime', which is required for 'opt=4' (default is $\texttt{seq}(1e-4, 1e-2, \texttt{length.out}=50)$).
fold	Number of folds used in cross validations (default is 5).
S1	True contempaneous $p \times p$ covariance matrix of the data if it is known in advance. If provided, pre-transformation will use S1 instead of options in 'opt'.
cv_opt	Specify which tuning parameter and the corresponding estimated contempenous correlation (and the precision) matrix to be used for the pre-transformation. For example, 'cv_opt = 2' will choose λ and the estimated contempenous correlation (and the precision) matrix with the second smallest cross validation error (default value is 1, the minimun error).

Details

For a p-dimensional weakly stationary time series ε_t with mean zero, denote by $\Sigma(k) = \text{cov}(\varepsilon_{t+k}, \varepsilon_t)$ and $\Gamma(k) = \text{diag}\{\Sigma(0)\}^{-1/2}\Sigma(k)\text{diag}\{\Sigma(0)\}^{-1/2}$, respectively, the autocovariance and the autocorrelation of at lag k. With the available observations $\varepsilon_1, \ldots, \varepsilon_n$, let

$$\widehat{\Gamma}(k) \equiv \{\widehat{\rho}_{ij}(k)\}_{1\le i,j\le p} = \text{diag}\{\widehat{\Sigma}(0)\}^{-1/2}\widehat{\Sigma}(k)\text{diag}\{\widehat{\Sigma}(0)\}^{-1/2}$$

be the sample autocorrelation matrix at lag k, where $\widehat{\Sigma}(k)$ is the sample autocovariance matrix. Consider the hypothesis testing problem

$$H_0 : \{\varepsilon_t\} \text{ is white noise} \quad \text{versus} \quad H_1 : \{\varepsilon_t\} \text{ is not white noise.}$$

To test the above hypothesis of multivariate or high dimensional white noise, we include the traditional portmanteau tests with test statistics: $Q_1 = n\sum_{k=1}^{K} \text{tr}\{\widehat{\Gamma}(k)^T\widehat{\Gamma}(k)\}$, $Q_2 = n^2\sum_{k=1}^{K} \text{tr}\{\widehat{\Gamma}(k)^T\widehat{\Gamma}(k)\}/(n-k)$, and $Q_3 = n\sum_{k=1}^{K} \text{tr}\{\widehat{\Gamma}(k)^T\widehat{\Gamma}(k)\} + p^2K(K+1)/(2n)$. Also, we include the Lagranage multiplier test as well as the Tiao-Box likelihood ratio test. For the portmanteau tests, both χ^2-approximation and normal approximation are reported.

Since $\Gamma(k) \equiv 0$ for any $k \ge 1$ under H_0, the newly proposed maximum cross-correlation-based test uses statistic

$$T_n = \max_{1\le k\le K} T_{n,k},$$

where $T_{n,k} = \max_{1\le i,j\le p} n^{1/2}|\widehat{\rho}_{ij}(k)|$ and $K \ge 1$ is prescribed. Null is rejected whenever $T_n > \text{cv}_\alpha$, where $\text{cv}_\alpha > 0$ is the critical value determined by novel bootstrap method proposed by Chang, Yao and Zhou (2017) with no further assumptions on the data structures.

Value

res	Test output: fail to reject (coded as 0) or reject (coded as 1).
p_value	p-values or approximated p-value.
M1	Square root of the estimated contempenous precision matrix if pre-transfermation was applied.

Author(s)

Meng Cao, Wen Zhou

References

Chang, J., Yao, Q. and Zhou, W., 2017. Testing for high-dimensional white noise using maximum cross-correlations. Biometrika, 104(1): 111-127.

Cai, T.T., Liu, W., and Luo, X., 2011. A constrained L1 minimization approach for sparse precision matrix estimation. Journal of the American Statistical Association 106(494): 594-607.

Lutkepohl, H., 2005. New introduction to multiple time series analysis. Springer Science & Business Media.

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 = 500
k0 = 10
delta = 1.5
alpha = 0.05
wntest(X, M, k_max, kk, type = 1, opt = 0)
## Not run:
wntest(X, M, k_max, kk, type = 1, opt = 4, cv_opt = 1)

## End(Not run)
```

Index