Package 'trc'

April 24, 2025

Type Package

Title Truncated Rank Correlation

Version 0.2

Date 2025-04-15

Maintainer Donghyeon Yu `<dyu@inha.ac.kr>`

Description

A new measure of similarity between a pair of mass spectrometry (MS) experiments, called truncated rank correlation (TRC). To provide a robust metric of similarity in noisy high-dimensional data, TRC uses truncated top ranks (or top m-ranks) for calculating correlation. Truncated rank correlation as a robust measure of test-retest reliability in mass spectrometry data. For more details see Lim et al. (2019) <doi:10.1515/sagmb-2018-0056>.

License GPL (>= 2)

URL `https://sites.google.com/site/dhyeonyu/software`

NeedsCompilation yes

Repository CRAN

Date/Publication 2025-04-24 17:20:02 UTC

Author Johan Lim [aut],
Donghyeon Yu [aut, cre],
Hsun-chih Kuo [aut],
Hyungwon Choi [aut],
Scott Walmsley [aut]

Contents

`k_tau` *Kendall's tau for two vector observations*

Description

This function calculates the Kendall's tau for two vector observations for the purpose of checking inner calculation.

Usage

```
k_tau(X,Y)
```

Arguments

`X`	An observed data vector from the first condition.
`Y`	An observed data vector from the second condition.

Details

Kendall's tau for two vector observations.

Value

`tau`	A calculated Kendall's tau value.

References

Lim, J., Yu, D., Kuo, H., Choi, H., and Walmsely, S. (2019). Truncated Rank Correlation as a robust measure of test-retest reliability in mass spectrometry data. Statistical Applications in Genetics and Molecular Biology, 18(4).

Examples

```
p = 100
sig_z = 1.15
sig_e = 1
mu_z = 2
mu_e = 8
m0 = 30

S1 = rnorm(p,mean=mu_e,sd=sig_e)
S2 = rnorm(p,mean=mu_e,sd=sig_e)

if(m0!=0)
{
   X = mu_z + rnorm(m0,mean=0,sd=sig_z)
   indx = 1:p
   s_indx = sort(sample(indx,m0))
   S1[s_indx] = S1[s_indx] + X
```

```
   S2[s_indx] = S2[s_indx] + X
}

S1 = exp(S1)
S2 = exp(S2)

# Kendall's tau
ktau <- k_tau(S1,S2)
ktau
```

---

null_perm *Procedure for estimating the null distribution of the TRC tau with the m value chosen by the proposed rule.*

Description

Procedure for estimating the null distribution of the TRC tau with the m value chosen by the proposed rule.

Usage

```
null_perm(X,Y,nperm=1000,start=3,range_m=0.5,span=0.5,seed=21,all_m=FALSE)
```

Arguments

`X`	An observed data vector from the first condition.
`Y`	An observed data vector from the second condition.
`nperm`	the number of permutations to estimate the null distribution (default: 1000).
`start`	A lower bound of a search region for the threshold rank m (default: 3).
`range_m`	A proportion of length of X for specifying the end of the search region for m (default: 0.8).
`span`	A parameter alpha which controls the degree of smoothing in loess function.
`seed`	An initial seed for the permutation.
`all_m`	a logical flag for returning permuted TRC tau values for all m values (default: FALSE).

Details

Null distributions of the TRC tau with a given m value, the Kendall’s tau, and Pearson’s correlation are estimated by the permuted samples.

Value

`perm_trc`	A vector of TRC tau values from the permuted samples with the m value chosen by the proposed rule.
`hist_m`	A vector of the chosen m values for permutations.
`perm_ktau`	A vector of Kendall's tau values from the permuted samples.
`perm_rho`	A vector of Pearson's correlation values from the permuted samples.
`perm_trc_all_m`	A matrix of permuted TRC tau values for all m values, in which each column stores the permuted TRC tau values for corresponding m value.

References

Lim, J., Yu, D., Kuo, H., Choi, H., and Walmsely, S. (2019). Truncated Rank Correlation as a robust measure of test-retest reliability in mass spectrometry data. Statistical Applications in Genetics and Molecular Biology, 18(4).

Examples

```
p = 100
sig_z = 1.15
sig_e = 1
mu_z = 2
mu_e = 8
m0 = 30

S1 = rnorm(p,mean=mu_e,sd=sig_e)
S2 = rnorm(p,mean=mu_e,sd=sig_e)

if(m0!=0)
{
   X = mu_z + rnorm(m0,mean=0,sd=sig_z)
   indx = 1:p
   s_indx = sort(sample(indx,m0))
   S1[s_indx] = S1[s_indx] + X
   S2[s_indx] = S2[s_indx] + X
}

S1 = exp(S1)
S2 = exp(S2)

null_res = null_perm(S1,S2,nperm=1000,start=3,range_m=0.5,span=0.2,seed=21,all_m=FALSE)
```

`null_perm_m0` *Procedure for estimating the null distribution of the TRC tau with a given m value*

Description

Procedure for estimating the null distribution of the TRC tau with a given m value.

Usage

```
null_perm_m0(X,Y,nperm=1000,m=5,seed=21)
```

Arguments

`X`	An observed data vector from the first condition.
`Y`	An observed data vector from the second condition.
`nperm`	the number of permutations to estimate the null distribution (default: 1000).
`m`	A rank threshold for the calculation of TRC tau (default: 5).
`seed`	An initial seed for the permutation.

Details

Null distribution of the TRC tau with a given m value is estimated by the permuted samples.

Value

`perm_tau`	A vector of calculated TRC tau values from the permuted samples

References

Lim, J., Yu, D., Kuo, H., Choi, H., and Walmsely, S. (2019). Truncated Rank Correlation as a robust measure of test-retest reliability in mass spectrometry data. Statistical Applications in Genetics and Molecular Biology, 18(4).

Examples

```
p = 100
sig_z = 1.15
sig_e = 1
mu_z = 2
mu_e = 8
m0 = 30

S1 = rnorm(p,mean=mu_e,sd=sig_e)
S2 = rnorm(p,mean=mu_e,sd=sig_e)

if(m0!=0)
{
   X = mu_z + rnorm(m0,mean=0,sd=sig_z)
   indx = 1:p
   s_indx = sort(sample(indx,m0))
   S1[s_indx] = S1[s_indx] + X
   S2[s_indx] = S2[s_indx] + X
}
```

```
S1 = exp(S1)
S2 = exp(S2)

null_res = null_perm_m0(S1,S2,nperm=1000,m=5,seed=21)
```

`rho` *Pearson's correlation for two vector observations*

Description

This function calculates the Pearson's correlation for two vector observations for the purpose of checking inner calculation.

Usage

```
rho(X,Y)
```

Arguments

`X`	An observed data vector from the first condition.
`Y`	An observed data vector from the second condition.

Details

Pearson's correlation for two vector observations.

Value

`rho`	A calculated Pearson's correlation value.

References

Lim, J., Yu, D., Kuo, H., Choi, H., and Walmsely, S. (2019). Truncated Rank Correlation as a robust measure of test-retest reliability in mass spectrometry data. Statistical Applications in Genetics and Molecular Biology, 18(4).

Examples

```
p = 100
sig_z = 1.15
sig_e = 1
mu_z = 2
mu_e = 8
m0 = 30

S1 = rnorm(p,mean=mu_e,sd=sig_e)
S2 = rnorm(p,mean=mu_e,sd=sig_e)
```

```
if(m0!=0)
{
   X = mu_z + rnorm(m0,mean=0,sd=sig_z)
   indx = 1:p
   s_indx = sort(sample(indx,m0))
   S1[s_indx] = S1[s_indx] + X
   S2[s_indx] = S2[s_indx] + X
}

S1 = exp(S1)
S2 = exp(S2)

# Pearson's correlation
pcor = rho(S1,S2)
pcor
```

`trc_cor_test` *Procedure for calculating p-values*

Description

Procedure for calculating p-values of Pearson's rho, Kendall's tau, TRC tau for two-sided test for the null hypothesis correaltion is equal to 0 based on the estimated null distribution by permutation.

Usage

```
trc_cor_test(X,Y, nperm=10000,start=3,range_m=0.8, span=0.5, seed=21, m0=NULL)
```

Arguments

`X`	An observed data vector from the first condition.
`Y`	An observed data vector from the second condition.
`nperm`	the number of permutations to estimate the null distribution (default: 10000).
`start`	A lower bound of a search region for the threshold rank m (default: 3).
`range_m`	A proportion of length of X for specifying the end of the search region for m (default: 0.8).
`span`	A parameter alpha which controls the degree of smoothing in loess function (default: 0.5).
`seed`	An initial seed for the permutation (default: 21).
`m0`	a specific m value for p-value of the TRC tau with m (defalut: NULL (not reported)).

Details

The p-values are caculated based on the estimated null distributions of the TRC tau with a given m value, the Kendall's tau, and Pearson's correlation with the permuted samples, respectively.

Value

`measure`	a vector of calculated Pearson's rho, Kendall's tau, and TRC tau with m chosen by the proposed rule if m0 = NULL; a vector of calculated Pearson's rho, Kendall's tau, TRC tau with m0, TRC tau with m chosen by the proposed rule if m0 is specified.
`p_val`	a vector of p-values for Pearson's rho, Kendall's tau, and TRC tau with m chosen by the proposed rule if m0 = NULL; a vector of p-values for Pearson's rho, Kendall's tau, TRC tau with m0, TRC tau with m chosen by the proposed rule if m0 is specified.
`chs_m`	the chosen m value by the proposed procedure.
`mean_perm_trc`	a mean value of the estimated null distribution of TRC tau by permutation.

References

Lim, J., Yu, D., Kuo, H., Choi, H., and Walmsely, S. (2019). Truncated Rank Correlation as a robust measure of test-retest reliability in mass spectrometry data. Statistical Applications in Genetics and Molecular Biology, 18(4).

Examples

```
p = 100
sig_z = 1.15
sig_e = 1
mu_z = 2
mu_e = 8
m0 = 30

S1 = rnorm(p,mean=mu_e,sd=sig_e)
S2 = rnorm(p,mean=mu_e,sd=sig_e)

if(m0!=0)
{
   X = mu_z + rnorm(m0,mean=0,sd=sig_z)
   indx = 1:p
   s_indx = sort(sample(indx,m0))
   S1[s_indx] = S1[s_indx] + X
   S2[s_indx] = S2[s_indx] + X
}

S1 = exp(S1)
S2 = exp(S2)

trc_cor_test(S1,S2, nperm=1000,start=3,range_m=0.8, span=0.2, seed=21, m0=NULL)
```

`trc_m_search` *Procedure for the choice of m for the TRC tau*

Description

Procedure for the choice of m for the TRC tau.

Usage

```
trc_m_search(X,Y,start=3,range_m=0.8,span=0.3)
```

Arguments

`X`	An observed data vector from the first condition.
`Y`	An observed data vector from the second condition.
`start`	A lower bound of a search region for the threshold rank m (default: 3).
`range_m`	A proportion of length of X for specifying the end of the search region for m (default: 0.8).
`span`	A parameter alpha which controls the degree of smoothing in loess function.

Details

The thresholding rank m is chosen by the proposed procedure in Lim et al. (2019).

Value

`tau`	A calculated TRC tau value with the chosen m value (chs_m).
`chs_m`	the chosen m value.
`km_tau_vec`	A vector of calculated k_m * TRC tau values for the given values of m [start, floor(range_m*n)]
`km_tau_loess`	A fitted values by the local regression with loess function for km_tau_vec .

References

Lim, J., Yu, D., Kuo, H., Choi, H., and Walmsely, S. (2019). Truncated Rank Correlation as a robust measure of test-retest reliability in mass spectrometry data. Statistical Applications in Genetics and Molecular Biology, 18(4).

Examples

```
p = 100
sig_z = 1.15
sig_e = 1
mu_z = 2
mu_e = 8
m0 = 30
```

```
S1 = rnorm(p,mean=mu_e,sd=sig_e)
S2 = rnorm(p,mean=mu_e,sd=sig_e)

if(m0!=0)
{
   X = mu_z + rnorm(m0,mean=0,sd=sig_z)
   indx = 1:p
   s_indx = sort(sample(indx,m0))
   S1[s_indx] = S1[s_indx] + X
   S2[s_indx] = S2[s_indx] + X
}

S1 = exp(S1)
S2 = exp(S2)

# tau_m
trc_res = trc_m_search(S1,S2,start=3,range_m=0.8,span=0.2)
trc_res$tau
trc_res$chs_m
```

`trc_tau` *Trucated Rank Correlation*

Description

TRC tau is a robust corrleation measure based on the truncated rank values.

Usage

```
trc_tau(X,Y,m=5)
```

Arguments

`X`	An observed data vector from the first condition.
`Y`	An observed data vector from the second condition.
`m`	A rank threshold for the calculation of TRC tau.

Details

Given a rank threshold m, trc_tau calculates the TRC tau value.

Value

`tau`	A calculated TRC tau value.

References

Lim, J., Yu, D., Kuo, H., Choi, H., and Walmsely, S. (2019). Truncated Rank Correlation as a robust measure of test-retest reliability in mass spectrometry data. Statistical Applications in Genetics and Molecular Biology, 18(4).

Examples

```
p = 100
sig_z = 1.15
sig_e = 1
mu_z = 2
mu_e = 8
m0 = 30

S1 = rnorm(p,mean=mu_e,sd=sig_e)
S2 = rnorm(p,mean=mu_e,sd=sig_e)

if(m0!=0)
{
   X = mu_z + rnorm(m0,mean=0,sd=sig_z)
   indx = 1:p
   s_indx = sort(sample(indx,m0))
   S1[s_indx] = S1[s_indx] + X
   S2[s_indx] = S2[s_indx] + X
}

S1 = exp(S1)
S2 = exp(S2)

tau0 = trc_tau(S1,S2,m=m0)
tau0
```

Index