

Package ‘HDMedian’

January 13, 2024

Type Package
Title High-dimensional inference based on geometric median
Version 0.1.0
Author Liuhua Peng, Guanghui Cheng
Maintainer Liuhua Peng <liuhua.peng@unimelb.edu.au>
Description This package implements several approaches for high-dimensional inference based on the spatial median.
License N/A
Encoding UTF-8
LazyData true
RoxygenNote 7.2.3

R topics documented:

| | |
|----------------------------------------------|-----------|
| MANOVA_CaiXia_Ip | 1 |
| MANOVA_Mean_Lin2021 | 2 |
| MANOVA_Median | 4 |
| SCI_TwoSample_HD_Mean_Konietschke | 5 |
| SCI_TwoSample_HD_Mean_XueYao | 6 |
| SCI_TwoSample_HD_Median | 7 |
| Test_TwoSample_HD_Mean_Konietschke | 8 |
| Test_TwoSample_HD_Mean_XueYao | 10 |
| Test_TwoSample_HD_Median | 11 |
| Index | 13 |

| | |
|------------------|-----------------------------------------------------------------------------------------------------------|
| MANOVA_CaiXia_Ip | <i>Cai and Xia (2014)’s MANOVA test for high-dimensional means with identity transformed observations</i> |
|------------------|-----------------------------------------------------------------------------------------------------------|

Description

Let μ_1, \dots, μ_K be the population means of K groups. The hypothesis testing problem is

$$H_0 : \mu_1 = \dots = \mu_K \text{ versus } H_1 : H_0 \text{ is not true.}$$

Usage

```
MANOVA_CaiXia_Ip(dat, K, N, p, alpha = 0.05)
```

Arguments

| | |
|-------|------------------------------------------------------------|
| dat | a list of data matrices, with each element from each group |
| K | the number of groups |
| N | a vector contains the number of observation in each group |
| p | dimensions of the data |
| alpha | significance level of the test |

Details

This function implements the MANOVA test in Cai and Xia (2014) with identity transformed observations

Value

A list containing the following objects:

- alpha - the pre-specified significance level
- test_stat_Ip - value of the test statistic
- crit_val - value of the critical value
- test_res - test result, 1 (reject) or 0 (not reject)

Examples

```
X1 <- matrix(rnorm(5000, 0, 1), 50, 100)
X2 <- matrix(rnorm(5000, 0.2, 1), 50, 100)
X3 <- matrix(rnorm(5000, 0.4, 1), 50, 100)
dat <- list(X1, X2, X3)
N <- c(50, 50, 50)
MANOVA_CaiXia_Ip(dat, K=3, N, p=100, alpha=0.05)
```

| | |
|---------------------|-----------------------------------------------------------------------------------------------------|
| MANOVA_Mean_Lin2021 | <i>Lin, Lopes and Muller (2021)'s MANOVA test for high-dimensional means based on sample means.</i> |
|---------------------|-----------------------------------------------------------------------------------------------------|

Description

Let μ_1, \dots, μ_K be the population means of K groups. The hypothesis testing problem is

$$H_0 : \mu_1 = \dots = \mu_K \text{ versus } H_1 : H_0 \text{ is not true.}$$

The p-value of the test is approximated by bootstrapping centralized normal random vectors with estimated covariance matrices. The results include both approaches with and without standardization.

Usage

```
MANOVA_Mean_Lin2021(dat, K, N, p, n_boot = 400)
```

Arguments

| | |
|---------------------|------------------------------------------------------------|
| <code>dat</code> | a list of data matrices, with each element from each group |
| <code>K</code> | the number of groups |
| <code>N</code> | a vector contains the number of observation in each group |
| <code>p</code> | dimensions of the data |
| <code>n_boot</code> | the number of bootstrap iterations |

Details

The results include both approaches with and without standardization. The approach with standardization refers to the case of $\tau = 0$ in Lin, Lopes and Muller (2021), while the approach without standardization refers to the case of $\tau = 1$ in the paper.

Value

A list containing the following objects:

- `test_p_value_non_std` - p-value of the test without standardization
- `test_p_value_std` - p-value of the test with standardization
- `test_stat_mean_non_std` - value of the test statistic without standardization
- `test_stat_mean_std` - value of the test statistic with standardization
- `res_mean_boot_non_std` - a vector of length `n_boot` containing values of the bootstrap version of the test statistic without standardization
- `res_mean_std_boot` - a vector of length `n_boot` containing values of the bootstrap version of the test statistic with standardization

Examples

```
X1 <- matrix(rnorm(5000, 0, 1), 50, 100)
X2 <- matrix(rnorm(5000, 0.2, 1), 50, 100)
X3 <- matrix(rnorm(5000, 0.4, 1), 50, 100)
dat <- list(X1, X2, X3)
N <- c(50, 50, 50)
MANOVA_Mean_Lin2021(dat, K=3, N, p=100, n_boot=400)
```

| | |
|---------------|-----------------------------------------------------------------------------|
| MANOVA_Median | <i>MANOVA test for high-dimensional medians based on geometric medians.</i> |
|---------------|-----------------------------------------------------------------------------|

Description

Let m_1, \dots, m_K be the population medians of K groups. The hypothesis testing problem is

$$H_0 : m_1 = \dots = m_K \text{ versus } H_1 : H_0 \text{ is not true.}$$

The p-value of the test is approximated by the multiplier bootstrap. The results include both approaches with and without standardization. There are two different bootstrap approaches for the standardized statistic, one standardization is based on the original data (std_ver1), the other standardization method is based on the bootstrap sample (std_ver2).

Usage

```
MANOVA_Median(dat, K, N, p, n_boot = 400)
```

Arguments

| | |
|--------|------------------------------------------------------------|
| dat | a list of data matrices, with each element from each group |
| K | the number of groups |
| N | a vector contains the number of observation in each group |
| p | dimensions of the data |
| n_boot | the number of bootstrap iterations |

Details

The results include both approaches with and without standardization. The approach with standardization refers to the case of $\tau = 0$ in Cheng, Lin and Peng (2023), while the approach without standardization refers to the case of $\tau = 1$ in the paper. There are two different bootstrap approaches for the standardized statistic, one standardization is based on the original data (std_ver1), the other standardization method is based on the bootstrap sample (std_ver2).

Value

A list containing the following objects:

- test_p_value_non_std - p-value of the test without standardization
- test_p_value_std_ver1 - p-value of the test with standardization based on the original data
- test_p_value_std_ver2 - p-value of the test with standardization based on the the bootstrap sample
- test_stat_median_non_std - value of the test statistic without standardization
- test_stat_median_std - value of the test statistic with standardization
- res_median_boot_non_std - a vector of length n_boot containing values of the bootstrap version of the test statistic without standardization
- res_median_boot_std_ver1 - a vector of length n_boot containing values of the bootstrap version of the test statistic with standardization based on the original data
- res_median_boot_std_ver2 - a vector of length n_boot containing values of the bootstrap version of the test statistic with standardization based on the the bootstrap sample

Examples

```

X1 <- matrix(rnorm(5000, 0, 1), 50, 100)
X2 <- matrix(rnorm(5000, 0.2, 1), 50, 100)
X3 <- matrix(rnorm(5000, 0.4, 1), 50, 100)
dat <- list(X1, X2, X3)
N <- c(50, 50, 50)
MANOVA_Median(dat, K=3, N, p=100, n_boot=400)

```

SCI_TwoSample_HD_Mean_Konietschke

Konietschke et al.(2021)'s approach for constructing simultaneous confidence intervals for the difference between two-sample high-dimensional means

Description

Let μ_1 and μ_2 be the population medians of the two groups. The simultaneous confidence intervals is valid for all components of $\mu_1 - \mu_2$.

Usage

```

SCI_TwoSample_HD_Mean_Konietschke(
  X1,
  X2,
  n1,
  n2,
  p,
  n_boot = 1000,
  level = 0.95
)

```

Arguments

| | |
|--------|---------------------------------------------------------------|
| X1 | a $n_1 \times p$ matrix containing data from the first group |
| X2 | a $n_2 \times p$ matrix containing data from the second group |
| n1 | the sample size of the first group |
| n2 | the sample size of the second group |
| p | dimensions of the data |
| n_boot | number of bootstrap iterations |
| level | confidence level of the confidence intervals |

Details

As in Konietschke et al. (2021), the confidence intervals are built upon standardized mean difference. The limits of the intervals are approximated by multiplier bootstrap with Rademacher weights.

Value

A list containing the following objects:

- level - the pre-specified confidence level
- res_SCI_mean_std - a $2 \times p$ matrix containing the simultaneous confidence intervals based on standardized means, the first row contains the lower bounds and the second contains the upper bounds

Examples

```
X1 <- matrix(rnorm(5000, 0, 1), 50, 100)
X2 <- matrix(rnorm(5000, 0.2, 1), 50, 100)
SCI_TwoSample_HD_Mean_Konietschke(X1, X2, 50, 50, 100)
```

SCI_TwoSample_HD_Mean_XueYao

Xue and Yao (2020)'s approach for constructing simultaneous confidence intervals for the difference between two-sample high-dimensional means

Description

Let μ_1 and μ_2 be the population medians of the two groups. The simultaneous confidence intervals is valid for all components of $\mu_1 - \mu_2$.

Usage

```
SCI_TwoSample_HD_Mean_XueYao(X1, X2, n1, n2, p, n_boot = 1000, level = 0.95)
```

Arguments

| | |
|--------|---------------------------------------------------------------|
| X1 | a $n_1 \times p$ matrix containing data from the first group |
| X2 | a $n_2 \times p$ matrix containing data from the second group |
| n1 | the sample size of the first group |
| n2 | the sample size of the second group |
| p | dimensions of the data |
| n_boot | number of bootstrap iterations |
| level | confidence level of the confidence intervals |

Details

As in Xue and Yao (2020), the confidence intervals are built upon non-standardized mean difference. The limits of the intervals are approximated by multiplier bootstrap with normal weights.

Value

A list containing the following objects:

- level - the pre-specified confidence level
- res_SCI_mean_non_std - a $2 \times p$ matrix containing the simultaneous confidence intervals based on non-standardized means, the first row contains the lower bounds and the second contains the upper bounds

Examples

```
X1 <- matrix(rnorm(5000, 0, 1), 50, 100)
X2 <- matrix(rnorm(5000, 0.2, 1), 50, 100)
SCI_TwoSample_HD_Mean_XueYao(X1, X2, 50, 50, 100)
```

SCI_TwoSample_HD_Median

Simultaneous confidence intervals for the difference between two-sample high-dimensional medians based on geometric medians

Description

Let m_1 and m_2 be the population medians of the two groups. The simultaneous confidence intervals is valid for all components of $m_1 - m_2$. The limits of the intervals are approximated by multiplier bootstrap. The results include both approaches with and without standardization. There are two different bootstrap approaches for the standardized statistic, one standardization is based on the original data (std_ver1), the other standardization method is based on the bootstrap sample (std_ver2).

Usage

```
SCI_TwoSample_HD_Median(
  X1,
  X2,
  n1,
  n2,
  p,
  weight = "Rademacher",
  n_boot = 1000,
  level = 0.95
)
```

Arguments

| | |
|----|---------------------------------------------------------------|
| X1 | a $n_1 \times p$ matrix containing data from the first group |
| X2 | a $n_2 \times p$ matrix containing data from the second group |
| n1 | the sample size of the first group |
| n2 | the sample size of the second group |
| p | dimensions of the data |

| | |
|--------|------------------------------------------------------------------------------------------------|
| weight | specify the type of weight variable in the multiplier bootstrap, a common choice is Rademacher |
| n_boot | number of bootstrap iterations |
| level | confidence level of the confidence intervals |

Details

The results include both approaches with and without standardization. The approach with standardization refers to the case of $\tau = 0$ in Cheng, Lin and Peng (2023), while the approach without standardization refers to the case of $\tau = 1$ in the paper. There are two different bootstrap approaches for the standardized statistic, one standardization is based on the original data (std_ver1), the other standardization method is based on the bootstrap sample (std_ver2).

Value

A list containing the following objects:

- level - the pre-specified confidence level
- res_SCI_non_std - a $2 \times p$ matrix containing the simultaneous confidence intervals based on non-standardized medians, the first row contains the lower bounds and the second contains the upper bounds
- res_SCI_std_ver1 - a $2 \times p$ matrix containing the simultaneous confidence intervals based on standardized medians, the first row contains the lower bounds and the second contains the upper bounds. The standardization is based on the original data.
- res_SCI_std_ver2 - a $2 \times p$ matrix containing the simultaneous confidence intervals based on standardized medians, the first row contains the lower bounds and the second contains the upper bounds. The standardization is based on the bootstrap sample.

Examples

```
X1 <- matrix(rnorm(5000, 0, 1), 50, 100)
X2 <- matrix(rnorm(5000, 0.2, 1), 50, 100)
SCI_TwoSample_HD_Median(X1, X2, 50, 50, 100)
```

Test_TwoSample_HD_Mean_Konietschke

Konietschke et al.(2021)'s two-sample test for high-dimensional means

Description

Let μ_1 and μ_2 be the population means of the two groups. The hypothesis testing problem is

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 \neq \mu_2.$$

The p-value of the test is approximated by the multiplier bootstrap.

Usage

```
Test_TwoSample_HD_Mean_Konietschke(
  X1,
  X2,
  n1,
  n2,
  p,
  n_boot = 1000,
  alpha = 0.05
)
```

Arguments

| | |
|--------|---------------------------------------------------------------|
| X1 | a $n_1 \times p$ matrix containing data from the first group |
| X2 | a $n_2 \times p$ matrix containing data from the second group |
| n1 | the sample size of the first group |
| n2 | the sample size of the second group |
| p | dimensions of the data |
| n_boot | number of bootstrap iterations |
| alpha | significance level of the test |

Details

As in Konietschke et al. (2021), the test statistic is standardized, and the p-value of the test is approximated by the multiplier bootstrap with Rademacher weights.

Value

A list containing the following objects:

- alpha - the pre-specified significance level
- res_test_mean_std - result of the test
- p_val_mean_std - p-value of the test
- test_stat_mean_std - value of the test statistic
- res_mean_std_boot - a vector of length n_boot containing values of the bootstrap version of the test statistic

Examples

```
X1 <- matrix(rnorm(5000, 0, 1), 50, 100)
X2 <- matrix(rnorm(5000, 0.2, 1), 50, 100)
Test_TwoSample_HD_Mean_Konietschke(X1, X2, 50, 50, 100)
```

Test_TwoSample_HD_Mean_XueYao

Xue and Yao (2020)'s two-sample test for high-dimensional means

Description

Let μ_1 and μ_2 be the population means of the two groups. The hypothesis testing problem is

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 \neq \mu_2.$$

The p-value of the test is approximated by the multiplier bootstrap.

Usage

```
Test_TwoSample_HD_Mean_XueYao(X1, X2, n1, n2, p, n_boot = 1000, alpha = 0.05)
```

Arguments

| | |
|--------|---------------------------------------------------------------|
| X1 | a $n_1 \times p$ matrix containing data from the first group |
| X2 | a $n_2 \times p$ matrix containing data from the second group |
| n1 | the sample size of the first group |
| n2 | the sample size of the second group |
| p | dimensions of the data |
| n_boot | number of bootstrap iterations |
| alpha | significance level of the test |

Details

As in Xue and Yao (2020), the test statistic is not standardized, and the p-value of the test is approximated by the multiplier bootstrap with normal weights.

Value

A list containing the following objects:

- alpha - the pre-specified significance level
- res_test_mean_non_std - result of the test without standardization
- p_val_mean_non_std - p-value of the test without standardization
- test_stat_mean_non_std - value of the test statistic without standardization
- res_mean_non_std_boot - a vector of length n_boot containing values of the bootstrap version of the test statistic without standardization

Examples

```
X1 <- matrix(rnorm(5000, 0, 1), 50, 100)
X2 <- matrix(rnorm(5000, 0.2, 1), 50, 100)
Test_TwoSample_HD_Mean_XueYao(X1, X2, 50, 50, 100)
```

Test_TwoSample_HD_Median

Two-sample test for high-dimensional medians based on geometric medians.

Description

Let m_1 and m_2 be the population medians of the two groups. The hypothesis testing problem is

$$H_0 : m_1 = m_2 \text{ versus } H_1 : m_1 \neq m_2.$$

The p-value of the test is approximated by the multiplier bootstrap. The results include both approaches with and without standardization. There are two different bootstrap approaches for the standardized statistic, one standardization is based on the original data (std_ver1), the other standardization method is based on the bootstrap sample (std_ver2).

Usage

```
Test_TwoSample_HD_Median(
  X1,
  X2,
  n1,
  n2,
  p,
  weight = "Rademacher",
  n_boot = 1000,
  alpha = 0.05
)
```

Arguments

| | |
|--------|------------------------------------------------------------------------------------------------|
| X1 | a $n_1 \times p$ matrix containing data from the first group |
| X2 | a $n_2 \times p$ matrix containing data from the second group |
| n1 | the sample size of the first group |
| n2 | the sample size of the second group |
| p | dimensions of the data |
| weight | specify the type of weight variable in the multiplier bootstrap, a common choice is Rademacher |
| n_boot | number of bootstrap iterations |
| alpha | significance level of the test |

Details

The results include both approaches with and without standardization. The approach with standardization refers to the case of $\tau = 0$ in Cheng, Lin and Peng (2023), while the approach without standardization refers to the case of $\tau = 1$ in the paper. There are two different bootstrap approaches for the standardized statistic, one standardization is based on the original data (std_ver1), the other standardization method is based on the bootstrap sample (std_ver2).

Value

A list containing the following objects:

- `alpha` - the pre-specified significance level
- `res_test_non_std` - result of the test without standardization
- `res_test_std_ver1` - result of the test with standardization based on the original data
- `res_test_std_ver2` - result of the test with standardization based on the the bootstrap sample
- `p_val_non_std` - p-value of the test without standardization
- `p_val_std_ver1` - p-value of the test with standardization based on the original data
- `p_val_std_ver2` - p-value of the test with standardization based on the the bootstrap sample
- `test_stat_medn_non_std` - value of the test statistic without standardization
- `test_stat_medn_std` - value of the test statistic with standardization
- `res_medn_boot_non_std` - a vector of length `n_boot` containing values of the bootstrap version of the test statistic without standardization
- `res_medn_boot_std_ver1` - a vector of length `n_boot` containing values of the bootstrap version of the test statistic with standardization based on the original data
- `res_medn_boot_std_ver2` - a vector of length `n_boot` containing values of the bootstrap version of the test statistic with standardization based on the the bootstrap sample

Examples

```
X1 <- matrix(rnorm(5000, 0, 1), 50, 100)
X2 <- matrix(rnorm(5000, 0.2, 1), 50, 100)
Test_TwoSample_HD_Median(X1, X2, 50, 50, 100)
```

Index

MANOVA_CaiXia_Ip, [1](#)
MANOVA_Mean_Lin2021, [2](#)
MANOVA_Median, [4](#)

SCI_TwoSample_HD_Mean_Konietschke, [5](#)
SCI_TwoSample_HD_Mean_XueYao, [6](#)
SCI_TwoSample_HD_Median, [7](#)

Test_TwoSample_HD_Mean_Konietschke, [8](#)
Test_TwoSample_HD_Mean_XueYao, [10](#)
Test_TwoSample_HD_Median, [11](#)