Package 'SetTest'

July 15, 2024

Type Package

Title Group Testing Procedures for Signal Detection and Goodness-of-Fit

Version 0.3.0

Author Hong Zhang and Zheyang Wu

Maintainer Hong Zhang <hzhang@wpi.edu>

Description It provides cumulative distribution function (CDF), quantile, p-value, statistical power calculator and random number generator for a collection of group-testing procedures, including the Higher Criticism tests, the one-sided Kolmogorov-Smirnov tests, the one-sided Berk-Jones tests, the one-sided phi-divergence tests, etc. The input are a group of p-values. The null hypothesis is that they are i.i.d. Uniform(0,1). In the context of signal detection, the null hypothesis means no signals. In the context of the goodness-of-fit testing, which contrasts a group of i.i.d. random variables to a given continuous distribution, the input p-values can be obtained by the CDF transformation. The null hypothesis means that these random variables follow the given distribution. For reference, see [1]Hong Zhang, Jiashun Jin and Zheyang Wu. ``Distributions and power of optimal signal-detection statistics in finite case'', IEEE Transactions on Signal Processing (2020) 68, 1021-1033; [2] Hong Zhang and Zheyang Wu. ``The general goodness-of-fit tests for correlated data'', Computational Statistics & Data Analysis (2022) 167, 107379.

License GPL-2

Encoding UTF-8

RoxygenNote 7.2.0

NeedsCompilation no

Repository CRAN

Date/Publication 2024-07-15 03:40:02 UTC

Contents

`pbj` *CDF of Berk-Jones statitic under the null hypothesis.*

Description

CDF of Berk-Jones statitic under the null hypothesis.

Usage

```
pbj(q, M, k0, k1, onesided = FALSE, method = "ecc", ei = NULL)
```

Arguments

`q`	- quantile, must be a scalar.
`M`	- correlation matrix of input statistics (of the input p-values).
`k0`	- search range starts from the k0th smallest p-value.
`k1`	- search range ends at the k1th smallest p-value.
`onesided`	- TRUE if the input p-values are one-sided.
`method`	- default = "ecc": the effective correlation coefficient method in reference 2. "ave": the average method in reference 3, which is an earlier version of reference 2. The "ecc" method is more accurate and numerically stable than "ave" method.
`ei`	- the eigenvalues of M if available.

Value

The left-tail probability of the null distribution of B-J statistic at the given quantile.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and power of optimal signal-detection statistics in finite case", IEEE Transactions on Signal Processing (2020) 68, 1021-1033 2. Hong Zhang and Zheyang Wu. "The general goodness-of-fit tests for correlated data", Computational Statistics & Data Analysis (2022) 167, 107379 3. Hong Zhang and Zheyang Wu. "Generalized Goodness-Of-Fit Tests for Correlated Data", arXiv:1806.03668.

See Also

`stat.bj` for the definition of the statistic.

Examples

```
pval <- runif(10)
bjstat <- stat.phi(pval, s=1, k0=1, k1=10)$value
pbj(q=bjstat, M=diag(10), k0=1, k1=10)
```

---

phc — *CDF of Higher Criticism statistic under the null hypothesis.*

Description

CDF of Higher Criticism statistic under the null hypothesis.

Usage

```
phc(q, M, k0, k1, onesided = FALSE, method = "ecc", ei = NULL)
```

Arguments

`q`	- quantile, must be a scalar.
`M`	- correlation matrix of input statistics (of the input p-values).
`k0`	- search range starts from the k0th smallest p-value.
`k1`	- search range ends at the k1th smallest p-value.
`onesided`	- TRUE if the input p-values are one-sided.
`method`	- default = "ecc": the effective correlation coefficient method in reference 2. "ave": the average method in reference 3, which is an earlier version of reference 2. The "ecc" method is more accurate and numerically stable than "ave" method.
`ei`	- the eigenvalues of M if available.

Value

The left-tail probability of the null distribution of HC statistic at the given quantile.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and power of optimal signal-detection statistics in finite case", IEEE Transactions on Signal Processing (2020) 68, 1021-1033 2. Hong Zhang and Zheyang Wu. "The general goodness-of-fit tests for correlated data", Computational Statistics & Data Analysis (2022) 167, 107379 3. Hong Zhang and Zheyang Wu. "Generalized Goodness-Of-Fit Tests for Correlated Data", arXiv:1806.03668.

See Also

`stat.hc` for the definition of the statistic.

Examples

```
pval <- runif(10)
hcstat <- stat.phi(pval, s=2, k0=1, k1=5)$value
phc(q=hcstat, M=diag(10), k0=1, k1=10)
```

---

power.bj — *Statistical power of Berk and Jones test.*

Description

Statistical power of Berk and Jones test.

Usage

```
power.bj(
  alpha,
  n,
  beta,
  method = "gaussian-gaussian",
  eps = 0,
  mu = 0,
  df = 1,
  delta = 0
)
```

Arguments

`alpha`	- type-I error rate.
`n`	- dimension parameter, i.e. the number of input statitics to construct B-J statistic.
`beta`	- search range parameter. Search range = (1, beta*n). Beta must be between 1/n and 1.
`method`	- different alternative hypothesis, including mixtures such as, "gaussian-gaussian", "gaussian-t", "t-t", "chisq-chisq", and "exp-chisq". By default, we use Gaussian mixture.

eps	- mixing parameter of the mixture.
mu	- mean of non standard Gaussian model.
df	- degree of freedom of t/Chisq distribution and exp distribution.
delta	- non-cenrality of t/Chisq distribution.

Details

We consider the following hypothesis test,

$$H_0 : X_i \sim F, H_a : X_i \sim G$$

Specifically, $F = F_0$ and $G = (1 - \epsilon)F_0 + \epsilon F_1$, where ϵ is the mixing parameter, F_0 and F_1 is speified by the "method" argument:

"gaussian-gaussian": F_0 is the standard normal CDF and $F = F_1$ is the CDF of normal distribution with μ defined by mu and $\sigma = 1$.

"gaussian-t": F_0 is the standard normal CDF and $F = F_1$ is the CDF of t distribution with degree of freedom defined by df.

"t-t": F_0 is the CDF of t distribution with degree of freedom defined by df and $F = F_1$ is the CDF of non-central t distribution with degree of freedom defined by df and non-centrality defined by delta.

"chisq-chisq": F_0 is the CDF of Chisquare distribution with degree of freedom defined by df and $F = F_1$ is the CDF of non-central Chisquare distribution with degree of freedom defined by df and non-centrality defined by delta.

"exp-chisq": F_0 is the CDF of exponential distribution with parameter defined by df and $F = F_1$ is the CDF of non-central Chisqaure distribution with degree of freedom defined by df and non-centrality defined by delta.

Value

Power of BJ test.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and Statistical Power of Optimal Signal-Detection Methods In Finite Cases", submitted.

2. Donoho, David; Jin, Jiashun. "Higher criticism for detecting sparse heterogeneous mixtures". Annals of Statistics 32 (2004).

3. Jager, Leah; Wellner, Jon A. "Goodness-of-fit tests via phi-divergences". Annals of Statistics 35 (2007).

4. Berk, R.H. & Jones, D.H. Z. "Goodness-of-fit test statistics that dominate the Kolmogorov statistics". Wahrscheinlichkeitstheorie verw Gebiete (1979) 47: 47.

See Also

`stat.bj` for the definition of the statistic.

Examples

```
power.bj(0.05, n=10, beta=0.5, eps = 0.1, mu = 1.2)
```

---

power.hc *Statistical power of Higher Criticism test.*

Description

Statistical power of Higher Criticism test.

Usage

```
power.hc(
  alpha,
  n,
  beta,
  method = "gaussian-gaussian",
  eps = 0,
  mu = 0,
  df = 1,
  delta = 0
)
```

Arguments

alpha	- type-I error rate.
n	- dimension parameter, i.e. the number of input statitics to construct Higher Criticism statistic.
beta	- search range parameter. Search range = (1, beta*n). Beta must be between 1/n and 1.
method	- different alternative hypothesis, including mixtures such as, "gaussian-gaussian", "gaussian-t", "t-t", "chisq-chisq", and "exp-chisq". By default, we use Gaussian mixture.
eps	- mixing parameter of the mixture.
mu	- mean of non standard Gaussian model.
df	- degree of freedom of t/Chisq distribution and exp distribution.
delta	- non-cenrality of t/Chisq distribution.

Details

We consider the following hypothesis test,

$$H_0 : X_i \sim F, H_a : X_i \sim G$$

Specifically, $F = F_0$ and $G = (1 - \epsilon)F_0 + \epsilon F_1$, where ϵ is the mixing parameter, F_0 and F_1 is speified by the "method" argument:

"gaussian-gaussian": F_0 is the standard normal CDF and $F = F_1$ is the CDF of normal distribution with μ defined by mu and $\sigma = 1$.

"gaussian-t": F_0 is the standard normal CDF and $F = F_1$ is the CDF of t distribution with degree of freedom defined by df.

"t-t": F_0 is the CDF of t distribution with degree of freedom defined by df and $F = F_1$ is the CDF of non-central t distribution with degree of freedom defined by df and non-centrality defined by delta. "chisq-chisq": F_0 is the CDF of Chisquare distribution with degree of freedom defined by df and $F = F_1$ is the CDF of non-central Chisquare distribution with degree of freedom defined by df and non-centrality defined by delta.

"exp-chisq": F_0 is the CDF of exponential distribution with parameter defined by df and $F = F_1$ is the CDF of non-central Chisqaure distribution with degree of freedom defined by df and non-centrality defined by delta.

Value

Power of HC test.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and Statistical Power of Optimal Signal-Detection Methods In Finite Cases", submitted.

2. Donoho, David; Jin, Jiashun. "Higher criticism for detecting sparse heterogeneous mixtures". Annals of Statistics 32 (2004).

See Also

`stat.hc` for the definition of the statistic.

Examples

```
power.hc(0.05, n=10, beta=0.5, eps = 0.1, mu = 1.2)
```

---

power.phi *Statistical power of phi-divergence test.*

Description

Statistical power of phi-divergence test.

Usage

```
power.phi(
  alpha,
  n,
  s,
  beta,
  method = "gaussian-gaussian",
```

```
  eps = 0,
  mu = 0,
  df = 1,
  delta = 0
)
```

Arguments

`alpha`	- type-I error rate.
`n`	- dimension parameter, i.e. the number of input statitics to construct phi-divergence statistic.
`s`	- phi-divergence parameter. s = 2 is the higher criticism statitic.s = 1 is the Berk and Jones statistic.
`beta`	- search range parameter. Search range = (1, beta*n). Beta must be between 1/n and 1.
`method`	- different alternative hypothesis, including mixtures such as, "gaussian-gaussian", "gaussian-t", "t-t", "chisq-chisq", and "exp-chisq". By default, we use Gaussian mixture.
`eps`	- mixing parameter of the mixture.
`mu`	- mean of non standard Gaussian model.
`df`	- degree of freedom of t/Chisq distribution and exp distribution.
`delta`	- non-cenrality of t/Chisq distribution.

Details

We consider the following hypothesis test,

$$H_0 : X_i \sim F, H_a : X_i \sim G$$

Specifically, $F = F_0$ and $G = (1 - \epsilon)F_0 + \epsilon F_1$, where ϵ is the mixing parameter, F_0 and F_1 is speified by the "method" argument:

"gaussian-gaussian": F_0 is the standard normal CDF and $F = F_1$ is the CDF of normal distribution with μ defined by mu and $\sigma = 1$.

"gaussian-t": F_0 is the standard normal CDF and $F = F_1$ is the CDF of t distribution with degree of freedom defined by df.

"t-t": F_0 is the CDF of t distribution with degree of freedom defined by df and $F = F_1$ is the CDF of non-central t distribution with degree of freedom defined by df and non-centrality defined by delta.

"chisq-chisq": F_0 is the CDF of Chisquare distribution with degree of freedom defined by df and $F = F_1$ is the CDF of non-central Chisquare distribution with degree of freedom defined by df and non-centrality defined by delta.

"exp-chisq": F_0 is the CDF of exponential distribution with parameter defined by df and $F = F_1$ is the CDF of non-central Chisqaure distribution with degree of freedom defined by df and non-centrality defined by delta.

Value

Power of phi-divergence test.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and Statistical Power of Optimal Signal-Detection Methods In Finite Cases", submitted.

2. Donoho, David; Jin, Jiashun. "Higher criticism for detecting sparse heterogeneous mixtures". Annals of Statistics 32 (2004).

See Also

`stat.phi` for the definition of the statistic.

Examples

```
#If the alternative hypothesis Gaussian mixture with eps = 0.1 and mu = 1.2:#
power.phi(0.05, n=10, s=2, beta=0.5, eps = 0.1, mu = 1.2)
```

---

pphi — *calculate the left-tail probability of phi-divergence under general correlation matrix.*

Description

calculate the left-tail probability of phi-divergence under general correlation matrix.

Usage

```
pphi(q, M, k0, k1, s = 2, t = 30, onesided = FALSE, method = "ecc", ei = NULL)
```

Arguments

`q`	- quantile, must be a scalar.
`M`	- correlation matrix of input statistics (of the input p-values).
`k0`	- search range starts from the k0th smallest p-value.
`k1`	- search range ends at the k1th smallest p-value.
`s`	- the phi-divergence test parameter.
`t`	- numerical truncation parameter.
`onesided`	- TRUE if the input p-values are one-sided.
`method`	- default = "ecc": the effective correlation coefficient method in reference 2. "ave": the average method in reference 3, which is an earlier version of reference 2. The "ecc" method is more accurate and numerically stable than "ave" method.
`ei`	- the eigenvalues of M if available.

Value

Left-tail probability of the phi-divergence statistics.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and power of optimal signal-detection statistics in finite case", IEEE Transactions on Signal Processing (2020) 68, 1021-1033 2. Hong Zhang and Zheyang Wu. "The general goodness-of-fit tests for correlated data", Computational Statistics & Data Analysis (2022) 167, 107379 3. Hong Zhang and Zheyang Wu. "Generalized Goodness-Of-Fit Tests for Correlated Data", arXiv:1806.03668.

Examples

```
M = toeplitz(1/(1:10)*(-1)^(0:9)) #alternating polynomial decaying correlation matrix
pphi(q=2, M=M, k0=1, k1=5, s=2)
pphi(q=2, M=M, k0=1, k1=5, s=2, method = "ecc")
pphi(q=2, M=M, k0=1, k1=5, s=2, method = "ave")
pphi(q=2, M=diag(10), k0=1, k1=5, s=2)
```

---

pphi.omni — *calculate the left-tail probability of omnibus phi-divergence statistics under general correlation matrix.*

Description

calculate the left-tail probability of omnibus phi-divergence statistics under general correlation matrix.

Usage

```
pphi.omni(q, M, K0, K1, S, t = 30, onesided = FALSE, method = "ecc", ei = NULL)
```

Arguments

`q`	- quantile, must be a scalar.
`M`	- correlation matrix of input statistics (of the input p-values).
`K0`	- vector of search range starts (from the k0th smallest p-value).
`K1`	- vector of search range ends (at the k1th smallest p-value).
`S`	- vector of the phi-divergence test parameters.
`t`	- numerical truncation parameter.
`onesided`	- TRUE if the input p-values are one-sided.
`method`	- default = "ecc": the effective correlation coefficient method in reference 2. "ave": the average method in reference 3, which is an earlier version of reference 2. The "ecc" method is more accurate and numerically stable than "ave" method.
`ei`	- the eigenvalues of M if available.

Value

Left-tail probability of omnibus phi-divergence statistics.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and power of optimal signal-detection statistics in finite case", IEEE Transactions on Signal Processing (2020) 68, 1021-1033 2. Hong Zhang and Zheyang Wu. "The general goodness-of-fit tests for correlated data", Computational Statistics & Data Analysis (2022) 167, 107379 3. Hong Zhang and Zheyang Wu. "Generalized Goodness-Of-Fit Tests for Correlated Data", arXiv:1806.03668.

Examples

```
M = matrix(0.3,10,10) + diag(1-0.3, 10)
pphi.omni(0.05, M=M, K0=rep(1,2), K1=rep(5,2), S=c(1,2))
```

`qbj` *Quantile of Berk-Jones statistic under the null hypothesis.*

Description

Quantile of Berk-Jones statistic under the null hypothesis.

Usage

```
qbj(p, M, k0, k1, onesided = FALSE, method = "ecc", ei = NULL, err_thr = 1e-04)
```

Arguments

`p`	- a scalar left probability that defines the quantile.
`M`	- correlation matrix of input statistics (of the input p-values).
`k0`	- search range starts from the k0th smallest p-value.
`k1`	- search range ends at the k1th smallest p-value.
`onesided`	- TRUE if the input p-values are one-sided.
`method`	- default = "ecc": the effective correlation coefficient method in reference 2. "ave": the average method in reference 3, which is an earlier version of reference 2. The "ecc" method is more accurate and numerically stable than "ave" method.
`ei`	- the eigenvalues of M if available.
`err_thr`	- the error threshold. The default value is 1e-4.

Value

Quantile of BJ statistics.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and power of optimal signal-detection statistics in finite case", IEEE Transactions on Signal Processing (2020) 68, 1021-1033 2. Hong Zhang and Zheyang Wu. "The general goodness-of-fit tests for correlated data", Computational Statistics & Data Analysis (2022) 167, 107379 3. Hong Zhang and Zheyang Wu. "Generalized Goodness-Of-Fit Tests for Correlated Data", arXiv:1806.03668.

See Also

stat.bj for the definition of the statistic.

Examples

```
## The 0.05 critical value of BJ statistic when n = 10:
qbj(p=.95, M=diag(10), k0=1, k1=5, onesided=FALSE)
qbj(p=1-1e-5, M=diag(10), k0=1, k1=5, onesided=FALSE, err_thr=1e-8)
```

---

qhc *Quantile of Higher Criticism statistics under the null hypothesis.*

Description

Quantile of Higher Criticism statistics under the null hypothesis.

Usage

```
qhc(p, M, k0, k1, onesided = FALSE, method = "ecc", ei = NULL, err_thr = 1e-04)
```

Arguments

p	- a scalar left probability that defines the quantile.
M	- correlation matrix of input statistics (of the input p-values).
k0	- search range starts from the k0th smallest p-value.
k1	- search range ends at the k1th smallest p-value.
onesided	- TRUE if the input p-values are one-sided.
method	- default = "ecc": the effective correlation coefficient method in reference 2. "ave": the average method in reference 3, which is an earlier version of reference 2. The "ecc" method is more accurate and numerically stable than "ave" method.
ei	- the eigenvalues of M if available.
err_thr	- the error threshold. The default value is 1e-4.

Value

Quantile of HC statistics.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and power of optimal signal-detection statistics in finite case", IEEE Transactions on Signal Processing (2020) 68, 1021-1033 2. Hong Zhang and Zheyang Wu. "The general goodness-of-fit tests for correlated data", Computational Statistics & Data Analysis (2022) 167, 107379 3. Hong Zhang and Zheyang Wu. "Generalized Goodness-Of-Fit Tests for Correlated Data", arXiv:1806.03668.

See Also

`stat.hc` for the definition of the statistic.

Examples

```
## The 0.05 critical value of HC statistic when n = 10:
qhc(p=.95, M=diag(10), k0=1, k1=5, onesided=FALSE)
qhc(p=1-1e-5, M=diag(10), k0=1, k1=5, onesided=FALSE, err_thr=1e-8)
```

---

qphi — *Quantile of phi-divergence statistic under the null hypothesis.*

Description

Quantile of phi-divergence statistic under the null hypothesis.

Usage

```
qphi(
  p,
  M,
  k0,
  k1,
  s = 2,
  t = 30,
  onesided = FALSE,
  method = "ecc",
  ei = NULL,
  err_thr = 1e-04
)
```

Arguments

`p`	- a scalar left probability that defines the quantile.
`M`	- correlation matrix of input statistics (of the input p-values).
`k0`	- search range starts from the k0th smallest p-value.
`k1`	- search range ends at the k1th smallest p-value.
`s`	- the phi-divergence test parameter.

t	- numerical truncation parameter.
onesided	- TRUE if the input p-values are one-sided.
method	- default = "ecc": the effective correlation coefficient method in reference 2. "ave": the average method in reference 3, which is an earlier version of reference 2. The "ecc" method is more accurate and numerically stable than "ave" method.
ei	- the eigenvalues of M if available.
err_thr	- the error threshold. The default value is 1e-4.

Value

Quantile of the phi-divergence statistics.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and power of optimal signal-detection statistics in finite case", IEEE Transactions on Signal Processing (2020) 68, 1021-1033 2. Hong Zhang and Zheyang Wu. "The general goodness-of-fit tests for correlated data", Computational Statistics & Data Analysis (2022) 167, 107379 3. Hong Zhang and Zheyang Wu. "Generalized Goodness-Of-Fit Tests for Correlated Data", arXiv:1806.03668.

See Also

stat.phi for the definition of the statistic.

Examples

```
qphi(p=.95, M=diag(10), k0=1, k1=5, s=2, onesided=FALSE)
qphi(p=1-1e-3, M=diag(10), k0=1, k1=5, s=2, onesided=FALSE)
qphi(p=1-1e-3, M=diag(10), k0=1, k1=5, s=2, onesided=FALSE, err_thr = 1e-6)
qphi(p=1-1e-5, M=diag(10), k0=1, k1=5, s=2, onesided=FALSE)
qphi(p=1-1e-5, M=diag(10), k0=1, k1=5, s=2, onesided=FALSE, err_thr = 1e-6)
qphi(p=1-1e-5, M=diag(10), k0=1, k1=5, s=2, onesided=FALSE, err_thr = 1e-8)
```

---

stat.bj *Construct Berk and Jones (BJ) statistics.*

Description

Construct Berk and Jones (BJ) statistics.

Usage

```
stat.bj(p, k0 = 1, k1 = NA)
```

Arguments

`p`	- vector of input p-values.
`k0`	- search range left end parameter. Default k0 = 1.
`k1`	- search range right end parameter. Default k1 = 0.5*number of input p-values.

Details

Let $p_{(i)}, i = 1, ..., n$ be a sequence of ordered p-values, the Berk and Jones statistic

$$BJ = \sqrt{2n} \max_{1 \leq i \leq \lfloor \beta n \rfloor} (-1)^j \sqrt{i/n * \log(i/n/p_{(i)}) + (1 - i/n) * \log((1 - i/n)/(1 - p_{(i)}))}$$

and when $p_{(i)} > i/n$, $j = 1$, otherwise $j = 0$.

Value

value - BJ statistic constructed from a vector of p-values.

location - the order of the p-values to obtain BJ statistic.

stat - vector of marginal BJ statistics.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and Statistical Power of Optimal Signal-Detection Methods In Finite Cases", submitted.

2. Jager, Leah; Wellner, Jon A. "Goodness-of-fit tests via phi-divergences". Annals of Statistics 35 (2007).

3. Berk, R.H. & Jones, D.H. Z. "Goodness-of-fit test statistics that dominate the Kolmogorov statistics". Wahrscheinlichkeitstheorie verw Gebiete (1979) 47: 47.

Examples

```
stat.bj(runif(10))
#When the input are statistics#
stat.test = rnorm(20)
p.test = 1 - pnorm(stat.test)
stat.bj(p.test, k0 = 2, k1 = 20)
```

`stat.hc` *Construct Higher Criticism (HC) statistics.*

Description

Construct Higher Criticism (HC) statistics.

Usage

```
stat.hc(p, k0 = 1, k1 = NA)
```

Arguments

`p`	- vector of input p-values.
`k0`	- search range left end parameter. Default k0 = 1.
`k1`	- search range right end parameter. Default k1 = 0.5*number of input p-values.

Details

Let $p_{(i)}$, $i = 1, ..., n$ be a sequence of ordered p-values, the higher criticism statistic

$$HC = \sqrt{n} \max_{1 \leq i \leq \lfloor \beta n \rfloor} [i/n - p_{(i)}]/\sqrt{p_{(i)}(1 - p_{(i)})}$$

Value

value - HC statistic constructed from a vector of p-values.

location - the order of the p-values to obtain HC statistic.

stat - vector of marginal HC statistics.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and Statistical Power of Optimal Signal-Detection Methods In Finite Cases", submitted.

2. Donoho, David; Jin, Jiashun. "Higher criticism for detecting sparse heterogeneous mixtures". Annals of Statistics 32 (2004).

Examples

```
stat.hc(runif(10))
#When the input are statistics#
stat.test = rnorm(20)
p.test = 1 - pnorm(stat.test)
stat.hc(p.test, k0 = 1, k1 = 10)
```

`stat.phi` *Construct phi-divergence statistics.*

Description

Construct phi-divergence statistics.

Usage

```
stat.phi(p, s, k0 = 1, k1 = NA)
```

Arguments

`p`	- vector of input p-values.
`s`	- phi-divergence parameter. s = 2 is the higher criticism statitic.s = 1 is the Berk and Jones statistic.
`k0`	- search range left end parameter. Default k0 = 1.
`k1`	- search range right end parameter. Default k1 = 0.5*number of input p-values.

Details

Let $p_{(i)}$, $i = 1, ..., n$ be a sequence of ordered p-values, the phi-divergence statistic

$$PHI = \sqrt{2n}/(s - s^2) \max_{1 \leq i \leq \lfloor \beta n \rfloor} (-1)^j \sqrt{1 - (i/n)^s (p_{(i)})^s - (1 - i/n)^{(1-s)} * (1 - p_{(i)})^{(1-s)}}$$

and when $p_{(i)} > i/n$, $j = 1$, otherwise $j = 0$.

Value

value - phi-divergence statistic constructed from a vector of p-values.

location - the order of the p-values to obtain phi-divergence statistic.

stat - vector of marginal phi-divergence statistics.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and Statistical Power of Optimal Signal-Detection Methods In Finite Cases", submitted.

2. Jager, Leah; Wellner, Jon A. "Goodness-of-fit tests via phi-divergences". Annals of Statistics 35 (2007).

Examples

```
stat.phi(runif(10), s = 2)
#When the input are statistics#
stat.test = rnorm(20)
p.test = 1 - pnorm(stat.test)
stat.phi(p.test, s = 0.5, k0 = 2, k1 = 5)
```

`stat.phi.omni` *calculate the omnibus phi-divergence statistics under general correlation matrix.*

Description

calculate the omnibus phi-divergence statistics under general correlation matrix.

Usage

```
stat.phi.omni(
  p,
  M,
  K0 = rep(1, 2),
  K1 = rep(length(M[1, ]), 2),
  S = c(1, 2),
  t = 30,
  onesided = FALSE,
  method = "ecc",
  ei = NULL
)
```

Arguments

`p`	- input pvalues.
`M`	- correlation matrix of input statistics (of the input p-values).
`K0`	- vector of search range starts (from the k0th smallest p-value).
`K1`	- vector of search range ends (at the k1th smallest p-value).
`S`	- vector of the phi-divergence test parameters.
`t`	- numerical truncation parameter.
`onesided`	- TRUE if the input p-values are one-sided.
`method`	- default = "ecc": the effective correlation coefficient method in reference 2. "ave": the average method in reference 3, which is an earlier version of reference 2. The "ecc" method is more accurate and numerically stable than "ave" method.
`ei`	- the eigenvalues of M if available.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and power of optimal signal-detection statistics in finite case", IEEE Transactions on Signal Processing (2020) 68, 1021-1033 2. Hong Zhang and Zheyang Wu. "The general goodness-of-fit tests for correlated data", Computational Statistics & Data Analysis (2022) 167, 107379 3. Hong Zhang and Zheyang Wu. "Generalized Goodness-Of-Fit Tests for Correlated Data", arXiv:1806.03668.

Examples

```
p.test = runif(10)
M = toeplitz(1/(1:10)*(-1)^(0:9)) #alternating polynomial decaying correlation matrix
stat.phi.omni(p.test, M=M, K0=rep(1,2), K1=rep(5,2), S=c(1,2))
```

`test.bj` *Multiple comparison test using Berk and Jones (BJ) statitics.*

Description

Multiple comparison test using Berk and Jones (BJ) statitics.

Usage

```
test.bj(prob, M, k0, k1, onesided = FALSE, method = "ecc", ei = NULL)
```

Arguments

`prob`	- vector of input p-values.
`M`	- correlation matrix of input statistics (of the input p-values).
`k0`	- search range starts from the k0th smallest p-value.
`k1`	- search range ends at the k1th smallest p-value.
`onesided`	- TRUE if the input p-values are one-sided.
`method`	- default = "ecc": the effective correlation coefficient method in reference 2. "ave": the average method in reference 3, which is an earlier version of reference 2. The "ecc" method is more accurate and numerically stable than "ave" method.
`ei`	- the eigenvalues of M if available.

Value

pvalue - the p-value of the Berk-Jones test.

bjstat - the Berk-Jones statistic.

location - the order of the input p-values to obtain BJ statistic.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and power of optimal signal-detection statistics in finite case", IEEE Transactions on Signal Processing (2020) 68, 1021-1033 2. Hong Zhang and Zheyang Wu. "The general goodness-of-fit tests for correlated data", Computational Statistics & Data Analysis (2022) 167, 107379 3. Hong Zhang and Zheyang Wu. "Generalized Goodness-Of-Fit Tests for Correlated Data", arXiv:1806.03668. 4. Leah Jager and Jon Wellner. "Goodness-of-fit tests via phi-divergences". Annals of Statistics 35 (2007).

See Also

`stat.bj` for the definition of the statistic.

Examples

```
test.bj(runif(10), M=diag(10), k0=1, k1=10)
#When the input are statistics#
stat.test = rnorm(20)
p.test = 2*(1 - pnorm(abs(stat.test)))
test.bj(p.test, M=diag(20), k0=1, k1=10)
```

`test.hc` *Multiple comparison test using Higher Criticism (HC) statitics.*

Description

Multiple comparison test using Higher Criticism (HC) statitics.

Usage

```
test.hc(prob, M, k0, k1, onesided = FALSE, method = "ecc", ei = NULL)
```

Arguments

`prob`	- vector of input p-values.
`M`	- correlation matrix of input statistics (of the input p-values).
`k0`	- search range starts from the k0th smallest p-value.
`k1`	- search range ends at the k1th smallest p-value.
`onesided`	- TRUE if the input p-values are one-sided.
`method`	- default = "ecc": the effective correlation coefficient method in reference 2. "ave": the average method in reference 3, which is an earlier version of reference 2. The "ecc" method is more accurate and numerically stable than "ave" method.
`ei`	- the eigenvalues of M if available.

Value

pvalue - The p-value of the HC test.

hcstat - HC statistic.

location - the order of the input p-values to obtain HC statistic.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and power of optimal signal-detection statistics in finite case", IEEE Transactions on Signal Processing (2020) 68, 1021-1033 2. Hong Zhang and Zheyang Wu. "The general goodness-of-fit tests for correlated data", Computational Statistics & Data Analysis (2022) 167, 107379 3. Hong Zhang and Zheyang Wu. "Generalized Goodness-Of-Fit Tests for Correlated Data", arXiv:1806.03668. 4. Donoho, David; Jin, Jiashun. "Higher criticism for detecting sparse heterogeneous mixtures". Annals of Statistics 32 (2004).

See Also

stat.hc for the definition of the statistic.

Examples

```
pval.test = runif(10)
test.hc(pval.test, M=diag(10), k0=1, k1=10)
#When the input are statistics#
stat.test = rnorm(20)
p.test = 2*(1 - pnorm(abs(stat.test)))
test.hc(p.test, M=diag(20), k0=1, k1=10)
```

---

test.phi — *Multiple comparison test using phi-divergence statistics.*

Description

Multiple comparison test using phi-divergence statistics.

Usage

```
test.phi(prob, M, k0, k1, s = 2, onesided = FALSE, method = "ecc", ei = NULL)
```

Arguments

prob	- vector of input p-values.
M	- correlation matrix of input statistics (of the input p-values).
k0	- search range starts from the k0th smallest p-value.
k1	- search range ends at the k1th smallest p-value.
s	- phi-divergence parameter. s = 2 is the higher criticism statitic.s = 1 is the Berk and Jones statistic.
onesided	- TRUE if the input p-values are one-sided.
method	- default = "ecc": the effective correlation coefficient method in reference 2. "ave": the average method in reference 3, which is an earlier version of reference 2. The "ecc" method is more accurate and numerically stable than "ave" method.
ei	- the eigenvalues of M if available.

Value

pvalue - The p-value of the phi-divergence test.

phistat - phi-diergence statistic.

location - the order of the input p-values to obtain phi-divergence statistic.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and power of optimal signal-detection statistics in finite case", IEEE Transactions on Signal Processing (2020) 68, 1021-1033 2. Hong Zhang and Zheyang Wu. "The general goodness-of-fit tests for correlated data", Computational Statistics & Data Analysis (2022) 167, 107379 3. Hong Zhang and Zheyang Wu. "Generalized Goodness-Of-Fit Tests for Correlated Data", arXiv:1806.03668. 4. Leah Jager and Jon Wellner. "Goodness-of-fit tests via phi-divergences". Annals of Statistics 35 (2007).

See Also

`stat.phi` for the definition of the statistic.v

Examples

```
stat.test = rnorm(20) # Z-scores
p.test = 2*(1 - pnorm(abs(stat.test)))
test.phi(p.test, M=diag(20), s = 0.5, k0=1, k1=10)
test.phi(p.test, M=diag(20), s = 1, k0=1, k1=10)
test.phi(p.test, M=diag(20), s = 2, k0=1, k1=10)
```

---

test.phi.omni — *calculate the right-tail probability of omnibus phi-divergence statistics under general correlation matrix.*

Description

calculate the right-tail probability of omnibus phi-divergence statistics under general correlation matrix.

Usage

```
test.phi.omni(prob, M, K0, K1, S, onesided = FALSE, method = "ecc", ei = NULL)
```

Arguments

`prob`	- vector of input p-values.
`M`	- correlation matrix of input statistics (of the input p-values).
`K0`	- vector of search range starts (from the k0th smallest p-value).
`K1`	- vector of search range ends (at the k1th smallest p-value).
`S`	- vector of the phi-divergence test parameters.
`onesided`	- TRUE if the input p-values are one-sided.
`method`	- default = "ecc": the effective correlation coefficient method in reference 2. "ave": the average method in reference 3, which is an earlier version of reference 2. The "ecc" method is more accurate and numerically stable than "ave" method.
`ei`	- the eigenvalues of M if available.

Value

p-value of the omnibus test.

p-values of the individual phi-divergence test.

References

1. Hong Zhang, Jiashun Jin and Zheyang Wu. "Distributions and power of optimal signal-detection statistics in finite case", IEEE Transactions on Signal Processing (2020) 68, 1021-1033 2. Hong Zhang and Zheyang Wu. "The general goodness-of-fit tests for correlated data", Computational Statistics & Data Analysis (2022) 167, 107379 3. Hong Zhang and Zheyang Wu. "Generalized Goodness-Of-Fit Tests for Correlated Data", arXiv:1806.03668.

Examples

```
M = matrix(0.3,10,10) + diag(1-0.3, 10)
test.phi.omni(runif(10), M=M, K0=rep(1,2), K1=rep(5,2), S=c(1,2))
```

Index