Package 'Docovt'

June 30, 2025

Title Distributed Online Covariance Matrix Tests

Date 2025-06-29

Version 0.2

Description Distributed Online Covariance Matrix Tests 'Docovt' is a powerful tool designed to efficiently process and analyze distributed datasets. It enables users to perform covariance matrix tests in an online, distributed manner, making it highly suitable for large-scale data analysis. By leveraging advanced computational techniques, 'Docovt' ensures robust and scalable solutions for statistical analysis, particularly in scenarios where data is dispersed across multiple nodes or sources. This package is ideal for researchers and practitioners working with high-dimensional data, providing a flexible and efficient framework for covariance matrix estimation and hypothesis testing. The philosophy of 'Docovt' is described in Guo G.(2025) <doi:10.1016/j.physa.2024.130308>.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.2

Imports stats

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Guangbao Guo [aut, cre] (ORCID:
<https://orcid.org/0000-0002-4115-6218>),
Congfan Zhang [aut]

Maintainer Guangbao Guo <ggb11111111@163.com>

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2025-06-30 02:10:02 UTC

Contents

CLX *Two-Sample Covariance Test by Cai, Liu and Xia (2013)*

Description

Given two sets of data matrices X and Y, where X is an n1 rows and p cols matrix and Y is an n2 rows and p cols matrix, we conduct hypothesis testing of the covariance matrix between two samples. The null hypothesis is:

$$H_0 : \Sigma_1 = \Sigma_2$$

Σ_1 and Σ_2 are the sample covariance matrices of X and Y respectively. This test method is based on the test method proposed by Cai, Liu and Xia (2013). When the pval value is less than the significance coefficient (generally 0.05), the null hypothesis is rejected.

Usage

```
CLX(X,Y)
```

Arguments

X	A matrix of n1 by p
Y	A matrix of n2 by p

Value

stat	a test statistic value.
pval	a test p_value.

References

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(501):265-277.

Examples

```
## generate X and Y.
p= 500;  n1 = 100; n2 = 150
X=matrix(rnorm(n1*p), ncol=p)
Y=matrix(rnorm(n2*p), ncol=p)
## run test
CLX(X,Y)
```

---

cm13 *One-Sample Covariance Test by Cai and Ma (2013)*

Description

Given data, it performs 1-sample test for Covariance where the null hypothesis is

$$H_0 : \Sigma_n = \Sigma_0$$

where Σ_n is the covariance of data model and Σ_0 is a hypothesized covariance based on a procedure proposed by Cai and Ma (2013).

Usage

```
cm13(X,Sigma0, alpha)
```

Arguments

X	an $(n \times p)$ data matrix where each row is an observation.
Sigma0	a $(p \times p)$ given covariance matrix.
alpha	level of significance.

Value

a named list containing:

statistic a test statistic value.

threshold rejection criterion to be compared against test statistic.

reject a logical; TRUE to reject null hypothesis, FALSE otherwise.

Examples

```
## generate data from multivariate normal with trivial covariance.
p = 5;n=10
X=data = matrix(rnorm(n*p), ncol=p)
alpha=0.05
Sigma0=diag(ncol(X))
cm13(X,Sigma0, alpha)
```

cmtwo *Two-Sample Covariance Test by Cai and Ma (2013)*

Description

Given two sets of data, it performs 2-sample test for equality of covariance matrices where the null hypothesis is

$$H_0 : \Sigma_1 = \Sigma_2$$

where Σ_1 and Σ_2 represent true (unknown) covariance for each dataset based on a procedure proposed by Cai and Ma (2013). If `statistic` $>$ `threshold`, it rejects null hypothesis.

Usage

```
cmtwo(X, Y, alpha)
```

Arguments

X	an $(m \times p)$ matrix where each row is an observation from the first dataset.
Y	an $(n \times p)$ matrix where each row is an observation from the second dataset.
alpha	level of significance.

Value

a named list containing

statistic a test statistic value.

threshold rejection criterion to be compared against test statistic.

reject a logical; TRUE to reject null hypothesis, FALSE otherwise.

Examples

```
## generate 2 datasets from multivariate normal with identical covariance.
p= 5;  n1 = 100; n2 = 150; alpha=0.05
X=data1 = matrix(rnorm(n1*p), ncol=p)
Y=data2 = matrix(rnorm(n2*p), ncol=p)

# run test
cmtwo(X, Y, alpha)
```

LC *Two-Sample Covariance Test by Li and Chen (2012)*

Description

Given two sets of data matrices X and Y, where X is an n1 rows and p cols matrix and Y is an n2 rows and p cols matrix, we conduct hypothesis testing of the covariance matrix between two samples. The null hypothesis is:

$$H_0 : \Sigma_1 = \Sigma_2$$

Σ_1 and Σ_2 are the sample covariance matrices of X and Y respectively. This test method is based on the test method proposed by Li and Chen (2012). When the pval value is less than the significance coefficient (generally 0.05), the null hypothesis is rejected.

Usage

```
LC(X,Y)
```

Arguments

X	A matrix of n1 by p
Y	A matrix of n2 by p

Value

stat	a test statistic value.
pval	a test p_value.

References

Li, J. and Chen, S. X. (2012). Two sample tests for high-dimensional covariance matrices. The Annals of Statistics, 40(2):908-940.

Examples

```
## generate X and Y.
p= 500;  n1 = 100; n2 = 150
X=matrix(rnorm(n1*p), ncol=p)
Y=matrix(rnorm(n2*p), ncol=p)
## run test
LC(X,Y)
```

PEC *Two-Sample Covariance Test by Yu, Li and Xue (2022)*

Description

Given two sets of data matrices X and Y, where X is an n1 rows and p cols matrix and Y is an n2 rows and p cols matrix,, we conduct hypothesis testing of the covariance matrix between two samples. The null hypothesis is:

$$H_0 : \Sigma_1 = \Sigma_2$$

Σ_1 and Σ_2 are the sample covariance matrices of X and Y respectively. This test method is based on the test method proposed by Yu, Li and Xue (2022). When the pval value is less than the significance coefficient (generally 0.05), the null hypothesis is rejected.

Usage

```
PEC(X,Y)
```

Arguments

X	A matrix of n1 by p
Y	A matrix of n2 by p

Value

stat	a test statistic value.
pval	a test p_value.

References

Yu, X., Li, D., and Xue, L. (2022). Fisher's combined probability test for high-dimensional covariance matrices. Journal of the American Statistical Association, (in press):1-14.

Examples

```
## generate X and Y.
p= 500;  n1 = 100; n2 = 150
X=matrix(rnorm(n1*p), ncol=p)
Y=matrix(rnorm(n2*p), ncol=p)
## run test
PEC(X,Y)
```

PECO *Two-Sample Covariance Test by Yu, Li, Xue and Li(2022)*

Description

Given two sets of data matrices X and Y, where X is an n1 rows and p cols matrix and Y is an n2 rows and p cols matrix, we conduct hypothesis testing of the covariance matrix between two samples. The null hypothesis is:

$$H_0 : \Sigma_1 = \Sigma_2$$

Σ_1 and Σ_2 are the sample covariance matrices of X and Y respectively. This test method is based on the test method proposed by Yu, Li, Xue and Li (2022). When the pval value is less than the significance coefficient (generally 0.05), the null hypothesis is rejected.

Usage

```
PECO(X,Y,delta = NULL)
```

Arguments

X	A matrix of n1 by p
Y	A matrix of n2 by p
delta	A scalar used as the threshold for building PE components, usually the default value.

Value

stat	a test statistic value.
pval	a test p_value.

References

Yu, X., Li, D., Xue, L., and Li, R. (2022). Power-enhanced simultaneous test of high-dimensional mean vectors and covariance matrices with application to gene-set testing. Journal of the American Statistical Association, (in press):1-14.

Examples

```
## generate X and Y.
p= 500;  n1 = 100; n2 = 150
X=matrix(rnorm(n1*p), ncol=p)
Y=matrix(rnorm(n2*p), ncol=p)
## run test
PECO(X,Y)
```

PEF *Two-Sample Covariance Test by Yu, Li and Xue (2022)*

Description

Given two sets of data matrices X and Y, where X is an n1 rows and p cols matrix and Y is an n2 rows and p cols matrix,, we conduct hypothesis testing of the covariance matrix between two samples. The null hypothesis is:

$$H_0 : \Sigma_1 = \Sigma_2$$

Σ_1 and Σ_2 are the sample covariance matrices of X and Y respectively. This test method is based on the test method proposed by Yu, Li and Xue (2022). When the pval value is less than the significance coefficient (generally 0.05), the null hypothesis is rejected.

Usage

```
PEF(X,Y)
```

Arguments

X	A matrix of n1 by p
Y	A matrix of n2 by p

Value

stat	a test statistic value.
pval	a test p_value.

References

Yu, X., Li, D., and Xue, L. (2022). Fisher's combined probability test for high-dimensional covariance matrices. Journal of the American Statistical Association, (in press):1-14.

Examples

```
## generate X and Y.
p= 500;  n1 = 100; n2 = 150
X=matrix(rnorm(n1*p), ncol=p)
Y=matrix(rnorm(n2*p), ncol=p)
## run test
PEF(X,Y)
```

syk *One-Sample Covariance Test by Srivastava, Yanagihara, and Kubokawa (2014)*

Description

Given data, it performs 1-sample test for Covariance where the null hypothesis is

$$H_0 : \Sigma_n = \Sigma_0$$

where Σ_n is the covariance of data model and Σ_0 is a hypothesized covariance based on a procedure proposed by Srivastava, Yanagihara, and Kubokawa (2014).

Usage

```
syk(data, Sigma0, alpha)
```

Arguments

`data`	an $(n \times p)$ data matrix where each row is an observation.
`Sigma0`	a $(p \times p)$ given covariance matrix.
`alpha`	level of significance.

Value

a named list containing

statistic a test statistic value.

threshold rejection criterion to be compared against test statistic.

reject a logical; `TRUE` to reject null hypothesis, `FALSE` otherwise.

Examples

```
## generate data from multivariate normal with trivial covariance.
p = 5;n=10
data = matrix(rnorm(n*p), ncol=p)
alpha=0.05
Sigma0=diag(ncol(data))
## run the test
syk(data, Sigma0, alpha)
```

Index