Package: fasano.franceschini.test (via r-universe)

September 13, 2024

Type Package

Title Fasano-Franceschini Test: A Multivariate Kolmogorov-Smirnov Two-Sample Test

Version 2.2.3

Description An implementation of the two-sample multivariate Kolmogorov-Smirnov test described by Fasano and Franceschini (1987) <doi:10.1093/mnras/225.1.155>. This test evaluates the null hypothesis that two i.i.d. random samples were drawn from the same underlying probability distribution. The data can be of any dimension, and can be of any type (continuous, discrete, or mixed).

License MIT + file LICENSE

URL `https://github.com/braunlab-nu/fasano.franceschini.test`

BugReports `https://github.com/braunlab-nu/fasano.franceschini.test/issues`

Depends R (>= 3.6.2)

Imports Rcpp (>= 1.0.0), RcppParallel (>= 5.0.1)

Suggests testthat (>= 3.0.0)

LinkingTo Rcpp (>= 1.0.0), RcppParallel (>= 5.0.1)

Config/testthat/edition 3

Encoding UTF-8

LazyData true

RoxygenNote 7.3.1

SystemRequirements GNU make

Repository https://braunlab-nu.r-universe.dev

RemoteUrl https://github.com/braunlab-nu/fasano.franceschini.test

RemoteRef HEAD

RemoteSha 4fdf12b009745cd996cf1371564b636ccaf627ed

Contents

`fasano.franceschini.test` *Fasano-Franceschini Test*

Description

Performs a two-sample multivariate Kolmogorov-Smirnov test as described by Fasano and Franceschini (1987). This test evaluates the null hypothesis that two i.i.d. random samples were drawn from the same underlying probability distribution. The data can be of any dimension and of any type (continuous, discrete, or mixed).

Usage

```
fasano.franceschini.test(
  S1,
  S2,
  nPermute = 100,
  threads = 1,
  seed = NULL,
  verbose = TRUE,
  method = c("r", "b")
)
```

Arguments

`S1`	A `matrix` or `data.frame`. Each row represents one observation.
`S2`	A `matrix` or `data.frame`. Each row represents one observation.
`nPermute`	A nonnegative `integer` setting the number of permutations to use for performing the permutation test. Default is `100`. If set to `0`, only the test statistic is computed.
`threads`	A positive `integer` or `"auto"` setting the number of threads to use during the permutation test. If set to `"auto"`, the number of threads is determined by `RcppParallel::defaultNumThreads()`. Default is `1`.
`seed`	An optional integer to seed the PRNG used for the permutation test. A seed must be passed to reproducibly compute p-values.
`verbose`	A `boolean` indicating whether to display a progress bar. Default is `TRUE`. Only available when `threads = 1`.
`method`	An optional `character` indicating which method to use to compute the test statistic. The two methods are `'r'` (range tree) and `'b'` (brute force). Both methods return the same results but may vary in computation speed. If this argument is not passed, the sample sizes and dimension of the data are used to infer which method is likely faster. See the Details section for more information.

Details

The test statistic can be computed using two different methods. Both methods return identical results, but have different time complexities:

- Range tree method: This method has a time complexity of *O(N*log(N)^(d-1))*, where *N* is the size of the larger sample and *d* is the dimension of the data.
- Brute force method: This method has a time complexity of *O(N^2)*.

The range tree method tends to be faster for low dimensional data or large sample sizes, while the brute force method tends to be faster for high dimensional data or small sample sizes. When `method` is not passed, the sample sizes and dimension of the data are used to infer which method will likely be faster. However, as the geometry of the samples can influence computation time, the method inferred to be faster may not actually be faster. To perform more comprehensive benchmarking for a specific dataset, `nPermute` can be set equal to `0`, which bypasses the permutation test and only computes the test statistic.

Value

A list of class `htest` containing the following components:

`statistic`	The value of the test statistic.
`p.value`	The permutation test p-value.
`method`	The name of the test.
`data.name`	The names of the original data objects.

References

- Fasano, G. & Franceschini, A. (1987). A multidimensional version of the Kolmogorov-Smirnov test. *Monthly Notices of the Royal Astronomical Society*, 225:155-170. doi:10.1093/mnras/225.1.155.

Examples

```
set.seed(0)

# create 2-D samples
S1 <- data.frame(x = rnorm(n = 20, mean = 0, sd = 1),
                 y = rnorm(n = 20, mean = 1, sd = 2))
S2 <- data.frame(x = rnorm(n = 40, mean = 0, sd = 1),
                 y = rnorm(n = 40, mean = 1, sd = 2))

# perform test
fasano.franceschini.test(S1, S2)

# perform test with more permutations
fasano.franceschini.test(S1, S2, nPermute = 150)

# set seed for reproducible p-value
fasano.franceschini.test(S1, S2, seed = 0)$p.value
fasano.franceschini.test(S1, S2, seed = 0)$p.value
```

```
# perform test using range tree method
fasano.franceschini.test(S1, S2, method = 'r')

# perform test using brute force method
fasano.franceschini.test(S1, S2, method = 'b')

# perform test using multiple threads to speed up p-value computation
## Not run:
fasano.franceschini.test(S1, S2, threads = 2)

## End(Not run)
```

Index